An ensemble wavelet-based stochastic data-driven framework for addressing nonlinearity, multiscale change, and uncertainty in water resources forecasting

John Martin Quilty

Department of Bioresource Engineering

McGill University

Submitted August 2018

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Abstract

Data-driven forecasting (i.e., regression, machine learning, artificial intelligence, etc.) has become a popular and very useful alternative to physically-based and conceptual forecasting approaches in the water resources domain since such methods solely rely on statistical relationships between explanatory variables and the target process, require no explicit physical knowledge of the processes under study, are rapid to develop, have low-costs, and are easy to implement in real-time. However, similar to physically-based and conceptual forecasting approaches, the nonlinear, multiscale, and uncertain nature of water resources provide challenges in the development of accurate and reliable data-driven forecasts.

To address the nonlinear, multiscale, and uncertain nature of water resources this research develops an innovative ensemble wavelet-based stochastic data-driven forecasting framework (EW-SDDFF) that results in forecasts of a target process in the form of a probability density function. EW-SDDFF is developed, tested, and applied to a real-world daily urban water demand forecasting experiment in Montreal, Quebec where it is shown to produce accurate and reliable forecasts at multiple lead times, outperforming numerous benchmarks, and performing especially well during the July, 2010 heatwave that affected Montreal (and many other parts of Quebec).

EW-SDDFF addresses the nonlinear, multiscale, and uncertain nature of water resources in three main ways: 1) it uses nonlinear information-theoretic input variable selection and nonlinear data-driven forecasting methods; 2) it uses wavelet transforms to address multiscale changes in explanatory variables and the target process; and 3) it adopts stochastics for the uncertainty assessment of input data, input variable selection, parameters, and model output. The end result of EW-SDDFF is a stochastic forecast that holistically addresses nonlinearity, multiscale change, and uncertainty.

The main innovations behind the EW-SDDFF are contained in its development which takes place in four key stages: 1) new computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods are developed to provide the most important input variables to nonlinear data-driven methods to forecast the target process; 2) a set of best (correct) practices are developed for using wavelet transforms correctly in wavelet-based forecasting models and formed into a new wavelet-based forecasting framework (WDDFF) that can be used with multiple wavelet transforms, different input variable selection methods, and data-driven forecasting models and that may be applied for the correct development of wavelet-based forecasting models for real-world applications; 3) uncertainty assessment is included in WDDFF by adopting a stochastic framework, resulting in a new stochastic wavelet-based

forecasting framework (SWDDFF); and 4) to take advantage of the strengths of multiple wavelet transforms, different input variable selection methods and data-driven models, the single-wavelet SWDDFF is transformed into an ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) by using multiple WDDFF forecasts as input data, improving forecast accuracy and reliability when compared to its single-wavelet counterparts (SWDDFF). EW-SDDFF includes both ensemble member selection and weighting uncertainties, using input variable selection and data-driven modeling, respectively, and also accounts for input data and ensemble model output uncertainties. Both SWDDFF and EW-SDDFF represent the most advanced single- and multi-wavelet data-driven forecasting frameworks in the literature.

Since EW-SDDFF quantifies forecast uncertainty (in the form of a probability density function), it may serve as a useful tool for operational, planning, and management tasks faced by water resources managers, especially during decision-making stages.

Résumé

Devenue, dans le domaine des ressources en eau, une alternative à la fois populaire et utile aux approches prévisionnelles conceptuelles ou fondées sur des critères physiques, les prévisions axées sur les données (c.à.d. régression, apprentissage machine, intelligence artificielle, etc.) sont privilégiées parce qu'elles reposent uniquement sur les relations statistiques entre les variables explicatives et le processus cible, n'exigent aucune connaissance physique explicite du processus étudié, sont rapidement élaborés, de faible coût, et faciles à mettre en œuvre en temps réel. Cependant, tout comme avec les approches prévisionnelles conceptuelles ou fondées sur des critères physiques, le caractère non-linéaire, multi-échelle et incertain des ressources hydriques pose un défi au développement de prévisions axées sur les données qui soient à la fois précises et fiables.

Afin de s'adresser au caractère non-linéaire, multi-échelle et incertain des ressources hydriques, la présente étude visa à élaborer un innovant ensemble prévisionnel stochastique axé sur les données invoquant l'utilisation d'ondelettes (EW-SDDFF), permettant ainsi de prévoir un processus cible sous forme de fonction de densité de probabilité. Élaboré, mis à l'épreuve et mis en pratique dans une situation réelle de prévision de la demande urbaine d'eau quotidienne de la ville de Montréal (Québec), le EW-SDDFF offrit des prévisions précises et fiables sous différents délais, surpassant plusieurs indices de référence, et s'avérant particulièrement performant lors de la canicule de juillet 2010 qui frappa Montréal et plusieurs autres régions du Québec.

L'EW-SDDFF aborde le caractère non-linéaire, multi-échelle et incertain des ressources hydriques de trois manières innovant, il emploi (i) un critère puisé de la théorie de l'information pour choisir les variables d'entrée, ainsi que des méthodes de prévision non-linéaires axées sur les données, (*ii*) une transformée en ondelettes pour s'adresser aux variations multi-échelles des variables explicatives et du processus cible, et (*iii*) la stochastique pour évaluer l'incertitude dans les données d'entrée, la sélection des données d'entrée, des paramètres et données de sortie des modèles. En fin de compte, l'EW-SDDFF offre une prévision stochastique qui s'adresse globalement à la non-linéarité, les variations multi-échelle et incertitude.

L'élaboration de l'EW-SDDFF s'opéra en quatre étapes principales: (*i*) élaborer de nouvelles méthodes de modélisation dont le traitement informatique est efficace, non-linéaire et non-paramétrique, et qui permettent de sélectionner de données d'entrée pertinentes à la prévision du processus cible par un

modèle non-linéaire axé sur les données, selon des critères puisés dans théorie de l'information; (*ii*) élaborer à la fois un ensemble de pratiques visant le bon usage des transformées en ondelettes dans les modèles prévisionnels à base d'ondelettes, et un réseau prévisionnel à base d'ondelettes (WDDFF) permettant de mettre en œuvre et comparer plusieurs types de transformées en ondelettes, différentes méthodes de sélection des données d'entrée, et différents modèles prévisionnels axées sur les données, tout cela pour créer de tels modèles pouvant être appliqués en situation réelle; (*iii*) en adoptant un cadre stochastique, une évaluation de l'incertitude est incluse dans le WDDFF, donnant lieu a un réseau prévisionnel à base d'ondelettes stochastique (SWDDFF); et (*iv*) en vue de tirer parti de la puissance des transformées en ondelettes multiples, des différentes méthodes de sélection des données d'entrée, et des modèles axées sur les données, un SWDDFF mono-ondelette fut transformé en un cadre stochastique multi-ondelettes d'ensemble axée sur les données (EW-SDDFF) et puisant sur de multiples prévisions WDDFF comme données d'entrée. En comparaison à son homologue mono-ondelette (SWDDFF), le EW-SDDFF améliora la précision et la fiabilité des prévisions. Le EW-SDDFF inclus à la fois une sélection d'ensemble des membres et une pondération d'incertitudes, de manière qu'un choix entre données d'entrée variables et une modélisation axée sur les données sont, respectivement, mis en œuvre. Cette approche tient également compte des incertitudes des données d'entrée et des données de sortie des modèles d'ensemble. Ensemble, SWDDFF et EW-SDDFF représentent les réseaux de prévision axées sur les données mono- et multi-ondelette les plus avancés rapportés dans la littérature.

Comme l'EW-SDDFF quantifie l'incertitude des prévisions (sous forme de fonction de densité de probabilité), il peut représenter un important outil pour les tâches opérationnelles, de planification et de gestion auxquels sont confrontés — particulièrement lors d'étapes décisionnelles — les gestionnaires des ressources en eau.

Acknowledgements

With all of my being I wish to thank the Lord for blessing me with the strength, perseverance, and patience to complete this work. Not a day has gone by without a struggle to stay honest and true to achieving the goal of finishing this thesis and I can say that without the Lord's help I would not have been able to complete this research.

Through these last several years I have come to see how blessed I am to have such supportive family, friends, mentors, colleagues, and acquaintances that have provided me with endless inspiration. Without listing the names of all who helped, I want to sincerely thank each and every person that has assisted me along this journey.

In my personal life, I am honored to have a beautiful, nurturing, and supportive wife who has, from the moment we met, until now, been a light supporting me in achieving my goals, sharing in my joys and hardships, and sacrificing much to help me. Erin, I love you with all of my heart and soul. This year has been a busy one, preparing for our wedding and finishing this doctorate; knowing that you always have my best interests in mind has provided me with the freedom to dedicate my mind, heart, and will to this research, in hopes that this leads to a brighter and more complete future. Thank you Erin: your love, support, and sacrifice has meant everything to me.

My family has always given me endless support in following my goals and aspirations. Whether in sports or academics, they have invested not only their time and money but also shared in my achievements and failures. Mom, Dad, and Jennifer: thank you for loving me unconditionally and always being there for me. Each of you has helped shape my understanding of family and how to give selflessly to others.

I want to thank my beautiful grandmothers, May and Jean, for all of their support and encouragement. Each and every visit I was asked how my studies were coming along and how close I was to finishing. Nanny and Grandma, I am pleased to now have the opportunity to say that my doctorate has been completed! It is with great sorrow that my grandmother, Jean Burrill, passed away a day after my defence on August 24, 2018. I was blessed with the opportunity to be at my grandmother's bedside with my family at the time of her passing. Grandma, may God grant your soul the peace it deserves.

Papa and Grandpa, I pray each day that God rests your souls. Both of you are in my thoughts and I am thankful to have had such strong grandfathers that demonstrated how to provide for their families and the effort required to establish strong careers and work habits.

To my truly amazing friends: Jesse, Mike, Paul, Alex, Dan, Rocco, Dom, Shannon, Catherine, and Lisa, thank you for blessing me with your friendship and for supporting me well before and especially through my studies.

In terms of my professional life, I can barely come up with words to express my sincere gratitude for the mentorship I have received from Prof. Adamowski. From my first e-mail exchange to the current moment, Prof. Adamowski has been the most humble, professional, and inspiring mentor I have ever been blessed to receive. Prof. Adamowski has helped me to grow both professionally and personally, seeing in me aptitudes I never knew I had and allowing me to convert some of what I once thought were weaknesses, into strengths. I want to thank you Prof. Adamowski for being an excellent mentor, I hope we are always able to find ways to continue our collaborations.

I would further like to thank others that have supported me by giving me their time or financial support: Marie-Amélie Boucher for allowing me to work on interesting projects in data assimilation and providing help and support with my research; Anteneh Belayneh and Bahaa Khalil for helping me learn the basics of statistical hydrology and coding in Matlab; Maheswaran Rathinasamy for helping me understand some of the finer details of wavelet transforms; Azhar, Julien, and Deasy for helping me with proposals, poster and oral presentations, and advice in preparing my thesis; Profs. Adamowski, Nguyen, Orsat, and Prasher for sitting on my comprehensive exam committee; Profs. Adamowski, Burn, Madramootoo, and Prasher for sitting on my doctoral defence committee; and administrative staff that have helped me over the years in the organization and completion of different milestones, such as Ms. Gregus and Ms. Trudeau. Financial support is recognized from Natural Sciences and Engineering Research Council of Canada (NSERC) Engage, Discovery, and Accelerator grants held by Prof. Adamowski as well as Graduate Excellence Awards from the Department of Bioresource Engineering.

Format of Thesis

The thesis must include the following:

(a) A table of contents;

(b) An abstract in English and French;

(c) An introduction which clearly states the rational and objectives of the research;

(d) A comprehensive review of the literature (in addition to that covered in the introduction to each paper);

(e) A final conclusion and summary;

1. As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g. in appendices) in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported in the thesis.

2. In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled "Contributions of Authors" as a preface to the thesis. The supervisor must attest to the accuracy of this statement at the doctoral oral defense. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to clearly specify the responsibilities of all the authors of the “coauthored papers”.

Contributions of the Authors

Chapters 3-6 of this thesis have been prepared and submitted for publication in peer-review journals and have been presented at various scientific conferences. The author of this thesis was responsible for the development, testing, and application of the different methods discussed in this research along with preparation of manuscripts submitted to peer-review journals as well as poster and oral presentations presented at scientific conferences. Prof. Adamowski is the supervisor of this thesis and provided valuable advice on all aspects of the research and contributed to the review and editing of each manuscript, poster, and oral presentation.

Dr. Bahaa Khalil, Post-doctoral Researcher in the Bioresource Engineering Department at McGill and Dr. Maheswaran Rathinasamy, Humboldt Fellow at the Potsdam Institute for Climate Impact Research, Germany, provided advice on the methodological aspects and contributed to the review and editing of the first manuscript published in Water Resources Research. Prof. Marie-Amélie Boucher at the Civil Engineering Department, Université de Sherbrooke, provided valuable advice on ensemble and probabilistic forecasting model evaluation and contributed to the review and editing of the fourth manuscript submitted for publication in Water Resources Research. Miss Deasy Nalley, PhD Candidate in the Bioresource Engineering Department at McGill, Dr. Azhar Inam Baig, Post-doctoral Researcher in the Bioresource Engineering Department at McGill, and Mr. Julien Mallard, PhD Candidate in the Bioresource Engineering Department at McGill each provided assistance for this research in terms of advice on preparing the PhD proposal, poster and oral presentations, and formatting of this PhD thesis.

List of publications and scientific presentations related to the thesis:

A. Thesis components that have been published or submitted for publication in peer-review journals:

Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. Water Resour. Res. 52, 2299–2326. doi:10.1002/2015WR016959

Quilty, J., Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J. Hydrol. 563, 336 – 353. doi:10.1016/j.jhydrol.2018.05.003

Quilty, J., Adamowski, J., 2018. A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes. J. Hydrol. (Under review: Manuscript Number, HYDROL28223)

Quilty, J., Adamowski, J., Boucher, M.-A., 2018. A stochastic data-driven ensemble forecasting framework for water resources: A case study using ensemble members derived from a database of deterministic wavelet-based models. Water Resour. Res. (Accepted pending revision: Manuscript Number, 2018WR023205)

B. Thesis components that have been presented at scientific conferences:

Quilty J. and Adamowski J., 2018. Nonlinear input variable selection techniques for multivariate data-driven modelling. Computational Methods in Water Resources (CMWR) XXII. June 3- 7, 2018. Saint-Malo, France. (Oral Presentation)

Quilty J. and Adamowski J., 2018. A wavelet-based stochastic data-driven forecasting framework for forecasting uncertain water resources processes. Computational Methods in Water Resources (CMWR) XXII. June 3- 7, 2018. Saint-Malo, France. (Oral Presentation)

Quilty J. and Adamowski J., 2018. Information-theoretic-based input variable selection for hydrology and water resources. European Geophysical Union (EGU) General Assembly. April 8- 13, 2018. Vienna, Austria. (Oral Presentation)

Quilty J. and Adamowski J., 2018. A stochastic data-driven forecasting framework using wavelets for forecasting uncertain hydrological and water resources processes. European Geophysical Union (EGU) General Assembly. April 8- 13, 2018. Vienna, Austria. (Poster Presentation)

Quilty J, Adamowski J, Khalil B, Rathinasamy, M., 2016. Nonlinear input variable selection techniques for multivariate water resources modeling. Statistics in Hydrology (STAHY) 2016 Workshop. September 26-27, 2016. Quebec City, Canada. (Oral Presentation)

Quilty J, Adamowski J, Khalil B, Rathinasamy, M., 2016. Nonlinear input variable selection techniques using information theory for water resources modeling. 69th National Conference of the Canadian Water Resources Association (CWRA). May 25- 27, 2016. Montreal, Canada. (Oral Presentation)

Quilty J, Adamowski J, Khalil B, Rathinasamy, M., 2015. Comparison of different time delay embedding strategies for urban water demand (UWD) forecasting using machine learning techniques. European

Geophysical Union (EGU) 10th Alexander von Humboldt Conference. November 18- 20, 2015. Addis Ababa, Ethiopia. (Oral Presentation)

Quilty J, Adamowski J. 2015. Machine learning based urban water demand (UWD) forecasts incorporating stochastic weather inputs. European Geophysical Union (EGU) 10th Alexander von Humboldt Conference. November 18- 20, 2015. Addis Ababa, Ethiopia. (Oral Presentation)

Table of Contents

Abstract ii
Résumé iv
Acknowledgements vi
Format of Thesis viii
Contributions of the Authors ix
Table of Contents xii
List of Figures xxi
List of Tables xxiv
List of Abbreviations xxvi
Chapter 1: Introduction 1
1.1. Water Resources Forecasting Approaches 1
1.1.1. Why are Data-Driven Forecasting Methods Useful in Water Resources? 2
1.2. Nonlinearity, Multiscale Change, and Uncertainty as Challenges in the Development of Accurate and Reliable Data-Driven Water Resources Forecasting Models 3
1.2.1. Challenge 1: Nonlinearity 3
1.2.1.1. Solution to Challenge 1 4
1.2.2. Challenges 2: Multiscale Change 5
1.2.2.1. Solution to Challenge 2 5
1.2.3. Challenge 3: Uncertainty Assessment 6
1.2.3.1. Solution to Challenge 3 6
1.3. Research Objectives 7
1.4. Contributions 9
1.5. Thesis Outline 10
1.6. References 11
Chapter 2: Literature Review 19
2.1. Nonlinearity – Data-Driven Forecasting and Input Variable Selection Methods 20
2.1.1. Different Data-Driven Methods 20
2.1.2. Popular Nonlinear Data-Driven Methods in Water Resources Forecasting 20
2.1.2.1. Drawbacks of Popular Nonlinear Data-Driven Methods in Water Resources Forecasting 21
2.1.2.2. Overcoming Drawbacks of Popular Nonlinear Data-Driven Methods in Water Resources Forecasting with Easy to Use and Computationally Efficient Methods 22

2.1.2.2.1. Extreme Learning Machine 22
2.1.2.2.2. Volterra Series Model 23
2.1.3. Nonlinearity – Input Variable Selection Methods 24
2.1.3.1. Overview 24
2.1.3.2. Popular Filter-based Input Variable Selection Methods 24
2.1.3.3. Methods that Account for Redundancy amongst Already Selected Inputs 25
2.1.3.4. Non-Heuristic Methods for the Estimation of Partial (Conditional) Mutual Information . 26
2.1.3.5. Non-Heuristic and Computationally Efficient Conditional Mutual Information-based Input Variable Selection 26
2.1.3.6. Input Variable Selection Uncertainty 27
2.2. Multiscale Change – Wavelet-based Forecasting 28
2.2.1. Overview of Different Methods used for Capturing Multiscale Changes in Data-Driven Water Resources Forecasting Models 28
2.2.2. The Use and Misuse of the Wavelet Transform in Wavelet-based Data-Driven Forecasting .. 30
2.2.2.1. Overview 30
2.2.2.2. Contrasting Different Wavelet Transforms 32
2.2.3. Best (Correct) Practices for Wavelet-based Forecasting of Real-World Processes 34
2.2.3.1. Best Practice 1: Using Causal Wavelet Transforms 35
2.2.3.2. Best Practice 2: Careful Selection of Decomposition Levels and Wavelet Filters 35
2.2.3.3. Best Practice 3: Proper Partitioning of Calibration and Validation Data 36
2.2.4. Best Practices and the Wavelet Data-Driven Forecasting Framework 37
2.3. Uncertainty – Using Stochastics for Uncertainty Assessment 37
2.3.1. Uncertainty Assessment in Water Resources Forecasting 38
2.3.2. The Blueprint for Converting Deterministic Forecasting Models into their Stochastic Counterparts 39
2.3.3. Stochastic Wavelet Data-Driven Forecasting Framework 40
2.3.4. Ensemble Wavelet - Data-Driven Forecasting Framework 41
2.4. Summary 43
2.5. References 44
Connecting Text to Chapter 3 62

Chapter 3: Bootstrap Rank-Ordered Conditional Mutual Information (broCMI): A Nonlinear Input Variable Selection Method for Water Resources Modeling 63
3.1. Abstract 63
3.2. Introduction 63
3.3. Theoretical Background 70
3.3.1. Entropy, Mutual Information, and Conditional Mutual Information 70
3.3.2. Input Variable Selection via Conditional/Partial Mutual Information 72
3.3.3. Input Variable Selection via Bootstrap Rank-Ordered Conditional Mutual Information 72
3.3.3.1. Bootstrap Rank-Ordered Conditional Mutual Information Overview 72
3.3.3.2. Bootstrap Rank-Ordered Conditional Mutual Information Details 74
3.4. Input Variable Selection Algorithms 77
3.4.1. Partial Correlation Input Selection 77
3.4.2. Partial Mutual Information Selection 77
3.4.3. *k* Nearest-Neighbours based Conditional Mutual Information 78
3.4.4. Kernel Density Estimation based Conditional Mutual Information 78
3.4.5. Edgeworth Approximations based Conditional Mutual Information 79
3.4.6. Bootstrap Rank-Ordered Conditional Mutual Information 79
3.4.7. Stopping-Criterion for Input Variable Selection 79
3.5. Experimental Set-Up 80
3.5.1. Experiment I – Input Variable Selection Comparisons using Synthetic Datasets 80
3.5.1.1. Selection Accuracy Evaluation Metrics for Input Variable Selection 80
3.5.1.2. Synthetic Test Problems 81
3.5.1.2.1. Linear and Nonlinear Autoregressive Datasets 82
3.5.1.2.2 Nonlinear Datasets with Exogenous Covariates 83
3.5.1.2.3 Kentucky River Basin Dataset 83
3.5.1.3. Procedure for Comparing Input Variable Selection Algorithms on Synthetic Datasets 84
3.5.2. Experiment II – Input Variable Selection Applied in the Context of Urban Water Demand Forecasting 85
3.5.2.1. Study Site and Datasets 85
3.5.2.2. Procedure for Comparing Input Variable Selection Algorithms on the Urban Water Demand Forecasting Datasets 87

3.6. Results and Discussion 88
3.6.1. Experiment I – Synthetic Datasets 88
3.6.1.1. Quantitative Performance Comparisons across Input Variable Selection Algorithms 89
3.6.1.1.1. Selection Accuracy 89
3.6.1.1.1.1. Linear and Nonlinear Autoregressive Datasets 89
3.6.1.1.1.2. Nonlinear Datasets with Exogenous Covariates 89
3.6.1.1.1.3. Kentucky River basin Dataset 90
3.6.1.1.2. Computational Run-Time 91
3.6.1.2. Qualitative Comparisons across Input Variable Selection Algorithms 95
3.6.1.2.1. Ease of Use and Robustness 95
3.6.1.2.2. Explanation Capability 95
3.6.1.2.3. Flexibility 96
3.6.2. Experiment II – Urban Water Demand Datasets 96
3.6.2.1 Different machine learning models (ELM and ANN) paired with broCMI 97
3.6.2.2. Different Input Variable Selection Algorithms Paired with ELM 100
3.6.2.3. Out-of-Sample ELM Forecasts 103
3.7. Closure 106
3.8. Acknowledgements 108
3.9. References 108
Appendix 118
3. A.1. Introduction 118
3. A.2. Generation of Candidate Input Variables 118
3. A.2.1. Phase Space Reconstruction 119
3. A.3. Dataset Partitioning 122
3. A.4. Selecting Input Variables 122
3. A.5. Forecasting Selected Input Variable Sets using Regression Models 122
3. A.5.1. Artificial Neural Network Theoretical Background 123
3. A.5.2. Artificial Neural Network Model Development 123

3. A.5.3. Extreme Learning Machine Theoretical Background 124
3. A.5.4. Extreme Learning Machine Development 126
3. A.6. Forecast Performance Evaluation 126
References 127
Connecting Text to Chapter 4 132
Chapter 4: Addressing the Incorrect Usage of Wavelet-based Hydrological and Water Resources Forecasting Models for Real-World Applications with Best Practices and a New Forecasting Framework 133
4.1. Abstract 133
4.2. Introduction 134
4.3. Theoretical Basis for Best (Correct) Practices in Wavelet-based Forecasting 139
4.3.1. An Overview of the Different Wavelet Transforms used in Hydrology and Water Resources for Wavelet-based Forecasting 139
4.3.2. Use and Misuse of Wavelet and Scaling Coefficients in Hydrology and Water resources Wavelet-based Forecasting 142
4.3.3. Boundary Condition-related Issues and Solutions in Wavelet-based Forecasting 144
4.3.3.1. The 'Future Data' Issue 144
4.3.3.1.1. The 'Future Data' Solution 147
4.3.3.2. The Improper Selection of Decomposition Level(s) and Wavelet Filter(s) Issue 147
4.3.3.2.1. The Solution for Properly Selecting Decomposition Level(s) and Wavelet Filter(s) . 149
4.3.3.3. The Dataset Partitioning Issue 149
4.3.3.3.1. The Solution for Correctly Partitioning a Dataset 150
4.4. Best (Correct) Practices and the Wavelet Data-Driven Forecasting Framework 151
4.4.1. Best Practices Adopted by the Wavelet Data-Driven Forecasting Framework 152
4.4.2. The 'Building Blocks' of the Wavelet Data-Driven Forecasting Framework 152
4.4.2.1. Pre-Processing Data 153
4.4.2.1.1. Different Wavelet-based Forecasting Methods 153
4.4.2.2. Input variable selection for wavelet and scaling coefficients 155
4.4.2.3. Data-Driven Model Selection 156
4.4.3. Summary of the Wavelet Data-Driven Forecasting Framework 157
4.5. Experimental Setup 158

4.5.1. Methodology for Wavelet Data-Driven Forecasting Framework 159
4.5.1.1. Quantitative Performance Comparisons across Input Variable Selection Algorithms 159
4.5.1.2. Input Variable Selection for Wavelet and Scaling Coefficients 160
4.5.1.3. Data-Driven Models 160
4.5.1.4. Forecast Calibration and Evaluation 160
4.5.2. Experimental Details 161
4.6. Results and Discussion 162
4.6.1 Best Performing Models 164
4.6.2. Best Performing Decomposition Levels and Wavelet Filters 166
4.6.3. Comparing the Wavelet Data-Driven Forecasting Framework against a Forecasting Model using the Maximal Overlap Discrete Wavelet Transform-based Multiresolution Analysis 169
4.7. Summary and Conclusions 171
4.8. Acknowledgements 174
4.9. References 174
Appendix 187
4. A.1. Details on Wavelet Decomposition Wavelet and Scaling Coefficients, and Wavelet and Scaling Filters 187
4. B.1. Study Area Details 192
4. C.1. Experimental Setup for the Wavelet Data-Driven Forecasting Framework 195
4. C.1.1. Target, Forecast Lead Times, and Explanatory Variables 195
4. C.1.2. Wavelet Decomposition: Decomposition Level and Wavelet Filter Selection and Wavelet-based Forecasting Approaches 196
4. C.1.3. Dataset Partitioning 196
4. C.1.4. Input Variable Selection, Forecast Calibration, and Assessment 198
4. C.1.5. Summary of the Different Experiments 199
4. D.1. Experiment Results 200
References 203
Connecting Text to Chapter 5 208
Chapter 5: A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes 209
5.1. Abstract 209

5.2. Introduction 209
5.3. Theory 215
5.3.1. A Brief Overview of the Blueprint (Montanari and Koutsoyiannis, 2012) 215
5.3.2. From the Blueprint to the Stochastic Data-Driven Forecasting Framework 217
5.3.3. Stochastic Wavelet Data-Driven Forecasting Framework 218
5.4. Applying the Stochastic Wavelet Data-Driven Forecasting Framework 219
5.5. Experiment Settings, Case Study, and Forecast Evaluation 221
5.5.1. Estimation of Probability Density Functions 222
5.5.1.1. Input Variable Selection Uncertainty 222
5.5.1.2. Parameter Uncertainty 223
5.5.1.3. Model Error Uncertainty 223
5.5.2. Sampling from the Probability Density Functions 224
5.5.3. Case Study 225
5.5.3.1. Model Settings 227
5.5.4. Forecast Evaluation Metrics 229
5.6. Results and Discussion 230
5.7. Summary and Conclusions 237
5.8. Acknowledgements 239
5.9. References 239
Connecting Text to Chapter 6 250
Chapter 6: A Stochastic Data-Driven Ensemble Forecasting Framework for Water Resources: A Case Study using Ensemble Members Derived from a Database of Deterministic Wavelet-based Models 251
6.1. Abstract 251
6.2. Introduction 251
6.2.1. A Blueprint for Converting Deterministic to Stochastic Forecasts 252
6.2.2. From a Blueprint for Stochastic Process-based Models to a Stochastic Data-Driven Forecasting Framework Involving Wavelets 253
6.2.3. The Case for a Multi-Wavelet Stochastic Data-Driven Ensemble Forecasting Framework 253
6.2.4. How to Make a Computationally Efficient Ensemble Multi-Wavelet Stochastic Data-Driven Forecasting Framework 255
6.2.5. Contributions and Benefits of the Proposed Framework 255
6.3. Methods 257
6.3.1. Wavelet Data-Driven Forecasting Framework 257

6.3.2. Stochastic Data-Driven Forecasting Framework 258
6.3.3. Stochastic Wavelet Data-Driven Forecasting Framework 260
6.3.4. Ensemble Stochastic Data-Driven Forecasting Framework using Wavelet-based Forecasts as Input Data 262
6.4. Experiment Details 264
6.4.1. Estimation and Sampling from the Different Probability Density Functions 264
6.4.2. Case Study 266
6.4.2.1. Study Site Overview 266
6.4.2.2. Input Data for the Ensemble Wavelet – Stochastic Data-Driven Forecasting Framework 267
6.4.2.3. Forecast Evaluation Metrics 268
6.4.2.4. Developing our Proposed Method and its Comparison with Benchmarks 268
6.5. Results and Discussion 270
6.5.1. Selecting Input Data for the Ensemble Wavelet-Stochastic Data-Driven Forecasting Framework and Deterministic Benchmark Results 270
6.5.2. Comparing the Proposed Model against its Benchmarks 271
6.5.2.1. Comparing the Proposed Method against its Deterministic Version 275
6.5.2.2. Comparing the Proposed Method against its Stochastic Benchmark 276
6.5.2.2.1. Using Coverage Probability Plots to Supplement Probabilistic Forecast Evaluation Metrics 276
6.5.2.2.2. The Effect of Ensemble Size on Performance 278
6.5.3. Stochastic Urban Water Demand Forecasting During a Heatwave 283
6.6. Summary and Conclusions 287
6.7. Acknowledgements 289
6.8. References 289
Appendix 298
6. A.1. Different Variants of Figure 6.8 in Text 298
Chapter 7: Summary and Conclusions 302
7.1. Computationally Efficient, Non-parametric, Nonlinear Information-Theoretic Input Variable Selection Methods 303
7.2. Best (Correct) Practices for Wavelet-based Forecasting and the Wavelet Data-Driven Forecasting Framework 303

7.3. Stochastic Wavelet Data-Driven Forecasting Framework 304
7.4. Ensemble Wavelet – Data-Driven Forecasting Framework 305
Chapter 8: Contributions to Knowledge, Limitations, and Recommendations for Further Research 307
8.1. Contributions to Knowledge 307
8.2. Limitations 308
8.3. Recommendations for Further Research 309

List of Figures

Figure 3.1. Input variable selection procedure for (a) the standard and (b) bootstrap rank-ordered approaches 73
Figure 3.2. Time series plots for the urban water demand time series used in Experiment II 87
Figure 3.3. Selection accuracy results for the synthetic test problems (Experiment I); note: the square brackets indicate the number of nearest-neighbours used in KNN and the rounded brackets indicate the number of bootstrap resamples used in broCMI 91
Figure 3.4. Average seasonal hydrograph generated via ELM and ANN for the 1 day lead time considering each urban water demand time series using inputs derived from broCMI 100
Figure 3.5. Comparing broCMI against EA, KDE, KNN, PMIS, and PCIS using their mean absolute residuals (forecasts versus observations) for each calendar date during the summer demand period generated via ELM for the 1 day lead time considering each urban water demand time series (note: square brackets indicate the number of nearest-neighbours used in KNN while the rounded brackets represents the model rank (i.e. a score of 1 indicates the best model)) 103
Figure 3.6. Scatter plots for the out-of-sample forecasts for each urban water demand time series and lead time using ELM with inputs selected via broCMI 104
Figure 4.1. Depiction of the 'future data' issue boundary condition (see Table 4.1) 146
Figure 4.2. Wavelet Data-Driven Forecasting Framework flow chart 153
Figure 4.3. Scatter plots for the best wavelet-based and non-wavelet-based forecasts for each lead time (1, 3, 5, 7, and 14 days) 164
Figure 4.4. Average urban water demand time series 'U' (measured in megalitres per day) decomposed by the maximal overlap discrete wavelet transform (light colored line) and the *à trous* algorithm (dark colored line) using the la14 wavelet filter and a decomposition level of six; where each 'W' time series represents the wavelet coefficients at a particular scale and 'V' represents the scaling coefficients; the grayed out area represents the number of coefficients (820) that are affected by the boundary condition for the given wavelet filter and decomposition level 167
Figure 4.5. Comparison between the best wavelet-based (gray) and non-wavelet-based (red) 1 day lead time forecasts for the last 366 records in the validation set: a) represent the hydrograph while b) represents the residual, measured in megalitres per day (ML/D) 168
Figure 4.6. Plot showing the differences in 1 day lead time forecast residuals (in megalitres per day (ML/D)) for the validation partition where a) represents the maximal overlap discrete wavelet transform (valid for real-world forecasting problems) with a NASH score of 0.916 and b) the maximal overlap discrete

wavelet transform-based multiresolution analysis (that should not be used for real-world forecasting problems due to its use of 'future data', see section 4.4.2.1., 4.4.2.3., and 4.6.1.) with a NASH score of 0.940 171
Figure 5.1. Workflow for the Stochastic Wavelet Data-Driven Forecasting Framework (modified from Montanari and Koutsoyiannis (2012) and Sikorska et al. (2015)) 221
Figure 5.2. Average urban water demand time series, U, (measured in megalitres per day) decomposed by the maximal overlap discrete wavelet transform using the la14 wavelet filter and a decomposition level of six; where each 'W' time series represents the wavelet coefficients at a particular scale and 'V' represents the scaling coefficients; wavelet and scaling coefficients are also measured in megalitres per day 225
Figure 5.3. CRPS versus the number of resamples (n) for different lead times 234
Figure 5.4. Scatter plots for the mean of the n different forecasts defining $\boldsymbol{f}(\boldsymbol{Q})$ for different lead times a) 1 day ahead, b) 7 day ahead, and c) 14 day ahead 235
Figure 5.5. Scatter plots for the n different forecasts defining $\boldsymbol{f}(\boldsymbol{Q})$ for different lead times a) 1 day ahead, b) 7 day ahead, and c) 14 day ahead 235
Figure 5.6. Time series plots for the mean of the $\boldsymbol{f}(\boldsymbol{Q})$ forecasts (at a 1 day lead time (a), 7 day lead time (c), and 14 day lead time (e)) and for the 0.025 and 0.975 quantiles of the $\boldsymbol{f}(\boldsymbol{Q})$ forecasts (at a 1 day lead time (b), 7 day lead time (d), and 14 day lead time (f)) versus the observations 237
Figure 6.1. Workflow for the Stochastic Wavelet Data-Driven Forecasting Framework (according to Quilty and Adamowski (2018a) and originally modified from Montanari and Koutsoyiannis (2012) and Sikorska et al. (2015)) 262
Figure 6.2. Time series plots for the benchmark (SWDDFF), EW-SDDFF variants (MLR (without input variable selection), PCIS-MLR, EA-ELM, and EA-SOV), and RW against the observed time series for a) 1, b) 7, and c) day lead time(s); the variable $\boldsymbol{U}$ denotes urban water demand measured in megalitres per day [ML/D] 275
Figure 6.3. Coverage Probability Plot for the benchmark (SWDDFF) and EW-SDDFF variants (MLR (without input variable selection), PCIS-MLR, EA-ELM, and EA-SOV) for a) 1, b) 7, and c) 14 day lead time(s) 277
Figure 6.4. CRPS versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead time(s) 279
Figure 6.5. AW versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead time(s) 280
Figure 6.6. IS versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead time(s) 281
Figure 6.7. CPP_mse versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead times 282

Figure 6.8. Mean forecasts and their 95 % prediction intervals for the SWDDFF and the EW-SDDFF for the a) 1 (EA-ELM), b) 7 (EA-SOV), and c) 14 (EA-SOV) day lead time(s) 283
Figure 6.9. Mean forecasts and their 95 % prediction intervals for the SWDDFF and the EW-SDDFF for the a) 1 (MLR) and b) 7 (MLR) day lead time(s) 284

List of Tables

Table 3.1. Pseudo-code for bootstrap rank-ordered conditional mutual information (broCMI) 76
Table 3.2. Synthetic datasets used for comparing input variable selection algorithms 82
Table 3.3. Descriptive statistics for the time series used in Experiment II 86
Table 3.4. Response variables and candidate input variable sets for Experiment II 88
Table 3.5. Computational run-times for the various input variable selection algorithms on the synthetic datasets (the brackets proceeding broCMI indicates the number of bootstrap resamples used in the algorithm) 94
Table 3.6. Input variable sets selected via broCMI and used to generate 1 and 3 day lead time UWD forecasts (note: N^H represents the number of hidden layer neurons used in ELM or ANN) 99
Table 3.7. Comparison of 1 day lead time forecasts using different input variable selection algorithms and ELM for the *a3w* time series (note: the square brackets indicate the number of nearest-neighbours used in KNN; the number of unique input variable sets for each dataset are listed in the rounded brackets beside Model Rank) 101
Table 3.8. Comparison of 1 day lead time forecasts using different input variable selection algorithms and ELM for the *amg* time series (note: the square brackets indicate the number of nearest-neighbours used in KNN; the number of unique input variable sets for each dataset are listed in the rounded brackets beside Model Rank) 101
Table 3.9. ELM out-of-sample performance for 1 and 3 day lead time UWD forecasts (note: N^H represents the number of hidden layer neurons used in ELM) 104
Table 4.1. Wavelet decomposition formulae for different wavelet transforms 141
Table 4.2. Example of wavelet decomposition using the different wavelet transforms (in Table 4.1) 145
Table 4.3. 1 day lead time forecast results 163
Table 5.1. Description of models 227
Table 5.2. Test set performance for different models using the deterministic (NASH, RMSE, and MAE) and probabilistic (CRPS, PICP, AW, and IS) forecast evaluation metrics 232
Table 6.1. Random walk (RW) benchmark results used for identifying input data for the EW-SDDFF 270
Table 6.2. Information and results for the best WDDFF models 271
Table 6.3. Test set performance for the EW-SDDFF, EW-DDFF, SWDDFF, and WDDFF models (Note: the results for best EW-SDDFF and EW-DDFF models for a particular forecast evaluation metric are bolded while the results for the SWDDFF and WDDFF models that outperformed their EW-SDDFF and EW-DDFF counterparts are highlighted in grey) 273

Table 6.4. Test set performance for the EW-SDDFF, EW-DDFF, SWDDFF, and WDDFF models leading up to, during, and after the July 2010 heatwave (July 3 – 17, 2010) (Note: the results for best forecast models for a particular forecast evaluation metric are bolded) 286

List of Abbreviations

a3w	Average urban water demand in 3W pressure zone (Ottawa, Ontario)
ACF	Autocorrelation function
AIC	Akaike Information Criterion
amg	Average urban water demand in Morgan's Grant pressure zone (Ottawa, Ontario)
ANN	Artificial neural network
api	Antecedent Precipitation Index
AR	Autoregressive
ARIMA	Autoregressive integrated moving averages
ARIMAX	Autoregressive integrated moving averages with exogenous variables
AT	*À trous* algorithm
AW	Average prediction interval width
BANN	Bootstrap artificial neural networks
BC	Boundary condition
BIC	Bayesian Information Criterion
broCMI	Bootstrap rank-ordered conditional mutual information
BPF	Bayesian processor of forecasts
CMI	Conditional mutual information
CPP	Coverage Probability Plot
CPP_mse	Coverage Probability Plot mean square error
CRPS	Continuous ranked probability score
CWT	Continuous wavelet transform

DDFF	Data-Driven Forecasting Framework
DWT	Discrete wavelet transform
DWT-MRA	Discrete wavelet transform multiresolution analysis
EA	Edgeworth Approximations-based conditional mutual information
EA-ELM	Edgeworth Approximations extreme learning machine
EA-SOV	Edgeworth Approximations second order Volterra series model
EA-WELM	Edgeworth Approximations wavelet-based extreme learning machine
EA-WSOV	Edgeworth Approximations wavelet-based second order Volterra series model
EDI	Effective Drought Index
ELM	Extreme learning machine
EMD	Empirical mode decomposition
EW-DDFF	Ensemble Wavelet – Data-Driven Forecasting Framework
EW-SDDFF	Ensemble Wavelet – Stochastic Data-Driven Forecasting Framework
FFBP-ANN	Feedforward backpropagation artificial neural network
FT	Fourier transform
GCM	General circulation models
GLUE	Generalized Likelihood Uncertainty Estimation
GRNN	General regression neural networks
HKp	Hurst-Kolmogorov processes
IQR	Interquartile range
IS	Interval score
IVS	Input variable selection
IVS4EM	Input variable selection for environmental modeling

KDE	Kernel density estimation-based conditional mutual information
KNN	K nearest-neighbours-based conditional mutual information
LSSVR	Least-squares support vector regression
MAE	Mean absolute error
MAPE	Mean absolute percentage error
mat	Maximum air temperature
MI	Mutual information
MLR	Multiple linear regression
MODWT	Maximal overlap discrete wavelet transform
MODWT-MRA	Maximal overlap discrete wavelet transform multiresolution analysis
MRA	Multiresolution analysis
MSE	Mean square error
NASH	Nash-Sutcliffe Efficiency Index
NL	Nonlinear
NSERC	Natural Sciences and Engineering Research Council of Canada
PACF	Partial autocorrelation function
PCIS	Partial correlation input selection
PCIS-MLR	Partial correlation input selection multiple linear regression
PCIS-WMLR	Partial correlation input selection wavelet-based multiple linear regression
PDF	Probability density function
PI	Partial Information
PICP	Prediction interval coverage probability
PMI	Partial mutual information

PMIS	Partial mutual information input selection
rain	Rainfall
RMSE	Root mean squared error
RW	Random walk
SDDFF	Stochastic Data-Driven Forecasting Framework
SLFN	Single layer feed-forward network
SOV	Second order Volterra series model
SSA	Singular spectrum analysis
SVR	Support vector regression
SWDDFF	Stochastic Wavelet Data-Driven Forecasting Framework
TAR	Threshold autoregressive
TDMI	Time-delayed mutual information
UWD	Urban water demand
WDDFF	Wavelet Data-Driven Forecasting Framework
WELM	Wavelet-based extreme learning machine
WMLR	Wavelet-based multiple linear regression
WT	Wavelet transform
WT-ANN	Wavelet transform-based artificial neural network
WT-BANN	Wavelet transform-based bootstrap artificial neural network

Chapter 1: Introduction

Water resources forecasts are often used to support the design, operation, management, and planning of water resources systems (Loucks and van Beek, 2017). However, the nonlinear, multiscale, and uncertain nature of water resources provide challenges in the development of accurate and reliable forecasting models (Bogner and Pappenberger, 2011; House-Peters and Chang, 2011; Maier et al., 2010). This research attempts to address these three issues by developing, testing, and applying a novel ensemble multi-wavelet stochastic data-driven (e.g., time series, machine learning, artificial intelligence, etc. (Solomatine and Ostfeld, 2008)) forecasting framework. The method proposed in this research is named the Ensemble Wavelet – Data-Driven Forecasting Framework (EW-SDDFF).

First, a brief description of different water resources forecasting approaches (e.g., physically-based, conceptual, and data-driven) is given in order to provide background on available methods and to justify the use of a data-driven forecasting framework. Next, discussion is provided with reference to the different challenges that hinder the development of accurate and reliable water resources forecasts (i.e., nonlinearity, multiscale change, and uncertainty). To address each of these challenges a number of solutions are identified and then stated in terms of the main research objective and several specific (supporting) objectives. Afterwards, the contributions of this research are stated before an outline is given for the remainder of the thesis.

1.1. Water Resources Forecasting Approaches

In the water resources domain there are three main approaches to forecasting: physically (process)-based, conceptual, and data-driven methods (Devia et al., 2015; Remesan and Mathew, 2015). In simplified terms, the first approach relies on physical laws to relate observable processes to the target process and is based on the theory of mass-balance; conceptual models seek to mimic physical processes through various simplifications in order to transform external factors (predictors, explanatory variables, etc.) into the target process (predictand, response variable, etc.) (Jajarmizadeh et al., 2012). In contrast to both physically-based and conceptual approaches, data-driven models seek to establish statistical relationships between explanatory variables and the target process and do not require any physical knowledge of the processes under study (albeit, physical knowledge of the interactions between explanatory variables and the target process is essential in understanding the benefits and limitations of data-driven forecasts). This research is concerned with data-driven forecasting methods. Justification for the use of data-driven forecasting approaches is given in the sub-section below.

1.1.1. Why are Data-Driven Forecasting Methods Useful in Water Resources?

Different than physically-based or conceptual models, data-driven methods rely only on time series records that have been collected at regular (or even irregular (Sun and Trevor, 2017)) time intervals in order to estimate statistical relationships (in the form of model parameters) that can map a set of explanatory variables into forecasts of the target process. In general, depending on the complexity of a given physically-based or conceptual model, time series records of significant length for a large number of different explanatory variables are needed to forecast the target process at a high computational cost (Clark et al., 2017; Fatichi et al., 2016). In contrast, data-driven methods can be developed by solely considering previous time series records of the given target process (i.e., they do not require an extensive set of model inputs outside of historical measurements of the target process) while their parameters may be solved using simple and computationally inexpensive methods such as least squares. Thus, data-driven methods may be used as a flexible and inexpensive alternative to physically-based or conceptual models that require (potentially, site-specific) data that is difficult and/or expensive to collect (Zaier et al., 2010). A further benefit of data-driven models is that accurate forecasts can be obtained even with relatively short time series records that may be corrupted by noise or measurement errors and that contain complex nonlinear relationships (Mishra and Desai, 2006; Sun and Trevor, 2018; Zhang et al., 2011). Once a data-driven model has been calibrated, it is simple to use in real-time applications since it has a mathematical structure that does not require specialized software in order to generate forecasts. In addition to their minimal information requirements, flexibility, computational efficiency, strong performance in cases of poor data quality and availability, and ease of real-time use, data-driven models (such as artificial neural networks) have been shown to provide similar or better performance than their physically-based or conceptual counterparts for hydrological and water resources modeling and forecasting applications such as predicting ice growth on lakes (Seidou et al., 2006), estimating daily evaporation rates (Antonopoulos et al., 2016), rainfall-runoff forecasting (Daliakopoulos and Tsanis, 2016), streamflow forecasting (Demirel et al., 2009), and real-time flood forecasting (Napolitano et al., 2010). In summary, data-driven methods are ideal alternatives to physically-based and conceptual models for water resources forecasting due to their rapid development times, minimal information requirements, and ease of real-time implementation (Adamowski, 2008) alongside their ability to provide similar or better performance than their physically-based and conceptual counterparts (Banihabib, 2016). Due to these useful properties, data-driven forecasting methods have been adopted in this research. In the next sub-section, some challenges facing the development of accurate and reliable data-driven water resources forecasts are described and

solutions for overcoming these challenges are proposed before a more in-depth discussion of the objectives of this research.

1.2. Nonlinearity, Multiscale Change, and Uncertainty as Challenges in the Development of Accurate and Reliable Data-Driven Water Resources Forecasting Models

Notwithstanding the attractiveness of using data-driven methods for forecasting water resources, they are not without their challenges due to the nonlinear, multiscale, and uncertain nature of water resources. The nonlinear and multiscale nature of water resources causes difficulties in providing accurate data-driven forecasts since the intra- and inter-scale changes in explanatory and target processes often interact in non-trivial (nonlinear) ways (Adamowski et al., 2012; Peters-Lidard et al., 2017; Schwenk and Foufoula-Georgiou, 2017). The identification of which explanatory variables to choose and which scales of change are most important for including in a forecasting model of a given a target process (i.e., input variable selection), greatly influences the calibration of data-driven models and their forecasting accuracy (Galelli and Castelletti, 2013; Rathinasamy et al., 2013; Sang et al., 2015; Tran et al., 2015). Furthermore, uncertainty seems to be an inherent feature of any water resources process and might not be completely resolved even with better models, abundant data, or sophisticated uncertainty estimation methods (Montanari and Koutsoyiannis, 2012). In cases where one has access to numerous forecasting models, the uncertainty in deciding which forecasting model(s) (i.e., an ensemble forecast) to use for a given process, including the weight assigned to each model's forecast, is also an important factor to consider when evaluating water resources forecasting uncertainty (Herger et al., 2018; Sivillo et al., 1997; Weijs and van de Giesen, 2013). It is for these reasons that each stage in the design of data-driven forecasting models (e.g., input variable selection, parameters, model selection, model output, etc.) should be considered and explicitly estimated (Tyralis and Koutsoyiannis, 2017) if a high level of forecast accuracy and reliability is to be achieved. The EW-SDDFF embodies such considerations in several novel ways. It is shown below (as well as in Chapter 3-6) how each of these issues are innovatively addressed and considered in EW-SDDFF. A brief discussion follows concerning the different challenges and their solutions (as proposed in this research).

1.2.1. Challenge 1: Nonlinearity

The practice of adopting nonlinear data-driven methods is well-established in the water resources forecasting domain with over two decades of research on this topic (Abrahart et al., 2012; Remesan and Mathew, 2015). In order to address nonlinearity in water resources, nonlinear data-driven methods (e.g., artificial neural networks, support vector regression, etc.) have been adopted to provide more accurate

forecasts than linear methods (e.g., multiple linear regression, autoregressive integrated moving averages, etc.) (Li et al., 2010; Valipour et al., 2013; Zeynoddin et al., 2018). One main drawback to using nonlinear data-driven forecasting methods is the large amount of parameters and hyper-parameters that need to be carefully selected in order to calibrate accurate forecasting models that provide strong performance out-of-sample. The careful calibration of many parameters and identification of suitable hyper-parameter settings often lead to computationally demanding calibration times. Since many real-world water resources forecasting applications (e.g., water quality warning systems (Shi et al., 2018)) require accurate forecasts that can address nonlinearities between the target process and explanatory variables as well as issue forecasts in a timely manner, nonlinear data-driven methods that can generate accurate and computationally efficient forecasts should be sought (Deo and Şahin, 2015).

A problem intimately connected to the development of nonlinear data-driven forecasts but less often investigated, is the careful selection of input variables (Maier et al., 2010), a pre-requisite for the development of any accurate data-driven model (Guyon and Elisseeff, 2003). If one selects redundant or irrelevant input (explanatory) variables, then the performance of a given data-driven forecast will (potentially) be severely compromised (Šindelář and Babuška, 2004). While linear-based input variable selection methods are more common in water resources forecasting, nonlinear methods are seldom used, and those that are adopted are generally very computationally demanding (Galelli et al., 2014). Therefore, in order to provide nonlinear data-driven water resources forecasting models with appropriate input variables, nonlinear input variable selection methods are required - methods which can identify useful input variables in a computationally efficient manner are highly valuable.

1.2.1.1. Solution to Challenge 1

This research investigates nonlinear data-driven models and input variable selection methods that together produce accurate forecasts that are computationally efficient to generate and which effectively account for nonlinear and complex interactions between target and explanatory variables. The nonlinear data-driven methods estimate model parameters using a computationally efficient least squares solution. The first method is based on a new type of artificial neural network, named extreme learning machine (ELM), which is much faster to calibrate than traditional artificial neural networks and provides similar or better performance (Huang et al., 2006). The second method is based on a second order Volterra series model (SOV) which accounts for second order (nonlinear) interactions between explanatory variables and the target process through a Taylor series expansion (Labat et al., 1999). A useful quality of the SOV is

that its parameters can be interpreted similarly to multiple linear regression (Maheswaran and Khosa, 2012). Both methods are described in more detail in Chapters 2-4.

The nonlinear input variable selection methods explored in this research are based on information theory. Information theory provides a framework that permits the estimation of general (linear and nonlinear) forms of dependencies amongst two or more variables. In particular, this thesis proposes the first computationally efficient non-parametric information-theoretic input variable selection methods that act as alternatives to current parametric approaches (that are computationally demanding). The first method allows for the estimation of information-theoretic quantities (such as entropy and mutual information) by using Taylor series expansions. An extremely useful innovation is developed that extends the first method to account for the estimation of uncertainty in the input variable selection process, creating another new input variable selection method. The two methods are named Edgeworth Approximations-based conditional mutual information (EA) and its variant that includes uncertainty assessment, bootstrap rank-ordered conditional mutual information (broCMI) (see Chapter 3 for more details). The nonlinear data-driven methods and input variable selection approaches mentioned above are considered in EW-SDDFF for generating forecasting models and selection forecasting model inputs (i.e., input variable selection), respectively.

1.2.2. Challenges 2: Multiscale Change

To improve forecast accuracy when using data-driven methods, including information on multiscale changes in water resources is essential (Koutsoyiannis et al., 2010). Since the early 2000's, researchers have made use of the wavelet transform for uncovering transients, trends, periodicities, and other peculiar phenomena across multiple scales in water resources processes and have used these physically meaningful traits as input to data-driven forecasting models, realizing superior accuracy when compared to models that did not incorporate the same information (Fahimi et al., 2017; Nourani et al., 2014). However, since the literature lacks a set of best (correct) practices for wavelet-based forecasting, a widely overlooked aspect of the wavelet transform has led to the use of data that exists ahead of the forecast date (referred to as, 'future data') to be used as input to the forecast, resulting in a large number of incorrectly-developed wavelet-based data-driven forecasting models (Du et al., 2017) (please see Chapter 4 for an in-depth treatment).

1.2.2.1. Solution to Challenge 2

To overcome this deficiency, this research develops a set of best (correct) practices for wavelet-based forecasting. The best practices ensure that the wavelet transform is applied correctly to the model inputs,

primarily by addressing the 'future data' issue. By adopting this set of best practices, one can use the wavelet transform to correctly develop wavelet-based data-driven forecasting models. The best practices are incorporated in the new Wavelet Data-Driven Forecasting Framework (WDDFF) (see Chapter 4) by using the wavelet transform to (correctly) extract multiscale information from the target and explanatory variables. Input variable selection is then used to select scale-based information from the explanatory variables, which is used as input to a particular data-driven model to forecast the target process. WDDFF is flexible and can be used with any (linear or nonlinear) input variable selection and data-driven method. Thus, WDDFF is a very useful tool that can be used for forecasting nonlinear and multiscale water resources. WDDFF is used as the base data-driven method for EW-SDDFF, which allows different wavelet-based forecasting models to be combined in a multi-wavelet ensemble stochastic data-driven forecasting framework.

1.2.3. Challenge 3: Uncertainty Assessment

Estimation of uncertainty is an important task often required in optimizing, managing, and planning water resources systems (Ajami et al., 2008; Wei, 2012). In order to estimate uncertainty in each stage of the forecast design and to derive a probability-based forecast as the end result (a very useful tool for operational, managerial, and decision-making tasks in water resources systems), one is faced with numerous options of varying complexity: Bayesian methods (Krzysztofowicz, 2014), evolutionary approaches (Dotto et al., 2012), Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 2014), and stochastics (Farmer and Vogel, 2016). Each uncertainty estimation method uses different means for characterizing uncertainty and most methods often require tuning several parameters, assuming distributional properties, or estimating a likelihood function, e.g., particle swarm optimization (Zambrano-Bigiarini and Rojas, 2013), Bayesian methods (Herr and Krzysztofowicz, 2015), and GLUE (Khoi and Thom, 2015), respectively.

1.2.3.1. Solution to Challenge 3

This research adopts a stochastics-based approach for uncertainty estimation referred to as the blueprint (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). This stochastics-based approach relies on the use of probability density functions to characterize uncertainty in different facets of the model design (e.g., input variable selection, model parameters, model output, etc.) and is different from other prominent methods. For example, it does not require the estimation of a likelihood function and can therefore estimate probability density functions using simple methods, such as the bootstrap (Efron and Tibshirani, 1993). By using a stochastic approach based on the bootstrap, uncertainty can be estimated

without any distributional assumptions, parametric settings, or likelihood estimation. The bootstrap simply relies on sampling with replacement from a given dataset. This allows the estimation of input data, input variable selection, parameter, and model output uncertainties. By considering these uncertainties, WDDFF is transformed into its stochastic version (SWDDFF) and its non-wavelet-based version, SDDFF (i.e., by not performing wavelet analysis). Both SDDFF and SWDDFF characterize the uncertainty in these respective aspects of the data-driven forecasting model design (i.e., input data, input variable selection, parameters, and model output) and can therefore be used for generating probability-based forecasts in a holistic manner. SWDDFF (and SDDFF) are the first wavelet- (and non-wavelet) based stochastic data-driven forecasting models to include input data, input variable selection, parameter, and model output uncertainty.

By feeding individual WDDFF forecasts into SDDFF, the proposed EW-SDDFF is generated. EW-SDDFF is an ensemble version of the SWDDFF and the first ensemble multi-wavelet stochastic forecasting framework for probability-based forecasting and the first ensemble stochastic data-driven water resources forecasting model to consider ensemble model selection and weighting uncertainties, alongside input data and ensemble model output uncertainties. The EW-SDDFF is the most advanced ensemble multi-wavelet forecasting framework proposed in the literature.

1.3. Research Objectives

Due to the complex nature of water resources that exhibit nonlinearity, multiscale change, and uncertainty, one wishing to adopt a data-driven forecasting framework must seek a paradigm that can address *each* of these features in all facets of the model design in order to enable accurate and reliable water resources forecasts. While there have been numerous studies in the data-driven water resources forecasting literature that have sought to address these issues (which will be discussed in more detail within Chapter 2), most proposed frameworks choose to address *only* one of these issues (e.g., nonlinearity (Laio et al., 2003; Wu et al., 2014)), others will focus on addressing two of these issues (e.g., nonlinearity and multiscale changes (Patil and Deka, 2017; Yadav and Eliza, 2017)), and very seldom, some will attempt to address each issue (Barzegar et al., 2017; Kasiviswanathan et al., 2016; Liu et al., 2015; Wang et al., 2013) but fall short of including uncertainty assessment in each stage of the data-driven forecast design (e.g., input variable selection, estimating uncertainties in model output, etc.).

Many of the most advanced data-driven water resources forecasting methods that seek to account for nonlinearity, multiscale change, and uncertainty assessment are based on the wavelet transform (Nourani et al., 2014). However, the vast majority of wavelet-based data-driven forecasting models that consider

uncertainty assessment use the wavelet transform incorrectly (e.g., Bachour et al. (2016); Kasiviswanathan et al. (2016); Sang et al. (2013); Tiwari and Chatterjee (2010)), which results in forecasting models that cannot be used properly for real-world applications.

It is clear that the literature is lacking a holistic data-driven forecasting framework that can properly address nonlinearity, multiscale change, and uncertainty assessment in the water resources domain. Therefore, the main objective of this research is to develop, test, and apply an ensemble multi-wavelet stochastic data-driven framework (EW-SDDFF) for real-world water resources forecasting applications. The purpose behind the development of EW-SDDFF is to address the short-comings of the many current water resources forecasting models that do not properly or holistically account for the nonlinear, multiscale, and uncertain nature of water resources. The ultimate goal of this research is to demonstrate that EW-SDDFF can be used to generate accurate and reliable (probability-based) forecasts of nonlinear, multiscale, and uncertain water resources. In order to meet this goal, the main objective of this research is split into several supporting objectives that explore specific components of EW-SDDFF. The specific (supporting) objectives of this research are to:

1. Develop, test, and apply two new nonlinear input variable selection methods (EA and broCMI) for synthetic and partially-synthetic input variable selection problems by comparing them against benchmark methods in terms of input variable selection accuracy, then couple the new nonlinear input variable selection methods and their benchmarks with nonlinear data-driven models and forecast real-world nonlinear water resources processes and assess the different methods in terms of forecast accuracy (Objective 1).
2. Develop a set of best (correct) practices for wavelet-based data-driven forecasting, then coalesce the best practices into a new wavelet-based data-driven forecasting framework (WDDFF) and test and apply the WDDFF for forecasting a real-world (nonlinear and) multiscale water resources process by comparing it against benchmark methods in terms of forecast accuracy (Objective 2).
3. Develop, test, and apply the new stochastic WDDFF, and its non-wavelet-based counterpart (SDDFF) on a real-world (nonlinear) and multiscale water resources process by comparing it against benchmark methods (that consider varying sources of uncertainty; e.g., input variable selection, parameter, and/or model output) (Objective 3) in terms of forecast accuracy and reliability.
4. Develop, test, and apply the new ensemble multi-wavelet version of SWDDFF (EW-SDDFF) on a real-world (nonlinear) and multiscale water resources process by comparing it against benchmark

methods (i.e., SWDDFF, WDDFF, and its non-stochastic version, EW-DDFF) in terms of forecast accuracy and reliability (Objective 4).

While the first objective is concerned with addressing nonlinearity in water resources forecasting, the second objective builds from the first to address (nonlinearity) and multiscale change. The third objective builds from the first and second objectives and accounts for (nonlinearity, multiscale change, and) uncertainty assessment in water resources forecasting. The fourth objective extends the new method developed in the third objective by considering an ensemble of multiple models in order to improve forecast accuracy and reliability. The goal of each specific objective is to show how forecast accuracy and/or reliability can be increased through the progressive developments of each component included in the EW-SDDFF.

The main study area considered in this research is Montreal, Quebec, where the focus is on forecasting average daily demand for the City's urban water supply system which supplies drinking water to over a million residents (Tiwari and Adamowski, 2013). The city experiences peak demands in the summer seasons primarily due to outdoor water use related to warmer weather and periods of low rainfall. Further information on this study site can be found in Chapter 4 and is used in the case studies contained in Chapters 4-6. In Chapter 3, several synthetic and partially-synthetic datasets (Galelli et al., 2014) are used for comparing the proposed EA and broCMI input variable selection methods against existing methods. In the same chapter, two real-world daily urban water demand time series from Ottawa, Ontario are also used for assessing nonlinear data-driven forecasting model performance based on explanatory variables selected by the proposed (EA and broCMI) methods and their competitors – detailed information on these datasets are available in Chapter 3. Therefore, Chapter 3 is the only chapter that uses different datasets than the remaining chapters. The next sub-section highlights the contributions of this research before giving an outline of for the remainder of this thesis.

1.4. Contributions

The research contained in this thesis is innovative in four main ways:

1. It develops, tests, and applies the first non-parametric and computationally efficient nonlinear input variable selection method based on information theory (i.e., the EA method) and advances upon this method by including uncertainty in the input variable selection procedure (i.e., the broCMI method).

2. It develops, tests, and applies the first set of best (correct) practices for wavelet-based data-driven forecasting for real-world applications (to address the fact that many current wavelet-based forecasting methods are incorrect and cannot be used for real-world applications); these best practices are adopted within a new wavelet-based data-driven forecasting framework (i.e., WDDFF) that can be used with any input variable selection method and data-driven model.
3. It develops, tests, and applies the first wavelet- (and non-wavelet-) based stochastic data-driven forecasting framework that considers input data, input variable selection, parameter, and model output uncertainty (i.e., SWDDFF and SDDFF, respectively).
4. It develops, tests, and applies the first ensemble multi-wavelet stochastic forecasting framework for probability-based forecasting and the first ensemble stochastic data-driven water resources forecasting model to consider ensemble model selection and weighting uncertainties, alongside input data and ensemble model output uncertainties (i.e., EW-SDDFF).

1.5. Thesis Outline

The objectives and contributions listed in the last two sub-sections are contained within this thesis and expounded as a series of manuscripts that flow from one objective to the next.

Chapter 2 reviews literature related to data-driven forecasting methods, input variable selection, multiscale time series approaches, wavelet-based forecasting, and uncertainty assessment methods in the water resources domain. The literature review is followed by four integrated manuscripts.

The first manuscript (Chapter 3) reviews input variable selection methods in general, introduces and develops the proposed EA and broCMI methods, provides their theoretical underpinnings, and then proceeds to test and apply the methods on synthetic, partially-synthetic, and real-world input variable selection problems by comparing them against benchmark methods.

The second manuscript (Chapter 4) provides an in-depth review on the use of wavelet-based data-driven forecasting for real-world applications and exposes three main issues that have led to the incorrect development of wavelet-based forecasts. A set of best (correct) practices are proposed (to address these three issues) and adopted in a new wavelet-based data-driven forecasting framework (WDDFF) that uses input variable selection methods from Chapter 3 (e.g., EA) and nonlinear data-driven methods to forecast a real-world water resources process.

The third manuscript (Chapter 5) explores the use of a blueprint for converting a deterministic forecast into a stochastic one and adapts this framework to be used with WDDFF (from Chapter 4) by considering

uncertainty in input data, input variable selection, parameter, and model output, resulting in the SWDDFF (and its non-wavelet-based counterpart SDDFF). SWDDFF is compared against benchmark methods on a real-world water resources forecasting problem.

The fourth manuscript (Chapter 6) discusses how to develop EW-SDDFF, an ensemble multi-wavelet version of SWDDFF (from Chapter 5) and demonstrates for the same study area in Chapter 4 and 5 that it provides substantially better performance than SWDDFF and other benchmark methods.

Chapter 7 discusses the most significant results of this research and provides concluding remarks.

Chapter 8 recounts the primary contributions of this research to the literature, discusses some limitations of the work covered in this thesis, and indicates several avenues for future research.

1.6. References

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L., 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog. Phys. Geogr. 36, 480–513. doi:10.1177/0309133312444943

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Adamowski, J.F., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. J. Hydrol. 353, 247–266. doi:10.1016/j.jhydrol.2008.02.013

Ajami, N.K., Hornberger, G.M., Sunding, D.L., 2008. Sustainable water resource management under hydrological uncertainty. Water Resour. Res. 44, 1–10. doi:10.1029/2007WR006736

Antonopoulos, V.Z., Gianniou, S.K., Athanasios, V., 2016. Artificial neural networks and empirical equations to estimate daily evaporation : application to Lake. Hydrol. Sci. J. 61, 2590–2599. doi:10.1080/02626667.2016.1142667

Bachour, R., Maslova, I., Ticlavilca, A.M., Walker, W.R., McKee, M., 2016. Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. Stoch. Environ. Res. Risk

Assess. 30, 103–117. doi:10.1007/s00477-015-1039-z

Banihabib, M.E., 2016. Performance of conceptual and black-box models in flood warning systems. Cogent Eng. 3. doi:10.1080/23311916.2015.1127798

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189

Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrol. Process. 28, 5897–5918. doi:10.1002/hyp.10082

Bogner, K., Pappenberger, F., 2011. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. Water Resour. Res. doi:10.1029/2010WR009137

Clark, M.P., Bierkens, M.F.P., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K.E., Pauwels, V.R.N., Cai, X., Wood, A.W., Peters-Lidard, C.D., 2017. The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. Hydrol. Earth Syst. Sci. 21, 3427–3440. doi:10.5194/hess-21-3427-2017

Daliakopoulos, I.N., Tsanis, I.K., 2016. Comparison of an artificial neural network and a conceptual rainfall – runoff model in the simulation of ephemeral streamflow. Hydrol. Sci. J. 61, 2763–2774. doi:10.1080/02626667.2016.1154151

Demirel, M.C., Venancio, A., Kahya, E., 2009. Flow forecast by SWAT model and ANN in Pracana basin , Portugal. Adv. Eng. Softw. 40, 467–473. doi:10.1016/j.advengsoft.2008.08.002

Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmos. Res. 153, 512–525. doi:10.1016/j.atmosres.2014.10.016

Devia, G.K., Ganasri, B.P., Dwarakish, G.S., 2015. A Review on Hydrological Models. Aquat. Procedia 4, 1001–1007. doi:10.1016/j.aqpro.2015.02.126

Dotto, C.B.S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D.T., Freni, G., Rauch, W., Deletic, A., 2012. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Res. 46, 2545–2558. doi:10.1016/j.watres.2012.02.009

Du, K., Zhao, Y., Lei, J., 2017. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. J. Hydrol. 552, 44–51.

doi:10.1016/j.jhydrol.2017.06.019

Efron, B., Tibshirani, R.J.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, London, U.K.

Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor. Appl. Climatol. 128, 875–903. doi:10.1007/s00704-016-1735-8

Farmer, W.H., Vogel, R.M., 2016. On the deterministic and stochastic use of hydrologic models. Water Resour. Res. 52, 5619–5633. doi:10.1002/2016WR019129

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., Tarboton, D., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J. Hydrol. 537, 45–60. doi:10.1016/j.jhydrol.2016.03.026

Galelli, S., Castelletti, A., 2013. Tree-based iterative input variable selection for hydrological modeling. Water Resour. Res. 49, 4295–4310. doi:10.1002/wrcr.20339

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. doi:10.1016/j.envsoft.2014.08.015

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Herger, N., Abramowitz, G., Knutti, R., Angélil, O., Lehmann, K., Sanderson, B.M., 2018. Selecting a climate model subset to optimise key ensemble properties. Earth Syst. Dyn. 9, 135–151. doi:10.5194/esd-9-135-2018

Herr, H.D., Krzysztofowicz, R., 2015. Ensemble Bayesian forecasting system Part I: Theory and algorithms. J. Hydrol. 524, 789–802. doi:10.1016/j.jhydrol.2014.11.072

House-Peters, L.A., Chang, H., 2011. Urban water demand modeling: Review of concepts, methods, and organizing principles. Water Resour. Res. 47, W05401. doi:10.1029/2010WR009624

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. doi:10.1016/j.neucom.2005.12.126

Jajarmizadeh, M., Harun, S., Salarpour, M., Jajarmizad, M., Harun, S., Salarpour, M., 2012. A Review of Theoretical Consideration and Types of Models In Hydrology. J. Environ. Sci. Technol. doi:10.39231jest.2012.249.261

Kasiviswanathan, K.S., He, J., Sudheer, K.P., Tay, J.-H., 2016. Potential application of wavelet neural network ensemble to forecast streamflow for flood management. J. Hydrol. 536, 161–173. doi:10.1016/j.jhydrol.2016.02.044

Khoi, D.N., Thom, V.T., 2015. Parameter uncertainty analysis for simulating streamflow in a river catchment of Vietnam. Glob. Ecol. Conserv. 4, 538–548. doi:10.1016/j.gecco.2015.10.007

Koutsoyiannis, D., Yao, H., Georgakakos, A., 2010. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrol. Sci. J. 53, 142–164. doi:10.1623/hysj.53.1.142

Krzysztofowicz, R., 2014. Probabilistic flood forecast: Exact and approximate predictive distributions. J. Hydrol. 517, 643–651. doi:10.1016/j.jhydrol.2014.04.050

Labat, D., Ababou, R., Mangin, A., 1999. Linear and nonlinear input/output models for karstic springflow and flood prediction at different time scales. Stoch. Environ. Res. Risk Assess. 13, 337–364. doi:10.1007/s004770050055

Laio, F., Porporato, A., Revelli, R., Ridolfi, L., 2003. A comparison of nonlinear flood forecasting methods. Water Resour. Res. 39, 1–4. doi:10.1029/2002WR001551

Li, P.H., Kwon, H.H., Sun, L., Lall, U., Kao, J.J., 2010. A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan. Int. J. Climatol. 30, 1256–1268. doi:10.1002/joc.1954

Liu, Z., Zhou, P., Zhang, Y., 2015. A Probabilistic Wavelet–Support Vector Regression Model for Streamflow Forecasting with Rainfall and Climate Information Input. J. Hydrometeorol. 16, 2209–2229. doi:10.1175/JHM-D-14-0210.1

Loucks, D.P., van Beek, E., 2017. Water Resource Systems Modeling: Its Role in Planning and Management BT - Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications, in: Loucks, D.P., van Beek, E. (Eds.), . Springer International Publishing, Cham, pp. 51–72. doi:10.1007/978-3-319-44234-1_2

Maheswaran, R., Khosa, R., 2012. Wavelet-Volterra coupled model for monthly stream flow forecasting. J. Hydrol. 450–451, 320–335. doi:10.1016/j.jhydrol.2012.04.017

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ. Model. Softw. 25, 891–909. doi:10.1016/j.envsoft.2010.02.003

Mishra, A.K., Desai, V.R., 2006. Drought forecasting using feed-forward recursive neural network. Ecol. Modell. 198, 127–138. doi:10.1016/j.ecolmodel.2006.04.017

Montanari, A., Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. Water Resour. Res. 48. doi:10.1029/2011WR011412

Napolitano, G., See, L., Calvo, B., Savi, F., Heppenstall, A., 2010. A conceptual and neural network model for real-time flood forecasting of the Tiber River in Rome. Phys. Chem. Earth 35, 187–194. doi:10.1016/j.pce.2009.12.004

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. J. Hydrol. 514, 358–377. doi:10.1016/j.jhydrol.2014.03.057

Patil, A.P., Deka, P.C., 2017. Performance evaluation of hybrid Wavelet-ANN and Wavelet-ANFIS models for estimating evapotranspiration in arid regions of India. Neural Comput. Appl. 28, 275–285. doi:10.1007/s00521-015-2055-0

Peters-Lidard, C.D., Clark, M., Samaniego, L., Verhoest, N.E.C., Van Emmerik, T., Uijlenhoet, R., Achieng, K., Franz, T.E., Woods, R., 2017. Scaling, similarity, and the fourth paradigm for hydrology. Hydrol. Earth Syst. Sci. 21, 3701–3713. doi:10.5194/hess-21-3701-2017

Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J. Hydrol. doi:10.1016/j.jhydrol.2013.09.025

Remesan, R., Mathew, J., 2015. Hydrological data driven modelling: A case study approach, Hydrological Data Driven Modelling: A Case Study Approach. Springer International Publishing.

Sang, Y.-F., Wang, Z., Liu, C., 2015. Wavelet Neural Modeling for Hydrologic Time Series Forecasting with Uncertainty Evaluation. Water Resour. Manag. 29, 1789–1801. doi:10.1007/s11269-014-0911-9

Sang, Y.F., Shang, L.Y., Wang, Z.G., Liu, C.M., Yang, M.G., 2013. Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting. Chinese Sci. Bull. 58, 3796–3805. doi:10.1007/s11434-013-5903-2

Schwenk, J., Foufoula-Georgiou, E., 2017. Are process nonlinearities encoded in meandering river planform morphology? J. Geophys. Res. Earth Surf. 122, 1534–1552. doi:10.1002/2016JF003929

Seidou, O., Ouarda, T.B.M.J., Bilodeau, L., Hessami, M., Bruneau, P., 2006. Modeling ice growth on Canadian lakes using artificial neural networks. Water Resour. Res. 42, 1–15. doi:10.1029/2005WR004622

Shi, B., Wang, P., Jiang, J., Liu, R., 2018. Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. Sci. Total Environ. 610–611, 1390–1399. doi:10.1016/j.scitotenv.2017.08.232

Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2015. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. J. Hydrol. Eng. 20.

Šindelář, R., Babuška, R., 2004. Input selection for nonlinear regression models. IEEE Trans. Fuzzy Syst. 12, 688–696. doi:10.1109/TFUZZ.2004.834810

Sivillo, J.K., Ahlquist, J.E., Toth, Z., 1997. An Ensemble Forecasting Primer. Weather Forecast. 12, 809–818. doi:10.1175/1520-0434(1997)012<0809:AEFP>2.0.CO;2

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinformatics 10, 3–22. doi:10.2166/hydro.2008.015

Sun, W., Trevor, B., 2018. Multiple model combination methods for annual maximum water level prediction during river ice breakup. Hydrol. Process. 32, 421–435. doi:10.1002/hyp.11429

Sun, W., Trevor, B., 2017. Combining k-nearest-neighbor models for annual peak breakup flow forecasting. Cold Reg. Sci. Technol. 143, 59–69. doi:10.1016/j.coldregions.2017.08.009

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. J. Hydrol. 394, 458–470.

doi:10.1016/j.jhydrol.2010.10.001

Tran, H.D., Muttil, N., Perera, B.J.C., 2015. Selection of significant input variables for time series forecasting. Environ. Model. Softw. 64, 156–163. doi:10.1016/j.envsoft.2014.11.018

Tyralis, H., Koutsoyiannis, D., 2017. On the prediction of persistent processes using the output of deterministic models. Hydrol. Sci. J. 62, 2083–2102. doi:10.1080/02626667.2017.1361535

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J. Hydrol. 476, 433–441. doi:10.1016/j.jhydrol.2012.11.017

Wang, Y., Zheng, T., Zhao, Y., Jiang, J., Wang, Y., Guo, L., Wang, P., 2013. Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. Environ. Sci. Pollut. Res. 20, 8909–8923. doi:10.1007/s11356-013-1874-8

Wei, S.-K., 2012. Estimating water deficit and its uncertainties in water-scarce area using integrated modeling approach. Water Sci. Eng. 5, 450–463. doi:10.1007/s11707- 009-0047-z

Weijs, S. V., van de Giesen, N., 2013. An information-theoretical perspective on weighted ensemble forecasts. J. Hydrol. 498, 177–190. doi:10.1016/j.jhydrol.2013.06.033

Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. Environ. Model. Softw. 54, 108–127. doi:10.1016/j.envsoft.2013.12.016

Yadav, B., Eliza, K., 2017. A hybrid wavelet-support vector machine model for prediction of Lake water level fluctuations using hydro-meteorological data. Meas. J. Int. Meas. Confed. 103, 294–301. doi:10.1016/j.measurement.2017.03.003

Zaier, I., Shu, C., Ouarda, T.B.M.J., Seidou, O., Chebana, F., 2010. Estimation of ice thickness on lakes using artificial neural network ensembles. J. Hydrol. 383, 330–340. doi:10.1016/j.jhydrol.2010.01.006

Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent Particle Swarm Optimisation software for model calibration. Environ. Model. Softw. 43, 5–25. doi:10.1016/j.envsoft.2013.01.004

Zeynoddin, M., Bonakdari, H., Azari, A., Ebtehaj, I., Gharabaghi, B., Riahi Madavar, H., 2018. Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate. J. Environ. Manage. 222, 190–206. doi:10.1016/j.jenvman.2018.05.072

Zhang, Q., Wang, B.-D., He, B., Peng, Y., Ren, M.-L., 2011. Singular Spectrum Analysis and ARIMA Hybrid Model for Annual Runoff Forecasting. Water Resour. Manag. 25, 2683–2703. doi:10.1007/s11269-011-9833-y

Chapter 2: Literature Review

The main focus of this research is to develop, test, and apply a new ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) and to demonstrate how it can be used for probability-based forecasting of any (nonlinear, multiscale, and uncertain) water resources process. The EW-SDDFF is very useful for operational, management, and decision-making tasks routinely faced by water resources managers. The EW-SDDFF contains three main components:

1. Addressing nonlinearity via nonlinear information-theoretic-based input variable selection methods (EA and broCMI) that identify which explanatory variables to use as input to a nonlinear data-driven model for forecasting the target water resources process.
2. Addressing multiple scales of change via a set of best (correct) practices for wavelet-based forecasts that are used to capture and extract multiscale properties of model inputs that were selected by the input variable selection methods (step 1) and that are subsequently fed to the wavelet-based data-driven framework (WDDFF) for forecasting the target process.
3. Addressing uncertainty via a stochastic framework used for quantifying uncertainty in the framework developed in step 2. This is accomplished in two main ways through:
    a. A single wavelet-based forecasting model (SWDDFF) that accounts for uncertainty in input data, input variable selection, parameter, and model output uncertainty from step 2.
    b. An ensemble multi-wavelet stochastic framework (EW-SDDFF) used for combining multiple wavelet-based forecasting models from step 2 in order to provide a computationally efficient version of step 3a (which only considers a single wavelet-based forecasting model). The EW-SDDFF includes ensemble model selection and weighting uncertainties, alongside input data and ensemble model output uncertainties.

The three steps are explored in this order within the literature review, which has been divided into three main sections that covers each of these components. The first sub-section covers the topic of addressing nonlinearity by giving a brief review of different data-driven forecasting methods before a more in-depth treatment of input variable selection methods that includes the new novel approaches developed in this research. To address multiscale change, the second sub-section gives a high-level overview of different methods for analysing multiscale change in water resources before honing in on wavelet-based methods, with a key focus on how wavelets have been incorrectly used for wavelet-based water resources forecasting and how this is resolved by the proposed set of best practices and their use in a new wavelet-based data-driven forecasting framework (WDDFF). The third sub-section focuses on uncertainty

assessment in water resources forecasting applications. This sub-section discusses how uncertainty can be estimated via stochastics and how uncertainty assessment can be incorporated in WDDFF, leading to two new stochastic wavelet-based forecasting frameworks (SWDDFF and its ensemble multi-wavelet version, EW-SDDFF) for the probability-based forecasting of water resources.

2.1. Nonlinearity – Data-Driven Forecasting and Input Variable Selection Methods

To generate accurate water resources forecasts it is often necessary to use nonlinear methods that can model such behaviour. Both linear and nonlinear data-driven approaches for water resources forecasting are briefly discussed before a more detailed analysis of (linear and nonlinear) input variable selection methods. This sub-section is split in two to cover both topics. While data-driven methods are a central component of the proposed EW-SDDFF, it is important to stress than any data-driven model may be used within EW-SDDFF. However, since the novelty of this research is tied to input variable selection, (the correct development of) wavelet-based forecasting models, and uncertainty assessment, higher importance is placed upon these topics in the literature review that ensues.

2.1.1. Different Data-Driven Methods

There are a wide range of data-driven methods used in water resources forecasting applications (Solomatine and Ostfeld, 2008): time series methods such as autoregressive integrated moving averages (ARIMA) (Mohammadi et al., 2006; Valipour et al., 2013) and Hurst-Kolmogorov processes (HKp) (Koutsoyiannis et al., 2010; Tyralis and Koutsoyiannis, 2017), multiple linear and nonlinear regression (Adamowski et al., 2012; Ghanbarpour et al., 2009), artificial neural networks (ANN) (Bougadis et al., 2005; Maier et al., 2010), support vector regression (SVR) (Baydaroğlu and Koçak, 2014; Matos et al., 2018), genetic programming (Danandeh et al., 2013; Elshorbagy et al., 2010), fuzzy logic (Goyal et al., 2014; Marvuglia et al., 2014), multivariate adaptive regression splines (Deo et al., 2017a; Kisi, 2016), Volterra series models (Hasanpour Kashani et al., 2014; Maheswaran and Khosa, 2012a), regression trees (Kisi, 2016; Rahimikhoob, 2016), non-parametric methods, such as nearest neighbours (Lee et al., 2017; Sharma et al., 2016), response surface method (Keshtegar et al., 2017, 2016), deep learning approaches (Li et al., 2016; Wu et al., 2015), and many others.

2.1.2. Popular Nonlinear Data-Driven Methods in Water Resources Forecasting

In recent years, the most popular of these data-driven methods in water resources forecasting are nonlinear methods that bear the traits of universal approximation such as ANN and SVR (Abrahart et al., 2012), see Afan et al. (2016); Remesan and Mathew (2015); Yaseen et al. (2015) for numerous applications

of the methods mentioned above. The benefit of using universal approximation techniques (such as ANN) for water resources forecasting, is that these methods have the ability to represent any (nonlinear) function to any arbitrary degree of accuracy, given enough parameters (Ince, 2006), which avoids the strict parametric forms afforded by traditional methods (such as multiple linear regression (MLR) or ARIMA) and addresses the inherent nonlinearity of the target process through a flexible mathematical expression (Maier et al., 2010). It is worth clarifying that a model parameter is a scalar value that functionally maps the explanatory variables of a given model to the output (within a given error tolerance). Hyper-parameters exert influence over how the model parameters are selected. To make this idea more explicit, imagine an autoregressive model that explains a certain water resources process: the parameters of the model would be the autoregressive coefficient(s) and the bias term (if any). An example of a hyper-parameter of this model would be the calibration algorithm used to select the (bias, if any, and the) autoregressive coefficient(s). If a particular calibration algorithm has settings that need to be adjusted to determine an optimal solution (i.e. the smallest prediction error), then these settings are also hyper-parameters of the model. Considering the ANN paradigm: parameters include the weights and biases, while hyper-parameters include the number of hidden layers, type of hidden neurons, learning rates, number of training epochs, etc.

2.1.2.1. Drawbacks of Popular Nonlinear Data-Driven Methods in Water Resources Forecasting

One of the largest drawbacks to using nonlinear data-driven methods such as ANN, SVR, and others is the large amount of parameters to tune and hyper-parameters that require careful selection if the resulting model is to provide accurate predictions of the target process (Luo, 2016). In short, these data-driven methods can be computationally intensive to calibrate and require expert intervention, hampering their development and operational usage. Another potential issue with data-driven methods (such as ANN) is that these methods are “black-boxes”, meaning that it is not straightforward to understand why/how a set of optimized parameters lead to a given model output (Mount et al., 2013), hampering the ability to associate meaning between data-driven model output(s) and the physical process under study, except in the case of very simple models (Giustolisi and Savic, 2006; Koutsoyiannis et al., 2010; Solomatine and Dulal, 2003).

To overcome these two main drawbacks of popular nonlinear data-driven methods used in many water resources forecasting studies, two different methods are explored in this research that are easy to use (fewer hyper-parameter settings), are computationally efficient to calibrate, and one of which has interpretational qualities similar to MLR.

2.1.2.2. Overcoming Drawbacks of Popular Nonlinear Data-Driven Methods in Water Resources Forecasting with Easy to Use and Computationally Efficient Methods

2.1.2.2.1. Extreme Learning Machine

The first method is the extreme learning machine (ELM). ELM was recently introduced to the literature (Huang et al., 2006) as a computationally efficient ANN variant that possesses two attractive qualities different from the traditional ANN: a randomized component setting input weights and biases to values drawn from a given distribution (e.g. normal, uniform, triangular, etc.) and the output weights are determined directly using linear systems of equations (Huang et al., 2012). These two strengths of the ELM have significantly improved the computational efficiency and generalization qualities when compared to the ANN and SVR methods (Huang et al., 2012), mainly because there is no requirement for a parameter optimization routine since the input weights are randomly generated and the output weights form a least-squares solution that can be solved directly. The only hyper-parameters of the ELM are the number of hidden nodes and their activation function type. This advantage possessed by ELM makes it more suitable for this research as it overcomes the need to tune model hyper-parameters in a careful manner, greatly reducing computational efforts and as will be demonstrated below, ELM has exhibited better performance than ANN and SVR methods in forecasting different water resources processes. As mentioned earlier, the ELM framework is very new to the water resources domain with very few applications. The earliest water resources study compared the performance of Echo State Networks, ELM, and ANN for seasonal stream flow prediction (Siqueira et al., 2012) and found both the Echo State Networks and ELM frameworks to provide good predictive ability and noted the significant improvement in simplicity for the ELM model construction and the superior performance when compared to ANN. The second study compared forecasts of fresh algae blooms in drinking water storage reservoirs at multiple lead times using SVR and ELM (Lou et al., 2016). The authors noted the superiority of the ELM forecasts to that of the SVR and recommend the ELM as a useful forecasting method. Deo and Şahin (2015) used ELM to forecast the Effective Drought Index (EDI) over eastern Australia and found ELM to outperform ANN. The authors obtained more computationally efficient forecasts with ELM over ANN and declared ELM as a promising new approach for determining the onset of future drought events. More recent studies have confirmed the high level of accuracy and computational efficiency of ELM when compared against traditional data-driven methods in water resources forecasting (Alizamir et al., 2018; Barzegar et al., 2018b; Dou and Yang, 2018; Heddam and Kisi, 2017; Lima et al., 2016; Rezaie-Balf and Kisi, 2017; Yaseen et al., 2016). Therefore, ELM is a very new and accurate forecasting method that overcomes the issues related to hyper-parameter selection (since it contains only two) and provides a simple model that

avoids computationally demanding calibration of model parameters as in data-driven methods such as ANN, SVR, fuzzy logic, etc. and is not restricted to assumptions of linearity as in ARIMA, MLR, and HKp.

2.1.2.2.2. Volterra Series Model

The second data-driven method that is considered in this research is the Volterra series model and specifically the second order Volterra series model (SOV) (Maheswaran and Khosa, 2014; Wu and Kareem, 2014). Essentially, the Volterra series model is a multiple-input single-output system of equations that is constructed as a Taylor series expansion and can therefore model any system to an arbitrary degree of accuracy, which depends on the order of the Taylor series expansion adopted for the problem (Labat et al., 1999). A first-order Volterra series model reduces to MLR.

In hydrology and water resources, a second order expansion (i.e., a second order Volterra series model) is most often suitable for the target processes studied in this domain (Diskin et al., 1984; Labat et al., 1999; Maheswaran and Khosa, 2012b). This means that in an SOV model, only second order interactions between explanatory variables and the target process are considered, which can be likened to a second order polynomial regression. What makes the SOV a powerful data-driven forecasting tool in water resources (and elsewhere) is that its model parameters can be solved as a least-squares solution (Wu and Kareem, 2014), providing an analytical expression for mapping explanatory variables to the target process in a manner akin to MLR. Although Volterra series models (and the SOV in particular) have been seldom studied in the water resources forecasting literature, there are a number of cases where the Volterra series models have outperformed or provided similar performance to other data-driven methods based on ANN, random forest, M5 model trees, and ELM (Maheswaran and Khosa, 2014, 2013a, 2013b, 2012a; Prasad et al., 2018; Rathinasamy and Khosa, 2012) with the added benefit that they provide an analytical forecasting equation that has interpretational value.

Due to the ease of use, computational efficiency, and interpretational value of the SOV it is adopted in this research alongside ELM as the nonlinear data-driven methods used for the purpose of mapping explanatory variables into target forecasts. MLR is also adopted as a linear data-driven benchmark method. The next sub-section focusses on input variable selection as a means for identifying useful explanatory variables that are subsequently fed into the data-driven forecasting methods (mentioned in this sub-section) to produce forecasts of a target process.

2.1.3. Nonlinearity – Input Variable Selection Methods

A brief overview of input variable selection is given before issues with current approaches are highlighted. Solutions to the highlighted problems are discussed according to the merits of the new input variable selection methods proposed in this research.

2.1.3.1. Overview

Input variable selection is an approach to identifying important explanatory variables from a large database of variables that may contain relevant, irrelevant, or redundant information in terms of a target process. Choosing an appropriate input variable selection method helps against selecting irrelevant inputs that can significantly influence data-driven model accuracy and add unnecessary complexity impacting model reliability, while also improving computational efficiency and potentially improving model interpretability (Back and Trappenberg, 2001; Hejazi and Cai, 2009; Šindelář and Babuška, 2004). In input variable selection procedures the user selects either a predefined number of input variables to choose or sets a threshold on the measure of dependence that halts the input variable selection procedure when the threshold is met (returning all input variables selected up to the particular point).

There are three main classes of input variable selection methods: filter (the focus of this research), wrapper, and embedded methods. Filters are most typically used for input variable selection since they may be applied independent of a model (e.g., ANN) while wrapper and embedded methods rely on a modeling mechanism to decide on the most important input variables (May et al., 2011). Wrapper and embedded methods will not be discussed in this research since they are computationally demanding and require some form of additional modeling mechanism for the selection of input variables (unlike filter methods).

2.1.3.2. Popular Filter-based Input Variable Selection Methods

The most popular input variable selection technique is the linear correlation coefficient ranking method, which essentially selects the explanatory variables that have the highest magnitude linear (Pearson) correlation with the target (May et al., 2011). A nonlinear analogue, and the most popular nonlinear input variable selection method, is based on the information-theoretic concept of mutual information (MI) (Shannon, 1948) and is known as the MI ranking method (Guyon and Elisseeff, 2003). This approach provides a significantly improved ability to identify dependencies between water resources variables since it can estimate nonlinear interactions (Khan et al., 2007). The linear and MI ranking approaches seek to select relevant variables and follow what is known as 'greedy' selection since they iteratively add one input variable at a time to the selected input variable set until a stopping criterion is met (i.e., either a

tolerance or pre-determined number of inputs), which is common with most filter-based input variable selection methods. However, a significant issue with these methods is that they do not account for redundancy amongst previously selected inputs and can lead to overly large input variable sets (especially for highly autocorrelated processes, which are commonly studied in water resources) that can cause poor model performance (May et al., 2011). At this point, the most basic linear and nonlinear approaches to input variable selection have been given; both methods seek to identify relevant variables but do not account for redundancy. Next, methods proposed to address redundancy (in addition to relevancy) will be discussed.

2.1.3.3. Methods that Account for Redundancy amongst Already Selected Inputs

To address the issue of (relevancy and) redundancy input variable methods based on partial (conditional) relevancy, that only recognizes a variable as being important if it shares significant correlation (information) with the target variables given all previously selected input variables, are considered. Using the concept of conditional relevancy, the partial correlation input selection (PCIS) and partial mutual information selection (PMIS) approaches were developed (Bowden et al., 2005b, 2005a, May et al., 2008b, 2008a, Sharma, 2000a, 2000b). For water resources forecasting problems, PMIS is preferred over PCIS since it accounts for nonlinear dependencies and has been shown to outperform PCIS in input variable selection problems in this domain (Galelli et al., 2014; May et al., 2008b).

PCIS uses the same concepts as partial least-squares regression for determining partial correlation between an input and the target, conditional on all other inputs. Similarly, PMIS involves the estimation of partial mutual information (PMI), also known as conditional MI (CMI), which is mutual information between an input and the response, conditional on all other inputs. While partial correlation in PCIS is derived analytically (Gustafson, 1961), the PMI (CMI) in PMIS is derived heuristically using two regression models (e.g., neural networks or kernel regression (Galelli et al., 2014; Sharma, 2000a)) to account for the conditional dependence of the previously selected inputs on the target and the input under consideration (for input selection) (May et al., 2008b; Sharma, 2000a). Therefore, input variable selection using PMIS is heavily dependent on conditional regression models and the optimization of their parameters, leading to studies that have sought optimal methods to estimate such parameters (Li et al., 2015a, 2015b). The downside to using PMIS, aside from its heuristic approach to the estimation of CMI, is the computational burden associated with the estimation of optimal parameters for the conditional models (Galelli et al., 2014). Therefore, while PMIS is a useful tool for nonlinear information-theoretic-based input variable selection and has found success in water resources forecasting studies (Chen et al., 2013; Fang et al., 2018;

He et al., 2011, 2015; Tran et al., 2015), more efficient methods that avoid heuristics are needed to reduce the computational burden and provide a direct means for estimating CMI.

2.1.3.4. Non-Heuristic Methods for the Estimation of Partial (Conditional) Mutual Information

More recently, a kernel density estimation approach called Partial Information (PI) was proposed for estimating CMI directly and therefore performing PMIS without heuristics (Sharma et al., 2016; Sharma and Mehrotra, 2014). Similarly to partial correlation, CMI has an analytical form that can be estimated using either joint entropy or mutual information terms based on the chain rule of entropy (Brown et al., 2012; Cover and Thomas, 2006) and it is this manner in which CMI is estimated by the authors in their PI approach (Sharma and Mehrotra, 2014). However, given the nature of kernel density estimation, which requires the specification of a bandwidth parameter (analogous to the bin width for histogram-based density estimation (Fernando et al., 2009)), the PI approach is also subject to careful bandwidth parameter selection in order to accurately estimate CMI. A similar approach to the kernel density estimation-based input variable selection relies on k nearest neighbours' statistics (Tsimpiris et al., 2012). While this method is more straightforward for parameter estimation (since its parameter, the number of nearest neighbours, is an integer rather than a continuous variable), it is still computationally demanding (Evans, 2008; Póczos and Schneider, 2012). Although non-heuristic approaches to CMI estimation and their related input variable selection methods improve upon the heuristic PMIS, they are still subject to high computation times. To overcome the computational issues with the methods mentioned in this sub-section, a very useful CMI-based input variable selection approach that does not require parametric tuning and is computationally efficient, is described in what follows.

2.1.3.5. Non-Heuristic and Computationally Efficient Conditional Mutual Information-based Input Variable Selection

In Van Hulle (2005), an Edgeworth approximation-based approach to MI estimation was defined and shown to be more computationally efficient and provide similar or better performance than kernel density estimation and nearest neighbour approaches. The approach derives MI via a Taylor series expansion around a reference distribution (taken as a zero mean Normal distribution with unit variance) and does not require any parameters to be set (Van Hulle, 2005). The Edgeworth approximations approach was used to estimate MI for short datasets, such as between the river Nile and the El-Nino Southern Oscillation Index (Khan et al., 2007, 2006). In this research, the Edgeworth-approximations approach was modified from MI estimation used in earlier studies (Khan et al., 2007, 2006; Van Hulle, 2005) for the non-parametric estimation of CMI (see Chapter 3 for details). The Edgeworth-approximations estimation of

CMI was then used in the same manner as earlier filter-based approaches, such as PCIS, PMIS, PI, etc. and formulated into an input variable selection routine that is named EA. The EA method is a computationally efficient, non-parametric, nonlinear information-theoretic method for input variable selection that does not require heuristic approaches for calculating CMI (such as PMIS) or require careful parametric tuning (such as PMIS and PI). It is therefore an attractive option for selecting input variables for nonlinear water resources forecasting problems. In the next sub-section, an approach is introduced based on the EA method that can be used to quantify input variable selection uncertainty.

2.1.3.6. Input Variable Selection Uncertainty

Filter-based input variable selection methods (e.g., PCIS, PMIS, PI, EA, etc.) focus on selecting a single set of input variables. However, since water resources are uncertain in nature it is reasonable to expect that there is likely to be uncertainty in the input variable selection procedure. Interestingly, the topic of input variable selection uncertainty is very new to the literature and only considered in a single study (outside of Chapter 3) that considers wrapper methods (and not filter methods, such as in this research) (Taormina et al., 2016). Essentially, input variable selection uncertainty supposes that for a given dataset if multiple samples are drawn and input variable selection performed, different selected input variable sets will be obtained. Therefore, this research proposed a bootstrap resampling approach to draw samples from a given dataset and uses EA to select input variables for each resample. The variables selected for each resample are stored. Afterwards, to assess the variation (uncertainty) across the selected input variable sets, a rank-ordering approach is used to identify those input variables which are selected often, seldom, or not at all. The combined bootstrap and rank-ordering procedure attached to the EA method results in the bootstrap rank-ordered CMI or broCMI. Since broCMI is based on the EA it is also computationally efficient and non-parametric. The broCMI approach is useful as it can be used for identifying a single input variable selection set that contains the inputs that are most often selected over a number of resamples or can be used to build an empirical input variable selection probability distribution that can be incorporated into uncertainty estimation of a data-driven forecasting model. Details on the broCMI method are given in Chapter 3 and the concept of estimating input variable selection uncertainty is explored using stochastics in Chapter 5 and 6.

In this sub-section, methods for addressing nonlinearity in water resources forecasting were discussed. Computationally efficient and easy to use nonlinear data-driven forecasting methods (i.e., ELM and SOV) that overcome limitations of traditional methods were highlighted. To provide useful model inputs to these data-driven techniques, two computationally efficient, non-parametric, nonlinear information-

theoretic-based input variable selections were introduced (EA and broCMI), the latter of which accounts for input variable selection uncertainty, a new topic in water resources forecasting. In the next sub-section, discussion is focussed on wavelet-based forecasting for addressing the multiscale nature of water resources.

2.2. Multiscale Change – Wavelet-based Forecasting

Addressing multiscale change in water resources is essential in improving forecasting accuracy when using data-driven methods (Hadi and Tombul, 2018). This sub-section briefly reviews some of the different approaches that have been paired with data-driven methods for assessing multiple scales of change in water resources forecasting and includes comparisons of these methods with wavelet-based forecasting methods, which are the focus of this sub-section. Most of the sub-section is devoted to a discussion on the incorrect usage of wavelets in the water resources forecasting literature followed by the proposed best (correct) practices for addressing this issue alongside the new wavelet-based data-driven forecasting framework (WDDFF) that adopts these best practices, conjoining the data-driven forecasting and input variable selection methods mentioned in the last sub-section.

2.2.1. Overview of Different Methods used for Capturing Multiscale Changes in Data-Driven Water Resources Forecasting Models

One of the traditional methods for capturing multiscale information and using it within water resources forecasts is through the Fourier Transform (FT) since it is able to extract periodic patterns from time series data (e.g., seasonality). For instance, Caiado (2010) used Fourier transforms to develop a Holt-Winters exponential smoothing function for one through seven day ahead forecasts and found it provided accurate urban water demand forecasts by extracting the seven and 365 day cycles embedded within the data. A study on river flow forecasting found that pre-processing input data by the FT and using this information as input to ANN resulted in worse performance than directly using the original input data (Atiya et al., 1999). The authors in Brentan et al. (2014) compared FT and the average reconstruction method for real-time urban water demand forecasting in Sao Paulo, Brazil using data measured every 20 minutes. The authors found the FT method the most appropriate for real-time forecasting as it was able to replicate the diurnal pattern; however, the authors noted that the FT had difficulty forecasting extreme demands. Typically, extreme demands are difficult to forecast due to their infrequent occurrence in the observed time series. Inherently, the FT assumes stationarity of a particular time series and it is expected to perform poorly at extracting the time period in which extreme values occur, as these values are outliers (i.e. they do not occur at a high frequency) and do not form a 'regular' pattern.

To overcome the inability of the FT to adequately capture infrequent extreme values typical of water resources time series (Hao and Singh, 2016; Katz et al., 2002), the water resources forecasting domain has more recently explored wavelet transforms (WT) (Abdollahi et al., 2017; Kalteh, 2013; Moosavi et al., 2013; Nanda et al., 2016; Sang, 2013) since the WT is known for its time-frequency localization properties (Percival and Walden, 2000; Percival et al., 2011) which allows the time-frequency analysis of (non-constant) variation through time for a given time series process. The ability of the WT to deal with multiscale time series makes it an ideal candidate for isolating low-frequency events. Therefore, since WT are apt at determining the period in which extreme values occur, use of the WT with nonlinear data-driven methods has improved the ability to forecast extreme value events common amongst water resources (Nourani et al., 2014), overcoming two prominent barriers (nonlinearity and multiscale changes) in providing accurate water resources forecasts. For instance, Adamowski et al. (2012) applied a coupled WT-ANN model and compared it against ANN, multiple linear regression, nonlinear regression, and autoregressive integrated moving average models for short-term urban water demand forecasting in Montreal, Canada. The authors found the WT-ANN model to perform the best as it was able to adequately address the multiscale behaviour of the urban water demand and explanatory variables, improving upon the ANN model which did not use time series pre-processed by WT. In Pammar and Deka (2017), the authors combined WT with SVR and compared it against SVR (without wavelet analysis) for forecasting daily pan evaporation finding that the WT-based SVR was better able to forecast the trends, seasonal patterns, and other complex behaviours, which the SVR could not forecast effectively. In a study on daily rainfall-runoff forecasting (Shoaib et al., 2018), a large number of ANN variants were coupled with WT and shown to provide substantially better performance than the standalone ANN and MLR models, with the authors noting that the WT-based forecasting methods were better able to capture different watershed signatures than the standalone models. While in Shi et al. (2018), WT coupled with ANN were found useful for uncovering high-frequency events in surface water quality prediction for anomaly detection which was subsequently used in an early warning system. Many other instances demonstrating the strong ability of WT coupled with data-driven methods to capture nonlinear and multiscale changes in water resources forecasting problems can be found in several review articles (Afan et al., 2016; Dixit et al., 2016; Fahimi et al., 2017; Nourani et al., 2014; Sang, 2013).

Aside from the WT, other approaches that can capture time-frequency localized properties of time series and that have been used with data-driven approaches for water resources forecasting include singular spectrum analysis (SSA) (Baratta et al., 2003) and empirical mode decomposition (EMD) (Karthikeyan and Kumar, 2013). SSA is related to principal component analysis and relies on the time-delay embedding of

time series in phase space. Therefore, SSA requires specification of a time delay and embedding dimension (Alexandrov et al., 2012) which can be difficult to estimate (Koutsoyiannis, 2006). Although SSA has been used in conjunction with data-driven methods for water resources forecasting (Latifoğlu et al., 2015; Unnikrishnan and Jothiprakash, 2018; Zhang et al., 2011), SSA can be computationally expensive and requires careful procedures to ensure that it provides meaningful results (Du et al., 2017; Zhang et al., 2015). Opposite to WT and SSA, which are rigorously grounded in mathematical principles (Allen, 1997; Hassani, 2010; Hassani et al., 2011; Heil and Walnut, 1989; Percival and Walden, 2000), EMD is a completely empirical method that relies solely on the dataset properties (mainly their local maxima and minima) (Huang et al., 1998). Although EMD has become popular for coupling with data-driven water resources forecasting models (Karthikeyan and Kumar, 2013; Xu et al., 2017; Zhang et al., 2016), it cannot be used correctly for forecasting applications. This is mainly because it suffers from boundary-conditions that make it difficult to incorporate new data points without having to recalibrate a given forecasting model for each new observation (Wang and Wu, 2016; Xiong et al., 2014; Zhang et al., 2015) which is time-consuming and non-optimal for real-time forecasting applications common in operational water resources systems (Bakshi, 1999; Maheswaran and Khosa, 2012b).

The WT is used in this research for capturing multiscale changes in water resources instead of the SSA and EMD since it can be updated in real-time (i.e., when new data is received) without a high computational cost (unlike SSA and EMD) and also because it is based on concrete mathematical properties that do not require empirical procedures for calculating multiscale changes such as EMD (which results in errors). Notwithstanding, the WT requires careful attention when it is adopted in forecasting applications in order to ensure that it used correctly and is thus applicable for real-world forecasting problems. Therefore, discussion on the use and misuse of WT in real-world wavelet-based forecasting is the subject of the next sub-section before identifying a set of best practices that shows how wavelet-based forecasts can be used correctly for real-world forecasting problems.

2.2.2. The Use and Misuse of the Wavelet Transform in Wavelet-based Data-Driven Forecasting

Before reviewing the use and misuse of WT in wavelet-based forecasting, some very brief properties of the WT are discussed. Chapter 4 provides a more in-depth treatment of these properties.

2.2.2.1. Overview

Wavelet transforms (like the Fourier transform) are calculated through convolution of a basis function with the original time series. Thus, just as the sine wave is the basis function for the FT, the WT also has a variety of possible basis functions. The different basis functions come in a variety of shapes (families)

and sizes (lengths or widths). A given wavelet family may be better at capturing transient behaviour while another may be better at capturing peaks. Wider wavelet basis functions are better suited to capturing long-period events and polynomial behaviour and have better frequency localization properties while narrower wavelet families have better time localization properties (which is why they are better at capturing transient events) (Maheswaran and Khosa, 2012b). Therefore, it is common within WT-based forecasting studies to match a given wavelet basis function to the type of properties contained within a target process and its explanatory variables (Barzegar et al., 2018a, 2017; Maheswaran and Khosa, 2012b; Rathinasamy et al., 2013; Shoaib et al., 2014). A second step in developing accurate WT-based forecasts is determining the scales of change that represent high variability and extracting these properties into sub-time series (Khokhlov et al., 2006; Percival et al., 2011) that can be included in the forecasting model (Karbasi, 2017; Kişi, 2011; Maheswaran and Khosa, 2012a; Murtagh et al., 2004; Nanda et al., 2016; Nourani et al., 2009; Özger et al., 2012). This step is referred to as wavelet decomposition, which requires one to not only specify which wavelet basis function to use but also the depth (level) of decomposition. The level of decomposition is usually selected to be no larger than half the number of time series records (due to the Nyquist frequency) and also such that the lowest-frequency information extracted from the time series is smooth and free of unnatural artifacts such as shark fins, triangles, etc. (Percival and Walden, 2000).

The procedure for using the WT for wavelet-based data-driven forecasting proceeds as follows: 1) a wavelet transform is selected, 2) a wavelet basis function and decomposition level is selected; 3) the WT using these 'settings' extracts the low- and high-frequency information (also known as scaling and wavelet coefficients) from the explanatory variables and/or target, i.e., wavelet decomposition); and 4) the wavelet and scaling coefficients are used as input to data-driven forecasting methods to forecast the target process.

There are several different WT that can be used and each WT have subtle differences. While the process of selecting a wavelet basis function and decomposition level is the same for each WT, the way in which low-and high-frequency information is extracted from the explanatory and/or target process (wavelet decomposition) is different and plays a key role in how a given WT can/should be used for wavelet-based forecasting. In the next sub-section, the different WT used in water resources forecasting studies are discussed before moving on to the implications of each method in real-world wavelet-based forecasting applications.

2.2.2.2. Contrasting Different Wavelet Transforms

Different WT used in water resources forecasting studies include the Continuous WT (CWT) (Hadi and Tombul, 2018), the discrete WT (DWT) multiresolution analysis (DWT-MRA) (Kişi, 2011), the maximal overlap DWT (MODWT) (see Chapter 4), the MODWT-MRA (Prasad et al., 2017), and the *à trous* algorithm (AT) (Rajaee et al., 2018).

The CWT allows one to extract time-frequency information for any scale of change (e.g., at a frequency (f) related to 7 days, 3 months, 365 days, 11 years, etc.). However, the CWT is more useful for continuous time signals and is often avoided in wavelet-based forecasting studies mainly due to its computational inefficiency and ad-hoc corrections required to address the fact that it is a continuous method applied to a discrete time series (Adamowski, 2008; Rathinasamy et al., 2014).

The DWT-MRA is the most popular WT used in wavelet-based forecasting and its numerous applications in water resources are covered in recent review articles (Afan et al., 2016; Dixit et al., 2016; Fahimi et al., 2017; Nourani et al., 2014; Sang, 2013; Yaseen et al., 2015). The DWT-MRA differs from the CWT in that it can be applied to discrete time series and its decomposition process is dyadic meaning that wavelet decomposition is carried out in powers of two (therefore, in the frequency domain, information can only be extracted within neighbouring frequency bands of the form $2^{-j} \leq f < 2^{-(j+1)}$, where j is a (integer) scale). Therefore, its decomposition process has a substantially reduced computational complexity when compared to the CWT. However, one of the main drawbacks of the DWT-MRA is that during decomposition it sub-samples the time series at powers of two and as the decomposition level is increased, fewer time series records are sampled in the calculation of the low-and high-frequency components. Another drawback is that it requires the time series to be a power of two and if it is not, requires alteration of the time series by appending values to the beginning of the time series to increase its sample size to the next highest power of two. This creates severe issues for real-time forecasting applications when new data is constantly being collected and used to update the forecasting model since the appending of new values changes the decomposed low- and high-frequency values at a given time point each time the DWT-MRA is re-computed. Therefore, it is easy to realize that the DWT-MRA is not ideal for real-world forecasting problems (Du et al., 2017; Maheswaran and Khosa, 2012b).

To mitigate this important issue concerning the DWT-MRA, the MODWT and MODWT-MRA were proposed to remove the need to sub-sample in powers of two during wavelet decomposition and as a result reduces the loss of information in the calculation of low-and high-frequency components (Percival and Walden, 2000). For a modest increase in computational cost (compared to the DWT-MRA), the

MODWT and MODWT-MRA do not require the time series to be a power of two and adding new data points to the time series does not change the low- and high-frequency values at a given time point each time the MODWT or MODWT-MRA are computed. The main difference between the two methods (MODWT and MODWT-MRA) is that the MODWT is a causal filter, meaning that only information from the present and past is used to calculate low- and high-frequency content of a time series at a given time point, while the MODWT-MRA is a non-causal filter, meaning that information before and after a given point in time are used in the calculation of low- and high-frequency information (Percival and Walden, 2000). For this reason, the MODWT is better suited to forecasting applications since it does not require any information from future times in the calculation of low- and high-frequency information for a given time point, while the MODWT-MRA cannot be used correctly for forecasting since any given point in time relies on information from the future. It is interesting to note that the DWT-MRA is also a non-causal filter, alike the MODWT-MRA and therefore requires information from the future to determine low-and high-frequency information at a given point in time and as a result cannot be used correctly for real-world forecasting applications.

The AT, alike the MODWT, is also a causal filter and can therefore be used for forecasting applications (Aussem et al., 1998). The difference between the AT and the MODWT is in how each method calculates coefficients in the high-frequency components. This difference is highlighted very clearly in Chapter 4 alongside mathematical and graphical depictions showing how and why, in the calculation of low- and high-frequency information at a given time point, the DWT-MRA and MODWT-MRA require future information while the MODWT and AT only require information from the past and present. Due to the manner in which the MODWT calculates high-frequency information, it may only be used to decompose explanatory variables and use those explanatory variables to forecast the target process directly. However, given the nature in which the AT decomposes a given time series, it may be used to decompose both target and explanatory variables and use the decomposed explanatory variables to forecast the decomposed target. The resulting forecasts for each scale of the decomposed target can then be aggregated to achieve an overall forecast of the target (i.e., a forecast of the original target prior to wavelet decomposition). Interestingly, the literature has not discussed this important distinction between the MODWT and AT and neither have both algorithms been compared against one another in terms of theory or forecasting performance outside of this research. More detailed information on the differences in wavelet-based forecasting models that use the MODWT or AT are given in Chapter 4.

Notwithstanding the fact that the MODWT and AT have properties that support their use in forecasting real-world (water resources) processes, there are some important best (correct) practices, developed in this research, that must be taken into account if they are to be used correctly. This is the subject of the next sub-section. Before proceeding, the important takeaways at this point are:

1. The DWT-MRA and the MODWT-MRA cannot be used correctly for forecasting real-world processes since they require information from the future in order to calculate low- and high-frequency coefficients at a particular point in time (i.e., they are non-causal filters). This is significant as the vast majority (about 90%, see Chapter 4) of wavelet-based forecasting studies, especially in water resources, adopt these methods and therefore have resulted in the development of incorrect forecasting models.
2. The MODWT and AT may be used correctly for wavelet-based forecasting of real-world (water resources) processes since they only consider past and present observations in the calculation of low- and high-frequency information at a given point in time (i.e., they are causal filters). However, best (correct) practices must be followed in order to use these methods correctly for real-world forecasting problems (see the next sub-section and Chapter 4).

2.2.3. Best (Correct) Practices for Wavelet-based Forecasting of Real-World Processes

This research contributes a set of best (correct) practices for the development of wavelet-based data-driven forecasting models, a key component that is missing in the wavelet-based forecasting literature. Since this the first set of best practices for wavelet-based forecasting proposed in the literature, the focus of this sub-section will be to briefly recount the different components of the best practices and highlight some areas where the literature has deviated from these best practices. A more thorough discussion of the ensuing topics can be found in Chapter 4.

The best practices for wavelet-based forecasting developed in this research are based on three key principles: 1) use of WT that are causal filters (e.g., that do not use future information), such as the MODWT and AT; 2) careful selection of decomposition levels and wavelet basis functions (also known as wavelet filters); and 3) proper partitioning of calibration and validation datasets. Following the first two principles results in what are termed “boundary-corrected” wavelet and scaling coefficients (high- and low-frequency components) that are calculated by performing wavelet decomposition on a given time series using causal filters (MODWT and AT) and carefully selecting decomposition levels and wavelet filters. The third step makes sure that there are a suitable number of wavelet and scaling coefficients in

the calibration and validation datasets to ensure the development of an accurate wavelet-based forecasting model that can be used for real-world applications.

2.2.3.1. Best Practice 1: Using Causal Wavelet Transforms

Both the MODWT and AT algorithms may be used for forecasting real-world water resources processes since they only require past and present time series observations to calculate low- and high-frequency information (scaling and wavelet coefficients). However, due to the differences in how the wavelet coefficients are calculated by the two methods: 1) the MODWT is restricted to decomposition of only the explanatory variables and must forecast the target directly (direct approach) while 2) the AT can use decomposed explanatory variables to forecast the target directly or it may be used to decompose the target and forecast the decomposed target using the decomposed explanatory variables, and then aggregate the decomposed target forecasts to achieve the forecast of the target at its original scale (multicomponent approach). This finding has not been discussed in the literature and its implications in terms of different levels of forecasting accuracy that can be achieved by either method is significant and discussed in more detail in Chapter 4.

2.2.3.2. Best Practice 2: Careful Selection of Decomposition Levels and Wavelet Filters

Since WT requires convolution between a wavelet filter and the time series (in other words, quantifying the similarity between a wavelet filter and a time series), a certain number of wavelet and scaling coefficients (based on the decomposition level and wavelet filter length) are effected by what is known as the boundary condition. A more detailed discussion of this item is given in Chapter 4. However, what is important to takeaway is that the boundary condition causes errors to be introduced in the calculation of wavelet and scaling coefficients at certain points in time (dependent on the WT, decomposition level, and wavelet filter). A high (low) decomposition level and wide (narrow) wavelet filter results in a high (low) number of time series observations affected by the boundary condition (Percival and Walden, 2000). For the DWT-MRA and the MODWT-MRA, the boundary condition is present at both the beginning and end of the time series (Maslova et al., 2016; Percival and Walden, 2000), while for the MODWT and AT it exists only at the beginning of the time series (Aussem et al., 1998; Percival and Walden, 2000). In the literature, it is very seldom that the boundary condition due to wavelet type, decomposition level, and wavelet filter selection are discussed (Aussem et al., 1998; Maslova et al., 2016; Percival, 2008; Percival et al., 2011; Percival and Mofjeld, 1997). While Maslova et al. (2016) explored the effect of different methods to reduce the effect of boundary condition, they used the MODWT-MRA and therefore, the models developed in their study cannot be used correctly for real-world forecasting applications. With

the exception of Chapter 4, no studies exist that explore the effect of the boundary condition due to wavelet type, decomposition level, and wavelet filter selection in terms of real-world forecasting performance.

The simple solution to the boundary condition problem for real-world forecasting as proposed in this research is to adopt the MODWT and AT for wavelet decomposition and to remove any wavelet and scaling coefficients from the beginning of the time series effected by the boundary condition. This results in what this research terms, "boundary-corrected" wavelet and scaling coefficients. The "boundary-corrected" wavelet and scaling coefficients can then be used for calibrating and validating real-world water resources forecasting models.

2.2.3.3. Best Practice 3: Proper Partitioning of Calibration and Validation Data

Many studies in the literature have made the inadvertent mistake of including wavelet and scaling coefficients affected by the boundary condition in the calibration and validation of their forecasting models, especially those adopting the DWT-MRA and MODWT-MRA (e.g., Barzegar et al. (2017); Deo et al. (2017b); Nourani et al. (2009); Prasad et al. (2017)), resulting in incorrectly developed forecasting models that cannot be used for real-world problems (see Chapter 4 for further details). To overcome this issue, this research proposes that the "boundary-corrected" wavelet and scaling coefficients (from the first two best practices) be divided into calibration and validation sets of suitable size, and their forecasting accuracy assessed on the validation set to ensure suitable accuracy is achieved for the user's intended purpose. Therefore, this best practice is intimately connected to the second best practice, as the number of "boundary-corrected" wavelet and scaling coefficients are determined by the selected decomposition level and wavelet filter. This research recommends that one select an initial set of wavelet decomposition level(s) and wavelet filter(s) and assess performance on the validation set in order to choose an optimal decomposition level and wavelet filter. Outside of this research, there does not exist best practices that base the selection of calibration and validation partitioning on "boundary-corrected" wavelet and scaling coefficients.

The best practices discussed in the last three sub-sections are an integral component of the new wavelet-based data-driven forecasting framework (WDDFF) proposed in this research. The subject of the next sub-section is focussed on the different components making up the WDDFF before moving on to how uncertainty assessment can be incorporated into the WDDFF using stochastics.

2.2.4. Best Practices and the Wavelet Data-Driven Forecasting Framework

The new WDDFF proposed in this research incorporates the best practices discussed in the last three sub-sections and adopts input variable selection and data-driven forecasting methods from section 2.1 (see also Chapter 3) in the following ways:

1. A wavelet-based forecasting method is selected according to either the MODWT (direct approach) or the AT (direct or multicomponent approach).
2. Input variable selection is used for selecting the most important wavelet and scaling coefficients to include in the forecasting model for the target process.
3. Data-driven forecasting methods are used to map the selected wavelet and scaling coefficients into target forecasts.

The WDDFF is the first wavelet-based forecasting framework to account for best (correct) practices for real-world wavelet-based forecasting applications. WDDFF is a very flexible and widely applicable tool for water resources managers requiring nonlinear and multiscale forecasts since it can adopt: 1) different wavelet-based forecasting methods (e.g., direct or multicomponent approaches based either on the MODWT or AT, noting that the MODWT can only be used with the direct approach), different input variable selection methods (e.g., PCIS, EA, broCMI, etc.), and different data-driven methods (e.g., MLR, ELM, SOV, etc.). Therefore, a large number of WDDFF variants can be developed and explored to determine which variant provides the best forecast or each can be used within an ensemble method, such as the proposed EW-SDDFF, which will be discussed further in the next section.

2.3. Uncertainty – Using Stochastics for Uncertainty Assessment

Since the WDDFF introduced in the last sub-section is only suited to addressing nonlinearity and multiscale change, a suitable uncertainty assessment method needs to be coupled with WDDFF in order to develop a holistic framework that can account for the nonlinear, multiscale, and uncertain nature of water resources and generate probability-based forecasts that are accurate and reliable. In this sub-section, a stochastic approach for uncertainty assessment is discussed and used for creating the stochastic version of WDDFF (SWDDFF) and its ensemble multi-wavelet version (EW-SDDFF). The new SWDDFF, a single-wavelet stochastic data-driven forecasting framework can be used to quantify uncertainty in the WDDFF and its resulting forecasts, while the EW-SDDFF, an ensemble multi-wavelet version of SWDDFF can be used to quantify the uncertainty in multiple WDDFF models, providing superior performance than its single-wavelet version (i.e., SWDDFF). The SWDDFF and EW-SDDFF are useful tools for forecasting

nonlinear, multiscale, and uncertain water resources that may be used by water resources managers for operating, planning, and managing water resources systems. This sub-section begins by introducing the stochastic framework adopted by SWDDFF and EW-SDDFF and compares it to other approaches in the literature, justifying its use in this research.

2.3.1. Uncertainty Assessment in Water Resources Forecasting

Uncertainty assessment is a crucial task in water resources forecasting in order to effectively manage, plan, and operate water resources systems (Farmer and Vogel, 2016; Krzysztofowicz, 2001; Matte et al., 2017). Different types of uncertainty are often considered (not necessarily together) in forecasting models, some of which include: initial conditions, input data, parameters, model structure, model output, etc. (Beven, 2015). A large benefit of considering different sources of uncertainty is that this information may be translated into a probability-based forecast (i.e., a forecast in the form of a probability density function) that can then be used for decision-making purposes (Ramos et al., 2013; Serinaldi and Kilsby, 2016; Yung et al., 2011). Therefore, the development and application of probability-based forecasting is an important area of research in the water resources domain (Hemri et al., 2015; Montanari and Grossi, 2008; Ren et al., 2018; Wang et al., 2017). Some of the more prominent approaches to developing probability-based forecasting frameworks include those based on Bayesian methods (Han and Coulibaly, 2017), evolutionary methods (Dotto et al., 2012), Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 2014), and stochastics (Fan et al., 2016).

The focus of this research is to use a recent stochastic framework for process-based models (Montanari and Koutsoyiannis, 2012) and to adopt it for the case of data-driven forecasting. This stochastic framework is chosen over other potential choices such as Bayesian methods since it does not require the specification of a prior or distributional assumptions, or methods such as GLUE that require estimation of a likelihood function (which may not include a trivial solution), or evolutionary methods which involve the selection of many hyper-parameters to obtain optimal performance. Furthermore, the stochastic approach of Montanari and Koutsoyiannis (2012) is straightforward to implement and can be used with very simple tools, such as the bootstrap (Efron and Tibshirani, 1993), making it a very useful and practical approach for estimating the probability density functions of different uncertainty sources (Sikorska et al., 2015). A very interesting result of adopting this stochastic framework and combining it with the bootstrap, is that many earlier data-driven forecasting methods that have used the bootstrap are special cases of the framework proposed in this research. The framework proposed by Montanari and Koutsoyiannis (2012), hereafter referred to as the blueprint, may be used for converting any deterministic

forecasting model (i.e., physically-based, data-driven, etc.) into a stochastic one. The next section uses recent literature to briefly explain this process and how it is modified from the case of physically-based models to data-driven models (further details on this process can be found in Chapter 5 and 6).

2.3.2. The Blueprint for Converting Deterministic Forecasting Models into their Stochastic Counterparts

The blueprint treats uncertainty as stemming from randomness (Montanari and Koutsoyiannis, 2012) and does not seek to split uncertainty into different (often times, "philosophical") categories such as epistemic or aleatory (Beven, 2015; Gong et al., 2013; Nearing et al., 2016), but rather treats uncertainty as a whole - an inability to perfectly explain a process. However, the blueprint does acknowledge that uncertainty or randomness is present in the different components making up the forecasting process and allows for the explicit consideration of such uncertainties, e.g., initial conditions, input data, parameters, model structure, model output, etc. To date, the blueprint has been used to explicitly consider input data, parameter, and model output uncertainty in process-based models (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). This research also considers these uncertainties but extends the blueprint to the case of data-driven methods (instead of physically-based models) and further enhances the framework by considering input variable selection uncertainty and wavelets in order to holistically account for nonlinearity, multiscale change, and uncertainty in a single framework - extending the WDDFF from a deterministic to a stochastic forecasting framework. This is the topic of section 2.3.3 (as well as Chapter 5). Firstly, how the blueprint is used to account for forecasting uncertainty is given below.

Since stochastic uncertainty assessment relies on randomness, in order to incorporate its use in forecasting it requires: 1) the specification and estimation of probability density function(s) and 2) randomly drawing samples from the probability density function(s) during forecasting. Therefore, the first step in using the blueprint for forecast uncertainty assessment is to explicitly define which sources of uncertainty are to be considered (all other sources not explicitly considered are assumed to be contained within the model error) and then to estimate the probability density function of each uncertainty source. Finally, during forecasting, for a given input, one draws from each probability density function a large number of times in order to build a forecast probability density function for the target process (Montanari and Koutsoyiannis, 2012). This procedure is described in more detail in Chapter 5 and 6. This is a very flexible paradigm as one can explicitly consider different sources of uncertainty that may be important to the problem and neglect explicitly quantifying other sources that may be insignificant or that are difficult

to characterize. In this way, the framework can be used to identify and quantify the significant sources of uncertainty for a given forecasting problem.

As mentioned above, a very useful feature of this blueprint is that it can be applied to physically-based, conceptual, and data-driven forecasting, while the different probability density functions can be estimated using simple tools, such as the bootstrap - a widely used uncertainty assessment tool in water resources forecasting (Erkyihun et al., 2016; Srivastav et al., 2007; Trichakis et al., 2011). Therefore, stochastics offers a natural and holistic framework for extending the nonlinear and multiscale WDDFF framework to incorporate uncertainty assessment broadening its applicability to a wider range of water resources management tasks requiring probability-based forecasts to improve decision-making.

2.3.3. Stochastic Wavelet Data-Driven Forecasting Framework

The SWDDFF is different than the blueprint in three keys ways: 1) it adopts data-driven forecasting methods (e.g., MLR, ELM, etc.) instead of physically-based models; 2) it considers input variable selection uncertainty (in addition to input data, parameter, and model output uncertainty); and 3) it adopts wavelets to account for multiscale change. If wavelets are removed from the framework then it results in the Stochastic Data-Driven Forecasting Framework.

The SWDDFF is novel in that it is the first wavelet-based stochastic data-driven forecasting framework to consider input data, input variable selection, parameter, and model output uncertainties. Other wavelet-based forecasting approaches in the literature consider different sources of uncertainty (mainly parameter uncertainty) and few consider both parameter and model output uncertainty. Tiwari and Adamowski (2017) combined bootstrap resampling and wavelet-based ANN models for forecasting urban water demand in Calgary, Alberta. In their approach, bootstrap resampling was used for estimating parameter uncertainty; many other wavelet-based data-driven forecasting frameworks used the bootstrap for estimating parameter uncertainty (Belayneh et al., 2016; Kasiviswanathan et al., 2016; Khalil et al., 2015; Kumar et al., 2015; Sehgal et al., 2014; Tiwari and Chatterjee, 2011). Wang et al. (2013) proposed a wavelet-based ANN for monthly forecasting of water quality in Harbin (Northeast China) that accounts for both parameter and model output uncertainty. The above examples of bootstrap-based approaches for quantifying parameter or parameter and model output uncertainty in wavelet-based data-driven forecasting models are special cases of the SWDDFF. For example, by explicitly considering parameter and model output uncertainty, but neglecting input data and input variable selection uncertainty, the SWDDFF can be used to replicate the model described in Wang et al. (2013). Therefore, not only is the SWDDFF a new and useful forecasting method for holistically capturing nonlinearity,

multiscale change, and uncertainty but it also may be modified to take on the form of many different approaches studied earlier in the literature, supporting its use for a wide array of water resources forecasting problems. While other approaches to uncertainty assessment in wavelet-based forecasting have been proposed, such as Bayesian-based methods (Bachour et al., 2016; Liu et al., 2015; Maslova et al., 2016; Sang et al., 2015, 2013), these methods neglect input variable selection uncertainty. This research (Chapter 5) shows that input variable selection uncertainty plays a key role in improving forecasting accuracy and reliability within SWDDFF. Therefore, given that SWDDFF includes input data, input variable selection, parameter, and model output uncertainty it represents the most advanced wavelet-based forecasting framework to date.

While the SWDDFF represents the most advanced wavelet-based forecasting framework, it can still be improved upon. Several recent studies (Alizadeh et al., 2017; Barzegar et al., 2018a, 2017; Rathinasamy et al., 2013) have demonstrated that by creating an ensemble framework of multiple wavelet-based (multi-wavelet) forecasts one can improve forecasting performance over the case where only a single-wavelet-based forecasting model is used (i.e., a single wavelet-based forecasting model that considers only a single decomposition level and wavelet filter). In the next sub-section, it is shown how the single-wavelet SWDDFF can be extended to an ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) that uses multiple WDDFF-based forecasts as input data. This generates a more robust probability-based forecasting framework that has higher accuracy and reliability when compared against SWDDFF.

2.3.4. Ensemble Wavelet - Data-Driven Forecasting Framework

Decomposing a time series using different decomposition levels and wavelet filters can be likened to taking a picture of an object (time series) with a multitude of zoom lenses (wavelet filters) and zoom factors (decomposition level). Each lens and zoom factor can take a picture of the same object; however, each picture of the object would differ based on the qualities of the lens and its zoom setting(s). Therefore, different decomposition levels and wavelet filters can be used to extract different interpretations of behaviours (properties) from a given time series and it is the task of the modeller to identify the most appropriate decomposition level(s) and wavelet filter(s) for a particular task. However, in many cases multiple decomposition levels and wavelet filters may perform very well but capture different phenomena, such as base flow, seasonal peaks, or transient events (Maheswaran and Khosa, 2012b). Therefore, it is intuitive to try to utilize the strength of each model (based on different decomposition levels and wavelet filters) in an ensemble to improve forecast accuracy and reliability.

For instance, Rathinasamy et al. (2013) created an ensemble of wavelet-based second order Volterra series models using Bayesian Model Averaging for streamflow forecasting at different time scales and found their ensemble model to provide better performance than any of the single wavelet-based models. Barzegar et al. (2017) used an ensemble of wavelet-based data-driven forecasting models using least-squares boosting for monthly groundwater forecasting finding their ensemble model to also outperform the best single wavelet-based model. Similar findings have also been found in other ensemble wavelet-based forecasting studies for water quality (Barzegar et al., 2018a) and suspended sediment concentration (Alizadeh et al., 2017). However, in each of the studies the focus was on issuing deterministic forecasts based on an ensemble multi-wavelet approach and none of these studies considered the uncertainty in the resulting forecasts. This research addresses this gap by proposing an ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) that generates probability-based forecasts and estimates ensemble model selection and ensemble weighting uncertainties, in addition to input data and ensemble model output uncertainties.

The EW-SDDFF is the first ensemble stochastic data-driven forecasting framework that considers ensemble model selection and weighting uncertainties and is the first ensemble multi-wavelet stochastic data-driven forecasting framework. Ensemble model selection and weighting uncertainties are estimated using input variable selection and data-driven forecasting methods, respectively. The EW-SDDFF extends the SWDDFF from the case of a single-wavelet-based stochastic forecasting framework to an ensemble multi-wavelet-based stochastic forecasting framework. A key benefit of EW-SDDFF is that, due to its data-driven nature, it can incorporate as input data the output from different forecasting methods (such as different physically-based models or data-driven methods, or even numerical weather predictions).

In this research, EW-SDDFF takes as input data multiple wavelet-based forecasts generated by the WDDFF. Since EW-SDDFF implicitly considers ensemble model selection (using input variable selection) and weighting (using data-driven methods), it automatically determines which WDDFF forecasts to incorporate in the ensemble and assigns a weight (importance) to each forecast. This is a major advantage of EW-SDDFF as typically, ensemble forecasting methods in the literature separately perform ensemble model selection and weighting (Doycheva et al., 2017; Lee et al., 2012; Tapiador and Gallardo, 2006; Weijs and van de Giesen, 2013). The EW-SDDFF can also incorporate the uncertainty in each WDDFF forecast used as input data in addition to estimating ensemble model output uncertainty, resulting in a probability-based forecast. A major advantage of the EW-SDDFF over the SWDDFF is that it can include the strengths of multiple wavelet-based forecasting models that use different decomposition levels and wavelet filters

(i.e., based on different WDDFF forecasts) and can inherently select the most useful forecasts to use and appropriately weight their contribution to the overall forecast, all while incorporating uncertainty in input data, ensemble model selection, ensemble model weighting, and ensemble model output uncertainties. This feature of EW-SDDFF is shown to provide significant improvements in forecasting accuracy and reliability when compared to SWDDFF and other benchmarks in the case study used in Chapter 6. Therefore, EW-SDDFF represents a very useful probability-based forecasting tool that accounts for nonlinearity, multiscale change, and uncertainty assessment that can be adopted by water resources managers for a variety of important tasks related to the operation, planning, and management of water resources systems. The final sub-section of this chapter contains a brief summary of the novel tools discussed in this chapter that will be studied in more detail in Chapter 3-6.

2.4. Summary

This chapter focussed on a literature review that introduced and discussed several key facets of this research and compared their novelty to approaches existing in the literature. The new methods discussed in this chapter include:

1. New computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods have been proposed (EA and broCMI, the latter considers input variable selection uncertainty) for selection of explanatory variables for use in nonlinear and computationally efficient data-driven forecasting methods (ELM and SOV), allowing for the nonlinear nature of water resources to be forecasted.
2. A set of best (correct) practices for wavelet-based forecasting have been proposed and coalesced into a new wavelet-based data-driven forecasting framework (WDDFF) that can adopt different wavelet-based forecasting methods (e.g., MODWT and AT, and their direct and/or multicomponent approaches), input variable selection methods, and data-driven forecasting models, to enable capturing nonlinearity and multiscale changes in water resources forecasting problems. The best practices are very useful (and necessary) since the majority of wavelet-based forecasting models in the literature are incorrect and cannot be used for real-world forecasting applications.
3. A new holistic stochastic-based uncertainty assessment framework that can be coupled with WDDFF was proposed to generate:
    a. A single-wavelet stochastic data-driven forecasting framework (SWDDFF) that accounts for input data, input variable selection, parameter, and model output uncertainty.

b. An ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) - a multi-wavelet version of SWDDFF - that accounts for ensemble model selection and weighting uncertainties, as well as input data and model output uncertainty.

In the chapters to follow (3-6), each of these innovations will be developed, tested, and applied to real-world water resources forecasting problems and will be demonstrated to provide accurate and reliable forecasts when compared against existing methods in the literature, highlighting their usefulness for a number of challenging forecasting problems facing water resources managers.

2.5. References

Abdollahi, S., Raeisi, J., Khalilianpour, M., Ahmadi, F., Kisi, O., 2017. Daily Mean Streamflow Prediction in Perennial and Non-Perennial Rivers Using Four Data Driven Techniques. Water Resour. Manag. doi:10.1007/s11269-017-1782-7

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L., 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog. Phys. Geogr. 36, 480–513. doi:10.1177/0309133312444943

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Adamowski, J.F., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. J. Hydrol. 353, 247–266. doi:10.1016/j.jhydrol.2008.02.013

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. doi:10.1016/j.jhydrol.2016.07.048

Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T.S., 2012. A Review of Some Modern Approaches to the Problem of Trend Extraction. Econom. Rev. doi:10.1080/07474938.2011.608032

Alizadeh, M.J., Jafari Nodoushan, E., Kalarestaghi, N., Chau, K.W., 2017. Toward multi-day-ahead forecasting of suspended sediment concentration using ensemble models. Environ. Sci. Pollut. Res.

28017–28025. doi:10.1007/s11356-017-0405-4

Alizamir, M., Kisi, O., Zounemat-Kermani, M., 2018. Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. Hydrol. Sci. J. 63, 63–73. doi:10.1080/02626667.2017.1410891

Allen, M., 1997. Optimal filtering in singular spectrum analysis. Phys. Lett. A 234, 419–428. doi:10.1016/S0375-9601(97)00559-8

Atiya, A.F., El-Shoura, S.M., Shaheen, S.I., El-Sherif, M.S., 1999. A comparison between neural-network forecasting techniques--case study: river flow forecasting. IEEE Trans. Neural Netw. 10, 402–409. doi:10.1109/72.750569

Aussem, A., Campbell, J., Murtagh, F., 1998. Wavelet-based feature extraction and decomposition strategies for financial forecasting. J. Comput. Intell. Financ. 6, 5–12.

Bachour, R., Maslova, I., Ticlavilca, A.M., Walker, W.R., McKee, M., 2016. Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. Stoch. Environ. Res. Risk Assess. 30, 103–117. doi:10.1007/s00477-015-1039-z

Back, A.D., Trappenberg, T.P., 2001. Selecting inputs for modeling using normalized higher order statistics and independent component analysis. IEEE Trans. Neural Networks 12, 612–617. doi:10.1109/72.925564

Bakshi, B., 1999. Multiscale analysis and modeling using wavelets. J. Chemom. 434, 415–434.

Baratta, D., Cicioni, G., Masulli, F., Studer, L.E., 2003. Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting. Neural Networks 16, 375–387. doi:10.1016/S0893-6080(03)00022-4

Barzegar, R., Asghari Moghaddam, A., Adamowski, J., Ozga-Zielinski, B., 2018a. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Environ. Res. Risk Assess. 32, 799–813. doi:10.1007/s00477-017-1394-z

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189

Barzegar, R., Moghaddam, A.A., Deo, R., Fijani, E., Tziritis, E., 2018b. Mapping groundwater contamination

risk of multiple aquifers using multi-model ensemble of machine learning algorithms. Sci. Total Environ. 621, 697–712. doi:10.1016/j.scitotenv.2017.11.185

Baydaroğlu, Ö., Koçak, K., 2014. SVR-based prediction of evaporation combined with chaotic approach. J. Hydrol. 508, 356–363. doi:10.1016/j.jhydrol.2013.11.008

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172–173. doi:10.1016/j.atmosres.2015.12.017

Beven, K., 2015. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol. Sci. J. 6667, 150527103244004. doi:10.1080/02626667.2015.1031761

Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrol. Process. 28, 5897–5918. doi:10.1002/hyp.10082

Bougadis, J., Adamowski, K., Diduch, R., 2005. Short-term municipal water demand forecasting. Hydrol. Process. 19, 137–148. doi:10.1002/hyp.5763

Bowden, G.J., Dandy, G.C., Maier, H.R., 2005a. Input determination for neural network models in water resources applications. Part 1 - Background and methodology. J. Hydrol. 301, 75–92. doi:10.1016/j.jhydrol.2004.06.021

Bowden, G.J., Maier, H.R., Dandy, G.C., 2005b. Input determination for neural network models in water resources applications. Part 2. Case study: Forecasting salinity in a river. J. Hydrol. 301, 93–107. doi:10.1016/j.jhydrol.2004.06.020

Brentan, B.M., Ribeiro, L.C.L.J., Luvizotto-Jr, E., Mendonça, D.C., Guidi, J.M., 2014. Synthetic Reconstruction of Water Demand Time Series for Real Time Demand Forecasting. J. Water Resour. Prot. 6, 1437–1443.

Brown, G., Pocock, A., Zhao, M.-J., Lujan, M., 2012. Conditional Likelihood Maximisation: A Unifying Framework for Mutual Information Feature Selection. J. Mach. Learn. Res. 13, 27–66. doi:10.1016/j.patcog.2015.11.007

Caiado, J., 2010. Performance of Combined Double Seasonal Univariate Time Series Models for Forecasting Water Demand. J. Hydrol. Eng. 15, 215–222. doi:10.1061/(ASCE)HE.1943-5584.0000182

Chen, L., Singh, V.P., Guo, S., Zhou, J., Ye, L., 2013. Copula entropy coupled with artificial neural network

for rainfall–runoff simulation. Stoch. Environ. Res. Risk Assess. 28, 1755–1767. doi:10.1007/s00477-013-0838-3

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory. Wiley-Interscience, New York.

Danandeh, A., Kahya, E., Olyaie, E., 2013. Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. J. Hydrol. 505, 240–249. doi:10.1016/j.jhydrol.2013.10.003

Deo, R.C., Kisi, O., Singh, V.P., 2017a. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos. Res. 184, 149–175. doi:10.1016/j.atmosres.2016.10.004

Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmos. Res. 153, 512–525. doi:10.1016/j.atmosres.2014.10.016

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2017b. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stoch. Environ. Res. Risk Assess. 31, 1211–1240. doi:10.1007/s00477-016-1265-z

Diskin, M.H., Boneh, A., Golan, A., 1984. Identification of a Volterra series conceptual model based on a cascade of nonlinear reservoirs. J. Hydrol. 68, 231–245. doi:10.1016/0022-1694(84)90213-0

Dixit, P., Londhe, S., Deo, M.C., 2016. Review of Applications of Neuro-Wavelet Techniques in Water Flows. Ina. Lett. 1, 99–104. doi:10.1007/s41403-016-0015-3

Dotto, C.B.S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D.T., Freni, G., Rauch, W., Deletic, A., 2012. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Res. 46, 2545–2558. doi:10.1016/j.watres.2012.02.009

Dou, X., Yang, Y., 2018. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. Comput. Electron. Agric. 148, 95–106. doi:10.1016/j.compag.2018.03.010

Doycheva, K., Horn, G., Koch, C., Schumann, A., König, M., 2017. Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning. Adv. Eng. Informatics 33, 427–439.

doi:10.1016/j.aei.2016.11.001

Du, K., Zhao, Y., Lei, J., 2017. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. J. Hydrol. 552, 44–51. doi:https://doi.org/10.1016/j.jhydrol.2017.06.019

Efron, B., Tibshirani, R.J.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, London, U.K.

Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D.P., 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. Hydrol. Earth Syst. Sci. 14, 1943–1961. doi:10.5194/hess-14-1943-2010

Erkyihun, S.T., Rajagopalan, B., Zagona, E., Lall, U., Nowak, K., 2016. Wavelet-based time series bootstrap model for multidecadal streamflow simulation using climate indicators. Water Resour. Res. 52, 4061–4077. doi:10.1002/2016WR018696

Evans, D., 2008. A computationally efficient estimator for mutual information. Proc. R. Soc. A Math. Phys. Eng. Sci. 464, 1203–1215. doi:10.1098/rspa.2007.0196

Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor. Appl. Climatol. 128, 875–903. doi:10.1007/s00704-016-1735-8

Fan, F.M., Schwanenberg, D., Alvarado, R., Assis dos Reis, A., Collischonn, W., Naumman, S., 2016. Performance of Deterministic and Probabilistic Hydrological Forecasts for the Short-Term Optimization of a Tropical Hydropower Reservoir. Water Resour. Manag. 30, 3609–3625. doi:10.1007/s11269-016-1377-8

Fang, W., Huang, S., Huang, Q., Huang, G., Meng, E., Luan, J., 2018. Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information. J. Hydrol. 561, 764–779. doi:10.1016/j.jhydrol.2018.04.038

Farmer, W.H., Vogel, R.M., 2016. On the deterministic and stochastic use of hydrologic models. Water Resour. Res. 52, 5619–5633. doi:10.1002/2016WR019129

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176. doi:10.1016/j.jhydrol.2008.10.019

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. doi:10.1016/j.envsoft.2014.08.015

Ghanbarpour, M.R., Abbaspour, K.C., Hipel, K.W., 2009. A comparative study in long-term river flow forecasting models. Int. J. River Basin Manag. 7, 403–413. doi:10.1080/15715124.2009.9635398

Giustolisi, O., Savic, D.A., 2006. A symbolic data-driven technique based on evolutionary polynomial regression. J. Hydroinformatics 8, 207–222. doi:10.2166/hydro.2006.020

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., Hero, A.O., 2013. Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. Water Resour. Res. 49, 2253–2273. doi:10.1002/wrcr.20161

Goyal, M.K., Bharti, B., Quilty, J., Adamowski, J., Pandey, A., 2014. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst. Appl. 41. doi:10.1016/j.eswa.2014.02.047

Gustafson, R.L., 1961. Partial Correlations in Regression Computations. J. Am. Stat. Assoc. 56, 363–367. doi:10.1080/01621459.1961.10482120

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Hadi, S.J., Tombul, M., 2018. Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination. J. Hydrol. 561, 674–687. doi:10.1016/j.jhydrol.2018.04.036

Han, S., Coulibaly, P., 2017. Bayesian flood forecasting methods: A review. J. Hydrol. 551, 340–351. doi:10.1016/j.jhydrol.2017.06.004

Hao, Z., Singh, V.P., 2016. Review of dependence modeling in hydrology and water resources. Prog. Phys. Geogr. 40, 549–578. doi:10.1177/0309133316632460

Hasanpour Kashani, M., Ghorbani, M.A., Dinpashoh, Y., Shahmorad, S., 2014. Comparison of Volterra Model and Artificial Neural Networks for Rainfall-Runoff Simulation. Nat. Resour. Res. 23, 341–354. doi:10.1007/s11053-014-9235-y

Hassani, H., 2010. Singular spectrum analysis based on the minimum variance estimator. Nonlinear Anal. Real World Appl. 11, 2065–2077. doi:10.1016/j.nonrwa.2009.05.009

Hassani, H., Xu, Z., Zhigljavsky, A., 2011. Singular spectrum analysis based on the perturbation theory. Nonlinear Anal. Real World Appl. 12, 2752–2766. doi:10.1016/j.nonrwa.2011.03.020

He, J., Valeo, C., Chu, A., Neumann, N.F., 2011. Prediction of event-based stormwater runoff quantity and quality by ANNs developed using PMI-based input selection. J. Hydrol. 400, 10–23. doi:10.1016/j.jhydrol.2011.01.024

He, X., Guan, H., Qin, J., 2015. A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. J. Hydrol. 527, 88–100. doi:10.1016/j.jhydrol.2015.04.047

Heddam, S., Kisi, O., 2017. Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. Environ. Sci. Pollut. Res. 24, 16702–16724. doi:10.1007/s11356-017-9283-z

Heil, C., Walnut, D., 1989. Continuous and discrete wavelet transforms. Soc. Ind. Appl. Math. 31, 628–666. doi:10.1137/1031129

Hejazi, M.I., Cai, X., 2009. Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm. Adv. Water Resour. 32, 582–593. doi:10.1016/j.advwatres.2009.01.009

Hemri, S., Lisniak, D., Klein, B., 2015. Multivariate postprocessing techniques for probabilistic hydrological forecasting. Water Resour. Res. 51, 7436–7451. doi:10.1002/2014WR016473

Huang, G.-B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man, Cybern. Part B Cybern. 42, 513–529. doi:10.1109/TSMCB.2011.2168604

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. doi:10.1016/j.neucom.2005.12.126

Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., Liu, H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. A Math. Phys. Eng. Sci. 454, 995, 903. doi:10.1098/rspa.1998.0193

Ince, H., 2006. Non-parametric regression methods. Comput. Manag. Sci. 3, 161–174. doi:10.1007/s10287-005-0006-4

Kalteh, A.M., 2013. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Comput. Geosci. 54, 1–8. doi:10.1016/j.cageo.2012.11.015

Karbasi, M., 2017. Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet-Gaussian Process Regression Model. Water Resour. Manag. doi:10.1007/s11269-017-1853-9

Karthikeyan, L., Kumar, D.N., 2013. Predictability of nonstationary time series using wavelet and EMD based ARMA models. J. Hydrol. 502, 103–119. doi:10.1016/j.jhydrol.2013.08.030

Kasiviswanathan, K.S., He, J., Sudheer, K.P., Tay, J.-H., 2016. Potential application of wavelet neural network ensemble to forecast streamflow for flood management. J. Hydrol. 536, 161–173. doi:10.1016/j.jhydrol.2016.02.044

Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. Water Resour. 25, 1287–1304. doi:10.1016/S0309-1708(02)00056-8

Keshtegar, B., Allawi, M.F., Afan, H.A., El-Shafie, A., 2016. Optimized River Stream-Flow Forecasting Model Utilizing High-Order Response Surface Method. Water Resour. Manag. 30, 3899–3914. doi:10.1007/s11269-016-1397-4

Keshtegar, B., Kisi, O., Asce, M., 2017. Modified Response-Surface Method : New Approach for Modeling Pan Evaporation. J. Hydrol. Eng. 22, 1–14. doi:10.1061/(ASCE)HE.1943-5584.0001541.

Khalil, B., Broda, S., Adamowski, J., Ozga-Zielinski, B., Donohoe, A., 2015. Short-term forecasting of groundwater levels under conditions of mine-tailings recharge using wavelet ensemble neural network models. Hydrogeol. J. 23, 121–141. doi:10.1007/s10040-014-1204-3

Khan, S., Bandyopadhyay, S., Ganguly, A.R., Saigal, S., Erickson III, D.J., Protopopescu, V., Ostrouchov, G., 2007. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 76. doi:10.1103/PhysRevE.76.026209

Khan, S., Ganguly, A.R., Bandyopadhyay, S., Saigal, S., Erickson III, D.J., Protopopescu, V., Ostrouchov, G., 2006. Nonlinear statistics reveals stronger tie between ENSO and the tropical hydrological cycle. Geophys. Res. Lett. 33. doi:10.1029/2006GL027941

Khokhlov, V.N., Glushkov, A. V, Loboda, N.S., 2006. On the nonlinear interaction between global

teleconnection patterns 132, 447–465. doi:10.1256/qj.05.05

Kisi, O., 2016. Modeling reference evapotranspiration using three different heuristic regression approaches. Agric. Water Manag. 169, 162–172. doi:10.1016/j.agwat.2016.02.026

Kişi, Ö., 2011. A combined generalized regression neural network wavelet model for monthly streamflow prediction. KSCE J. Civ. Eng. 15, 1469–1479. doi:10.1007/s12205-011-1004-4

Koutsoyiannis, D., 2006. On the quest for chaotic attractors in hydrological processes. Hydrol. Sci. J. 51, 1065–1091. doi:10.1623/hysj.51.6.1065

Koutsoyiannis, D., Yao, H., Georgakakos, A., 2010. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrol. Sci. J. 53, 142–164. doi:10.1623/hysj.53.1.142

Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology. J. Hydrol. 249, 2–9. doi:10.1016/S0022-1694(01)00420-6

Kumar, S., Tiwari, M.K., Chatterjee, C., Mishra, A., 2015. Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method. Water Resour. Manag. 29, 4863–4883. doi:10.1007/s11269-015-1095-7

Labat, D., Ababou, R., Mangin, A., 1999. Linear and nonlinear input/output models for karstic springflow and flood prediction at different time scales. Stoch. Environ. Res. Risk Assess. 13, 337–364. doi:10.1007/s004770050055

Latifoğlu, L., Kişi, Ö., Latifoğlu, F., 2015. Importance of hybrid models for forecasting of hydrological variable. Neural Comput. Appl. 26, 1669–1680. doi:10.1007/s00521-015-1831-1

Lee, J.A., Kolczynski, W.C., McCandless, T.C., Haupt, S.E., 2012. An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Wind Prediction. Mon. Weather Rev. 140, 2270–2286. doi:10.1175/MWR-D-11-00065.1

Lee, T., Ouarda, T.B.M.J., Yoon, S., 2017. KNN-based local linear regression for the analysis and simulation of low flow extremes under climatic influence. Clim. Dyn. 49, 3493–3511. doi:10.1007/s00382-017-3525-0

Li, C., Bai, Y., Zeng, B., 2016. Deep Feature Learning Architectures for Daily Reservoir Inflow Forecasting. Water Resour. Manag. 30, 5145–5161. doi:10.1007/s11269-016-1474-8

Li, X., Maier, H.R., Zecchin, A.C., 2015a. Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. Environ. Model. Softw. 65, 15–29. doi:10.1016/j.envsoft.2014.11.028

Li, X., Zecchin, A.C., Maier, H.R., 2015b. Improving partial mutual information-based input variable selection by consideration of boundary issues associated with bandwidth estimation. Environ. Model. Softw. 71, 78–96. doi:10.1016/j.envsoft.2015.05.013

Lima, A.R., Cannon, A.J., Hsieh, W.W., 2016. Forecasting daily streamflow using online sequential extreme learning machines. J. Hydrol. 537, 431–443. doi:10.1016/j.jhydrol.2016.03.017

Liu, Z., Zhou, P., Zhang, Y., 2015. A Probabilistic Wavelet–Support Vector Regression Model for Streamflow Forecasting with Rainfall and Climate Information Input. J. Hydrometeorol. 16, 2209–2229. doi:10.1175/JHM-D-14-0210.1

Lou, I., Xie, Z., Ung, W.K., Mok, K.M., 2016. Freshwater algal bloom prediction by extreme learning machine in Macau Storage Reservoirs. Neural Comput. Appl. 27, 19–26. doi:10.1007/s00521-013-1538-0

Luo, G., 2016. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Netw. Model. Anal. Heal. Informatics Bioinforma. 5, 18. doi:10.1007/s13721-016-0125-6

Maheswaran, R., Khosa, R., 2014. A Wavelet-Based Second Order Nonlinear Model for Forecasting Monthly Rainfall. Water Resour. Manag. 28, 5411–5431. doi:10.1007/s11269-014-0809-6

Maheswaran, R., Khosa, R., 2013a. Long term forecasting of groundwater levels with evidence of non-stationary and nonlinear characteristics. Comput. Geosci. 52, 422–436. doi:10.1016/j.cageo.2012.09.030

Maheswaran, R., Khosa, R., 2013b. Wavelets-based non-linear model for real-time daily flow forecasting in Krishna River. J. Hydroinformatics 15, 1022. doi:10.2166/hydro.2013.135

Maheswaran, R., Khosa, R., 2012a. Wavelet-Volterra coupled model for monthly stream flow forecasting. J. Hydrol. 450–451, 320–335. doi:10.1016/j.jhydrol.2012.04.017

Maheswaran, R., Khosa, R., 2012b. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ. Model. Softw. 25, 891–909. doi:10.1016/j.envsoft.2010.02.003

Marvuglia, A., Messineo, A., Nicolosi, G., 2014. Coupling a neural network temperature predictor and a fuzzy logic controller to perform thermal comfort regulation in an office building. Build. Environ. 72, 287–299. doi:10.1016/j.buildenv.2013.10.020

Maslova, I., Ticlavilca, A.M., Mckee, M., 2016. Adjusting wavelet-based multiresolution analysis boundary conditions for long-term streamflow forecasting. Hydrol. Process. doi:10.1002/hyp.10564

Matos, J.P., Portela, M.M., Schleiss, A.J., 2018. Towards Safer Data-Driven Forecasting of Extreme Streamflows. Water Resour. Manag. 32, 701–720. doi:10.1007/s11269-017-1834-z

Matte, S., Boucher, M.A., Boucher, V., Fortier Filion, T.C., 2017. Moving beyond the cost-loss ratio: Economic assessment of streamflow forecasts for a risk-Averse decision maker. Hydrol. Earth Syst. Sci. 21, 2967–2986. doi:10.5194/hess-21-2967-2017

May, R.J., Dandy, G., Maier, H., 2011. Review of input variable selection methods for artificial neural networks. Artif. Neural Networks-Methodological Adv. Biomed. Appl. 19–44.

May, R.J., Dandy, G.C., Maier, H.R., Nixon, J.B., 2008a. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. Environ. Model. Softw. 23, 1289–1299. doi:10.1016/j.envsoft.2008.03.008

May, R.J., Maier, H.R., Dandy, G.C., Fernando, T.M.K.G., 2008b. Non-linear variable selection for artificial neural networks using partial mutual information. Environ. Model. Softw. 23, 1312–1326. doi:10.1016/j.envsoft.2008.03.007

Mohammadi, K., Eslami, H.R., Kahawita, R., 2006. Parameter estimation of an ARMA model for river flow forecasting using goal programming. J. Hydrol. 331, 293–299. doi:https://doi.org/10.1016/j.jhydrol.2006.05.017

Montanari, A., Grossi, G., 2008. Estimating the uncertainty of hydrological forecasts: A statistical approach. Water Resour. Res. 44. doi:10.1029/2008WR006897

Montanari, A., Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. Water Resour. Res. 48. doi:10.1029/2011WR011412

Moosavi, V., Vafakhah, M., Shirmohammadi, B., Behnia, N., 2013. A Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods. Water Resour. Manag. 27, 1301–1321. doi:10.1007/s11269-012-0239-2

Mount, N.J., Dawson, C.W., Abrahart, R.J., 2013. Legitimising data-driven models : exemplification of a new data-driven mechanistic modelling framework. Hydrol. Earth Syst. Sci. 17, 2827–2843. doi:10.5194/hess-17-2827-2013

Murtagh, F., Starck, J.L., Renaud, O., 2004. On neuro-wavelet modeling. Decis. Support Syst. 37, 475–484. doi:10.1016/S0167-9236(03)00092-7

Nanda, T., Sahoo, B., Beria, H., Chatterjee, C., 2016. A wavelet-based non-linear autoregressive with exogenous inputs (WNARX) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products. J. Hydrol. 539, 57–73. doi:10.1016/j.jhydrol.2016.05.014

Nearing, G.S., Tian, Y., Gupta, H. V., Clark, M.P., Harrison, K.W., Weijs, S. V., 2016. A Philosophical Basis for Hydrologic Uncertainty. Hydrol. Sci. J. 6667, 1–20. doi:10.1080/02626667.2016.1183009

Nourani, V., Alami, M.T., Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng. Appl. Artif. Intell. 22, 466–472. doi:10.1016/j.engappai.2008.09.003

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. J. Hydrol. 514, 358–377. doi:10.1016/j.jhydrol.2014.03.057

Özger, M., Mishra, A.K., Singh, V.P., 2012. Long Lead Time Drought Forecasting Using a Wavelet and Fuzzy Logic Combination Model: A Case Study in Texas. J. Hydrometeorol. 13, 284–297. doi:10.1175/JHM-D-10-05007.1

Pammar, L., Deka, P.C., 2017. Daily pan evaporation modeling in climatically contrasting zones with hybridization of wavelet transform and support vector machines. Paddy Water Environ. 15, 711–722. doi:10.1007/s10333-016-0571-x

Percival, D.B., 2008. Analysis of Geophysical Time Series Using Discrete Wavelet Transforms: An Overview, in: Donner, R. V, Barbosa, S.M. (Eds.), Nonlinear Time Series Analysis in the Geosciences. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 61–79. doi:10.1007/978-3-540-78938-3_4

Percival, D.B., Lennox, S.M., Wang, Y.G., Darnell, R.E., 2011. Wavelet-based multiresolution analysis of Wivenhoe Dam water temperatures. Water Resour. Res. doi:10.1029/2010WR009657

Percival, D.B., Mofjeld, H.O., 1997. Analysis of Subtidal Coastal Sea Level Fluctuations Using Wavelets. J. Am. Stat. Assoc. 92, 868. doi:10.2307/2965551

Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis, 1st ed, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Póczos, B., Schneider, J.G., 2012. Nonparametric estimation of conditional information and divergences. Int. Conf. Artif. Intell. Stat. 22, 914--923.

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2018. Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. Soil Tillage Res. 181, 63–81. doi:https://doi.org/10.1016/j.still.2018.03.021

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ANN-based stream flow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos. Res. 197, 42–63. doi:10.1016/j.atmosres.2017.06.014

Rahimikhoob, A., 2016. Comparison of M5 Model Tree and Artificial Neural Networks Methodologies in Modelling Daily Reference Evapotranspiration from NOAA Satellite Images. Water Resour. Manag. 1–13. doi:10.1007/s11269-016-1331-9

Rajaee, T., Ravansalar, M., Adamowski, J.F., Deo, R.C., 2018. A New Approach to Predict Daily pH in Rivers Based on the “à trous” Redundant Wavelet Transform Algorithm. Water, Air, Soil Pollut. 229, 85. doi:10.1007/s11270-018-3715-3

Ramos, M.H., Van Andel, S.J., Pappenberger, F., 2013. Do probabilistic forecasts lead to better decisions? Hydrol. Earth Syst. Sci. 17, 2219–2232. doi:10.5194/hess-17-2219-2013

Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J. Hydrol. doi:10.1016/j.jhydrol.2013.09.025

Rathinasamy, M., Khosa, R., 2012. Multiscale nonlinear model for monthly streamflow forecasting: a wavelet-based approach. J. Hydroinformatics 14, 424. doi:10.2166/hydro.2011.130

Rathinasamy, M., Khosa, R., Adamowski, J., Ch, S., Partheepan, G., Anand, J., Narsimlu, B., 2014. Wavelet-

based multiscale performance analysis: An approach to assess and improve hydrological models. Water Resour. Res. 50, 9721–9737. doi:10.1002/2013WR014650

Remesan, R., Mathew, J., 2015. Hydrological data driven modelling: A case study approach, Hydrological Data Driven Modelling: A Case Study Approach. Springer International Publishing.

Ren, W., Yang, T., Shi, P., Xu, C. yu, Zhang, K., Zhou, X., Shao, Q., Ciais, P., 2018. A probabilistic method for streamflow projection and associated uncertainty analysis in a data sparse alpine region. Glob. Planet. Change 165, 100–113. doi:10.1016/j.gloplacha.2018.03.011

Rezaie-Balf, M., Kisi, O., 2017. New formulation for forecasting streamflow: evolutionary polynomial regression vs. extreme learning machine. Hydrol. Res. doi:10.2166/nh.2017.283

Sang, Y.-F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. 122, 8–15. doi:10.1016/j.atmosres.2012.11.003

Sang, Y.-F., Wang, Z., Liu, C., 2015. Wavelet Neural Modeling for Hydrologic Time Series Forecasting with Uncertainty Evaluation. Water Resour. Manag. 29, 1789–1801. doi:10.1007/s11269-014-0911-9

Sang, Y.F., Shang, L.Y., Wang, Z.G., Liu, C.M., Yang, M.G., 2013. Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting. Chinese Sci. Bull. 58, 3796–3805. doi:10.1007/s11434-013-5903-2

Sehgal, V., Tiwari, M.K., Chatterjee, C., Sehgal, V., Tiwari, M.K., Chatterjee, C., 2014. Wavelet Bootstrap Multiple Linear Regression Based Hybrid Modeling for Daily River Discharge Forecasting. Water Resour Manag. 28, 2793–2811. doi:10.1007/s11269-014-0638-7

Serinaldi, F., Kilsby, C.G., 2016. A Blueprint for Full Collective Flood Risk Estimation: Demonstration for European River Flooding. Risk Anal. doi:10.1111/risa.12747

Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Sharma, A., 2000a. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 - A strategy for system predictor identification. J. Hydrol. 239, 232–239. doi:10.1016/S0022-1694(00)00346-2

Sharma, A., 2000b. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 - A nonparametric probabilistic forecast model. J. Hydrol. 239, 249–258.

doi:10.1016/S0022-1694(00)00348-6

Sharma, A., Mehrotra, R., 2014. An information theoretic alternative to model a natural system using observational information alone. Water Resour. Res. 50, 650–660. doi:10.1002/2013WR013845

Sharma, A., Mehrotra, R., Li, J., Jha, S., 2016. A programming tool for nonparametric system prediction using Partial Informational Correlation and Partial Weights. Environ. Model. Softw. 83, 271–275. doi:10.1016/j.envsoft.2016.05.021

Shi, B., Wang, P., Jiang, J., Liu, R., 2018. Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. Sci. Total Environ. 610–611, 1390–1399. doi:10.1016/j.scitotenv.2017.08.232

Shoaib, M., Shamseldin, A.Y., Khan, S., Khan, M.M., Khan, Z.M., Sultan, T., Melville, B.W., 2018. A Comparative Study of Various Hybrid Wavelet Feedforward Neural Network Models for Runoff Forecasting. Water Resour. Manag. 32, 83–103. doi:10.1007/s11269-017-1796-1

Shoaib, M., Shamseldin, A.Y., Melville, B.W., 2014. Comparative study of different wavelet based neural network models for rainfall-runoff modeling. J. Hydrol. 515, 47–58. doi:10.1016/j.jhydrol.2014.04.055

Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2015. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. J. Hydrol. Eng. 20.

Šindelář, R., Babuška, R., 2004. Input selection for nonlinear regression models. IEEE Trans. Fuzzy Syst. 12, 688–696. doi:10.1109/TFUZZ.2004.834810

Siqueira, H., Boccato, L., Attux, R.R.F., Lyra, C., 2012. Echo State Networks and Extreme Learning Machines: A Comparative Study on Seasonal Streamflow Series Prediction, in: Huang, T., Zeng, Z., Li, C., Leung, C.S. (Eds.), Proc. of the 19th International Conference on Neural Information Processing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 491–500. doi:10.1007/978-3-642-34481-7{_}60

Solomatine, D.P., Dulal, K.N., 2003. Model trees as an alternative to neural networks in rainfall—runoff modelling. Hydrol. Sci. J. 48, 399–411. doi:10.1623/hysj.48.3.399.45291

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinformatics 10, 3–22. doi:10.2166/hydro.2008.015

Srivastav, R.K., Sudheer, K.P., Chaubey, I., 2007. A simplified approach to quantifying predictive and

parametric uncertainty in artificial neural network hydrologic models. Water Resour. Res. 43, 1–12. doi:10.1029/2006WR005352

Taormina, R., Galelli, S., Karakaya, G., Ahipasaoglu, S.D., 2016. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models. J. Hydrol. 542, 18–34. doi:10.1016/j.jhydrol.2016.07.045

Tapiador, F.J., Gallardo, C., 2006. Entropy-based member selection in a GCM ensemble forecasting. Geophys. Res. Lett. 33, 2–5. doi:10.1029/2005GL024888

Tiwari, M.K., Adamowski, J.F., 2017. An ensemble wavelet bootstrap machine learning approach to water demand forecasting: a case study in the city of Calgary, Canada. Urban Water J. 14, 185–201. doi:10.1080/1573062X.2015.1084011

Tiwari, M.K., Chatterjee, C., 2011. A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. J. Hydroinformatics 13, 500. doi:10.2166/hydro.2010.142

Tran, H.D., Muttil, N., Perera, B.J.C., 2015. Selection of significant input variables for time series forecasting. Environ. Model. Softw. 64, 156–163. doi:10.1016/j.envsoft.2014.11.018

Trichakis, I., Nikolos, I., Karatzas, G.P., 2011. Comparison of bootstrap confidence intervals for an ANN model of a karstic aquifer response. Hydrol. Process. 25, 2827–2836. doi:10.1002/hyp.8044

Tsimpiris, A., Vlachos, I., Kugiumtzis, D., 2012. Nearest neighbor estimate of conditional mutual information in feature selection. Expert Syst. Appl. 39, 12697–12708. doi:10.1016/j.eswa.2012.05.014

Tyralis, H., Koutsoyiannis, D., 2017. On the prediction of persistent processes using the output of deterministic models. Hydrol. Sci. J. 62, 2083–2102. doi:10.1080/02626667.2017.1361535

Unnikrishnan, P., Jothiprakash, V., 2018. Daily rainfall forecasting for one year in a single run using Singular Spectrum Analysis. J. Hydrol. 561, 609–621. doi:https://doi.org/10.1016/j.jhydrol.2018.04.032

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J. Hydrol. 476, 433–441. doi:https://doi.org/10.1016/j.jhydrol.2012.11.017

Van Hulle, M.M., 2005. Edgeworth Approximation of Multivariate Differential Entropy. Neural Comput. 17, 1903–1910. doi:10.1162/0899766054323026

Wang, S., Huang, G.H., Baetz, B.W., Ancell, B.C., 2017. Towards robust quantification and reduction of uncertainty in hydrologic predictions: Integration of particle Markov chain Monte Carlo and factorial polynomial chaos expansion. J. Hydrol. 548, 484–497. doi:10.1016/j.jhydrol.2017.03.027

Wang, Y., Wu, L., 2016. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. Energy 112, 208–220. doi:10.1016/j.energy.2016.06.075

Wang, Y., Zheng, T., Zhao, Y., Jiang, J., Wang, Y., Guo, L., Wang, P., 2013. Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. Environ. Sci. Pollut. Res. 20, 8909–8923. doi:10.1007/s11356-013-1874-8

Weijs, S. V., van de Giesen, N., 2013. An information-theoretical perspective on weighted ensemble forecasts. J. Hydrol. 498, 177–190. doi:10.1016/j.jhydrol.2013.06.033

Wu, T., Kareem, A., 2014. Simulation of nonlinear bridge aerodynamics: A sparse third-order Volterra model. J. Sound Vib. 333, 178–188. doi:10.1016/j.jsv.2013.09.003

Wu, Z.Y., El-Maghraby, M., Pathak, S., 2015. Applications of deep learning for smart water networks. Procedia Eng. 119, 479–485. doi:10.1016/j.proeng.2015.08.870

Xiong, T., Bao, Y., Hu, Z., 2014. Does restraining end effect matter in EMD-based modeling framework for time series prediction? Some experimental evidences. Neurocomputing 123, 174–184. doi:10.1016/j.neucom.2013.07.004

Xu, X., Zhang, X., Fang, H., Lai, R., Zhang, Y., Huang, L., Liu, X., 2017. A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach. Environ. Model. Softw. 88, 151–167. doi:10.1016/j.envsoft.2016.11.010

Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. J. Hydrol. doi:10.1016/j.jhydrol.2015.10.038

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-Shafie, A., 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. J. Hydrol. 542, 603–614. doi:10.1016/j.jhydrol.2016.09.035

Yung, B.B., Tolson, B.A., Burn, D.H., 2011. Risk assessment of a water supply system under climate variability: a stochastic approach. Can. J. Civ. Eng. 38, 252–262. doi:10.1139/L10-132

Zhang, H., Singh, V.P., Wang, B., Yu, Y., 2016. CEREF: A hybrid data-driven model for forecasting annual

streamflow from a socio-hydrological system. J. Hydrol. 540, 246–256. doi:10.1016/j.jhydrol.2016.06.029

Zhang, Q., Wang, B.-D., He, B., Peng, Y., Ren, M.-L., 2011. Singular Spectrum Analysis and ARIMA Hybrid Model for Annual Runoff Forecasting. Water Resour. Manag. 25, 2683–2703. doi:10.1007/s11269-011-9833-y

Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J. Hydrol. 530, 137–152. doi:10.1016/j.jhydrol.2015.09.047

Connecting Text to Chapter 3

This chapter describes the development, testing, and application of two novel computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods. The first method, Edgeworth Approximations-based condition mutual information (EA) provides a useful alternative to existing computationally expensive, parametric, nonlinear information-theoretic input variable selection methods and provides similar or better performance. The second method, bootstrap rank-ordered conditional mutual (broCMI) is an extension of the EA method that accounts for input variable selection uncertainty, a new topic in the input variable selection literature. The EA and broCMI approaches are useful for identifying which input (explanatory) variables to include in the development of nonlinear data-driven water resources forecasting models.

This chapter was published in Water Resources Research (Quilty et al., 2016). The format of the paper has been modified to ensure consistency with the style of this thesis. A list of references cited in this paper is available at the end of the chapter.

The author of the thesis was responsible for the development, testing, and application of the different methods and wrote the manuscript presented here. Prof. Adamowski, the supervisor of this thesis, provided valuable advice on all aspects of the research and contributed to the review and editing of the manuscript. Dr. Bahaa Khalil, Post-doctoral Researcher in the Bioresource Engineering Department at McGill and Dr. Maheswaran Rathinasamy, Humboldt Fellow at the Potsdam Institute for Climate Impact Research, Germany, provided advice on the methodological aspects and contributed to the review and editing of the manuscript.

Chapter 3: Bootstrap Rank-Ordered Conditional Mutual Information (broCMI): A Nonlinear Input Variable Selection Method for Water Resources Modeling

3.1. Abstract

The input variable selection problem has recently garnered much interest in the time series modeling community, especially within water resources applications, demonstrating that information theoretic (nonlinear) based input variable selection algorithms such as partial mutual information (PMI) selection (PMIS) provide an improved representation of the modeled process when compared to linear alternatives such as partial correlation input selection (PCIS). PMIS is a popular algorithm for water resources modeling problems considering nonlinear input variable selection; however, this method requires the specification of two nonlinear regression models, each with parametric settings that greatly influence the selected input variables. Other attempts to develop input variable selection methods using conditional mutual information (CMI) (an analogue to PMI) have been formulated under different parametric pretenses such as *k* nearest-neighbour (KNN) statistics or kernel density estimates (KDE). In this paper we introduce a new input variable selection method based on CMI that uses a non-parametric multivariate continuous probability estimator based on Edgeworth approximations (EA). We improve the EA method by considering the uncertainty in the input variable selection procedure by introducing a bootstrap resampling procedure that uses rank statistics to order the selected input sets; we name our proposed method bootstrap rank-ordered CMI (broCMI). We demonstrate the superior performance of broCMI when compared to CMI based alternatives (EA, KDE, and KNN), PMIS, and PCIS input variable selection algorithms on a set of seven synthetic test problems and a real-world urban water demand (UWD) forecasting experiment in Ottawa, Canada.

Keywords: input variable selection, conditional mutual information, uncertainty, bootstrap, rank-ordering, regression models

3.2. Introduction

Regression models such as multiple linear regression (MLR), artificial neural networks (ANN), and support vector regression (SVR) are ubiquitous across many scientific and technological applications seeking to quantify relationships existing between a process (response variable) and a set of potential drivers (candidate input variables or explanatory variables) that may contain relevant, irrelevant, and/or

redundant information. The task of the data modeller is twofold: 1) to extract only the relevant input variables and separate them from those candidate inputs that are irrelevant and/or redundant, also known as the input variable selection task and 2) to build regression models using input variables from the selection task to explain the process. Choosing an appropriate input variable selection method reduces the amount of variables used within the model and this helps *against* selecting irrelevant inputs that can significantly influence model accuracy and add unnecessary complexity impacting model reliability (Hejazi and Cai, 2009) while also increasing computational efficiency, minimizing redundancy, reducing noise, and increasing the interpretability of the model (Back and Trappenberg, 2001; May et al., 2008a; Šindelář and Babuška, 2004).

There are three general types of input variable selection algorithms: filter (the focus of this paper), wrapper, and embedded methods. Filters are the most commonly used input variable selection method since they may be applied independent of the regression model (e.g. ANN) while both wrappers and embedded methods rely on a model in order to choose the final set of input variables (May et al., 2011). The most common filter based input variable selection algorithm is the linear (and partial linear) correlation ranking method (May et al., 2011) which selects input variables one at a time, choosing the input variable that has the largest dependence score with the response variable, which has also been referred to as 'greedy selection' (Brown et al., 2012). We focus on filters in this work due to their independence from and their ability to be coupled with any type of regression model and direct the interested reader to Guyon and Elisseeff (2003) and Galelli et al. (2014) for further information on wrapper and embedded methods.

The goal of this paper is to propose and assess a new filter based input variable selection method that may easily be coupled with regression models (we consider nonlinear neural network based machine learning models) for specifying (water resources) time series/processes (e.g. urban water demand (UWD), streamflow, rainfall-runoff, etc.). Our proposed algorithm is derived using concepts from information theory; specifically we consider mutual information (MI) (Shannon, 1948) based metrics since they capture general (nonlinear) dependence existing between (multivariate) variables by using their joint and marginal probability distribution functions (PDF) to compute the measure. Thus, MI is an improvement over the linear correlation statistic due to its ability to estimate general (linear and nonlinear) dependencies amongst (potentially) multivariate variables. Due to these useful theoretical qualities MI based metrics are popular in diverse fields such as: biology (Penner et al., 2011), chemistry (Hnizdo et al., 2007), medical practice (Lee and Maslove, 2015), engineering (Nichols et al., 2006), finance (You et al.,

2015), psychology (Harel et al., 2007), image and video analysis (Korshunov and Ooi, 2011; Liu and Karam, 2005), geophysics (Chen et al., 2008), agriculture (Mishra et al., 2013), and water resources (Gong et al., 2013; Maier et al., 2006; Sharma, 2000a) (among many fields of application).

In an experiment focused on input variable selection, Guyon and Elisseeff (2003) discussed linear correlation coefficient ranking and mutual information ranking input variable selection approaches. Both linear and MI ranking methods list in order (from first to last) the candidate input variables with the largest dependence score with the response variable. To help guide efforts in selecting an effective input variable selection strategy Guyon and Elisseeff (2003) concluded three main points from their experiments that should be considered when choosing a pre-existing, or designing a new, input variable selection routine. Their findings are paraphrased as follows: 1) presumably redundant variables (variables that score similar dependence measures) can be utilized to reduce noise and facilitate better (potentially nonlinear) mappings between response and explanatory variables, 2) perfectly correlated variables are truly redundant and cannot provide additional information to an augmented input variable set by including one or more perfectly correlated variable(s). Very large (positive or negative) correlation does not necessarily mean the absence of variable complementarity. Therefore, input variable selection methods that score variables individually and independently of each other cannot determine which combination of variables would give the best performance, and 3) two variables that are useless by themselves can be useful together.

From the above three statements it is realized that for an input variable selection method to be effective at choosing the best hypothetical input variables it must consider variable redundancy and complementarity simultaneously (which requires a multivariate approach to calculate such dependencies). This means that standard linear correlation and MI ranking techniques are not sufficient for an optimal input variable selection routine as linear correlation can only assess linear dependence and it is not a conditional measure - MI is also not a conditional measure.

Since we are focused on the problem of selecting appropriate input variables when working with datasets containing candidate input variables that may be considered as either relevant, redundant, or irrelevant in accordance with some process (response variable) and other side information (explanatory variables deemed as important from an earlier study or through expert knowledge, for example) it is of practical significance to choose an evaluation metric which can define a candidate input variable as either relevant, redundant, or irrelevant. To this end, we are motivated by the recent work of Brown et al. (2012) who derived the conditional mutual information metric (CMI) (Cover and Thomas, 2006) under the objective

function of the conditional likelihood maximization between the response and explanatory variables. (We note that CMI has also been referred to as partial mutual information (PMI) elsewhere in the literature (Frenzel and Pompe, 2007; Sharma and Mehrotra, 2014).) The authors derived the CMI metric under this pretense, compared CMI to other heuristic CMI alternatives developed earlier in the machine learning community, used each technique for input variable selection, and then applied the selected input variables towards multiclass classification problems (see Brown et al. (2012) for details). The caveat of the CMI formulation in Brown et al. (2012) is that the conditional mutual information breaks down into the following terms: *informational relevancy* – *informational redundancy* + *informational conditional redundancy*. (This statement is embodied by equation (12) in Brown et al. (2012).) We can appreciate that the CMI metric (by its very nature) explicitly quantifies a candidate variable's relevancy and conditional relevancy - even if the candidate variable seems to contain redundant information with pre-selected inputs (side information) it may still be useful if it can add any further information to the system (i.e. the difference between conditional redundancy and redundancy terms). Thus, the CMI embodies the qualities of an objective (nonlinear) input variable selection routine as outlined by Guyon and Elisseeff (2003) and for this reason CMI is considered in this work for the input variable selection task.

Our important modification to the method proposed by Brown et al. (2012) is that we develop CMI for use with multivariate datasets containing continuous variables which are used in time series modeling scenarios (rather than multivariate categorical variables which are focused on in Brown et al. (2012)). This modification is made possible via Edgeworth approximations enabling the estimation of arbitrary continuous multivariate probability distribution functions (Van Hulle, 2005). The benefit of formulating CMI via Edgeworth approximations is that it does not require any parameters to be set as it is based on Hermitian polynomials and standardized cumulants (please refer to Van Hulle (2005) for details). Formulating MI through Edgeworth approximations has been shown to outperform other MI estimators such as those based on kernel density estimation and *k* nearest-neighbours for estimating MI for multivariate Gaussian and exponential distributions (Van Hulle, 2005) and has also been used to estimate the MI for short and noisy data (Khan et al., 2007); for example, measuring nonlinear dependencies between hydrological variables such as the river Nile and the El Nino-Southern Oscillation cycle (Khan et al., 2006). Extending the logic applied to MI estimation (Khan et al., 2007, 2006; Van Hulle, 2005), we use Edgeworth approximations to formulate a new CMI based input variable selection method which we name Edgeworth approximations (EA) for short.

The EA input variable selection method was developed to offer a competitive non-parametric alternative to current nonlinear CMI based alternatives that require the specification of parameters which must be carefully selected to guarantee an optimal input variable selection performance. Existing parametric CMI/PMI (i.e. information theoretic) based input variable selection methods that we focus on include the popular partial mutual information selection (PMIS) method (May et al., 2008b; Sharma, 2000a, 2000b; Sharma et al., 2000), the kernel density estimation CMI method (KDE) (used as the basis for the Partial Information modeling approach (Sharma and Mehrotra, 2014)), and the *k* nearest-neighbour CMI method (KNN) (Frenzel and Pompe, 2007; Tsimpiris et al., 2012; Vejmelka and Paluš, 2008).

The PMIS method was introduced in Sharma (2000a) (and subsequently explored in Sharma (2000b) and Sharma et al. (2000)). The PMIS algorithm estimates partial mutual information/CMI by calculating the MI between residual information from two intermediate regression models (that each require parametric settings); one of which estimates the residual information between the pre-selected input variables (or side information) and the response and the other which estimates the residual information between the pre-selected input variables and the candidate input variable (please see details within Sharma (2000a)). The two main components of the PMIS algorithm require specification of: 1) the method used to estimate the MI between the residual information (pre-calculated by the intermediate regression models) and 2) the (conditional) regression models. For instance, the intermediate regression models have been specified in the literature through multivariate kernel density estimates (He et al., 2011; May et al., 2008a, 2008b, Sharma, 2000a, 2000b; Sharma et al., 2000) and general regression neural networks (GRNN) (Bowden et al., 2005b, 2005a, Chen et al., 2013, 2014). In regards to MI estimation within PMIS, this has been calculated through (multivariate) kernel density estimation (Sharma, 2000a, Bowden et al., 2005a; May et al., 2008a); average shifted histograms (Fernando et al., 2009), and copula entropy (Chen et al., 2013, 2014). (Entropy is known as a measure of uncertainty for a given (univariate or multivariate) variable computed through the variables' PDF and it is through the notion of entropy that MI is derived (Singh, 2013); while copula entropy (AghaKouchak, 2014; Hao and Singh, 2015) is analogous with the negative of MI (please see details within Calsaverini and Vicente (2009); Ma and Sun (2011); Zeng and Durrani (2011).)

The KDE (based CMI) method, recently introduced as part of the Partial Information modeling approach (Sharma and Mehrotra, 2014), relies on multivariate kernel density estimation to compute the CMI metric, requiring specification of the kernel type and the kernel's parameters; a common choice of which is the Gaussian kernel and Gaussian Reference Rule (for the kernel bandwidth), respectively.

The KNN (based CMI) method, introduced in Frenzel and Pompe (2007), computes the CMI metric by using distances in nearest-neighbours space (Gómez-Herrero et al., 2015; Kraskov et al., 2004). This method requires the specification of the number of nearest-neighbours (k_{nn}) to estimate CMI, which is (in general) dataset dependent and must be explored through trial-and-error (Tsimpiris et al., 2012).

In addition to introducing the non-parametric EA method for calculation of the CMI, we further extend the EA method by acknowledging uncertainty within the CMI based input variable selection method by adopting a resampling procedure based on the bootstrap (Efron and Tibshirani, 1993), allowing us to generate multiple selected input variable sets. To make use of the information contained across the multiple selected input variable sets we incorporate an idea from Kuncheva (2007) and use the rank-ordering procedure to encompass the variability across the selected input variable sets and generate new input variable sets by ordering the selected input variables based on the variation in their selection ranks across the resamples (this procedure is explained further in section 3.3.3.2 where we define our proposed input variable selection method). Essentially, the rank-ordering procedure uses the (selection) ranks of the selected input variables across multiple trials (bootstrap resamples) and returns a single list of ordered ranks (for the selected input variables) based on the variability of the (input variable) ranks across the various trials. We have named our proposed input variable selection method: bootstrap rank-ordered conditional mutual information (broCMI).

Both the EA and broCMI algorithms incorporate all desired facets of an objective input variable selection algorithm as outlined by Guyon and Elisseeff (2003) and include the added benefit of being non-parametric; this is a very attractive quality as it mitigates the need to search for optimal parameterization during the input variable selection procedure, such as required for PMIS, KDE, and KNN (CMI based methods,) and avoids heuristic approaches, such as the combination of kernel density estimation and intermediate regression models as within PMIS (May et al., 2008a). Our proposed broCMI further improves upon these favorable characteristics (of the EA method) by additionally considering the uncertainty in the input variable selection procedure (through bootstrap resampling and rank-ordering of the resampled trials).

Both Elshorbagy et al. (2010) and Galelli et al. (2014) have noted that any prospective modeling tool that is introduced to the modeling community for a particular application should be compared with other existing methods to assess its added value. Therefore, we adopt, in addition to our EA method, four previously established input variable selection algorithms for comparison with our newly proposed broCMI algorithm. The three nonlinear CMI based methods described above (i.e. KDE, KNN, and PMIS)

are included in addition to a linear alternative, partial correlation input selection (PCIS) (May et al., 2008a; Tran et al., 2015). Thus, we explore and compare three parametric (KDE, KNN, and PMIS) and three non-parametric (PCIS, EA, and broCMI) input variable selection algorithms in this study. Each of these algorithms are discussed further in section 3.4 including any required parametric settings or modifications that we make to an original method. While we acknowledge other nonlinear input variable selection algorithms such as the Gamma Test (Goyal et al., 2014; Stefánsson et al., 1997; Wan Jaafar et al., 2011), the Iterative Input Selection (Galelli and Castelletti, 2013), or the Genetic Programming (Tran et al., 2015) methods we do not consider these algorithms in this work to preserve brevity and to focus on information theoretic based methods; however, these algorithms may be considered in future studies - readers interested in other input variable selection algorithms within water resources and environmental modeling may consult Maier et al. (2010), May et al. (2011), Galelli et al. (2014), and Tran et al. (2015).

To compare broCMI against EA, KDE, KNN, PMIS, and PCIS input variable selection algorithms we carried out two experiments in this study that are relevant in demonstrating any gains that may be achieved by using broCMI for input variable selection. Our first experiment uses a set of synthetic datasets developed for the sole purpose of evaluating new (and/or existing) input variable selection methods in a standardized manner using selection accuracy metrics that can identify the degree to which a model is correctly/incorrectly specified (Galelli et al., 2014). Our second experiment focuses on an urban water supply system whereby candidate input variables (historical urban water demand and meteorological variables) are selected (using EA, broCMI, and existing methods) and used in forecasting urban water demand at multiple lead times using neural network based machine learning models (specifically, ANN and the extreme learning machine (ELM) (Huang et al., 2006)) whereby forecast performance evaluation is used to identify the best predictor set selected by the various input variable selection algorithms. We have adopted neural network based models in this work for mapping explanatory variables into process forecasts as these (nonlinear) methods are very common in water resources forecasting applications (Abrahart et al., 2012; Yaseen et al., 2015). We study ELM as an alternative to the traditional feedforward backpropagation (FFBP)-ANN (which requires iterative training and does not guarantee a global solution) since it is a more computationally efficient method for solving the network parameters (i.e. via a global least squares solution, please see the online supplementary material [appendix] for further details) enabling the data modeller to test a wide variety of network architectures in a smaller amount of time (with similar or better performance) (Huang et al., 2012) which becomes very important for large datasets (Akusok et al., 2015). The interested reader may refer to our supplementary material [appendix] for theoretical background on ELM.

The rest of this paper is organized as follows: section 3.3 provides theoretical background used for developing our broCMI algorithm; section 3.4 briefly describes the various input variable selection algorithms (including any required parametric settings); section 3.5 outlines the design of our experiments used to demonstrate the usefulness of broCMI on benchmark and real-world datasets; section 3.6 provides discussion on the experimental results; and section 3.7 concludes our study by summarizing our main findings and highlighting opportunities for future research.

3.3. Theoretical Background

In this section we describe our proposed bootstrap rank-ordered conditional mutual information input variable selection algorithm by first highlighting information theoretic background used in its development.

3.3.1. Entropy, Mutual Information, and Conditional Mutual Information

When dealing with continuous variables one must apply differential entropy to estimate the entropy of a random variable (as explained in Shannon (1948)). Given an independent and identically distributed (i.i.d.) random variable $V \in \mathbb{R}^d$ with PDF $p(v)$ and set theoretic support $\mathcal{V}$ (which is dropped for simplicity in sub-sequent entropy and MI formulations) the (differential) entropy for continuous data is defined by (Cover and Thomas, 1991):

$$H(V) = -\int_{\mathcal{V}} p(v) \ln p(v)\, dv \tag{3.1}$$

When the natural logarithm is used, the units of measure for entropy are in nats (natural units). A simple extension from single variable entropy to joint variable entropy considering variables of arbitrary dimension can be calculated as follows (Cover and Thomas, 1991):

$$H(V_1, V_2, \dots, V_n) = -\int \cdots \int p(v_1, \dots, v_n) \ln p(v_1, \dots, v_n)\, dv_1 \dots dv_n \tag{3.2}$$

From here, one can begin to frame marginal and joint entropies as conditional entropy terms. The conditional entropy between two random variables can be calculated using the chain rule (Cover and Thomas, 1991):

$$H(V_1|V_2) = H(V_1, V_2) - H(V_2) \tag{3.3}$$

Given the above formulae one may now estimate MI and CMI quantities. MI uses the joint and marginal PDFs of the considered variables to measure dependence between such variables and can be calculated from joint and marginal entropy terms as follows (Cover and Thomas, 1991):

$$\begin{aligned} MI(V_1;V_2) &= H(V_1) + H(V_2) - H(V_1,V_2) \\ &= \int\int p(v_1,v_2)\ln\frac{p(v_1,v_2)}{p(v_1)p(v_2)}dv_1dv_2 \end{aligned} \tag{3.4}$$

By introducing a third variable one may calculate the MI between the first two variables conditioned on the third variable to estimate the CMI. The CMI using three variables can be calculated as follows (Tsimpiris et al., 2012):

$$CMI(V_1;V_2|V_3) = MI\big(V_1;(V_2,V_3)\big) - MI(V_1;V_3) \tag{3.5}$$

Since MI may be calculated through marginal and joint entropy terms, CMI can also be decomposed into joint and marginal entropy terms as follows (Frenzel and Pompe, 2007):

$$CMI(V_1;V_2|V_3) = H(V_1,V_3) + H(V_2,V_3) - H(V_3) - H(V_1,V_2,V_3) \tag{3.6}$$

(If V_3 contains no information then Eq. 3.6 will reduce to Eq. 3.4.) We use Eq. 3.6 (i.e. multivariate differential entropy terms) to estimate CMI for KDE, along with our EA and broCMI algorithms. The KNN algorithm uses a different approach (i.e. nearest-neighbour distances) to estimate CMI and we direct the interested reader to Frenzel and Pompe (2007) for further details.

Since the CMI and PMI metrics are equivalent (Sharma and Mehrotra, 2014) the EA, broCMI, KDE, KNN, and PMIS algorithms implicitly consider variable redundancy and complementarity. Recalling that the PMIS algorithm (within each referenced article) estimates PMI in a number of heuristic ways, our proposed EA and broCMI (along with KDE) algorithms are more natural in that there are no heuristics used: the joint and marginal entropy terms are estimated and then aggregated. This avoids involving heuristic techniques to account for redundancy and complementarity that is accounted for with nonlinear regression models in recent PMI approaches (Bowden et al., 2005a; May et al., 2008a; Chen et al., 2014).

If one considers the variable V_3 to contain conditional information (i.e. the pre-selected input variables) the PMI metric introduced by Sharma (2000a) may be calculated as follows:

$$PMI(V_1;V_2|V_3) = MI(V_1';V_2') = H(V_1') + H(V_2') - H(V_1',V_2') \tag{3.7}$$

where

$$V_1' = V_1 - E[V_1|V_3] \text{ and } V_2' = V_2 - E[V_2|V_3] \tag{3.8}$$

and $E[]$ is the expectation operator. (The conditional expectation terms are generally estimated using kernel regression (Sharma, 2000a) or general regression neural networks (Bowden et al., 2005a).) The analogy between PMI and partial correlation is observed by replacing MI in Eq. 3.7 by the linear (Pearson) correlation statistic and the conditional expectation terms in Eq. 3.8 by multiple linear regression (May et al., 2008a).

3.3.2. Input Variable Selection via Conditional/Partial Mutual Information

For the CMI based input variable selection algorithms we study, each follows the 'greedy' selection principle (Brown et al., 2012) (choosing the best input variable) one input variable at a time and may be summarized as follows (noting that for the PCIS approach, linear correlation replaces MI estimation and multiple linear regression replaces the conditional expectation as shown by Eqs. 3.7 3.8, respectively) (Tsimpiris et al., 2012): 1) among all input variables find the most relevant to the response variable by the MI metric and add the variable to the selected input variable set; 2) to find the next optimal input variable to be added to the selected input variable set, compute for all remaining candidate input variables the CMI/PMI metric considering the already selected input variable(s); 3) asses if the candidate variable should be added to the selected input variable set (based either on a specific threshold being satisfied or a predefined number of input variables having been selected); and 4) repeat steps 2 and 3 until the termination criterion is met (please refer to Figure 3.1(a)).

3.3.3. Input Variable Selection via Bootstrap Rank-Ordered Conditional Mutual Information

3.3.3.1. Bootstrap Rank-Ordered Conditional Mutual Information Overview

The proposed broCMI input variable selection algorithm is built on three main components: 1) resampling with replacement (i.e. bootstrapping) from a given input-output dataset (for a specified number of bootstrap resamples); 2) input variable selection using CMI for each bootstrap resampled input-output

data pair (i.e. Figure 3.1(a)); and 3) rank-ordering the selected input variable sets returned from input variable selection for each bootstrap resample. In Figure 3.1(a) we have illustrated the standard input variable selection procedure (adapted from Galelli et al. (2014)) along with our proposed modifications (bootstrap rank-ordering) to the standard procedure (Figure 3.2(b)). We should state that although we have developed our bootstrap rank-ordering algorithm for use with CMI (in particular using Edgeworth approximations of multivariate entropy) it may easily be extended to any other input variable selection algorithm returning the order of importance of the selected input variable set, such as PMIS or PCIS. In future studies we will explore the coupling of bootstrap rank-ordering with a variety of input variable selection algorithms.

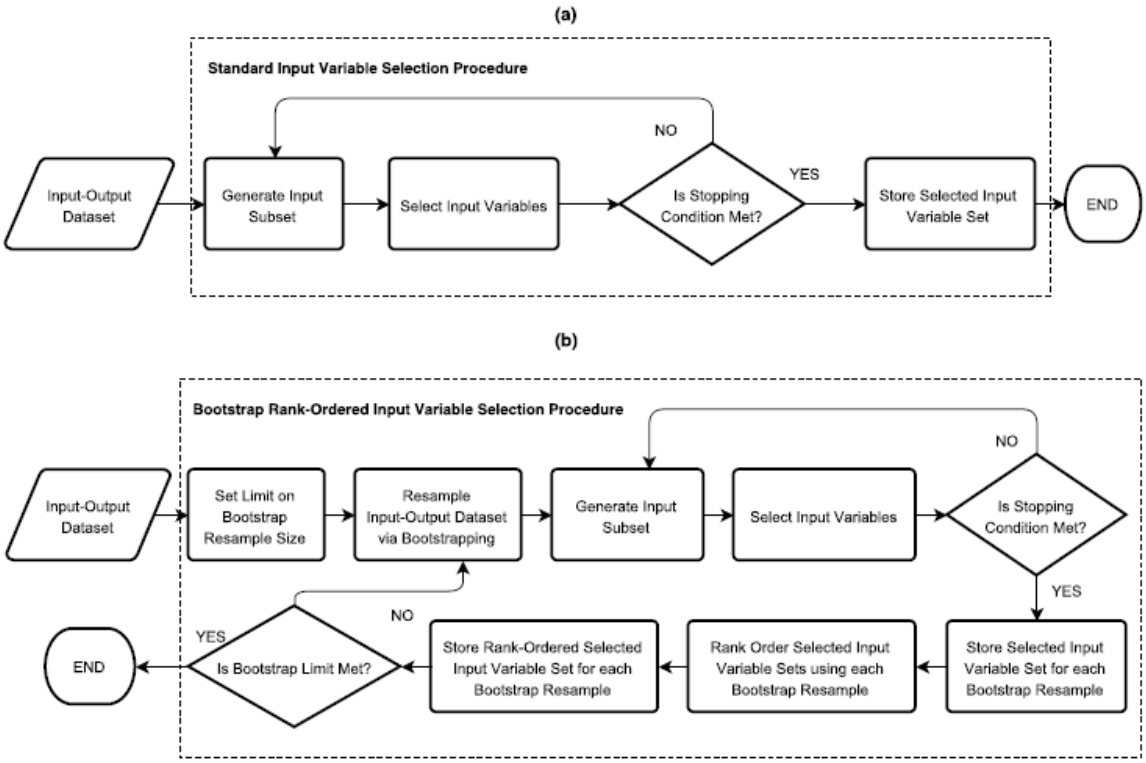

Figure 3.1. Input variable selection procedure for (a) the standard and (b) bootstrap rank-ordered approaches

If one opted to select all input variables in the candidate set then broCMI essentially orders the importance of each input variable based on its estimated dependency with the response variable (in addition to any previously selected input variables). In this regard, we describe two different methods that may be used to terminate and return information from the broCMI algorithm: 1) an exhaustive search through the whole input variable set and 2) a stopping/threshold-criterion. A stopping-criterion may be

used to define a significance threshold that must be met for a candidate input variable to be selected (else the algorithm is terminated and the inputs selected up to that point are returned) or one may simply request a particular number of inputs to be returned (i.e. the algorithm selects a specified number of inputs and ranks them by their order of importance), while the exhaustive search requires that each input variable has been selected by broCMI (i.e. ranking all candidate input variables by their order of importance).

3.3.3.2. Bootstrap Rank-Ordered Conditional Mutual Information Details

We begin defining broCMI by considering the input-output dataset: Y represents the response (target/output) variable and assumes an underlying i.i.d. process where X (explanatory variables/inputs) are mapped to Y using N observations. Each observation can be regarded as a pair (x, y) where $x \in \mathbb{R}^d$ represents the explanatory variable(s) vector (in *d*-dimensional space) and $y \in \mathbb{R}$ is the response scalar (target) drawn from the underlying random variables $X = \{X_1, \dots, X_d\}$ and Y. (Note: we do not discuss the case of information theoretic estimation for non i.i.d. processes as this is beyond the scope of this work, however, the recent (annual) Beyond i.i.d. in Information Theory (2015) conferences (and their proceedings), along with the November 2015 special issue (with the same title) in the peer-reviewed journal *Entropy*, may be of interest to the reader as this is an evolving area of research.)

We further develop broCMI as follows: introduce Ω a *d*-dimensional binary row vector used to keep track of which input variables have been selected up to step *i* $(1 \leq i \leq d)$ of the input variable selection procedure for each bootstrap resample. Then use x_Ω to indicate the vector of the selected input variables which is the vector x projected onto the dimensions of Ω. The complement $x_{\widetilde{\Omega}}$ indicates the unselected input variables, with the full input variable vector $x = \{x_\Omega, x_{\widetilde{\Omega}}\}$ (Brown et al., 2012). Defining (X_b, Y_b) as a particular bootstrap resample (b) from the input-output data pair (X, Y) we may now provide the CMI calculation that is used at step i $(1 \leq i \leq d)$ of the broCMI input variable selection procedure for the *c*-th $(1 \leq c \leq d)$ candidate input variable in $x_{\widetilde{\Omega}}$ (noting that the size of c is decreased by one each time i is incremented):

$$CMI\left(X^i_{b\,\widetilde{\Omega}_c}; Y_b | X^i_{b\,\Omega}\right) = H\left(X^i_{b\,\widetilde{\Omega}_c}, X^i_{b\,\Omega}\right) + H\left(Y_b, X^i_{b\,\Omega}\right) - H\left(X^i_{b\,\Omega}\right) - H\left(X^i_{b\,\widetilde{\Omega}_c}, Y_b, X^i_{b\,\Omega}\right) \qquad (3.9)$$

If broCMI is implemented with a stopping-criterion, then we may introduce the (*B*-dimensional) row vector θ which simply counts the number of input variables selected and returned by the CMI algorithm (i.e. the number of variables in x_Ω) for each bootstrap resample. (The variable θ is not necessary if

broCMI is setup as an exhaustive search.) The matrix $\eta \in \mathbb{Z}^{d\times B}$ can then be used to indicate the position in which each input variable is selected (i.e. once the stopping-criterion is met or each input variable has been selected) for each bootstrap resampled data pair. Depending on the two aforementioned methods (exhaustive search or stopping-criterion) in which broCMI may be used, the variable η can take on values differently. For the first case (exhaustive search) we illustrate the use of the position vector with an example: if one had four input variables $X = \{X_1, X_2, X_3, X_4\}$ and it was determined by a hypothetical input variable selection strategy that they should be ordered as $\{X_2, X_3, X_1, X_4\}$ then η would have the following values $\{3; 1; 2; 4\}$. For the second case (stopping-criterion) we need to specify two different scenarios that can potentially arise: 1) none of the input variables are selected (for a particular bootstrap resampled data pair) and 2) only some of the input variables are selected (for a particular bootstrap resampled data pair). If the case arises where an empty selected input variable set is returned (for a particular bootstrap resampled data pair) we simply ignore the instance (as is reasonable to do) and in the second case we rank (using the position vector) the unselected variables equally by assigning the "maximum" penalty to each unselected input (set as the dimension of the candidate input variable set, d) for that bootstrap resampled data pair. Assigning the maximum penalty (d) ensures that each unselected input variable is treated equally. Reusing our example above, consider the case where a hypothetical input variable selection algorithm (using a stopping-criterion) returns the selected input variables in the following order, $\{X_2, X_4\}$ (from the input dataset $X = \{X_1, X_2, X_3, X_4\}$) then η would take the values $\{4; 1; 4; 2\}$.

We now introduce the variable S which holds the selected input variable sets determined by CMI for each bootstrap resample (b). That is, S stores the results of the selected input variable sets from η_b (column b of η). The variable S contains all information necessary to perform the rank-ordering procedure mentioned above. The final variable to introduce is R and this variable holds the rank-ordered input variable sets. The variables S and R may take on two different forms depending on whether an exhaustive search or stopping-criterion strategy is employed when using broCMI: 1) in the exhaustive case ($S \in \mathbb{Z}^{d\times B}$ and $R \in \mathbb{Z}^{d\times B}$) R is determined simply by summing S across its columns for increasing bootstrap size (i.e. from $1{:}\,2, 1{:}\,b, \ldots, 1{:}\,B$) and sorting the result in ascending order and 2) if a stopping-criterion is used then we form S and R using the information contained in θ (a vector holding the selected input variable set size for each bootstrap resample). A row vector Ξ is then used to identify and sort in ascending order the unique selected input variable set sizes existing in θ. Thus, the variable Ξ can be no larger in dimension than θ (i.e. $\Xi \subseteq \theta$) and is said to be of size $N_\Xi \leq B$. The variable S is formed as a function of Ξ. Thus, $S(\Xi(j))$ gathers all selected input variable sets (from the bootstrap resamples) containing at least as many

selected input variables as the j^{th} ($j \in 1,2,\ldots,N_{\Xi}$) entry in Ξ and stores them for rank-ordering. The variable R is then formed as a function of Ξ (similar to S). Thus, $R(\Xi(j))$ is formed by rank-ordering $S(\Xi(j))$ across bootstrap resamples within $S(\Xi(j))$ (using the same approach as discussed for the exhaustive search scenario) for $j \in 1,2,\ldots,N_{\Xi}$. We provide pseudo-code for the broCMI algorithm considering both the exhaustive search (Table 3.1(a)) and stopping-criterion (Table 3.1(b)) scenarios in Table 3.1.

Table 3.1. Pseudo-code for bootstrap rank-ordered conditional mutual information (broCMI)

Table 3.1(a) – broCMI: exhaustive search scenario

1. Input: (X, Y) (input-output dataset) and B (bootstrap resample size); initialize: η (position matrix), S (selected input variable sets), and R (rank-ordered selected input variable sets)
2. **for** b = 1 to B (*resampling*)
3. Draw bootstrap resamples from $1{:}\,N$ creating resampled dataset $\{X_b, Y_b\}$
4. **for** i = 1 to *d* (*CMI estimation*)
5. **for** *c* = 1 to *d*-*i*+1
6. Compute $CMI\left(X^i_{b\widetilde{\Omega}_c}; Y_b | X^i_{b\Omega}\right)$ and retain the input variable X_c with the largest CMI
7. Determine η_b, S_b, and compute R_b (*rank-ordering*)
8. **return** S and R

Table 3.1(b) – broCMI: stopping-criterion scenario

1. Input: (X, Y) (input-output dataset) and B (bootstrap resample size); initialize: η (position matrix), S (selected input variable sets), and R (rank-ordered selected input variable sets)
2. **for** b = 1 to B (*resampling*)
3. Draw bootstrap resamples from $1{:}\,N$ creating resampled dataset $\{X_b, Y_b\}$
4. **for** i = 1 to *d* (*CMI estimation*)
5. **for** *c* = 1 to *d*-*i*+1
6. Compute $CMI\left(X^i_{b\widetilde{\Omega}_c}; Y_b | X^i_{b\Omega}\right)$ and retain the input variable X_c with the largest CMI
7. Determine η_b, θ_b, and Ξ

8. **for** j = 1 to N_{Ξ}

9. Determine $S(\Xi(j))$ and compute $R(\Xi(j))$ (*rank-ordering*)

10. **return** S and R

3.4. Input Variable Selection Algorithms

The six input variable selection algorithms we explore follow the procedures outlined in section 3.3.2 (EA, KDE, KNN, PMIS, and PCIS) and 3.3.3 (broCMI). We give a brief overview of each method in the sub-sections below, noting any parametric settings or modifications we make to an original algorithm; the stopping-criterion used to terminate each algorithm is also briefly described.

3.4.1. Partial Correlation Input Selection

Partial correlation input selection involves the use of partial correlation measures when assessing the dependence of a candidate input variable with a response variable by conditioning on pre-selected input variables and is performed through multiple linear regressions as described in May et al. (2008a). We developed the PCIS code in Matlab using custom scripts.

3.4.2. Partial Mutual Information Selection

The PMI selection algorithm is analogous to PCIS but instead of using linear correlation to assess dependence between the candidate input variable and the response, mutual information is used; instead of multiple linear regression to account for the conditional dependence of the pre-selected input variable(s) on the response and candidate input variable, nonlinear regressions are used (e.g. kernel regressions) (Sharma, 2000a). Thus, PMIS can be seen as a type of nonlinear analogue of PCIS. The PMIS method uses the PMI calculation described by Eq. 3.7. We have opted to use general regression neural networks within PMIS when accounting for (nonlinear) conditional dependencies as this method was found by the present authors to be more computationally efficient when compared to kernel regression. We used Matlab to develop GRNN models which automatically tuned the network's smoothing parameter (the only required parameter in GRNN). Although evaluating multiple criteria for the selection of the optimal smoothing parameter for the GRNN is a worthwhile consideration, it is outside the scope of this work; however, a guideline for the selection of optimal smoothing parameter selection for GRNN is provided in Li et al. (2014). To estimate the PMI of a candidate input variable one also needs to specify the method for computing MI between the candidate input and the response (after removing conditional

dependencies of pre-selected input variables via GRNNs) and we do so using kernel density estimation using multivariate Gaussian distributions and adopting the Gaussian Reference Rule for bandwidth selection (Moon et al., 1995; Sharma, 2000a) as is done in May et al. (2008a). The interested reader wishing to explore other methods for selecting the kernel bandwidth may consult Harrold et al. (2001) and Li et al. (2015a, 2015b) for such considerations. We developed the PMIS code in Matlab using custom scripts.

3.4.3. *k* Nearest-Neighbours based Conditional Mutual Information

The third input variable selection method that we considered is based on CMI derived through *k* nearest-neighbours estimation (Kraskov et al., 2004; Tsimpiris et al., 2012) which was recently programmed into the Information Theoretic Measures (TIM) Matlab toolbox (Gomez-Herrero et al., 2015). We use the TIM Matlab toolbox to compute this version of CMI (http://www.cs.tut.fi/~timhome/tim/tim.htm) and develop custom Matlab scripts to carry out input variable selection using this technique. The KNN method requires identification of the number of nearest-neighbours k_{nn} to use in the algorithm. Kraskov et al. (2004) recommend using between two and four nearest-neighbours as a guideline. We explored a slightly larger range of 1-15 nearest-neighbours based on an interest in assessing different input variable sets that may be selected by choosing a larger number of nearest-neighbours than usually necessary.

3.4.4. Kernel Density Estimation based Conditional Mutual Information

The fourth input variable selection method we consider is a slight modification of the Partial Information approach recently introduced in Sharma and Mehrotra (2014). Essentially Sharma and Mehrotra (2014) formulate a CMI based input variable selection (and modelling) algorithm using multivariate kernel density estimation and convert the CMI dependence measure on to a 0.0 - 1.0 scale, transforming the CMI statistic into the Partial Information statistic, using the partial informational correlation identity with CMI. They formulate the CMI statistic through multivariate MI terms (please refer to equation 4 in Sharma and Mehrotra (2014)) – our slight modification is that we estimated CMI using multivariate entropy terms instead of multivariate MI terms (please refer to Eq. 3.6). Like Sharma and Mehrotra (2014) we specify the multivariate kernel density estimates to take the form of multivariate Gaussian kernels and adopted the Gaussian Reference Rule for kernel bandwidth selection (Moon et al., 1995). The interested reader is directed to Li et al. (2015a, 2015b) to provide useful information regarding different approaches for specifying the multivariate kernel density estimation within the KDE algorithm. We developed the KDE method in Matlab using custom scripts.

3.4.5. Edgeworth Approximations based Conditional Mutual Information

We derived CMI for the fifth input variable selection algorithm through Edgeworth approximations of multivariate differential entropy (Van Hulle, 2005). For this method, CMI is calculated using multivariate entropy terms as in Eq. 3.6. However, unlike the PMIS, KNN, and KDE methods, EA is non-parametric. We developed the EA method using the Information Theoretic Estimators (ITE) Matlab toolbox described in Szabó et al. (2012, 2007) and Szabó (2014) along with implementing custom scripts in Matlab.

3.4.6. Bootstrap Rank-Ordered Conditional Mutual Information

The final input variable selection method used in this work is broCMI, the main contribution of this paper. The broCMI algorithm (as previously mentioned) uses the EA approach for determining which input variables should be selected for each bootstrap resample, thus broCMI is also non-parametric. One only need specify the number of bootstrap resamples to draw when using broCMI. The selection of an appropriate bootstrap resample size is discussed further in section 3.5.1.3. An advantage of using the broCMI method is that there are multiple input variable sets that are developed due to the resampling procedure. Thus, (for this study) each input variable set produced by broCMI during the rank-ordering step was considered when searching for an optimal input variable set for a given dataset.

The broCMI algorithm was developed in Matlab using custom scripts (along with the requirement of the ITE Matlab toolbox (Szabo, 2014) for the estimation of CMI via the EA method).

3.4.7. Stopping-Criterion for Input Variable Selection

Current research has shown that the stopping-criterion approach for input variable selection is the most popular in water resources modeling applications (see for instance, May et al. (2008a), Fernando et al. (2009), He et al. (2011), Galelli et al. (2014), and Chen et al. (2014)) and is adopted in this work to facilitate meaningful comparisons between our proposed input variable selection methods (EA and broCMI) and existing benchmark methods (e.g. PMIS, PCIS, etc.). To permit a fair comparison between input variable selection methods we adopted a single stopping-criterion (i.e. the Hampel test (May et al., 2008a)) and applied it to each method (EA, KDE, KNN, PMIS, PCIS, and broCMI). We believe that this is the most objective way in which we could have assessed the various input variable selection algorithms when using a stopping-criterion; however, other authors (e.g. Galelli et al. (2014)) have sometimes adopted different stopping-criteria for different algorithms. Again, we avoided this situation because we find it rather subjective. Some of the other stopping-criteria used for input variable selection within water resources modeling studies include the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), bootstrap tests, outlier detection methods, tabulated critical values, and t-tests based on partial weights

(among others) (Sharma, 2000a; May et al., 2008a; He et al., 2011; Galelli et al., 2014; Sharma and Mehrotra, 2014; Li et al., 2015a; Chen et al., 2014). In this study we have adopted the computationally efficient Hampel test stopping-criterion because it is a robust outlier detection method that is simple to implement and has been used in other studies focused on input variable selection for water resources modeling with success (May et al., 2008a; Fernando et al., 2009; Chen et al., 2014). The Hampel test is an efficient stopping-criterion because, at a particular selection step, it relies only on comparing the CMI for each candidate input variable and measuring its distance from the median CMI value (of all considered candidate input variables) and then determines if the Hampel distance for each candidate is greater than a significance threshold (generally chosen as 3 (which is also used in this study), identifying the variable as an outlier) – out of those candidate input variables meeting the significance threshold, the variable with the largest Hampel distance is added to the selected input variable set. We refer the interested reader to May et al. (2008a) for further details on input variable selection stopping-criteria (including the Hampel test criterion).

3.5. Experimental Set-Up

In this section we discuss details regarding the two input variable selection experiments that were designed to compare broCMI with other relevant input variable selection algorithms (EA, KDE, KNN, PMIS, and PCIS).

3.5.1. Experiment I – Input Variable Selection Comparisons using Synthetic Datasets

3.5.1.1. Selection Accuracy Evaluation Metrics for Input Variable Selection

In this study the input variable selection accuracy metrics developed in Galelli et al. (2014) are adopted for measuring the accuracy of each considered input variable selection algorithm using the synthetic datasets that are described in section 3.5.1.2. We give a very brief synopsis of these accuracy metrics and recommend the interested reader to review Galelli et al. (2014) for further details. The selection accuracy (SA) score measures the similarity between the selected input subset produced by a particular input variable selection algorithm and the true input subset (which must be known *a priori*) and is given as follows (Galelli et al., 2014):

$$SA = \gamma \frac{k}{K} + (1-\gamma)\left(1 - \frac{p}{P-K}\right) \tag{3.10}$$

where P is the total number of candidate input variables, K is the total number of relevant input variables, k is the number of relevant inputs selected by a given algorithm, p is the number of extraneous inputs selected, and γ is a weight (ranging between 0.0 and 1.0) that penalizes the selection of irrelevant inputs against those that are correctly selected. The SA metric can take values between 0.0 and 1.0: when $SA = 1.0$ the model has been correctly specified and when $SA = 0.0$ the model is completely mis-specified by only selecting extraneous inputs (Galelli et al., 2014). The authors (Galelli et al., 2014) recommend that γ be set such that it reflects the general preference for selecting some irrelevant variables if it allows the selection of the correct model inputs. That is to say, the weighting factor should be set such that an input variable selection algorithm that selects the correct inputs in addition to several irrelevant inputs should be (to a certain degree) preferable over an algorithm that can only select (for example) one or two relevant variables (from a list of many) and that does not include any irrelevant variables. Through a theoretical example Galelli et al. (2014) found $\gamma = 0.7$ to provide a suitable weighting strategy (please see (Galelli et al., 2014) for more details). We also adopt this weight when calculating the SA score in our work.

The SA score can further be divided into two terms describing the proportion of correct inputs that have been selected (SAc) and the proportion of extraneous inputs that have been selected (SAe) – the two terms are given as follows (Galelli et al., 2014):

$$SAc = \frac{k}{K} \tag{3.11}$$

$$SAe = 1 - \frac{p}{P - K} \tag{3.12}$$

where both terms take on values between 0.0 and 1.0 – values closer to 1.0 are indicative of a better model for the considered dataset.

3.5.1.2. Synthetic Test Problems

The seven synthetic datasets used in our first experiment comparing different input variable selection algorithms using selection accuracy evaluation metrics were selected from an online database with detailed dataset descriptions (http://ivs4em.deib.polimi.it) and which were discussed and utilized in the recent study of Galelli et al. (2014). These datasets are part of the input variable selection for environmental modeling (IVS4EM) project focused on introducing a new evaluation framework for input

variable selection methods using selection accuracy evaluation metrics (e.g. SA, SAc, and SAe) and qualitative criteria (e.g. computational feasibility, ease of use, etc.). We will only give a very brief account of the seven synthetic datasets (detailed in Table 3.2) that were used in our aforementioned comparison of input variable selection algorithms and direct the reader to Galelli et al. (2014) (and their online supplementary material [appendix] mentioned therein) for further information. The datasets that we consider as part of our first experiment have been used in other input variable selection studies within water resources modeling (see for instance, (Sharma, 2000a; Bowden et al., 2005a; May et al., 2008a; Fernando et al., 2009; Galelli and Castelletti, 2013)) and for these reasons the particular datasets considered in this work were adopted. The most valuable characteristic of these datasets is that there exists 30 different instances of each dataset generated through random sampling, allowing for the statistical assessment of selection accuracy results (e.g. SA, SAc, and SAe). The various datasets are described below – AR is taken as autoregressive; TAR as threshold-autoregressive; and NL as nonlinear.

Table 3.2. Synthetic datasets used for comparing input variable selection algorithms

Dataset	No. Observations	*K*	*P*	Fully/Partially Synthetic	Nonlinear	High Noise	High Collinearity
AR1	500	1	15	Fully		X	X
AR9	500	3	15	Fully		X	X
TAR1	500	1	15	Fully	X	X	X
TAR2	500	2	15	Fully	X	X	X
NL	500	3	15	Fully	X		
NL2	500	3	15	Fully	X	X	X
Kentucky	4739	4	21	Partially	X		X

3.5.1.2.1. Linear and Nonlinear Autoregressive Datasets

We utilized the AR1, AR9, TAR1, and TAR2 datasets (Sharma, 2000a; Hejazi and Cai, 2009; Sharma and Mehrotra, 2014; Galelli et al., 2014; Chen et al., 2015a) to examine how each input variable selection method performs in terms of high noise and collinearity (linear correlation between variable pairs considered to be larger than 0.7 for the purposes of this study) situations that are both linear (AR) and nonlinear (TAR) in nature. The datasets may be described as follows:

$$AR1: x_t = 0.9x_{t-1} + 0.866\epsilon_t \tag{3.13}$$

$$AR9: x_t = 0.3x_{t-1} - 0.6x_{t-4} - 0.5x_{t-9} + \epsilon_t \quad (3.14)$$

$$TAR1: x_t = \begin{cases} -0.9x_{t-3} + 0.1\epsilon_t & \text{if } x_{t-3} \leq 0 \\ 0.4x_{t-3} + 0.1\epsilon_t & \text{if } x_{t-3} > 0 \end{cases} \quad (3.15)$$

$$TAR2: x_t = \begin{cases} -0.5x_{t-6} + 0.5x_{t-10} + 0.1\epsilon_t & \text{if } x_{t-6} \leq 0 \\ 0.8x_{t-10} + 0.1\epsilon_t & \text{if } x_{t-6} > 0 \end{cases} \quad (3.16)$$

where ϵ_t is the noise component in the model (i.e. random Gaussian noise with zero mean and unit standard deviation). The number of relevant variables for each dataset are given in Table 3.2, however it is clear for each model which input variables are relevant to the response. A total of 15 candidate input variables (e.g. $x_{t-1}, x_{t-2}, \ldots, x_{t-15}$) were generated and there were 500 observations for each of the autoregressive models.

3.5.1.2.2 Nonlinear Datasets with Exogenous Covariates

The nonlinear datasets with exogenous covariates considered in this work were discussed and utilized in Bowden et al. (2005a) and are given as follows:

$$y = x_2^2 + \cos(x_6) + 0.35\sin(x_9) + s\epsilon \quad (3.17)$$

where s is a scaling factor that may be used to alter the noise level ϵ in the output variable. Again, there were 500 observations available and 15 candidate input variables were randomly sampled from x, which was drawn from a Gaussian distribution with zero mean and predefined covariance matrix $\sum$. For the NL dataset, s was set equal to zero and the covariance matrix set to the identity matrix. Therefore each input variable was sampled independently of one another (e.g. no collinearity was present amongst input variables). For the NL2 dataset s was set equal to one, introducing a high level of noise in the response variable. The covariance matrix was set such that 25 pairs of inputs were highly correlated (i.e. linear correlation > 0.7).

3.5.1.2.3 Kentucky River Basin Dataset

The Kentucky (River basin) dataset originates from the work of Jain and Srinivasulu (2006) whereby the data are derived from real-world rainfall-runoff data in the basin with an output (response) variable synthetically generated from an ANN model (hence the identification of this dataset in Table 3.2 as

partially synthetic). The dataset contains 26 years (1960-1972 and 1977-1989) of average daily streamflow data (Q_t) from the Kentucky River and includes an effective rainfall variable (Er) that incorporates rainfall data from five gauges within the basin. The average daily streamflow at time step t is set as the response variable and the candidate input (explanatory) variables include the 10 previous time steps of the average daily stream flow and the effective rainfall at time step t, including the 10 previous time steps. Thus, there were a total of 21 candidate input variables ($Q_{t-1}, Q_{t-2}, \dots, Q_{t-10}, Er_t, Er_{t-1}, \dots, Er_{t-10}$) and a total of 4739 observations (Jain and Srinivasulu, 2006; Galelli et al., 2014). The model used to generate the synthetic flow data is as follows (Galelli et al., 2014):

$$Q_t = f(Q_{t-1}, Q_{t-2}, Er_t, Er_{t-1}) + \epsilon \tag{3.18}$$

where $f(\cdot)$ represents a single hidden layer, four hidden node, fully connected FFBP-ANN. The 30 replicates of this dataset were generated by resampling the noise term ϵ (Galelli et al., 2014).

3.5.1.3. Procedure for Comparing Input Variable Selection Algorithms on Synthetic Datasets

For the first experiment each of the six input variable selection methods (EA, KDE, KNN, PMIS, PCIS, and broCMI) are compared against one another using selection accuracy metrics (SA, SAc, SAe) for the synthetic input variable selection datasets mentioned in the last sub-section. We used 30 different replications for each of the seven synthetic datasets (AR1, AR9, TAR1, TAR2, NL, NL2, and Kentucky) that were generated through random sampling (Galelli et al., 2014). These synthetic datasets share the commonality that each bears a specific input-output relationship and there are both relevant and exogenous (redundant or irrelevant) input variables contained within the input variable sets (allowing quantitative assessment using the selection accuracy evaluation metrics discussed in section 3.5.1.1). We performed input variable selection using each algorithm for all 30 replications for each of the seven synthetic datasets. After obtaining selection accuracy results for the 30 different instances for each dataset we computed the mean and interquartile range (IQR) of the selection accuracy metrics to enable a graphical comparison between each input variable selection method.

Since this is the first investigation evaluating the abilities of broCMI for input variable selection, we chose four different bootstrap resample sizes (5, 25, 50, and 100) for evaluation and for attempting to identify an optimal (or at least an appropriate) bootstrap resample size. These four resample sizes were adopted based on computational resources. This miniature experiment was considered to help identify an appropriate amount of bootstrap resamples that are needed to improve selection accuracy results. The

bootstrap resample size that provided the optimal selection accuracy was then used in our second experiment where we mapped input variables selected by the various input variable selection algorithms into urban water demand forecasts using machine learning methods (e.g. ELM).

3.5.2. Experiment II – Input Variable Selection Applied in the Context of Urban Water Demand Forecasting

3.5.2.1. Study Site and Datasets

The site location of the urban water supply system used in our second experiment focused on forecasting urban water demand (using inputs selected by the various input variable selection algorithms) is located at 45º19'N 75º40'W in Ottawa, the capital of Canada. The urban water supply system serves the City of Ottawa with nearly 871, 000 residents (City of Ottawa, 2013a). The urban water supply system is divided into pressure zones segregated throughout the City of Ottawa. The specific pressure zones used in this study are the 3W and Morgan's Grant (MG) pressure zones. The amount of estimated persons residing in 3W and MG pressure zones are 83, 615 and 4196, respectively (City of Ottawa, 2013b). The 3W pressure zone is subject to demands of residential, commercial, and industrial uses while the MG pressure zone is solely comprised of residential customers (City of Ottawa, 2009). Ottawa is 114 meters above sea level and has a humid continental climate that has an average annual daily air temperature of 6.0 °C; in July the average air temperature is 20.9 °C, and in January the average air temperature is -10.8 °C. The average annual rainfall amounts to 732.0 millimeters (City of Ottawa, 2013c). This study focuses on the summer demand period (since the winter demand period is insensitive to rainfall and air temperature due to the cool climate and very limited outdoor water use) which covers May 1 to September 30 (a total of 153 days for each summer demand period), inclusively.

In this study the response variables that were forecasted were average daily UWD for the 3W and MG pressure zones at lead times of 1 and 3 day(s) ahead between the dates of May 1, 1999 to September 30, 2012, and September 26, 2003 (since data within this pressure zone was unavailable for May 1 – September 25, 2003) to September 30, 2012, respectively. The lead times were chosen by the water utility (City of Ottawa) as useful lead times for planning water main maintenance during summer months. The water utility also considers daily UWD forecasts valuable since the hourly demand pattern scales to both average UWD; this can be used to define which hours of the day maintenance work should be initiated/completed.

In the rest of this work the following short forms are assigned to the pressure zone water demand time series that are forecasted (all measurements are in megalitres per day (ML/D)): average daily water demand for the 3W pressure zone (*a3w*) and average daily water demand for the MG pressure zone (*amg*).

Based on previous studies by one of the contributing authors (Adamowski et al., 2013, 2012; Adamowski, 2008; Karapataki, 2010; Tiwari and Adamowski, 2013, 2015) previous days' UWD, rainfall (*rain* (mm/day)), and maximum air temperature (*mat* (°C)) time series have all been reported as important explanatory variables to consider in building UWD forecast models and as such have been included in this study, too. Furthermore, an alternative variable that can potentially provide further explanatory information to an UWD forecast model is the daily Antecedent Precipitation Index (*api* (mm)) which was successfully used in the UWD models developed in Wong et al. (2010) and Cresswell and Naser (2013), and is also considered as a soil moisture index (Anctil et al., 2004) (which may influence lawn-watering, an influential factor in urban water demand during the summer demand period in Ottawa). Therefore, *api* was included as a potential explanatory variable in this study. All UWD and meteorological data used in this experiment were made available to the authors by the City of Ottawa (with the exception of *api* which was calculated from the daily rainfall data). Any dates that contained missing UWD or meteorological data were excluded from our analysis which amounted to 8% for *a3w*, 5% for *p3w*, 6% for *amg*, and 1% for *pmg*.

Descriptive statistics for each time series variable considered in this experiment are presented in Table 3.3 below while each UWD time series is presented in Figure 3.2. The procedure for comparing the input variable selection algorithms for this experiment (including details on the number of observations and the candidate input variable sets for each UWD dataset) are described in the sub-section below.

Table 3.3. Descriptive statistics for the time series used in Experiment II

Time Series	Units of Measure	Mean	Variance	Minimum	Maximum	Median
a3w	ML/D	27.730	30.409	11.889	51.387	26.421
amg		0.666	0.096	0.197	1.593	0.640
api	mm/D	19.168	196.553	0.345	136.367	15.412
rain	mm/D	2.937	47.459	0.000	108.600	0.000
mat	°C	23.722	27.534	4.600	37.300	24.300

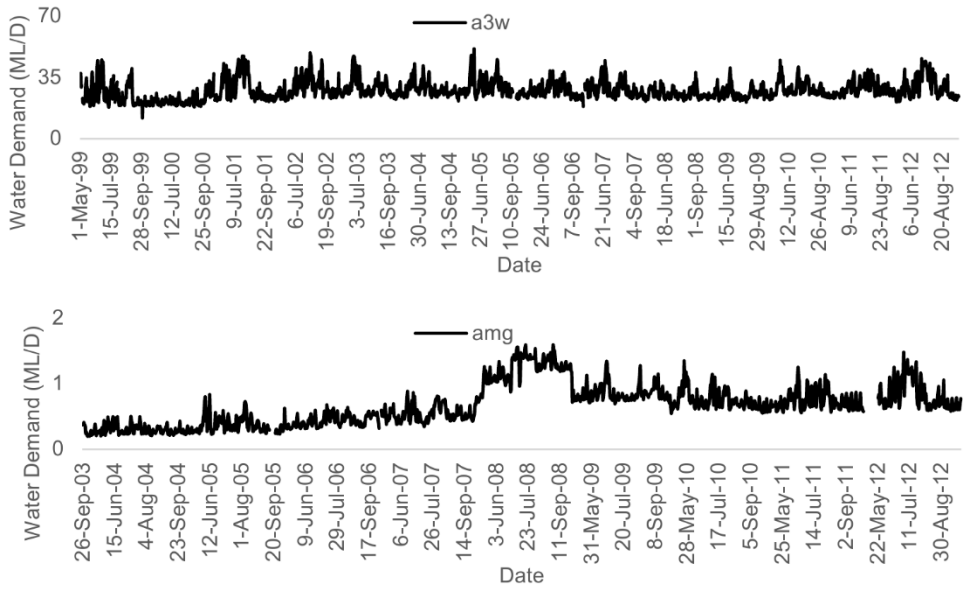

Figure 3.2. Time series plots for the urban water demand time series used in Experiment II

3.5.2.2. Procedure for Comparing Input Variable Selection Algorithms on the Urban Water Demand Forecasting Datasets

In our second experiment we explored the use of the six input variable selection algorithms (EA, KDE, KNN, PMIS, PCIS, and broCMI) for selecting relevant variables that were subsequently used in forecasting urban water demand time series from Ottawa, Canada at lead times of 1 and 3 day(s) ahead. We forecasted average UWD for the 3W and Morgan's Grant pressure zones within the City of Ottawa's water utility boundaries. Since the true function underlying the urban water demand processes in these pressure zones is unknown, it is impossible to identify which input variables are correct in specifying the system so one must produce forecasts using a suitable strategy (e.g. machine learning) and assess their forecasts using appropriate evaluation metrics in order to judge which input variable selection algorithms provided the best input variables. Thus, in order to carry out this experiment five basic steps were undertaken: 1) generation of candidate input variable sets (from potential explanatory variables); 2) dataset partitioning; 3) input variable selection using each algorithm; 4) mapping the selected input variable sets to forecasts using machine learning methods; and 5) evaluating the forecasts to identify the best input variable selection algorithm(s). Explanations regarding each of these steps have been placed in the online supplementary material [appendix] due to space limitations. The response and candidate input variable sets (along with the total number of available observations) used for input variable selection and UWD forecasting are provided in Table 3.4 and were generated through phase space reconstruction using the

time delay embedding window approach (Kim et al., 1999) as mentioned earlier (please see details within the online supplementary material [appendix]).

Table 3.4. Response variables and candidate input variable sets for Experiment II

Response (Target) Variables	Candidate Input Variable Set			
	Urban Water Demand Explanatory Variables	Meteorological Explanatory Variables	Number of Candidate Input Variables	Number of Available Observations
a3w(*t*+1, *t*+3)	*a3w*(*t*, *t*-1, …, *t*-15)	[*api*(*t*, *t*-1), *rain*(*t*, *t*-1, …, *t*-8), *mat*(*t*, *t*-1, …, *t*-21)]	49	1949
amg(*t*+1, *t*+3)	*amg*(*t*, *t*-1, …, *t*-19)		53	1260

The important idea to maintain when considering this case study is the identification of the best selected input variable set is solely based on the fitness (performance) of the machine learning model using such inputs; whereby each dataset was split into training (calibration of model parameters), cross-validation (model selection), and testing (independent evaluation) sets (Hastie et al., 2009) (please see details within the online supplementary material [appendix]). Different fitness functions may lead to different optimal selected input variable sets; the selection of the fitness function should always be chosen to meet the objective of the modeller. Here we choose to minimize the sum of squared errors by adopting the root mean square error (RMSE) as the fitness function (a popular choice in water resources modeling); however, the maximization of information theoretic terms (such as the mutual information between forecasts and observations) may be useful in different modeling problems such as the quantification of process modeling uncertainty (Gong et al., 2013). We believe it necessary to stress this point to benefit the interested reader in their own explorations of input variable selection algorithm comparisons on (real-world) time series modeling problems.

In the next section we provide the results of both experiments and provide relevant discussion regarding our findings.

3.6. Results and Discussion

3.6.1. Experiment I – Synthetic Datasets

In this sub-section we present our comparative results obtained from the EA, KDE, KNN, PMIS, PCIS, and broCMI algorithms on the synthetic input variable selection datasets and provide discussion regarding the quantitative (selection accuracy and computation time) and qualitative (ease of use and robustness,

explanation capability, and flexibility) performance of each method using the input variable selection evaluation procedures recently outlined in Galelli et al. (2014). We follow the guidelines for quantitative and qualitative input variable selection evaluation provided in Galelli et al. (2014) as we believe their framework to be the most robust protocol currently available for the comparison of input variable selection algorithms.

3.6.1.1. Quantitative Performance Comparisons across Input Variable Selection Algorithms

3.6.1.1.1. Selection Accuracy

The selection accuracy (SA, SAc, SAe) results for each input variable algorithm is provided in Figure 3.3. The following discussion covers the key results related to each synthetic dataset:

3.6.1.1.1.1. Linear and Nonlinear Autoregressive Datasets

The autoregressive datasets are marked by their high degrees of collinearity and noise. Interestingly, the only algorithm that performed poorly on the linear autoregressive datasets was the PCIS algorithm. The reason for the poor SA score is related to the Hampel test criterion providing too strict of a tolerance (as reflected in the SAc scores) in combination with high noise – with reduced noise on the nonlinear autoregressive datasets, performance increased for PCIS across each selection accuracy metric, even outperforming some of the nonlinear methods such as PMIS and KNN. It is clearly seen that broCMI outperforms the EA method (indicating that the bootstrap rank-ordering procedure (Figure 3.1(b)) is useful for increasing selection accuracy) with as few as 5 bootstrap resamples – with increasing bootstrap resample size the performance of broCMI consistently increased. It is seen that for this class of datasets broCMI and the KDE method are the most robust and accurate methods as reflected by their SA, SAc, SAe, and IQR scores. The KNN method exhibits the largest fluctuation in selection accuracy across the nonlinear algorithms - we explored different k_{nn} values ranging from 1 to 15 - it is of course possible that exploring a larger range of values may lead to improved performance for this algorithm, however such considerations are reserved for future research.

3.6.1.1.1.2. Nonlinear Datasets with Exogenous Covariates

The nonlinear datasets with exogenous covariates were specified for two cases: noiseless (NL) and noisy (NL2). It is clear that the linear PCIS algorithm performed very poorly for these datasets, which is more dramatically pronounced for the noisy NL2 dataset. The EA method is the worst-performing nonlinear method for the NL dataset but performs slightly better than PMIS for the NL2 dataset. In each case the broCMI algorithm can be seen to significantly outperform EA as the number of bootstrap resamples are increased. Similar to the autoregressive datasets, as few as 5 bootstrap resamples were needed to

significantly improve performance over the EA method, again indicating that the bootstrap rank-ordering procedure can improve the quality of returned input variable sets when compared to the standard approach. For the NL2 case, the increase in selection accuracy performance as a function of increased bootstrap resample size is clearly seen for the broCMI method. Again it is seen that the broCMI is the best performing method for the nonlinear datasets with exogenous covariates. A notable mention is the KNN method which performed just as well as PMIS and broCMI on the NL dataset and performed second best after broCMI for the NL2 dataset.

3.6.1.1.1.3. Kentucky River basin Dataset

This partially synthetic hydrological example is characterized by its high degree of nonlinearity and collinearity. The most striking result is the fact that PCIS outperformed PMIS. This is due to the Hampel test stopping-criterion placing too strict of a tolerance on the selection of candidate variables for PMIS as evidenced by the SAc score. The KNN method performed intermediately between all other methods and in general exhibited the largest fluctuations in selection accuracy as evidenced by the IQR score. The best performing method on this dataset is KDE, with perfect model specification. The broCMI method is ranked in second place – as the bootstrap resample size increases from 25-100 the selection accuracy stays constant around 0.95. The only algorithms that could identify the correct input variables for this dataset were EA, KDE, and broCMI. However, EA and broCMI had the tendency to include a few irrelevant variables in the input variable sets. It can be seen that broCMI, even with as few as 5 bootstrap resamples, was able to improve the SAe score when compared to EA, providing further evidence that the bootstrap rank-ordering procedure may be used to improve input variable selection accuracy.

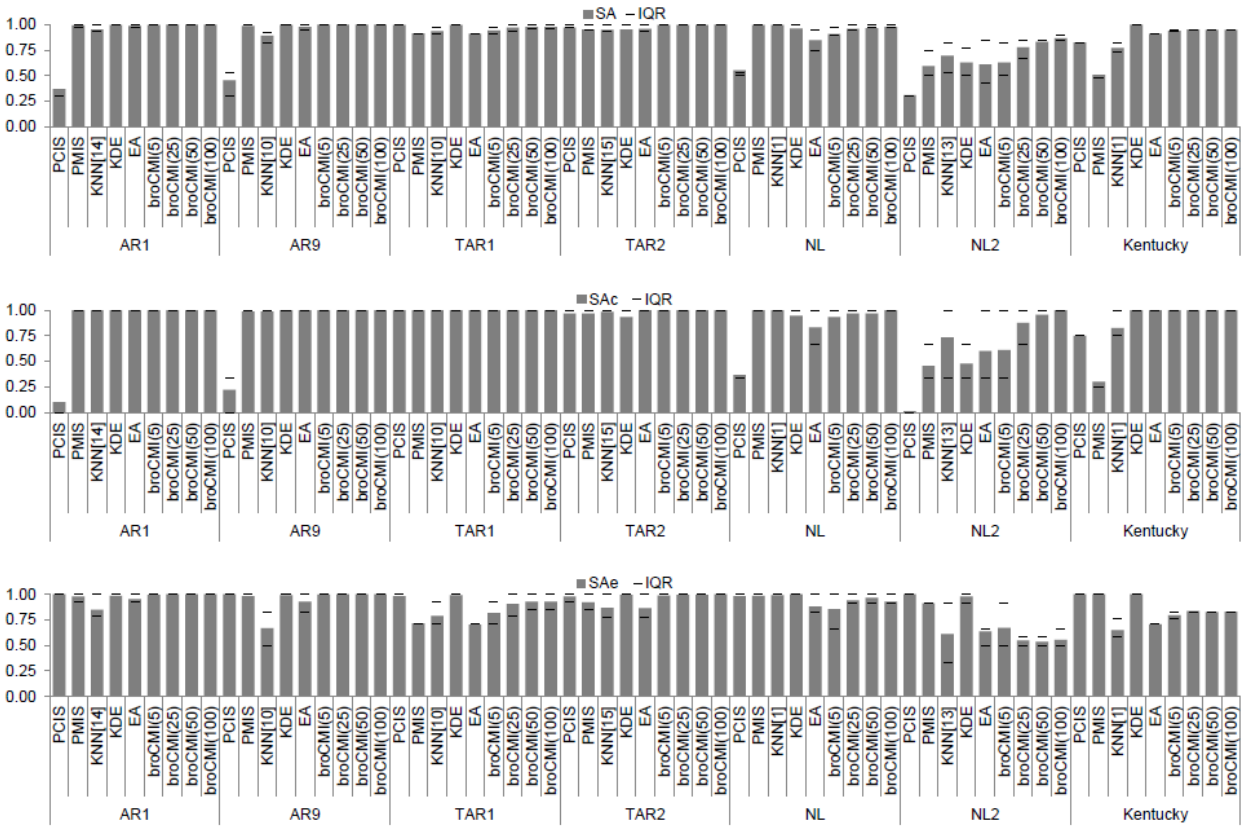

Figure 3.3. Selection accuracy results for the synthetic test problems (Experiment I); note: the square brackets indicate the number of nearest-neighbours used in KNN and the rounded brackets indicate the number of bootstrap resamples used in broCMI

It is clear for the synthetic datasets that the most consistent and best performing input variable selection algorithm is broCMI in terms of selection accuracy evaluation metrics (SA, SAc, SAe). Our proposed broCMI method was either the best (AR1, AR9, TAR2, and NL2) or second best (TAR1, NL, and Kentucky) performing algorithm across each synthetic dataset and also provided the most robust results (smallest IQR) – not a single other input variable selection algorithm displayed near the same amount of consistency as broCMI. Since using 100 bootstrap resamples within broCMI provided the best selection accuracy results for this experiment, the same number of bootstrap resamples were used within broCMI for Experiment II.

3.6.1.1.2. Computational Run-Time

In Table 3.5 we provide the computational run-times required to generate the results presented in the last sub-section. It can easily be seen that the most computationally efficient method is the linear PCIS. The most computationally efficient nonlinear method is EA, followed by KNN, KDE, broCMI (using 5

bootstrap resamples), PMIS, and then broCMI (for bootstrap resample sizes of 25, 50, and 100). Since a stopping-criterion is employed for all input variable selection algorithms in this work, the run-time order becomes a bit cumbersome to detail for a method such as broCMI (since the algorithm involves a search function for isolating input variable sets from the bootstrap resamples of a particular input variable set size), a similar difficulty was found in Galelli et al. (2014) when the authors considered the run-time order for the genetic algorithm-ANN based input variable selection approach and found it impractical to provide a general run-time order formula due to its complexity. Instead of considering the run-time order for the stopping-criterion approach, we consider the run-time order for the exhaustive search scenario (i.e. when all candidate input variables are selected) as is done by Galelli et al. (2014). This approach is more suitable to follow since it provides the worst-case run-time order for a given algorithm. We do not provide a step-by-step analysis of the run-time order calculations for each algorithm as a suitable procedure is already provided in Galelli et al. (2014) for the PMIS and PCIS algorithms used in our experiments (please see their Appendix B for details). Instead, we follow the exact same approach as taken by the authors and compute the run-time order for the KDE, EA, and broCMI algorithms, providing sufficient details on their various components attributing to their associated run-time orders. Since we were unable to locate a run-time order analysis for the KNN based CMI method in the literature we do not provide an explicit run-time order for this method and instead refer the interested reader to the works of Kraskov et al. (2004), Van Hulle (2005), Frenzel and Pompe (2007), and Evans (2008) as a starting point for such a study. We do note that Van Hulle (2005), who explored kernel density estimation, *k* nearest-neighbours, and Edgeworth approximations for multivariate entropy and MI estimation, found the *k* nearest-neighbours (multivariate) MI estimation approach to be more computationally efficient than that based on kernel density estimation, but less efficient when compared to Edgeworth approximations; with KDE and KNN having a time-complexity (i.e., as $N \to \infty$) of $\mathcal{O}(N^2 \cdot d)$ and Edgeworth approximations having a time-complexity of $\mathcal{O}(N \cdot d^3)$. Although the kernel density estimation and the *k* nearest-neighbours (multivariate entropy and MI estimation) methods have the same worst-case run-time orders, the *k* nearest-neighbours approach is usually faster since nearest-neighbours are searched for within an unsorted list while kernel density estimation, in all cases, requires one to compute a pairwise-distance calculation for each datum in the dataset which takes (for the best-case) as long or longer (in computation time) than sorting a list of nearest-neighbour distances (Van Hulle, 2005). (Of interest in improving the efficiency of the nearest-neighbour search routine, Evans (2008) discusses algorithms that may be employed to improve the computational efficiency (over the naive approach) for searching for nearest-neighbours in multivariate spaces (in the context of MI estimation).)

The PMIS (PCIS) algorithm requires one to estimate the MI (linear correlation) between two conditional model residuals (see Eq. 3.7 and Eq. 3.8), respectively). In our experiment, kernel density estimation (linear correlation) was used for computing Eq. 3.7 and general regression neural networks (multiple linear regression) were used for computing Eq. 3.8, each with a run-time order of $\mathcal{O}(N^2)$ ($\mathcal{O}(N)$) and $\mathcal{O}(d^2 \cdot N^2 + d^3)$ ($\mathcal{O}(d^2 \cdot N + d^3)$), respectively. Considering input variable selection for the worst-case scenario (i.e. selecting all candidate input variables, P) this amounts to a total run-time order of $\mathcal{O}(P^4 \cdot N^2 + P^5)$ ($\mathcal{O}(P^4 \cdot N + P^5)$) and a time-complexity of $\mathcal{O}(N^2)$ ($\mathcal{O}(N)$) for the PMIS (PCIS) algorithm (Galelli et al., 2014). The run-time order derivations for the KDE and EA algorithms follows directly from the PMIS algorithm except instead of estimating CMI/PMI using general regression neural networks (to compute the two conditional model residuals) and kernel density estimation (to compute MI), multivariate entropy terms Eq. 3.6 are used instead. Since Eq. 3.6 is used for the KDE and EA methods, their run-time orders are $\mathcal{O}(N^2 \cdot P^2)$ and $\mathcal{O}(N \cdot P^4)$, respectively. (The run-time order for KNN is thus less than or equal to KDE based on the analysis in Van Hulle (2005).) The run-time order for the worst-case input variable selection scenario for broCMI is determined by realizing that an (EA based) CMI estimate is made for each bootstrap resample. Drawing a bootstrap resample has a run-time order of $\mathcal{O}(N)$ (which is of negligible importance to the final run-time order since the EA based CMI estimation grows linearly); afterwards the returned selected input variable sets from each bootstrap resample are rank-ordered (following Table 3.1(a)) which has a run-time order of $\mathcal{O}(B! \cdot P + P^2)$; $\mathcal{O}(B! \cdot P)$ that represents the sum over each selected variable for an increasing bootstrap resample size (i.e. from $1{:}\,B$) and $\mathcal{O}(P^2)$ that represents the worst-case run-order time for ranking an unsorted list of P variables. Combining the run-time orders for the various components of the broCMI algorithm, we arrive at a final run-time order of $\mathcal{O}\big(B \cdot P \cdot (N \cdot P^3 + B! + P)\big)$.

Table 3.5. Computational run-times for the various input variable selection algorithms on the synthetic datasets (the brackets proceeding broCMI indicates the number of bootstrap resamples used in the algorithm)

	Input Variable Selection Algorithm								
	EA	KDE	KNN	PCIS	PMIS	broCMI (5)	broCMI (25)	broCMI (50)	broCMI (100)
Dataset	Computation Time (seconds)								
AR1	2.61	18.05	9.66	0.06	56.00	31.32	165.81	320.73	714.92
AR9	3.97	28.21	20.92	0.23	110.34	50.83	243.19	515.61	1075.86
TAR1	5.34	17.98	11.68	0.50	180.69	61.70	315.75	635.87	1342.10
TAR2	3.94	18.70	11.44	0.58	99.38	38.24	201.33	424.80	898.01
NL	4.14	25.85	8.99	0.43	108.79	45.35	225.28	495.15	1043.44
NL2	6.16	17.23	21.89	0.03	81.13	54.13	290.25	614.48	1304.60
Kentucky	35.71	4560.83	289.44	1.98	3682.89	332.71	1789.20	3746.79	7795.05

Of practical interest, it may be seen that for each dataset broCMI significantly increases the selection accuracy and input variable selection robustness when compared to EA - significant selection accuracy performance increases are obtained with as few as 5 bootstrap resamples, which demonstrates the exceptional performance and computational efficiency of broCMI for the synthetic datasets.

3.6.1.2. Qualitative Comparisons across Input Variable Selection Algorithms

To compare the qualitative characteristics of EA, KDE, KNN, PMIS, PCIS, and broCMI we use the following criteria from Galelli et al. (2014): ease of use and robustness, explanation capability, and flexibility – each of these criteria are listed below with appropriate reference to each input variable selection algorithm:

3.6.1.2.1. Ease of Use and Robustness

This metric relates to the settings one needs to employ to use a particular algorithm. The PCIS and EA methods are the easiest to use as they do not require any parametric settings. The broCMI algorithm is the next easiest to use, as one simply needs to increase the bootstrap resample size until sufficient results are obtained. If sufficient results are not achieved by increasing the bootstrap resample size within computational resource allowances then another algorithm should be considered. The KNN method is the fourth easiest to use as one simply needs to specify the number of nearest-neighbours k_{nn} to employ, however this parameter is problem dependent and needs to be optimized carefully to ensure optimal results. The KDE method requires a kernel type and associated parameters to be specified which require tuning for optimal performance. The PMIS algorithm is the most parametrically involved method of all the methods considered in this work. It requires one to specify a kernel type and associated parameters to estimate MI (or a specified copula distribution and associated parameters (Chen et al. (2015a))) and requires the specification of two intermediate regression models (each with tunable parameters).

3.6.1.2.2. Explanation Capability

Each method used in this work has the ability to express the level of dependence associated between each selected input variable and the response variable. Such dependencies between selected inputs and the process response have been exploited by Sharma and Mehrotra (2014) to enable predictive modeling by taking into account such dependencies in the model specification. The approach introduced in Sharma and Mehrotra (2014) may also be extended to include ensemble averaging of forecasts from a set of bootstrap resamples if bootstrap rank-ordering is coupled with their Partial Information approach – such considerations are set aside for future research.

3.6.1.2.3. Flexibility

This characteristic is related to the type of input variable selection employed (filter, wrapper, or embedded) and the ability to mitigate poor performance in a given method by improving a particular work flow process or removing time-consuming steps. Since each input variable selection algorithm, with the exception of broCMI, follows the standard input variable selection process (see Figure 3.1 (a)) we only briefly consider the flexibility offered by bootstrap rank-ordering (see Figure 3.1 (b)). Based on the results presented in Figure 3.3 one may clearly ascertain that the bootstrap rank-ordering approach greatly improves performance over the standard method (see differences between the EA method and broCMI). Since the bootstrap rank-ordering procedure is more or less a post-processing technique, it may be used with any other input variable selection algorithm used in this work. We plan to explore this in future studies.

3.6.2. Experiment II – Urban Water Demand Datasets

To explore broCMI for addressing real-world time series (water resources) modeling problems we compare its abilities against pre-existing methods (PCIS, PMIS, KNN, and KDE), in addition to the EA algorithm (used for selecting inputs within broCMI) for selecting input variables for use in machine learning models (in this case the new ELM approach) by using the selected variables from each algorithm to generate urban water demand forecasts in Ottawa, Canada and evaluating their forecast performance to identify the best input variable selection algorithm. We focus on demonstrating: 1) how the forecasts generated via inputs selected by broCMI perform when they are used with different machine learning models (i.e. ELM and ANN), 2) how forecasts using the same model type (e.g. ELM) but with inputs selected via broCMI compares to those selected by EA, KDE, KNN, PMIS, and PCIS, and 3) how 1 and 3 day lead time forecasts generated via machine learning models (i.e. ELM) using inputs selected via broCMI performs out-of-sample.

Before we begin our discussion of the results of this experiment a few important points must be mentioned. The broCMI algorithm used 100 bootstrap resamples as this was demonstrated to provide the best selection accuracy results in Experiment I. We only present the results (optimal selected input variable sets, machine learning model structure/parameters, and model performance) for the best developed UWD forecasts for the 1 and 3 day lead times. We also only present the results of ANN for 1 day lead times to try and keep the presentation of the results as brief as possible; we present the results of ELM for all lead times due to its computational efficiency and similar performance with ANN and also because it is a newer approach than ANN. It is worth reiterating that the fitness function was chosen as

the root mean square error and that optimal models were selected based on their cross-validation performance (fitness) (Hastie et al., 2009).

For this (UWD forecasting) experiment the broCMI algorithm provided the best input variable sets for each dataset and model type providing further evidence that the bootstrap rank-ordering procedure (Figure 3.1(b)) may be used to improve the quality of selected input variable sets when compared to the standard approach (Figure 3.1(a)). The run-time for the various algorithms is as follows: EA – 31 s; KDE – 22 s; KNN – 139 s; PMIS – 364 s; PCIS – 1 s; and broCMI – 2422 s. The reason for the longer run times for broCMI is because 100 bootstrap resamples were used in this experiment (since it provided the best results for the test problems in Experiment I). Future studies focused solely on the broCMI method (in contrast to this study which focuses on introducing and comparing broCMI against EA and existing methods) can compare the optimal bootstrap resample size for different time series modeling problems as it is has been suggested that the optimal bootstrap resample size is problem dependent (Tiwari and Adamowski, 2013). In the next paragraph we discuss results pertaining to the different machine learning models and their performances when using inputs selected via broCMI.

3.6.2.1 Different machine learning models (ELM and ANN) paired with broCMI

In Table 3.6 we present the results of the input variable selection process for each forecast lead time (it is important to remember that the model fitness in Table 3.6 refers to the cross-validation performance, i.e. how the optimal models were selected - we do not report the training or test performances for this comparison since the performances are similar and to maintain brevity). It is seen in general that ELM and ANN are very competitive in terms of performance. We found ELM to be the most computationally efficient method in producing forecasts – our investigations demonstrated that on average for each ANN model that was developed, approximately 50 ELM models could be developed in the same time. We found the best input sets used in ANN and ELM to be very similar, providing evidence that the broCMI algorithm was able to identify suitable process predictors that can provide forecasts of a similar nature even when using different modeling techniques (e.g. ELM and ANN) with different model structures (i.e. number of hidden neurons and different input variables); see for instance the ELM and ANN models for $a3w(t+1)$ (Figure 3.4 and Table 3.6) where the models differ by a single predictor (i.e. $a3w(t-5)$) but provide similar forecast results. To show the differences between the two model types we plot them in Figure 3.4 by taking the mean of the 1 day lead time forecasts produced by the ELM and ANN models for each day in the summer water demand period (i.e. the mean forecast for the calendar dates May 1 to September 30 considering all available observations, thereby grouping training, cross-validation, and testing data

together) and plotting it against the mean observation (target) for that calendar date; which may be likened to an average seasonal hydrograph. One may see that the ELM and ANN models produce quite similar hydrographs, however noticeable differences between the models are still present. We found in general that the ANN required less hidden neurons than ELM to produce an optimal forecast – this is very likely due to the randomization of the ELM hidden neuron parameters that generally requires successive trials to be averaged over to improve model performance (Lima et al., 2015) (which was also adopted in this study, please see the online supplementary material [appendix] for details). The differences exhibited by each of the models can be regarded as useful should one wish to build an ensemble model. Diversity amongst forecasts would be exploitable in ensemble methods to improve the overall forecast accuracy (Tiwari and Adamowski, 2015). We intend to explore using each input variable set produced by broCMI in a probabilistic forecasting framework in future publications – it would also be interesting to compare probabilistic forecasts developed through different machine learning models (ELM, ANN, SVR, multivariate adaptive regression splines, or Bayesian based regression methods (Ciupak et al., 2015), for example) using inputs selected via broCMI.

Table 3.6. Input variable sets selected via broCMI and used to generate 1 and 3 day lead time UWD forecasts (note: N^H represents the number of hidden layer neurons used in ELM or ANN)

Target	No. Inputs	Selected Input Variable Sets					Model	N^H	Fitness (ML/D)
a3w(*t*+1)	9	*a3w*(*t*)	*a3w*(*t*-1)	*a3w*(*t*-3)	*rain*(*t*)	*a3w*(*t*-2)	ELM	18	2.736
		a3w(*t*-4)	*a3w*(*t*-6)	*a3w*(*t*-5)	*a3w*(*t*-7)				
	8	*a3w*(*t*)	*a3w*(*t*-1)	*a3w*(*t*-3)	*a3w*(*t*-2)	*rain*(*t*)	ANN	3	2.558
		a3w(*t*-4)	*a3w*(*t*-6)	*a3w*(*t*-7)					
a3w(*t*+3)	14	*a3w*(*t*)	*a3w*(*t*-4)	*a3w*(*t*-1)	*a3w*(*t*-2)	*rain*(*t*)	ELM	24	4.213
		a3w(*t*-3)	*a3w*(*t*-5)	*rain*(*t*-1)	*api*(*t*)	*mat*(*t*)			
		mat(*t*-2)	*rain*(*t*-5)	*a3w*(*t*-6)	*rain*(*t*-8)				
amg(*t*+1)	23	*amg*(*t*)	*amg*(*t*-6)	*amg*(*t*-3)	*amg*(*t*-13)	*amg*(*t*-2)	ELM	40	0.098
		amg(*t*-1)	*amg*(*t*-4)	*amg*(*t*-5)	*amg*(*t*-12)	*amg*(*t*-7)			
		amg(*t*-14)	*amg*(*t*-19)	*amg*(*t*-10)	*amg*(*t*-11)	*amg*(*t*-17)			
		amg(*t*-16)	*amg*(*t*-18)	*amg*(*t*-15)	*amg*(*t*-8)	*amg*(*t*-9)			
		rain(*t*)	*rain*(*t*-7)	*rain*(*t*-1)					
	20	*amg*(*t*)	*amg*(*t*-6)	*amg*(*t*-13)	*amg*(*t*-2)	*amg*(*t*-3)	ANN	1	0.097
		amg(*t*-1)	*amg*(*t*-12)	*amg*(*t*-5)	*amg*(*t*-4)	*amg*(*t*-7)			
		amg(*t*-14)	*amg*(*t*-11)	*amg*(*t*-10)	*amg*(*t*-19)	*amg*(*t*-15)			
		amg(*t*-17)	*amg*(*t*-8)	*amg*(*t*-16)	*amg*(*t*-18)	*amg*(*t*-9)			
amg(*t*+3)	29	*amg*(*t*)	*amg*(*t*-4)	*amg*(*t*-1)	*amg*(*t*-5)	*amg*(*t*-3)	ELM	46	0.112
		amg(*t*-11)	*amg*(*t*-2)	*amg*(*t*-8)	*amg*(*t*-12)	*amg*(*t*-7)			
		amg(*t*-9)	*amg*(*t*-6)	*amg*(*t*-10)	*amg*(*t*-13)	*amg*(*t*-15)			
		amg(*t*-14)	*amg*(*t*-16)	*amg*(*t*-17)	*amg*(*t*-19)	*amg*(*t*-18)			
		rain(*t*-2)	*rain*(*t*)	*rain*(*t*-7)	*api*(*t*-1)	*api*(*t*)			
		rain(*t*-3)	*rain*(*t*-4)	*rain*(*t*-1)	*mat*(*t*-2)				
		rain(*t*-1)	*rain*(*t*-8)						

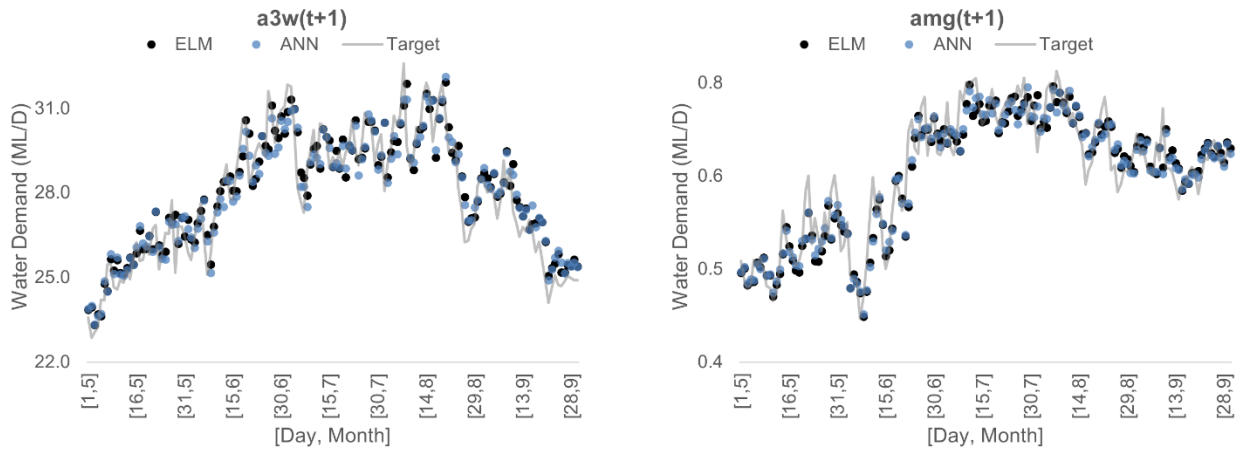

Figure 3.4. Average seasonal hydrograph generated via ELM and ANN for the 1 day lead time considering each urban water demand time series using inputs derived from broCMI

3.6.2.2. Different Input Variable Selection Algorithms Paired with ELM

We now explore ELM models created using the input variables produced by the various input variable selection algorithms by comparing broCMI against each algorithm. The inputs selected by the various input variable selection algorithms (EA, KDE, KNN, PMIS, PCIS, and broCMI) along with their model rank and (cross-validation) fitness is provided in Table 3.7 and 3.8. For the 1 day lead time forecasts we plot (please see Figure 3.5) the mean absolute errors (between forecast and observation) for each calendar date in the summer demand period, considering all available observations (i.e. combining training, cross-validation, and test data). Both Table 3.7 and 3.8 identify broCMI as the best performing algorithm which is corroborated by the plots in Figure 3.5, showing broCMI to exhibit smaller or equal mean absolute model residuals for the majority of the data points when compared to the other methods. For each dataset different mean absolute error patterns can easily be seen, providing great evidence to the diversity of models that may be developed by considering different input variable selection algorithms. The EA algorithm performs moderately well for each dataset in comparison to broCMI which provides further support that bootstrap rank-ordering (Figure 3.1(b)) can be used to improve the quality of selected input variable sets when applied instead of the standard method (Figure 3.1(a)). The KNN method seems to be the next best method for these datasets, followed by EA, PMIS, PCIS, and finally KDE. It is interesting to note that the KDE method performed the worst on this real-world example. This is very likely due to the combination of the Hampel test stopping-criterion providing too strict a tolerance and the improper specification of the kernel density bandwidth, which was set as the Gaussian Reference Rule for simplicity in this work (which has also been done in other similar studies, e.g. Sharma and Mehrotra (2014)). The

PMIS method also adopted the Gaussian Reference Rule for specifying the bandwidth for MI estimation. Carefully tuning the kernel bandwidth and exploration of other stopping-criteria would very likely result in much improved results for KDE and PMIS; however, this is outside the scope of this work and left for future research.

Table 3.7. Comparison of 1 day lead time forecasts using different input variable selection algorithms and ELM for the *a3w* time series (note: the square brackets indicate the number of nearest-neighbours used in KNN; the number of unique input variable sets for each dataset are listed in the rounded brackets beside Model Rank)

	a3w(*t*+1)					
	broCMI	EA	KDE	KNN[9]	PMIS	PCIS
Model Rank (/73)	1	44	73	8	72	68
Fitness (ML/D)	2.736	2.837	2.956	2.789	2.988	2.859
Input Order	Selected Input Variables					
1	*a3w*(*t*)	*a3w*(*t*)	*a3w*(*t*)	*a3w*(*t*)	*a3w*(*t*)	*a3w*(*t*)
2	*a3w*(*t*-1)	*a3w*(*t*-1)		*rain*(*t*)	*a3w*(*t*-1)	*rain*(*t*)
3	*a3w*(*t*-3)	*a3w*(*t*-2)		*a3w*(*t*-5)	*a3w*(*t*-4)	
4	*rain*(*t*)	*a3w*(*t*-3)				
5	*a3w*(*t*-2)	*rain*(*t*)				
6	*a3w*(*t*-4)	*a3w*(*t*-4)				
7	*a3w*(*t*-6)	*a3w*(*t*-6)				
8	*a3w*(*t*-5)	*a3w*(*t*-5)				
9	*a3w*(*t*-7)	*a3w*(*t*-10)				
10		*a3w*(*t*-9)				
11		*a3w*(*t*-8)				
12		*a3w*(*t*-11)				
13		*a3w*(*t*-7)				
14		*rain*(*t*-2)				
15		*rain*(*t*-1)				
16		*api*(*t*)				
17		*api*(*t*-1)				

Table 3.8. Comparison of 1 day lead time forecasts using different input variable selection algorithms and ELM for the *amg* time series (note: the square brackets indicate the number of nearest-neighbours used

in KNN; the number of unique input variable sets for each dataset are listed in the rounded brackets beside Model Rank)

	amg(*t*+1)					
	broCMI	EA	KDE	KNN[1]	PMIS	PCIS
Model Rank (/59)	1	21	59	34	20	46
Fitness (ML/D)	0.098	0.099	0.101	0.099	0.099	0.100
Input Order	Selected Input Variables					
1	*amg*(*t*)	*amg*(*t*)	*amg*(*t*)	*amg*(*t*)	*amg*(*t*)	*amg*(*t*)
2	*amg*(*t*-6)	*amg*(*t*-6)		*amg*(*t*-19)	*amg*(*t*-1)	*amg*(*t*-6)
3	*amg*(*t*-3)	*amg*(*t*-2)		*amg*(*t*-12)	*amg*(*t*-6)	
4	*amg*(*t*-13)	*amg*(*t*-13)		*amg*(*t*-6)	*amg*(*t*-5)	
5	*amg*(*t*-2)	*amg*(*t*-1)		*amg*(*t*-8)	*amg*(*t*-2)	
6	*amg*(*t*-1)	*amg*(*t*-3)		*mat*(*t*-8)	*amg*(*t*-3)	
7	*amg*(*t*-4)	*amg*(*t*-12)		*amg*(*t*-4)	*amg*(*t*-12)	
8	*amg*(*t*-5)	*amg*(*t*-7)		*rain*(*t*-8)	*amg*(*t*-13)	
9	*amg*(*t*-12)	*amg*(*t*-5)		*rain*(*t*-4)	*amg*(*t*-7)	
10	*amg*(*t*-7)	*amg*(*t*-4)		*amg*(*t*-5)	*amg*(*t*-4)	
11	*amg*(*t*-14)	*amg*(*t*-14)		*amg*(*t*-11)	*amg*(*t*-11)	
12	*amg*(*t*-19)	*amg*(*t*-11)		*amg*(*t*-13)	*amg*(*t*-19)	
13	*amg*(*t*-10)	*amg*(*t*-10)		*amg*(*t*-17)	*amg*(*t*-14)	
14	*amg*(*t*-11)	*amg*(*t*-9)		*amg*(*t*-7)	*amg*(*t*-8)	
15	*amg*(*t*-17)	*amg*(*t*-15)		*mat*(*t*-16)	*amg*(*t*-10)	
16	*amg*(*t*-16)	*amg*(*t*-19)		*mat*(*t*)	*amg*(*t*-18)	
17	*amg*(*t*-18)	*amg*(*t*-17)		*mat*(*t*-19)	*amg*(*t*-9)	
18	*amg*(*t*-15)	*amg*(*t*-16)			*amg*(*t*-15)	
19	*amg*(*t*-8)	*amg*(*t*-8)			*amg*(*t*-17)	
20	*amg*(*t*-9)	*amg*(*t*-18)			*amg*(*t*-16)	
21	*rain*(*t*)	*rain*(*t*-1)				
22	*rain*(*t*-7)	*rain*(*t*-4)				
23	*rain*(*t*-1)	*rain*(*t*-3)				
24		*rain*(*t*-2)				
25		*rain*(*t*-7)				
26		*rain*(*t*)				
27		*api*(*t*)				
28		*api*(*t*-1)				
29		*mat*(*t*)				
30		*mat*(*t*-11)				

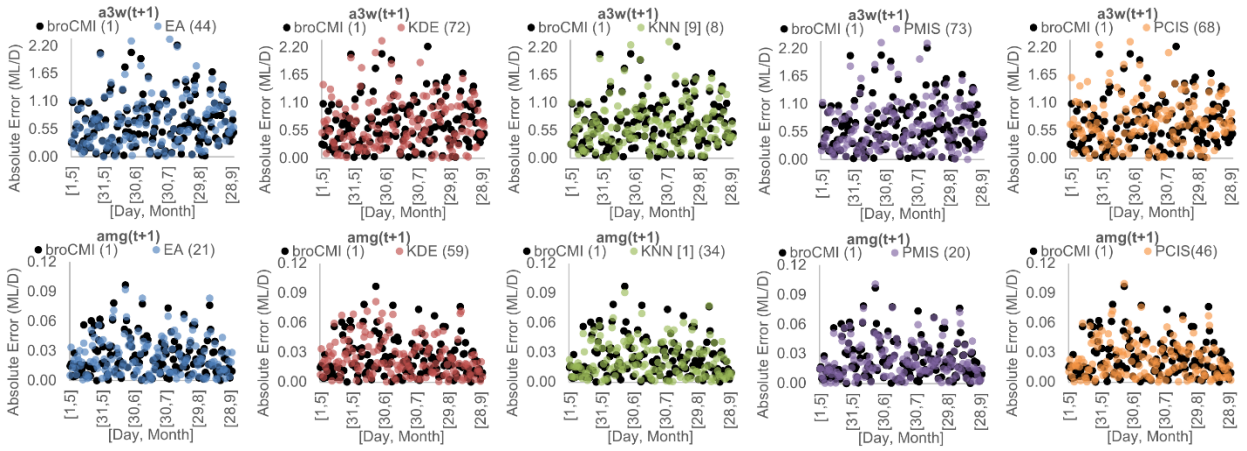

Figure 3.5. Comparing broCMI against EA, KDE, KNN, PMIS, and PCIS using their mean absolute residuals (forecasts versus observations) for each calendar date during the summer demand period generated via ELM for the 1 day lead time considering each urban water demand time series (note: square brackets indicate the number of nearest-neighbours used in KNN while the rounded brackets represents the model rank (i.e. a score of 1 indicates the best model))

3.6.2.3. Out-of-Sample ELM Forecasts

The out-of-sample (i.e. test set) 1 and 3 day lead time UWD forecasts using inputs selected via broCMI and forecasts generated via ELM were evaluated using the RMSE, Nash-Sutcliffe Efficiency Index (NASH) (Moriasi et al., 2007), and mean absolute percentage error (MAPE) (De Gooijer and Hyndman, 2006) performance metrics; their results are provided in Figure 3.6 and Table 3.9. One may notice that the forecasts for the *amg* dataset were the only models that provided satisfactory model efficiency scores (NASH > 0.5) across all lead times (Moriasi et al., 2007). The only other satisfactory forecast occurred for *a3w* at the 1 day lead time. The poor performance of the *a3w*(*t*+3) forecast is likely due to the average water demand processes responding to system drivers not available in the input variable selection sets (inadequately capturing the process variability) in addition to the increasing difficulty that goes along with forecasting chaotic time series at larger forecast horizons (Liu et al., 2004). As expected, the forecast performance decreases for increasing lead times across each time series.

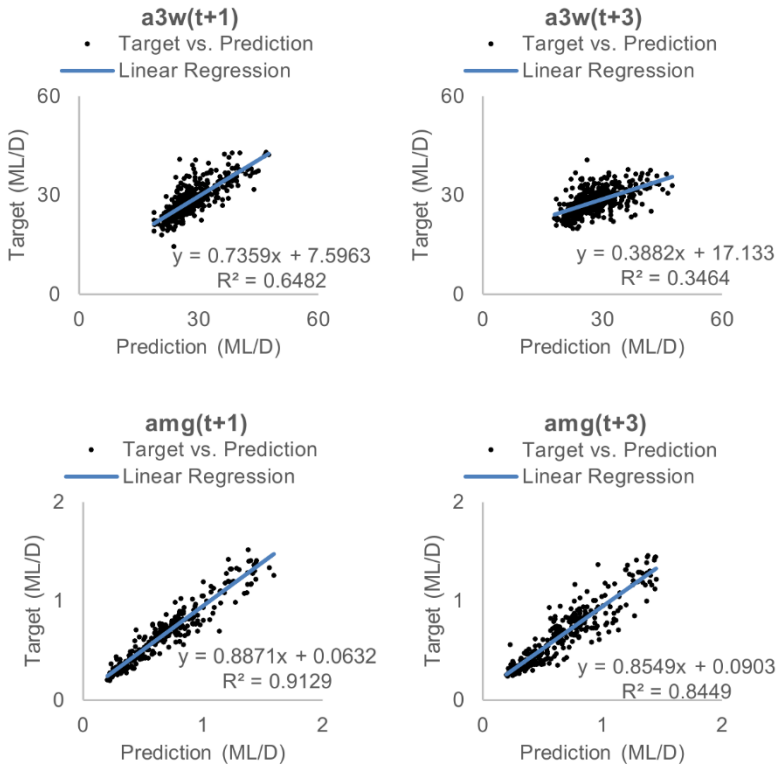

Figure 3.6. Scatter plots for the out-of-sample forecasts for each urban water demand time series and lead time using ELM with inputs selected via broCMI

Table 3.9. ELM out-of-sample performance for 1 and 3 day lead time UWD forecasts (note: N^H represents the number of hidden layer neurons used in ELM)

Dataset	Lead Time	ELM Model Structure (Inputs - N^H - Outputs)	NASH	RMSE (ML/D)	MAPE (%)
a3w	1	(9-18-1)	0.633	3.175	8.234
	3	(14-24-1)	0.339	4.402	11.794
amg	1	(23-40-1)	0.911	0.096	10.09
	3	(29-46-1)	0.844	0.126	15.028

Our observations on the selected input variable sets across the various UWD datasets can be summarized as follows:

1. Meteorological variables are valuable explanatory variables for forecasting UWD within the 3W and Morgan's Grant pressure zones in Ottawa, Canada;
2. The daily antecedent precipitation index and daily rainfall were found to be more prominent amongst the 1 day lead time forecasts while the maximum air temperature was found to be more

prominent amongst the 3 day lead time forecasts. This is likely due to the UWD process responding more acutely to rainfall or dry periods and responding more gradually to maximum air temperature, referred to as the "hysteresis" effect in Miaou (1990). It would be interesting to consider in future studies re-running the input variable selection procedure for these datasets by thresholding certain meteorological variables, for example maximum air temperature, to examine if the relationship between UWD and certain meteorological variables fluctuate within certain tolerance levels; and

3. In general, broCMI generated input variable sets that were found to be optimal when generating UWD forecasts in comparison with all other methods (EA, KDE, KNN, PMIS, and PCIS). In all cases the KDE, PMIS, and PCIS methods tended to choose too few input variables for each of the UWD forecast models (please refer to Table 3.7 and Table 3.8). The EA and KNN methods also tended to provide either too few input variables or selected input variable sets that were overly complex, resulting in poor quality UWD forecast models (for example, the EA method created an overly complex model for the *amg*(t+1) dataset while KNN selected too few inputs for the *amg*(t+1) dataset). The improved performance of broCMI (in comparison to the other input variable selection algorithms) relates to the multiple input variable sets it produces based on the bootstrap rank-ordering procedure which smooths variations in the resampled selected input variable sets by assessing the variability in their selection ranks and ordering the selected input variables by their magnitude of variability (from least to most variable). This procedure results in the selection of input variable sets that retain much of the training data variability, which is useful for improving the predictive capability of machine learning based forecasts (whose performance are highly dependent on the training data variability (LeBaron and Weigend, 1998; Singh et al., 2014)). For this reason, the bootstrap rank-ordering procedure would very likely improve the results obtained from, KDE, KNN, PMIS, and PCIS algorithms should they be coupled together.

Finally, we conclude this section by commenting on a trend within both experiments where the EA and broCMI algorithms selected a larger number of input variables when compared to the other algorithms. We provide the following observations:

1. PCIS lacks the ability to identify nonlinear dependencies existing between candidate inputs, pre-selected variables, and the response, which is a necessity for the proper specification of the processes in the more strongly nonlinear problems (i.e. NL, NL2, Kentucky, and UWD datasets). Because PCIS cannot detect significant nonlinear associations that exist in a number of the

datasets used in our experiments, the Hampel test, as expected, rejects the selection of significant variables that are useful in specifying these (nonlinear) processes;

2. EA and broCMI do not require parametric settings, thus the estimation of CMI for these methods is solely dataset dependent and not influenced by non-optimal parameter identification such as for the KDE, KNN, and PMIS methods. Therefore, the number of selected variables using KDE, KNN, and PMIS is dependent on both the chosen stopping-criterion and the parametric settings, which aside from the KNN method, were not varied in our experiments and depended solely on the dataset properties. This is the likely reason why EA (and broCMI) chose, on average, more input variables than the rest of the algorithms - the Hampel-test criterion did not provide a strict tolerance for these methods and, for certain problems, selected more input variables than were required to specify the process (see for instance the Kentucky dataset results). This is not necessarily a drawback of our proposed (EA and broCMI) algorithms as the addition of (a few) superfluous inputs, when those inputs pertinent to the problem are selected, is (in general) preferred to the opposite situation (Galelli et al., 2014); and
3. In general, the KDE and PMIS algorithms are mis-specified by the Gaussian Reference Rule assumption for the kernel bandwidth used in kernel density estimation, which insufficiently estimates the dependencies in the various nonlinear datasets (NL2, Kentucky, and the UWD datasets for PMIS and the UWD datasets for KDE). While for the KNN method, it is very likely that the optimal number of nearest-neighbours was not identified for the majority of the datasets as it is dataset dependent (Tsimpiris et al., 2012) and we only studied a relatively small range of nearest-neighbours (1-15) (since between two and four nearest-neighbours were suggested in the literature for MI estimation (Kraskov et al., 2004)). An investigation using a wider range of nearest-neighbours for the KNN method and different kernel types (and parameter optimization routines) for the KDE and PMIS algorithms can be more rigorously explored in further studies.

3.7. Closure

Through this study we developed, tested, and compared two new and non-parametric nonlinear input variable selection algorithms based on conditional mutual information, the EA and broCMI methods, on seven synthetic input variable selection datasets and a real-world time series (urban water demand) forecasting experiment (in Ottawa, Canada). The EA and broCMI input variable selection methods were validated against parametric algorithms already existing in the water resources domain (e.g. partial mutual information selection) demonstrating the comparable performance of the (non-parametric) EA

method and the superior performance of (non-parametric) broCMI when compared to each other method (EA, KDE, KNN, PMIS, and PCIS). The broCMI algorithm is a simple extension of the EA method in that it uses the EA method to select input variables over a number of bootstrap resamples of the dataset and then rank-orders the variability across their selection ranks to develop new input variable sets incorporating the variability across the resamples. We found the broCMI method to be the best overall method at selecting the best input variable sets across both synthetic and real-world experiments. The synthetic test problems showed broCMI to be the most accurate in terms of selection accuracy and it was also the most robust method (i.e. broCMI had the smallest selection accuracy interquartile range). For the real-world experiment, the urban water demand forecast models whose inputs were selected via broCMI were the best specified models, while all other input variable selection algorithms tended to mis-specify the input variables for each model, with no method providing comparable performance to broCMI. We also demonstrated the performance of broCMI for increasing bootstrap resample size but still believe much more testing on a wider variety of datasets is needed before making claims on appropriate bootstrap sizes which is in line with recent research that has suggested that the optimal bootstrap resample size is dataset dependent (Tiwari and Adamowski, 2013). However, this study found 100 bootstrap resamples to be appropriate for the datasets considered in this work; in many cases (please refer to the results for the synthetic test problems) as few as 5 bootstrap resamples provided a substantial increase in performance over the standard approach (e.g. the EA method).

The characteristic imbuing broCMI with the ability to outperform each other method presented in this paper lies in its ability to envelop variability across a number of selected input variable sets by simply using rank statistics to order the selected input variable sets for increasing resample size. An interesting metric that can be used to quantify the amount of variability present in the resampled selected input variable sets is the Kuncheva Stability Index, which provides a measure of uncertainty over the input variable selection procedure (Kuncheva, 2007). In future studies we plan to assess the relationship between forecast performance and the Kuncheva Stability Index for real-world time series (water resources) modeling problems as it may enable an automatic selection of the bootstrap resample size for the bootstrap rank-ordering procedure discussed in this paper. To improve the computational efficiency of the bootstrap resampling procedure new bootstrap methods designed for “big data” scenarios may be considered such as the new scalable bootstrap approach (Kleiner et al., 2014).

Future research will investigate the use of broCMI for creating ensemble probabilistic forecasts where each bootstrap resample will achieve a weight that determines its contribution to the ensemble.

Immediate extensions of our method to formulate a forecast model similar to Sharma and Mehrotra (2014) are indeed plausible too and may also be cast as an ensemble forecast scheme. We also have an interest in applying the bootstrap rank-ordering approach to the standard partial mutual information selection algorithm to assess its added value as it is currently the most popular nonlinear input variable selection method in water resources modeling (Maier et al., 2010). More so, the same algorithm may also be investigated by using Edgeworth approximations to compute mutual information instead of kernel density estimates (May et al., 2008a) or the copula-entropy approach (Chen et al., 2014). We also wish to extend the application of input variable selection to include time-frequency analysis (e.g. wavelet transforms, empirical mode decompositions, and singular spectrum analysis) by decomposing each time series into periodic and trend components whereby the complicated relationships between response and candidate inputs are more easily ascertained. Finally, pre-processing time series data before estimating conditional mutual information via methods such as independent component analysis is readily applicable to broCMI and may even reduce some of the computational burden by exploiting the independence property of entropy (please see section 3 in Gong et al. (2013)).

3.8. Acknowledgements

The authors would like to extend their deepest gratitude to the three anonymous reviewers, the associate editor, and the editor for their invaluable feedback that has helped to greatly improve the presentation of this paper. The authors would also like to thank the City of Ottawa for providing data utilized in this study. Requests for data used in this study should be made through opendata@ottawa.ca. The authors acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding part of the research presented in this paper under an NSERC Engage grant, as well as an NSERC Discovery Grant held by Prof. Adamowski. Special thanks are extended to Prof. Galelli at Singapore University along with R. Diduch, C. Rogers, J. Bougadis, and G. Reilly at the City of Ottawa for their time and helpful discussions.

3.9. References

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L., 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog. Phys. Geogr. 36, 480–513. doi:10.1177/0309133312444943

Adamowski, J., Adamowski, K., Prokoph, A., 2013. A Spectral Analysis Based Methodology to Detect Climatological Influences on Daily Urban Water Demand. Math. Geosci. doi:10.1007/s11004-012-

9427-0

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Adamowski, J.F., 2008. Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks. J. Water Resour. Plan. Manag. 134, 119–128. doi:10.1061/(ASCE)0733-9496(2008)134:2(119)

AghaKouchak, A., 2014. Entropy-copula in hydrology and climatology. J. Hydrometeorol. 15, 2176–2189. doi:10.1175/JHM-D-13-0207.1

Akusok, A., Bjork, K.M., Miche, Y., Lendasse, A., 2015. High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications. IEEE Access 3, 1011–1025. doi:10.1109/ACCESS.2015.2450498

Anctil, F., Michel, C., Perrin, C., Andréassian, V., 2004. A soil moisture index as an auxiliary ANN input for stream flow forecasting. J. Hydrol. 286, 155–167. doi:10.1016/j.jhydrol.2003.09.006

Back, A.D., Trappenberg, T.P., 2001. Selecting inputs for modeling using normalized higher order statistics and independent component analysis. IEEE Trans. Neural Networks 12, 612–617. doi:10.1109/72.925564

Beyond, i.i.d. in Information Theory, 2015. . Banff, Canada.

Bowden, G.J., Dandy, G.C., Maier, H.R., 2005a. Input determination for neural network models in water resources applications. Part 1 - Background and methodology. J. Hydrol. 301, 75–92. doi:10.1016/j.jhydrol.2004.06.021

Bowden, G.J., Maier, H.R., Dandy, G.C., 2005b. Input determination for neural network models in water resources applications. Part 2. Case study: Forecasting salinity in a river. J. Hydrol. 301, 93–107. doi:10.1016/j.jhydrol.2004.06.020

Brown, G., Pocock, A., Zhao, M.-J., Lujan, M., 2012. Conditional Likelihood Maximisation: A Unifying Framework for Mutual Information Feature Selection. J. Mach. Learn. Res. 13, 27–66.

Calsaverini, R.S., Vicente, R., 2009. An information-theoretic approach to statistical dependence: Copula information. EPL 88. doi:10.1209/0295-5075/88/68003

Chen, C., Huang, G.H., Li, Y.P., Zhou, Y., 2013. A robust risk analysis method for water resources allocation under uncertainty. Stoch. Environ. Res. Risk Assess. 27, 713–723. doi:10.1007/s00477-012-0634-5

Chen, J., Sharma, A.S., Edwards, J.W., Shao, X., Kamide, Y., 2008. Spatiotemporal dynamics of the magnetosphere during geospace storms: Mutual information analysis. J. Geophys. Res. Sp. Phys. 113. doi:10.1029/2007JA012310

Chen, L., Ye, L., Singh, V., Zhou, J., Guo, S., 2014. Determination of input for artificial neural networks for flood forecasting using the copula entropy method. J. Hydrol. Eng. 19. doi:10.1061/(ASCE)HE.1943-5584.0000932

City of Ottawa, 2013a. Population [Webpage]. URL: http://ottawa.ca/en/long-range-financial-plans/economy-and-demographics/population (accessed 1.14.14).

City of Ottawa, 2013b. Population Statistics for 3W and Morgan's Grant Pressure Zones [Interview]. (accessed 16.12.13).

City of Ottawa, 2013c. 55. Land and Climate Statistics, City of Ottawa and Area.

City of Ottawa, 2009. Infrastructure Master Plan.

Ciupak, M., Ozga-Zielinski, B., Adamowski, J., Quilty, J., Khalil, B., 2015. The application of Dynamic Linear Bayesian Models in hydrological forecasting: Varying Coefficient Regression and Discount Weighted Regression. J. Hydrol. 530, 762–784. doi:10.1016/j.jhydrol.2015.10.023

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory. Wiley-Interscience, New York.

Cresswell, M., Naser, G., 2013. A water resources management strategy for small water districts - a case study of the South East Kelowna irrigation district. Can. J. Civ. Eng. 40, 499–507. doi:10.1139/cjce-2012-0398

De Gooijer, J.G., Hyndman, R.J., 2006. 25 Years of Time Series Forecasting. Int. J. Forecast. 22, 443–473. doi:10.1016/j.ijforecast.2006.01.001

Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D.P., 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. Hydrol. Earth Syst. Sci. 14, 1931–1941. doi:10.5194/hess-14-1931-2010

Evans, D., 2008. A computationally efficient estimator for mutual information. Proc. R. Soc. A Math. Phys.

Eng. Sci. 464, 1203–1215. doi:10.1098/rspa.2007.0196

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176. doi:10.1016/j.jhydrol.2008.10.019

Frenzel, S., Pompe, B., 2007. Partial mutual information for coupling analysis of multivariate time series. Phys. Rev. Lett. 99. doi:10.1103/PhysRevLett.99.204101

Galelli, S., Castelletti, A., 2013. Tree-based iterative input variable selection for hydrological modeling. Water Resour. Res. 49, 4295–4310. doi:10.1002/wrcr.20339

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. doi:10.1016/j.envsoft.2014.08.015

Gómez-Herrero, G., Wu, W., Rutanen, K., Soriano, M.C., Pipa, G., Vicente, R., 2015. Assessing coupling dynamics from an ensemble of time series. Entropy 17, 1958–1970. doi:10.3390/e17041958

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., Hero, A.O., 2013. Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. Water Resour. Res. 49, 2253–2273. doi:10.1002/wrcr.20161

Goyal, M.K., Bharti, B., Quilty, J., Adamowski, J., Pandey, A., 2014. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst. Appl. 41. doi:10.1016/j.eswa.2014.02.047

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Hao, Z., Singh, V.P., 2015. Integrating entropy and copula theories for hydrologic modeling and analysis. Entropy 17, 2253–2280. doi:10.3390/e17042253

Harel, A., Ullman, S., Epshtein, B., Bentin, S., 2007. Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence. Vision Res. 47, 2010–2020. doi:10.1016/j.visres.2007.04.004

Harrold, T.I., Sharma, A., Sheather, S., 2001. Selection of a kernel bandwidth for measuring dependence in hydrologic time series using the mutual information criterion. Stoch. Environ. Res. Risk Assess. 15,

310–324. doi:10.1007/s004770100073

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed, Springer Series in Statistics. Springer New York, New York.

He, J., Valeo, C., Chu, A., Neumann, N.F., 2011. Prediction of event-based stormwater runoff quantity and quality by ANNs developed using PMI-based input selection. J. Hydrol. 400, 10–23. doi:10.1016/j.jhydrol.2011.01.024

Hejazi, M.I., Cai, X., 2009. Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm. Adv. Water Resour. 32, 582–593. doi:10.1016/j.advwatres.2009.01.009

Hnizdo, V., Darian, E.V.A., Fedorowicz, A., Demchuk, E., Li, S., Singh, H., 2007. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. J. Comput. Chem. 28, 655–668. doi:10.1002/jcc.20589

Huang, G.-B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man, Cybern. Part B Cybern. 42, 513–529. doi:10.1109/TSMCB.2011.2168604

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. doi: 10.1016/j.neucom.2005.12.126

Jain, A., Srinivasulu, S., 2006. Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. J. Hydrol. 317, 291–306. doi:10.1016/j.jhydrol.2005.05.022

Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. J. Hydrol. Eng. 729.

Khan, S., Bandyopadhyay, S., Ganguly, A.R., Saigal, S., Erickson III, D.J., Protopopescu, V., Ostrouchov, G., 2007. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 76. doi:10.1103/PhysRevE.76.026209

Khan, S., Ganguly, A.R., Bandyopadhyay, S., Saigal, S., Erickson III, D.J., Protopopescu, V., Ostrouchov, G., 2006. Nonlinear statistics reveals stronger tie between ENSO and the tropical hydrological cycle.

Geophys. Res. Lett. 33. doi:10.1029/2006GL027941

Kim, H.S., Eykholt, R., Salas, J.D., 1999. Nonlinear dynamics, delay times, and embedding windows. Phys. D Nonlinear Phenom. 127, 48–60.

Kleiner, A., Talwalkar, A., Sarkar, P., Jordan, M.I., 2014. A scalable bootstrap for massive data. J. R. Stat. Soc. Ser. B (Statistical Methodol. 76, 795–816. doi:10.1111/rssb.12050

Korshunov, P., Ooi, W.T., 2011. Video quality for face detection, recognition, and tracking. ACM Trans. Multimed. Comput. Commun. Appl. 7. doi:10.1145/2000486.2000488

Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 69, 66116–66138. doi:10.1103/PhysRevE.69.066138

Kuncheva, L.I., 2007. A stability index for feature selection, in: Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, AIA 2007. pp. 390–395.

LeBaron, B., Weigend, A.S., 1998. A bootstrap evaluation of the effect of data splitting on financial time series. IEEE Trans. Neural Networks 9, 213–220. doi:10.1109/72.655043

Lee, J., Maslove, D.M., 2015. Using information theory to identify redundancy in common laboratory tests in the intensive care unit Clinical decision-making, knowledge support systems, and theory. BMC Med. Inform. Decis. Mak. 15. doi:10.1186/s12911-015-0187-x

Li, X., Maier, H.R., Zecchin, A.C., 2015a. Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. Environ. Model. Softw. 65, 15–29. doi:10.1016/j.envsoft.2014.11.028

Li, X., Zecchin, A.C., Maier, H.R., 2015b. Improving partial mutual information-based input variable selection by consideration of boundary issues associated with bandwidth estimation. Environ. Model. Softw. 71, 78–96. doi:10.1016/j.envsoft.2015.05.013

Li, X., Zecchin, A.C., Maier, H.R., 2014. Selection of smoothing parameter estimators for general regression neural networks - Applications to hydrological and water resources modelling. Environ. Model. Softw. 59, 162–186. doi:10.1016/j.envsoft.2014.05.010

Lima, A.R., Cannon, A.J., Hsieh, W.W., 2015. Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation. Environ. Model. Softw. 73, 175–188. doi:10.1016/j.envsoft.2015.08.002

Liu, Z., Karam, L.J., 2005. Mutual information-based analysis of JPEG2000 contexts. IEEE Trans. Image Process. 14, 411–422. doi:10.1109/TIP.2004.841199

Ma, J., Sun, Z., 2011. Mutual information is copula entropy. Tsinghua Sci. Technol. 16, 51–54. doi:10.1016/S1007-0214(11)70008-6

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ. Model. Softw. 25, 891–909. doi:10.1016/j.envsoft.2010.02.003

Maier, H.R., Zecchin, A.C., Radbone, L., Goonan, P., 2006. Optimising the mutual information of ecological data clusters using evolutionary algorithms. Math. Comput. Model. 44, 439–450. doi:10.1016/j.mcm.2006.01.004

May, R.J., Dandy, G., Maier, H., 2011. Review of input variable selection methods for artificial neural networks. Artif. Neural Networks-Methodological Adv. Biomed. Appl. 19–44.

May, R.J., Dandy, G.C., Maier, H.R., Nixon, J.B., 2008a. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. Environ. Model. Softw. 23, 1289–1299. doi:10.1016/j.envsoft.2008.03.008

May, R.J., Maier, H.R., Dandy, G.C., Fernando, T.M.K.G., 2008b. Non-linear variable selection for artificial neural networks using partial mutual information. Environ. Model. Softw. 23, 1312–1326. doi:10.1016/j.envsoft.2008.03.007

Miaou, S. -P., 1990. A class of time series urban water demand models with nonlinear climatic effects. Water Resour. Res. 26, 169–178. doi:10.1029/WR026i002p00169

Mishra, A.K., Ines, A.V.M., Singh, V.P., Hansen, J.W., 2013. Extraction of information content from stochastic disaggregation and bias corrected downscaled precipitation variables for crop simulation. Stoch. Environ. Res. Risk Assess. 27, 449–457. doi:10.1007/s00477-012-0667-9

Moon, Y.-I., Rajagopalan, B., Lall, U., 1995. Estimation of mutual information using kernel density estimators. Phys. Rev. E 52, 2318–2321. doi:10.1103/PhysRevE.52.2318

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. doi:10.13031/2013.23153

Nichols, J.M., Seaver, M., Trickey, S.T., Salvino, L.W., Pecora, D.L., 2006. Detecting impact damage in experimental composite structures: An information-theoretic approach. Smart Mater. Struct. 15, 424–434. doi:10.1088/0964-1726/15/2/023

Penner, O., Grassberger, P., Paczuski, M., 2011. Sequence alignment, mutual information, and dissimilarity measures for constructing phylogenies. PLoS One 6. doi:10.1371/journal.pone.0014373

Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Sharma, A., 2000a. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 - A strategy for system predictor identification. J. Hydrol. 239, 232–239. doi:10.1016/S0022-1694(00)00346-2

Sharma, A., 2000b. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 - A nonparametric probabilistic forecast model. J. Hydrol. 239, 249–258. doi:10.1016/S0022-1694(00)00348-6

Sharma, A., Luk, K.C., Cordery, I., Lall, U., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 - Predictor identification of quarterly rainfall using ocean-atmosphere information. J. Hydrol. 239, 240–248. doi:10.1016/S0022-1694(00)00347-4

Sharma, A., Mehrotra, R., 2014. An information theoretic alternative to model a natural system using observational information alone. Water Resour. Res. 50, 650–660. doi:10.1002/2013WR013845

Šindelář, R., Babuška, R., 2004. Input selection for nonlinear regression models. IEEE Trans. Fuzzy Syst. 12, 688–696. doi:10.1109/TFUZZ.2004.834810

Singh, S.K., Jain, S.K., Bardossy, A., 2014. Training of artificial neural networks using information-rich data. Hydrology 1, 40–62.

Singh, V.P., 2013. Introduction, in: Entropy Theory and Its Application in Environmental and Water Engineering. Wiley-Blackwell, pp. 1–32. doi:10.1002/9781118428306.ch1

Stefánsson, A., Končar, N., Jones, A.J., 1997. A note on the gamma test. Neural Comput. Appl. 5, 131–133. doi:10.1007/BF01413858

Szabó, Z., 2014. Information theoretical estimators toolbox. J. Mach. Learn. Res. 15, 283–287.

Szabó, Z., Póczos, B., Lorincz, A., 2007. Undercomplete blind subspace deconvolution. J. Mach. Learn. Res. 8, 1063–1095.

Szabó, Z., Póczos, B., Lrincz, A., 2012. Separation theorem for independent subspace analysis and its consequences. Pattern Recognit. 45, 1782–1791. doi:10.1016/j.patcog.2011.09.007

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Tiwari, M.K., Adamowski, J.F., 2015. Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet–Bootstrap Machine-Learning Approach. J. Water Resour. Plan. Manag. 141. doi:10.1061/(ASCE)WR.1943-5452.0000454

Tran, H.D., Muttil, N., Perera, B.J.C., 2015. Selection of significant input variables for time series forecasting. Environ. Model. Softw. 64, 156–163. doi:10.1016/j.envsoft.2014.11.018

Tsimpiris, A., Vlachos, I., Kugiumtzis, D., 2012. Nearest neighbor estimate of conditional mutual information in feature selection. Expert Syst. Appl. 39, 12697–12708. doi:10.1016/j.eswa.2012.05.014

Van Hulle, M.M., 2005. Edgeworth Approximation of Multivariate Differential Entropy. Neural Comput. 17, 1903–1910. doi:10.1162/0899766054323026

Vejmelka, M., Paluš, M., 2008. Inferring the directionality of coupling with conditional mutual information. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 77. doi:10.1103/PhysRevE.77.026214

Wan Jaafar, W.Z., Liu, J., Han, D., 2011. Input variable selection for median flood regionalization. Water Resour. Res. 47. doi:10.1029/2011WR010436

Wong, J.S., Zhang, Q., Chen, Y.D., 2010. Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. Water Resour. Res. 46. doi:10.1029/2009WR008147

Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. J. Hydrol. doi:10.1016/j.jhydrol.2015.10.038

You, T., Fiedor, P., Hołda, A., 2015. Network analysis of the Shanghai stock exchange based on partial mutual information. J. Risk Financ. Manag. 8, 266–284.

Zeng, X., Durrani, T.S., 2011. Estimation of mutual information using copula density function. Electron. Lett. 47, 493–494. doi:10.1049/el.2011.0778

Appendix

This appendix contains online supplementary material related to Chapter 3 that may be accessed via: https://doi.org/10.1002/2015WR016959.

3. A.1. Introduction

This supporting information provides details regarding our second experiment on a real-world urban water demand (UWD) forecasting problem in Ottawa, Canada using kernel density estimation (KDE), k nearest-neighbors (KNN), partial mutual information selection (PMIS), partial correlation input selection (PCIS), and our proposed Edgeworth approximations (EA) and bootstrap rank-ordered conditional mutual information (broCMI) input variable selection algorithms. This experiment is carried out in the following steps: 1) generation of candidate input variable sets (from potentially useful explanatory variables); 2) dataset partitioning; 3) input variable selection using each algorithm; 4) mapping the selected input variable sets to forecasts using regression methods; and 5) evaluating the forecasts to identify the best input variable selection algorithm(s).

3. A.2. Generation of Candidate Input Variables

In order to generate candidate input variable sets for UWD forecasting at lead times of 1 and 3 days ahead we used historical UWD and available meteorological records provided to us by the City of Ottawa since these variables have been shown to have significant relationships with UWD at multiple forecast lead times (Adamowski et al., 2012; Adamowski, 2008; Bougadis et al., 2005). The explanatory variables used in this experiment included the daily antecedent precipitation index (*api*), rainfall (*rain*), maximum air temperature (*mat*), and historical UWD (average UWD in the 3W pressure zone, *a3w*, and average UWD in the Morgan's Grant pressure zone, *amg*). The identification of important explanatory variable time lags was carried out using nonlinear time series analysis since the relationship between UWD and its determinant factors are potentially nonlinear (House-Peters and Chang, 2011). Before we generated input variable sets we first confirmed the nonlinearity of each UWD and meteorological time series via the delay vector variance method (Gautama et al., 2003b, 2004a, 2004b) in order to justify the use of nonlinear time series analysis. Nonlinear time series analysis from chaos theory (please see the sub-section below) was used since UWD processes have been identified as a chaotic system within other urban water supply systems (Bai et al., 2014; Liu et al., 2004; Xu et al., 2015) providing a potentially useful nonlinear framework for describing the UWD processes considered in this work (which is an improvement over the classic autocorrelation and partial-autocorrelation analysis commonly applied in UWD forecasting (Tiwari and Adamowski, 2013, 2015)).

We found each of the UWD and meteorological time series considered in this work to have chaotic properties as positive largest Lyapunov exponents were estimated for each time series (an indicator of chaos, please see 3.10.2.1 for further details) using the methods suggested by Kodba et al. (2005) (results are not shown for brevity). We applied phase space reconstruction based on the time delay embedding window approach as described in sub-section 3.A.2.1 to each time series in this experiment producing the candidate input variable sets for the average UWD time series for both 3W and Morgan's Grant pressure zones. For example, if we consider the average UWD process in the 3W pressure zone the candidate input variable set contained the multivariate space spanned by the phase space reconstruction of *a3w*, *api*, rain, and *mat* time series combined into a single matrix. This candidate input variable set was then considered during input variable selection for the *a3w*(*t*+1) (1day), and *a3w*(t+3) (3 days) lead time forecast scenarios. The embedding window parameter determined via phase space reconstruction for each time series considered in this experiment were found to be: *a3w* – 15; *amg* – 19; *api* – 1; *rain* – 8; and *mat* – 21.

3. A.2.1. Phase Space Reconstruction

Phase space reconstruction stems from chaotic time series analysis (Packard et al., 1980; Takens, 1981). Chaotic time series analysis operates under the hypothesis that the studied system (e.g. urban water demand) is dynamic and sensitive to initial conditions. That is, the time evolution of a chaotic process may evolve along very different trajectories if the initial conditions specified for the system are slightly altered. The trajectories of the system are obtained through phase space reconstruction which maps a given univariate observational time series into an *m*-dimensional multivariate space, unfolding the dynamics of the system (which are considered the important time lags of the system – the number of important time lags specifies the dimension of the system, m). This new multivariate description of the time series may be used to forecast future trajectories or to uncover dependencies that may exist between different explanatory variables at different time lags (e.g. connections between urban water demand and rainfall at different time lags). Phase space reconstruction has been recently utilized for producing daily ensemble rainfall forecasts (Dhanya and Kumar, 2011), forecasting extreme air temperature events (Zhang et al., 2013), predicting daily evaporation (Baydaroğlu and Koçak, 2014), and forecasting urban water demand (Bai et al., 2014) (among other relevant applications).

In order to apply chaos based time series analysis, the time series of interest must first be identified as exhibiting chaotic characteristics. A chaotic system may be identified by transforming the observed time series into an *m*-dimensional space via phase space reconstruction and computing the largest Lyapunov exponent (Kodba et al., 2005; Wolf et al., 1985). Since the Lyapunov exponents measure the rate of

divergence (convergence) of two nearby initial points of a dynamical system, a positive (negative) Lyapunov exponent measures the average exponential divergence (convergence) of two nearby trajectories. By definition, a system that produces a positive largest Lyapunov exponent is dissipative and is considered chaotic (Kodba et al., 2005; Liu et al., 2004). We suggest the interested reader to review the work of Wolf et al. (1985), Liu et al. (2004), and Kodba et al. (2005) for necessary algorithms to implement the estimation of Lyapunov exponents and the work of Bradley and Kantz (2015) for general nonlinear time series analysis considerations. Once a time series has been identified as stemming from a chaotic system one may examine different phase space reconstructions to identify an optimal multivariate representation of the original time series. From a time series $y(t)$ with $t = 1, 2, \dots, N$ observations, one reconstructs a multivariate vector time series with evolution topologically equivalent to the original system via the transformation (Takens, 1981):

$$y(t) = y_t \rightarrow (y_t, y_{t-\tau}, \dots, y_{t-(m-1)\tau}) \tag{3.A.1}$$

where m is known as the embedding dimension and τ is known as the time delay and both together are known as the (time delay) embedding parameters. The embedding dimension determines the size of the phase space reconstruction while the time delay determines the dynamic components of the system. Generally, one chooses the time delay such that it induces the minimum level of dependence between the coordinates of the embedding vector $(y_t, y_{t-\tau}, \dots, y_{t-(m-1)\tau})$ (Bradley and Kantz, 2015). Often the phase space reconstruction step is carried out in two stages by: 1) determining the time delay (usually through autocorrelation or average mutual information) and 2) determining the embedding dimension (usually through the correlation sum or false nearest-neighbours) (Baydaroğlu and Koçak, 2014). However, since both the embedding dimension and the time delay are related to one another, the sequential estimation of both parameters can obscure the key dynamic features of the system which may be better captured by estimating both the embedding dimension and the time delay simultaneously (Kim et al., 1999). To this end, Gautama et al. (2003a) formulated the Entropy Ratio method based on multivariate differential entropy to compute the optimal time delay embedding parameters (τ_{opt}, m_{opt}) simultaneously and demonstrated this approach to outperform the sequential estimation of embedding parameters (using average mutual information and false nearest-neighbours) on a dynamical system benchmark (Henon Map) and real-world examples of physiological time series. The Entropy Ratio method has also proven useful in wind-profile forecasting using neural networks [Goh et al., 2006], in detecting signs of financial crisis (Addo et al., 2013), controlling the air temperature within

office buildings (Marvuglia et al., 2014), and detecting linear and nonlinear scaling laws in the Yangtze River flow (Wang et al., 2008). We also adopted the Entropy Ratio method for estimating the embedding parameters of time series in this study due to its simplicity and ability to select the time delay embedding parameters simultaneously. We direct the interested reader to Gautama et al. (2003a) for theoretical details.

To use the Entropy Ratio method for determining the optimal time delay embedding parameters (τ_{opt}, m_{opt}) of a time series one needs to provide a search-range over each embedding parameter as follows: $(\tau_{min}{:}\,\tau_{max}, m_{min}{:}\,m_{max})$. The issue of determining a suitable range for $\tau_{min}{:}\,\tau_{max}$ is carried out by three different methods in this study: autocorrelation function (ACF), partial autocorrelation function (PACF), and time-delayed mutual information (TDMI) (15 bins is used in this study) (Baydaroğlu and Koçak, 2014; Tsimpiris et al., 2012). One first computes the ACF, PACF, and TDMI of the time series which each return a set of significant time delays; the method (ACF, PACF, or TDMI) that produces the largest significant time delay is used for τ_{max}, and in all cases one uses $\tau_{min} = 1$. The range $(m_{min}{:}\,m_{max})$ is set at 1:5 since larger embedding dimensions require a large dataset (Maheswaran and Khosa, 2012). This was not available to the authors for this study.

Once the optimal time delay embedding parameters (τ_{opt}, m_{opt}) have been obtained the time delay embedding window (τ_w) may be determined from: $\tau_w = (m_{opt} - 1)\tau_{opt}$. There has been discussion, analysis, and results to demonstrate that one need only concern themselves with finding the embedding window of the time series in order to capture the system dynamics appropriately (Kim et al., 1999; Kugiumtzis, 1996). Gibson et al. (1992); Krakovská et al. (2015); Maus and Sprott (2011) (among others) conjecture that one only need determine τ_w as the isolation of the most important time delays reduces to a model dependent search (Small and Tse, 2004). For the present work, each time series $y(t)$ with $t = 1, 2, \ldots, N$ subjected to chaotic time series analysis (and characterized as chaotic) is transformed through phase space reconstruction using the embedding window strategy as follows:

$$y(t) = y_t \rightarrow (y_t, y_{t-1}, \ldots, y_{t-\tau_w}) \tag{3.A.2}$$

In this study the model dependent search (for important intermediate time delays, $t{:}\,t-1{:}\,t-\tau_w$) is accomplished via input variable selection (which is very well suited to the input variable selection problem since the larger set of time lags will make those that are significant easier to identify (Maus and Sprott,

2011)) and the suitability of the selected input variables are judged via evaluation of regression-based forecasts.

3. A.3. Dataset Partitioning

In this study training, cross-validation, and testing sets were used within all developed forecast models. Indices were chosen randomly based on the following percentages: 50% for training, 25% for cross-validation, and the remainder for testing. This partitioning was found to provide the best generalized models. The training set was used for calibrating the model parameters, the cross-validation set was used for choosing the best model and its parameters (using the root mean square error fitness function), and the test set was used to independently assess model performance (Hastie et al., 2009).

3. A.4. Selecting Input Variables

After generating the candidate input variable sets for each dataset and lead time forecast we selected input variable sets over the training indices using the six input variable selection algorithms: EA, KDE, KNN, PMIS, PCIS, and broCMI.

3. A.5. Forecasting Selected Input Variable Sets using Regression Models

Each UWD time series considered in this experiment was deemed nonlinear via the delay vector variance method (Gautama et al., 2003b, 2004a, 2004b), justifying the use of nonlinear regression models to produce forecasts for the time series in this experiment. The selected input variable sets chosen by each input variable selection algorithm were used to map the selected input variables into forecasts for each dataset and forecast lead time using regression methods. We note that it is possible for multiple input variable selection algorithms to select the same input variable set(s). In such cases, we only produced forecasts for the unique input variable sets and identified any cases where an optimal model was produced using selected input variables from multiple input variable selection algorithms.

In this experiment we considered the extreme learning machine (ELM) and artificial neural network (ANN) for producing UWD forecasts using inputs selected from the various input variable selection algorithms. Both methods are nonlinear regression paradigms with universal approximation capability (Ince, 2006) meaning that they are capable of approximating any continuous function within arbitrary precision given enough training data and/or model parameters. We considered the traditional feedforward backpropagation (FFBP)-ANN for comparison with the recently proposed extreme learning machine (ELM) (Huang et al., 2006) which shares the same topology as the FFBP-ANN but instead of iteratively training the network (as required by FFBP-ANN) the network parameters are determined through randomization

and solving a linear system of equations which provides a much more computationally feasible framework, especially for large scale forecasting tasks (Zhou et al., 2015). An advantage of ELM over ANN relates to the least-squares formulation of ELM which provides global solutions to the network parameters, unlike the FFBP-ANN which is based on iterative tuning of the network parameters in accordance with minimizing a fitness (objective) function (often leading to local minima solutions for the network parameters).

Since each modeling paradigm (ELM and ANN) requires initialization of their parameter space to begin the learning procedure, we randomly initialized each model five times and considered these five trials in identifying the best input variable sets (and network architecture). Since the ELM and ANN models were evaluated over numerous network architectures (due to varying the number of hidden layer neurons) the average fitness score over the five trials is used to identify the best network architecture. For each optimal model the five trials are then averaged to make the final forecast. Thus, each ELM and ANN forecast model is a combination of the forecasts produced by the five randomly initialized instances. We found this approach to smooth out random variations in the different model trials while still keeping the computation time at a feasible level. The best models were identified using the fitness function (root mean square error) on the cross-validation set.

3. A.5.1. Artificial Neural Network Theoretical Background

Since the application of the FFBP-ANN is very common within water resources forecasting applications (Abrahart et al., 2012) we direct the interested reader to Piotrowski et al. (2015) for theoretical development.

3. A.5.2. Artificial Neural Network Model Development

All ANN models developed in this study followed the best practices for ANN development outlined in Goyal and Burn (2012) and Adamowski and Karapataki (2010). The FFBP-ANN was used and its parameters (number of hidden neurons, network weights, and biases) were updated using the Levenberg-Marquardt backpropagation training algorithm. To further reduce the chance of poor generalization properties, early-stopping was used on a cross-validation set and set to six contiguous epochs. For the same reasons, the amount of training epochs were limited to 150 (which provided enough iterations to train each ANN model developed in this study). The initial step size of the network was set as 0.001 and the minimum performance gradient was set to 1e-7 (both default settings for the Levenberg-Marquardt algorithm in the Neural Network toolbox within Matlab). The ANN parameters were initialized using the Nguyen-Widrow method. Performing network parameter selection using this method provides more robust

estimates of the network parameters by taking into account over- and under-fitting. A three layer FFBP-ANN was used containing an input, hidden, and output layer. The hidden layer activation function was the tan-sigmoid and the output activation function was linear (default settings within the aforementioned Neural Network toolbox). All explanatory and response variable(s) were first normalized between [-1, 1] prior to model entry. Following recommendations made in Hecht-Nielsen (1989) we explored hidden neuron architectures spanning from 1 to $N^H \leq 2d + 1$ hidden neurons (Amiri et al., 2015). This upper limit for the number of ANN hidden layer neurons (along with slight variations, usually implementing a stricter upper limit, $N^H < 2d$) has been expressed by others in the literature (Shu and Burn, 2004; Shu and Ouarda, 2007). We explored different hidden neuron architectures for each selected input variable set considering between 1 and the upper limit of the hidden neurons as governed by each selected input variable set size. The fitness function was set as the root mean square error. The ANN models developed in this work were based on custom scripts incorporating functions from the Neural Network toolbox in Matlab.

3. A.5.3. Extreme Learning Machine Theoretical Background

The extreme learning machine is a recently proposed class of single layer feed-forward network (SLFN) with very similar performance to traditional ANN and least-squares support vector regression (LSSVR) models, with the potential to arrive at solutions in fractions of the time spent tuning ANN or LSSVR (Huang et al., 2012). The underlying differences with ELM (when compared to ANN) is that the input weights and hidden neuron biases are randomized (within the domain [-1, 1]) such that the output layer weights form a unique least-squares solution that may be solved by the Moore-Penrose generalized inverse technique, which yields a run-time improvement with respect to the gradient-based techniques that are generally employed for training ANN (Huang et al., 2006). The ELM method is regarded as a simple three-step procedure that requires no parameterization with the exception of determining a suitable amount of hidden neurons and their activation functions (which must be infinitely differentiable). Common activation function choices for the hidden neurons consist of sigmoid, sine, and hard limit (among many other choices).

The ELM SLFN with arbitrary distinct samples (x_t, y_t) for $t = 1, 2, \ldots, N$, where $x_t \in \mathbb{R}^d$ and $y_t \in \mathbb{R}$ (i.e. entries of the input/explanatory matrix X and the output/response vector Y at time t, respectively) with N^H randomly assigned hidden nodes and biases can be mathematically modeled as (Huang et al., 2006):

$$\sum_{i=1}^{N^H} \beta_i g_i(w_i \cdot x_t + z_i) = o_t \tag{3.A.3}$$

where β are the ELM output weights that are to be estimated from a given input-output dataset, g is the hidden layer activation function (taken as the sigmoid function in this work), w is the set of randomized input weights, z is the set of randomized bias/threshold parameters (such that $w_i \in \mathbb{R}^d$ and $z_i \in \mathbb{R}$), and o is the output of the network. The SLFN described in Eq. 3.A.3) can approximate the N dataset samples with zero error, i.e. $\sum_{t=1}^{N} \|o_t - y_t\| = 0$, suggesting that there exists w_i, z_i, and β_i such that (Huang et al., 2006):

$$\sum_{i=1}^{N^H} \beta_i g_i(w_i \cdot x_t + z_i) = y_t \tag{3.A.4}$$

for all $t = 1, 2, \ldots, N$. Thus, one may estimate β directly from the dataset by recasting the problem as a linear system of equations. The reformulation is written as follows [Huang et al., 2006]:

$$H\beta = Y \tag{3.A.5}$$

where:

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g_1(w_1 \cdot x_1 + z_1) & \cdots & g_{N^H}(w_{N^H} \cdot x_1 + z_{N^H}) \\ \vdots & \cdots & \vdots \\ g_1(w_1 \cdot x_N + z_1) & \cdots & g_{N^H}(w_{N^H} \cdot x_N + z_{N^H}) \end{bmatrix}_{N \times N^H} \tag{3.A.6}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{N^H}^T \end{bmatrix}_{N^H \times 1} \text{ and } Y = \begin{bmatrix} Y_1^T \\ \vdots \\ Y_N^T \end{bmatrix}_{N \times 1} \tag{3.A.7}$$

H is referred to as the hidden layer output matrix of the SLFN (and the T represents the transpose function). One may realize that β may be solved for by setting $\beta = H^{-1}Y$. This is accomplished in the ELM paradigm via the Moore-Penrose generalized inverse function [Huang et al., 2006]:

$$\hat{\beta} = H^{\dagger}Y \tag{3.A.8}$$

where † represents the Moore-Penrose generalized inverse and $\hat{\beta}$ represents the estimated output weights (the only parameters that need to be solved for in the SLFN ELM model). For newly acquired data (x_*) an ELM forecast ($\hat{y}$) may be achieved through (Akusok et al., 2015):

$$\hat{y} = \sum_{i=1}^{N^H} \hat{\beta}_i g_i(w_i \cdot x_* + z_i) \tag{3.A.9}$$

3. A.5.4. Extreme Learning Machine Development

The ELM method in comparison to FFBP-ANN does not require successive updates to model parameters (network weights and biases) and is solved as a linear system of equations using the Moore-Penrose generalized inverse technique (after randomly assigning the hidden layer neurons and biases with values between [-1, 1]). All ELM models designed in this study contain three layers: an input layer, hidden layer, and output layer. The hidden layer activation function was the sigmoid and the output activation function was linear. For consistency with the ANN models we used an upper limit (N^H) on the number of hidden neurons as suggested by Hecht-Nielsen (1989): $N^H \leq 2d + 1$ (Amiri et al., 2015). Thus, we explored different hidden neuron architectures for each selected input variable set considering between 1 and the upper limit of the hidden neurons as governed by each selected input variable set size. All explanatory and response variable(s) were first normalized between [0, 1] prior to model entry. The fitness function was set as the root mean square error. The ELM models used in this study were developed using custom Matlab scripts.

3. A.6. Forecast Performance Evaluation

In order to infer the adequacy of each input variable selection method in providing the best input variable sets for forecasting UWD we judge our forecasts using very common forecast performance metrics within the water resources domain: Nash-Sutcliffe Efficiency Index (NASH), root mean squared error (RMSE), and mean absolute percentage error (MAPE). (Since these performance metrics are very common within water resources forecasting we direct the reader to De Gooijer and Hyndman (2006); Moriasi et al. (2007) for these formulae.) In general, one seeks a NASH score as close to 1.0 as possible (indicating a perfect

fit), and one seeks RMSE and MAPE scores as close to 0 as possible (indicating null forecast error). If one obtains NASH > 0.5 then one can consider the forecast model satisfactory (Moriasi et al., 2007).

The optimal selected input variable sets are identified from the models with the best fitness performance on the cross-validation set as this set was used to judge the adequacy of each trained model's parametric settings. As usual, the test set is reserved for making out-of-sample inferences once a suitable model has been identified based on cross-validation performance (fitness) (Hastie et al., 2009).

References

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L., 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog. Phys. Geogr. 36, 480–513. doi:10.1177/0309133312444943

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Adamowski, J., Karapataki, C., 2010. Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms. J. Hydrol. Eng. 15, 729–743. doi:10.1061/(ASCE)HE.1943-5584.0000245

Adamowski, J.F., 2008. Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks. J. Water Resour. Plan. Manag. 134, 119–128. doi:10.1061/(ASCE)0733-9496(2008)134:2(119)

Addo, P.M., Billio, M., Guégan, D., 2013. Nonlinear dynamics and recurrence plots for detecting financial crisis. North Am. J. Econ. Financ. 26, 416–435. doi:10.1016/j.najef.2013.02.014

Akusok, A., Bjork, K.M., Miche, Y., Lendasse, A., 2015. High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications. IEEE Access 3, 1011–1025. doi:10.1109/ACCESS.2015.2450498

Amiri, B.J., Fatehi, I., Alizadeh, A., Adamowski, J., 2015. Modeling the relationship between catchment attributes and in-stream water quality. Water Resour. Manag. 29, 5055–5072. doi:10.1007/s11269-015-1103-y

Bai, Y., Wang, P., Li, C., Xie, J., Wang, Y., 2014. A multi-scale relevance vector regression approach for daily urban water demand forecasting. J. Hydrol. 517, 236–245. doi:10.1016/j.jhydrol.2014.05.033

Baydaroğlu, Ö., Koçak, K., 2014. SVR-based prediction of evaporation combined with chaotic approach. J. Hydrol. 508, 356–363. doi:10.1016/j.jhydrol.2013.11.008

Bougadis, J., Adamowski, K., Diduch, R., 2005. Short-term municipal water demand forecasting. Hydrol. Process. 19, 137–148. doi:10.1002/hyp.5763

Bradley, E., Kantz, H., 2015. Nonlinear time-series analysis revisited. Chaos 25. doi:10.1063/1.4917289

De Gooijer, J.G., Hyndman, R.J., 2006. 25 Years of Time Series Forecasting. Int. J. Forecast. 22, 443–473. doi:10.1016/j.ijforecast.2006.01.001

Dhanya, C.T., Kumar, D.N., 2011. Multivariate nonlinear ensemble prediction of daily chaotic rainfall with climate inputs. J. Hydrol. 403, 292–306. doi:10.1016/j.jhydrol.2011.04.009

Gautama, T., Mandic, D.P., Hulle, M.M. Van, 2003a. A differential entropy based method for determining the optimal embedding parameters of a signal, in: Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference On. p. VI-29-32 vol.6. doi:10.1109/ICASSP.2003.1201610

Gautama, T., Mandic, D.P., Van Hulle, M.M., 2004a. A novel method for determining the nature of time series. IEEE Trans. Biomed. Eng. Trans. Biomed. Eng. 51, 728–736. doi:10.1109/tbme.2004.824122

Gautama, T., Mandic, D.P., Van Hulle, M.M., 2004b. The delay vector variance method for detecting determinism and nonlinearity in time series. Phys. D Nonlinear Phenom. 190, 167–176. doi:10.1016/j.physd.2003.11.001

Gautama, T., Mandic, D.P., Van Hulle, M.M., 2003b. Indications of nonlinear structures in brain electrical activity. Phys. Rev. E 67, 046204. doi:10.1103/PhysRevE.67.046204

Gibson, J.F., Doyne Farmer, J., Casdagli, M., Eubank, S., 1992. An analytic approach to practical state space reconstruction. Phys. D Nonlinear Phenom. 57, 1–30. doi:10.1016/0167-2789(92)90085-2

Goyal, M.K., Burn, D.H., 2012. Discussion of "Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms" by Jan Adamowski and Christina Karapataki. J. Hydrol. Eng. 17, 833–834. doi:10.1061/(ASCE)HE.1943-5584.0000542

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed, Springer Series in Statistics. Springer New York, New York.

Hecht-Nielsen, R., 1989. Neurocomputing. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

House-Peters, L.A., Chang, H., 2011. Urban water demand modeling: Review of concepts, methods, and organizing principles. Water Resour. Res. 47, W05401. doi:10.1029/2010WR009624

Huang, G.-B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man, Cybern. Part B Cybern. 42, 513–529. doi:10.1109/TSMCB.2011.2168604

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications . Neurocomputing 70, 489–501. doi:10.1016/j.neucom.2005.12.126

Ince, H., 2006. Non-parametric regression methods. Comput. Manag. Sci. 3, 161–174. doi:10.1007/s10287-005-0006-4

Kim, H.S., Eykholt, R., Salas, J.D., 1999. Nonlinear dynamics, delay times, and embedding windows. Phys. D Nonlinear Phenom. 127, 48–60.

Kodba, S., Perc, M., Marhl, M., 2005. Detecting chaos from a time series. Eur. J. Phys. 26, 205–215. doi:10.1088/0143-0807/26/1/021

Krakovská, A., Mezeiová, K., Budáčová, H., 2015. Use of False Nearest Neighbours for Selecting Variables and Embedding Parameters for State Space Reconstruction. J. Complex Syst. 2015, 1–12. doi:10.1155/2015/932750

Kugiumtzis, D., 1996. State Space Reconstruction Parameters in the Analysis of Chaotic Time Series - the Role of the Time Window Length 95, 13–28. doi:10.1016/0167-2789(96)00054-1

Liu, J.-Q., Zhang, T.-Q., Yu, S.-K., 2004. Chaotic phenomenon and the maximum predictable time scale of observation series of urban hourly water consumption. J. Zhejiang Univ. Sci. 5, 1053–1059. doi:10.1631/jzus.2004.1053

Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

Marvuglia, A., Messineo, A., Nicolosi, G., 2014. Coupling a neural network temperature predictor and a fuzzy logic controller to perform thermal comfort regulation in an office building. Build. Environ. 72, 287–299. doi:10.1016/j.buildenv.2013.10.020

Maus, A., Sprott, J.C., 2011. Neural network method for determining embedding dimension of a time series. Commun. Nonlinear Sci. Numer. Simul. 16, 3294–3302. doi:10.1016/j.cnsns.2010.10.030

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. doi:10.13031/2013.23153

Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S., 1980. Geometry from a Time Series. Phys. Rev. Lett. 45, 712–716. doi:10.1103/PhysRevLett.45.712

Piotrowski, A.P., Napiorkowski, M.J., Napiorkowski, J.J., Osuch, M., 2015. Comparing various artificial neural network types for water temperature prediction in rivers. J. Hydrol. 529, 302–315. doi:10.1016/j.jhydrol.2015.07.044

Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resour. Res. 40, 1–10. doi:10.1029/2003WR002816

Shu, C., Ouarda, T.B.M.J., 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resour. Res. 43, 1–12. doi:10.1029/2006WR005142

Small, M., Tse, C.K., 2004. Optimal embedding parameters: A modelling paradigm. Phys. D Nonlinear Phenom. 194, 283–296. doi:10.1016/j.physd.2004.03.006

Takens, F., 1981. Detecting strange attractors in turbulence, in: Rand, D., Young, L.-S. (Eds.), Dynamical Systems and Turbulence, Warwick 1980. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 366–381.

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Tiwari, M.K., Adamowski, J.F., 2015. Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet–Bootstrap Machine-Learning Approach. J. Water Resour. Plan. Manag. 141. doi:10.1061/(ASCE)WR.1943-5452.0000454

Tsimpiris, A., Vlachos, I., Kugiumtzis, D., 2012. Nearest neighbor estimate of conditional mutual information in feature selection. Expert Syst. Appl. 39, 12697–12708. doi:10.1016/j.eswa.2012.05.014

Wang, G., Su, B., Kundzewicz, Z.W., Jiang, T., 2008. Linear and non-linear scaling of the Yangtze River flow. Hydrol. Process. 22, 1532–1536. doi:10.1002/hyp.6689

Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A., 1985. Determining Lyapunov exponents from a time series. Phys. D Nonlinear Phenom. 16, 285–317. doi:10.1016/0167-2789(85)90011-9

Xu, H., Ning, C., Li, S., Fan, X., 2015. Chaotic Dynamic Characteristics of an Integrated Water Management System. Br. J. Math. Comput. Sci. 10, 1–8. doi:10.9734/BJMCS/2015/19300

Zhang, Y., Yang, X., Zhang, L., Ma, W., Qiao, L., 2013. Set Pair Analysis Based on Phase Space Reconstruction Model and Its Application in Forecasting Extreme Temperature. Math. Probl. Eng. 2013, 1–7. doi:10.1155/2013/516150

Zhou, H., Huang, G.B., Lin, Z., Wang, H., Soh, Y.C., 2015. Stacked Extreme Learning Machines. IEEE Trans. Cybern. 45, 2013–2025. doi:10.1109/TCYB.2014.2363492

Connecting Text to Chapter 4

The new computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods from Chapter 3 form an integral component of the Wavelet Data-Driven Forecasting Framework (WDDFF) developed, tested, and applied in this chapter. The WDDFF is developed according to a set of best (correct) practices for wavelet-based forecasting for real-world applications. The new input variable selection methods from Chapter 3 are adopted to identify which wavelet and scaling coefficients (high- and low-frequency information extracted through the wavelet decomposition procedure) are to be used as input to a data-driven method for forecasting the target (water resources) process. The WDDFF is a useful framework for forecasting nonlinear and multiscale processes commonly encountered in the water resources domain.

This chapter was published in the Journal of Hydrology (Quilty and Adamowski, 2018). The format of the paper has been modified to ensure consistency with the style of this thesis. A list of references cited in this paper are available at the end of the chapter.

The author of the thesis was responsible for the development, testing, and application of the different methods and wrote the manuscript presented here. Prof. Adamowski, the supervisor of this thesis, provided valuable advice on all aspects of the research and contributed to the review and editing of the manuscript.

Chapter 4: Addressing the Incorrect Usage of Wavelet-based Hydrological and Water Resources Forecasting Models for Real-World Applications with Best Practices and a New Forecasting Framework

4.1. Abstract

Many recent studies propose wavelet-based hydrological and water resources forecasting models that have been incorrectly developed and that cannot properly be used for real-world forecasting problems. The incorrect development of these wavelet-based forecasting models occurs during wavelet decomposition (the process of extracting high- and low-frequency information into different sub-time series known as wavelet and scaling coefficients, respectively) and as a result introduces error into the forecast model inputs. The source of this error is due to the boundary condition that is associated with wavelet decomposition (and the wavelet and scaling coefficients) and is linked to three main issues: 1) using 'future data' (i.e., data from the future that is not available); 2) inappropriately selecting decomposition levels and wavelet filters; and 3) not carefully partitioning calibration and validation data. We identify that the discrete wavelet transform (DWT) multiresolution analysis (DWT-MRA) and maximal overlap discrete wavelet transform (MODWT) multiresolution analysis (MODWT-MRA), two commonly adopted methods in hydrological and water resources wavelet-based forecasting, suffer from these boundary conditions and cannot be used properly for real-world forecasting. However, by following a proposed set of best (correct) practices, we show that the MODWT and *à trous* algorithm (AT) can be used to correctly forecast target (e.g., hydrological and water resources) processes in real-world scenarios. In this vein, we contribute a set of best practices, which focusses on deriving "boundary-corrected" wavelet and scaling coefficients from time series data, overcoming the boundary condition issues and providing hydrological and water resources modellers with a justified and coherent strategy for developing wavelet-based forecasting models that may be used for real-world forecasting problems. We coalesce these best practices into a new forecasting framework named Wavelet Data-Driven Forecasting Framework (WDDFF) that uses a combination of input variable selection and data-driven models to convert "boundary-corrected" wavelet and scaling coefficients into forecasts of a target process. We demonstrate the superiority of WDDFF against benchmark forecasting models such as (non-wavelet-based) multiple linear regression, extreme learning machines, a second-order Volterra series model, and a recent wavelet-based forecasting model (that adopts the MODWT-MRA) through a real-world urban water demand forecasting experiment in Montreal, Canada.

Keywords: time series, forecasting, wavelets, best practices, hydrology, water resources

4.2. Introduction

The presence of multiple periodicities, transients, and/or trends in hydrological and water resources time series often present difficulties to traditional data-driven (time series, machine learning, computational intelligence, etc. (Solomatine and Ostfeld, 2008)) forecasting methods (e.g., autoregressive integrated moving average (ARIMA), multiple linear regression (MLR), artificial neural networks (ANN), etc.) (Adamowski et al., 2012). The fact that the wavelet transform (WT, or wavelets, for short) can decompose this information (wavelet decomposition) into separate (high- and low-frequency) sub-time series (wavelet and scaling coefficients) and provide a more coherent structure of the complex time series (and related variables) to a data-driven model, has often lead to significant improvements in forecast accuracy and for that reason has made it a popular tool in hydrological and water resources forecasting (Fahimi et al., 2017). Recent reviews (Afan et al., 2016; Dixit et al., 2016; Fahimi et al., 2017; Nourani et al., 2014; Sang, 2013; Yaseen et al., 2015) have highlighted the popularity of wavelet-based forecasting in hydrology and water resources; for example, using Scopus, Nourani et al. (2014) cited around 100 papers on this topic between 2003-2013 while at the time of writing this paper we identified around 110 papers between 2014-2018. However, despite the growing popularity of wavelet-based time series forecasts in hydrology and water resources, the proper design and interpretation of wavelet-based forecasts has not always been carefully scrutinized, often resulting in the design of forecasts that are invalid and that *cannot* be used in real-world scenarios (Du et al., 2017; Zhang et al., 2015) (see section 4.3 and 4.4 for details and section 4.6 for a real-world example). The largest misinterpretation regarding wavelet-based forecasts and a significant impediment to their use in real-world forecasting problems is related to wavelet decomposition (where isolation and extraction of relevant features from a given time series using WT occurs) and the inadvertent addition of error into the wavelet and scaling coefficients (sub-time series produced by the WT) – the main source responsible for generating this error in the wavelet and scaling coefficients is caused by what is known as the boundary condition (BC) (or boundary treatment) (Aussem et al., 1998; Bakshi, 1999; Maheswaran and Khosa, 2012a).

Below we briefly review basics of wavelet decomposition and then attempt a high-level discussion of the BC before a more detailed treatment in the next section. Essentially, wavelet decomposition can be understood as a convolution (which can be likened to quantifying the similarity) between a wavelet filter and a given time series, resulting in a set of wavelet and scaling coefficients (which contain high- and low-frequency information from the original time series). The high- and low-frequency content of the wavelet

and scaling coefficients depends on: 1) the decomposition level, which represents the different scales of change in the time series and 2) the wavelet filter, whose width and shape determines the type of features that can be extracted from the time series (e.g., trends, transients, periodicities, polynomial behaviour, etc.) (Maheswaran and Khosa, 2012a). Now, the BC arises during wavelet decomposition when one seeks to calculate wavelet or scaling coefficients at a particular time and scale that cannot be calculated *correctly* (i.e., without introducing error) because the entire range of time series observations that are required in the calculation are not available (we will refine this explanation later with some detailed examples). Different WT (e.g., Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), the Maximal Overlap DWT (MOWDT), and *à trous* algorithm (AT)) slightly vary in how wavelet decomposition is carried out and this (as will be discussed in detail in the next section) leads to important differences in how one is to evaluate and treat the BC associated with a particular WT. Before giving technical details, it is important to take note that the BC, if not properly treated, will add error into the wavelet and scaling coefficients in three distinct ways:

1. Through the inclusion of 'future data' (i.e., time series observations at future times ($> t$) are used in the calculation of wavelet or scaling coefficients at time t, see section 4.3.3.1.);
2. The improper selection of decomposition levels and wavelet filters (i.e., decomposition levels that are too high and/or wavelet filters that are too wide for a given time series – see section 4.3.3.2.); and
3. The incorrect partitioning of data into calibration and validation sets (i.e., due to issues 1 and 2 above, the calibration wavelet and scaling coefficients may include information from the time series observations that belong to the validation dataset and vice versa – see section 4.3.3.3.).

Before proceeding, we must make clear two key items:

1. We make an important distinction between error introduced to a wavelet-based forecasting model via the BC and errors encountered in general modeling problems. Error due to the BC is avoidable if proper precautions (which will be discussed in detail in the next section) are taken, while typical sources of error (e.g., due to model parameterization, initial conditions, model structure, etc.) are inherent in any modeling exercise (i.e., whether WT are used or not). This work is only focussed on the former.
2. Real-world wavelet-based forecasting is defined in this study as the use of a wavelet-based forecasting model that is calibrated on a sequential historical record (i.e., for $t \in 0,1,\ldots N_c - 1$, where N_c is the number of calibration records) and subsequently used to generate forecasts for a

validation record (i.e., for $t \in N_c, N_c + 1, \ldots N_c + N_v - 1$, where N_v is the number of validation records), where the validation record is unknown at the time of model calibration and becomes available to the forecaster only one record at a time (i.e., at $t = N_c + 1$ only time series records $t \in 0,1, \ldots N_c + 1$ are known to the forecaster, while time series records $t \in N_c + 2, N_c + 3, \ldots N_c + N_v - 1$ remain unknown (since they would not be available to the forecaster in a real-world scenario). By assuming this definition for real-world wavelet-based forecasting, wavelet decomposition must be carried out on the validation records sequentially, as would be the case in a real-world scenario, when new data becomes available to the forecaster only after it has been observed (measured), i.e., in real-time. It is the opinion of the authors that this is the correct way to treat the development of wavelet-based forecasting models for use in real-world scenarios.

In advance of our detailed exploration of each of the BC-related problems (i.e., the 'future data' issue, improper selection of decomposition levels and wavelet filters, and calibration and validation set partitioning) in the next section, it is important to note that when *any* of these sources of error are included in the wavelet coefficients and subsequently used within a wavelet-based forecasting model, error is introduced and carried throughout the entire model development and testing stages (model calibration and validation, respectively). This error, which is difficult to characterize (since it is related to many different factors, e.g., the type of WT, the decomposition level, wavelet filter, and dataset partitioning), can lead to both overly optimistic and pessimistic forecasts and is the result of the *incorrect* development of a wavelet-based forecasting model by neglecting to address the BC (Du et al., 2017). By not addressing *each* of the BC, one *cannot* use WT correctly for real-world forecasting problems (i.e., without introducing error due to the BC). Therefore, if one is to correctly develop a wavelet-based forecasting model it is important to understand how wavelet decomposition is related to the WT (e.g., DWT, MODWT, or AT) selected by the forecaster and how to adjust for the resulting BC. Of significance, we note (and will show in the next section) that the most popular WT, the DWT, used in hydrology and water resources wavelet-based forecasting studies (e.g., Altunkaynak and Nigussie (2017); Deo et al. (2017c); Dixit et al. (2016); Fahimi et al. (2017); Nourani et al. (2014); Rezaie-balf et al. (2017); Samadianfard et al. (2018)), *cannot* be corrected in terms of the above-mentioned BC and therefore *should not* be used for real-world forecasting applications (Du et al., 2017). As pointed out in Du et al. (2017), since the DWT does not address any of the boundary conditions during wavelet decomposition, the (incorrectly developed) DWT-based forecasting models often result in much better performance than what is realistically achievable. This is an important issue since an estimated 90% of papers (191/210) identified on Scopus between 2003 and 2018 use the DWT. However, we will show how the BC associated

with both the MODWT and AT can be straightforwardly addressed. Once the BC-related adjustments have been made, both the MODWT and AT can be used correctly for real-world forecasting problems. To illustrate the importance of the BC and its connection to these WT, we compare and contrast the DWT, MODWT, and AT algorithms showing where the BC arises for each method and how it can be addressed for the MODWT and AT (but not the DWT), allowing for the proper development of wavelet-based forecasting models for real-world use. Interestingly, the comparison of the BC-related issues arising from each of these WT (i.e., DWT, MODWT, and AT) in terms of (wavelet-based) forecasting applications has yet to be carried out in the literature.

Since a large majority of studies mentioned in numerous review articles on wavelet-based forecasting in hydrology and water resources (Afan et al., 2016; Dixit et al., 2016; Fahimi et al., 2017; Nourani et al., 2014; Yaseen et al., 2015) - including more recent studies (Bachour et al., 2016; Barzegar et al., 2018, 2017; Baydaroğlu et al., 2017; Maslova et al., 2016; Prasad et al., 2017; Rezaie-balf et al., 2017; Samadianfard et al., 2018; Seo et al., 2017; Shi et al., 2018; Yadav and Eliza, 2017; Yu et al., 2018) - propose wavelet-based forecasting models that *cannot* be used in real-world forecasting problems, due to (different combinations of) the BC-related issues mentioned above, we feel that it is very important to address this gap in the literature. Therefore, we make it our goal to explore and address the above-mentioned BC-related problems connected to DWT, MODWT, and AT, enabling the correct development of wavelet-based forecasts that may be used for real-world applications. To meet this goal, we contribute to the literature in two main ways:

1. By developing a set of best (and correct) practices for wavelet-based forecasting that addresses each of the three BC-related problems (e.g., 'future data', selection of decomposition levels and wavelet filters, and calibration and validation set partitioning) through careful analysis of different WT (i.e., the DWT, MODWT and AT) and their wavelet decomposition procedure and
2. Adopting our proposed best practices in a new wavelet-based forecasting framework that can be utilized in real-world forecasting scenarios.

Both contributions are important for the field of hydrology and water resources, as most wavelet-based forecasting studies in hydrology and water resources are incorrect and there currently does not exist: 1) a comprehensive study of the differences between the DWT, MODWT, and AT in terms of each BC; 2) best practices that addresses each of the BC; and 3) a general forecasting framework incorporating a set of best practices that can be applied to any type of data-driven model and be used for real-world (wavelet-based) forecasting scenarios. These important contributions are realized through the following steps:

1. We systematically review wavelet decomposition for the most popular WT used in hydrological and water resources forecasting - the DWT, MODWT, and AT - by focussing on the three BC-related issues (i.e., 'future data', selection of decomposition levels and wavelet filters, and calibration and validation set partitioning) in terms of both theory and practical application, showing how they are commonly misunderstood and misused, resulting in invalid forecasts;
2. We then introduce a set of best (correct) practices to address each of the BC-related issues that ultimately results in the calculation of what we term, "boundary-corrected" wavelet coefficients;
3. We show how "boundary-corrected" wavelet coefficients can be determined through two different WT, the MODWT and AT (that until now, have not been studied together in the wavelet-based forecasting literature);
4. We review the differences between the MODWT and AT, in terms of their wavelet decomposition strategies, discovering that for forecasting applications the MODWT can only be used for pre-processing input data, while the AT can be used for pre-processing both target and input data (to the best of our knowledge, this finding has been overlooked in the wavelet-based forecasting literature);
5. We develop a new wavelet-based forecasting framework, Wavelet Data-Driven Forecasting Framework (WDDFF), that adopts our best practices and uses the "boundary-corrected" wavelet coefficients, derived from either the MODWT or AT, and any data-driven model (e.g., ARIMA, MLR, ANN, etc.) to forecast a target process; and
6. We use WDDFF in a real-world daily urban water demand (UWD) forecasting exercise in Montreal, Canada, demonstrating its usefulness and potential for forecasting real-world hydrological and water resources time series.

To introduce our proposed best practices for wavelet-based forecasting and the WDDFF, the rest of this work is arranged in the following order. In section 4.3, we introduce the different wavelet decomposition approaches commonly used in wavelet-based hydrological and water resources forecasting models, discuss how the BC-related issues apply to each wavelet decomposition approach, and show how these issues can be solved. In section 4.4, we show how the solutions to each of the BC-related issues can be formulated as a set of best practices and then used within a new wavelet-based forecasting framework (WDDFF). Section 4.5 explains our real-world forecasting experiment used to demonstrate the usefulness of WDDFF. Section 4.6 focusses on the results of our experiment and discusses their significance. Finally, section 4.7 concludes the study by highlighting its main contributions, originality, and suggesting avenues for future research.

4.3. Theoretical Basis for Best (Correct) Practices in Wavelet-based Forecasting

4.3.1. An Overview of the Different Wavelet Transforms used in Hydrology and Water Resources for Wavelet-based Forecasting

Our focus in this sub-section is on the theoretical details of (and comparison between) the DWT, MODWT, and AT. We chose to focus on these three WT since the DWT is the most commonly used in hydrology and water resources wavelet-based forecasting (and, as we will show below, results in incorrect forecasts) and because the MODWT and AT are the only WT that can be used correctly for real-world forecasting.

In Table 4.1 we have provided the equations (Eqs. 4.1-4.9) needed to perform wavelet decomposition using the DWT (Eqs. 4.1-4.2), MODWT (Eqs. 4.4-4.5), and AT (Eqs. 4.7-4.8) (Aussem et al., 1998; Maheswaran and Khosa, 2012a; Percival and Walden, 2000). Equation 4.3, 4.6, and 4.9 can be used to reconstruct the time series (i.e., back to its original values) via the DWT, MODWT, and AT, respectively. We note that we have not included the CWT in our discussion at it suffers from the same BC-related issues as the DWT (less the dyadic condition mentioned below), typically requires many *ad-hoc* corrections when used for forecasting, and is computationally expensive (Adamowski, 2008a, 2008b; Nourani et al., 2009b; Rathinasamy et al., 2014). Previously studied details concerning the wavelet decomposition process that results in the (DWT, MODWT, and AT) wavelet (W_j, $\widetilde{W}_j$, and $\widetilde{W}_{j,t}^{a}$) and scaling coefficients (V_j, $\tilde{V}_j$, and $\tilde{V}_{j,t}^{a}$) (where $j \in 1,2,\dots,J$ and J is the decomposition level), including information on their properties and those of their wavelet and scaling filters, can be found in section 4.A of our Supplementary Material [appendix]. Instead, we focus on the key differences between the DWT, MODWT, and AT wavelet and scaling coefficients to prepare our discussion on how each of the BC-related issues (i.e., 'future data', selection of decomposition levels and wavelet filters, and calibration and validation set partitioning) are related to each WT.

The key difference between the DWT and the MODWT and AT is that the DWT involves decimation. Therefore, DWT wavelet and scaling coefficients at a given level j have 2^j fewer coefficients than at scale $j-1$. The decimation of the DWT causes several major issues for wavelet-based forecasting applications, thus making it invalid for real-world forecasting problems: 1) it requires the time series to be an integer multiple of 2^J, 2) it is shift-invariant; and 3) it is sensitive to adding new data points (Maheswaran and Khosa, 2012a; Percival and Walden, 2000; Walden, 2001) (each point is explained in more detail in section 4.A of our Supplementary Material [appendix]). The drawbacks of the DWT (mentioned in points 1-3) can be overcome through the MODWT and AT by simply rescaling the wavelet (h_l) and scaling filters (g_l) used by the DWT as follows: $\tilde{h}_l \equiv h_l/\sqrt{2}$ and $\tilde{g}_l \equiv g_l/\sqrt{2}$. The re-scaling of the

DWT filters conserves energy and obviates the need to sub-sample in the MODWT and AT (i.e., dyadic sub-sampling is avoided resulting in wavelet and scaling coefficients that are the same length as the original time series) (Walden, 2001).

From Table 4.1 (Eqs. 4.5 and 4.6), it is clear that the MODWT and AT share the same scaling coefficients (since they are calculated using the same low-pass filter $\tilde{g}$). However, their wavelet coefficients differ. The MODWT uses a high-pass filter ($\tilde{h}$) to calculate its wavelet coefficients, while the AT calculates its wavelet coefficients at scale j by differencing the scaling coefficients at levels j and $j-1$. Therefore, the AT permits an additive reconstruction of a given time series (Eq. 4.9) while the MODWT must apply an iterative reconstruction of the time series (through the reverse pyramid algorithm, Eq. 4.8) - this interesting result has important consequences for forecasting applications and is discussed later in this section.

Now that we have noted the key differences between the DWT, MODWT, and AT in terms of wavelet decomposition (and reconstruction), we can now look at how the wavelet and scaling coefficients produced by the different WT have been used and misused in hydrology and water resources wavelet-based forecasting studies. Afterwards, our discussion will focus on the implications of the BC related to each WT in terms of wavelet-based forecasting.

Table 4.1. Wavelet decomposition formulae for different wavelet transforms

	Wavelet Coefficients	Scaling Coefficients	Reconstruction Coefficients
DWT	$W_{j,t} = \sum_{l=0}^{L-1} h_l V_{j-1,2t+1-l \bmod N_{j-1}}$ (4.1)	$V_{j,t} = \sum_{l=0}^{L-1} g_l V_{j-1,2t+1-l \bmod N_{j-1}}$ (4.4)	$V_{j-1,t} = \sum_{l=0}^{L-1} h_l W^{\uparrow}_{j,t+l \bmod N_{j-1}} + \sum_{l=0}^{L-1} g_l V^{\uparrow}_{j,t+l \bmod N_{j-1}}$ (4.7)
MODWT	$\widetilde{W}_{j,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{V}_{j-1,t-2^{j-1}l \bmod N}$ (4.2)	$\tilde{V}_{j,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j-1,t-2^{j-1}l \bmod N}$ (4.5)	$\tilde{V}_{j-1,t} = \sum_{l=0}^{L-1} \tilde{h}_l \widetilde{W}_{j,t+2^{j-1}l \bmod N} + \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j,t+2^{j-1}l \bmod N}$ (4.8)
AT	$\widetilde{W}^a_{j,t} = \tilde{V}^a_{j-1,t} - \tilde{V}^a_{j,t}$ (4.3)	$\tilde{V}^a_{j,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}^a_{j-1,t-2^{j-1}l \bmod N}$ (4.6)	$\tilde{V}^a_{j-1,t} = \widetilde{W}^a_{j,t} + \tilde{V}^a_{j,t}$ (4.9)
Notes (DWT)	$t = 0, 1, \ldots, N_{j-1} - 1$; where: $N_j \equiv N/2^j$		
	$W^{\uparrow}_{j,t} \equiv \begin{cases} 0, & t = 0, 2, \ldots, N_{j-1} - 2 \\ W_{j,\frac{t-1}{2}}, & t = 1, 3, \ldots, N_{j-1} - 1 \end{cases}$ (note that the scaling coefficients are defined similarly)		
Notes (MODWT/ AT)	$t = 0, 1, \ldots, N - 1$		

t is a time index; $V_{0,t} \equiv \tilde{V}_{0,t} \equiv \tilde{V}^a_{0,t}$ represents the original time series at time t; $W_{j,t}$ ($V_{j,t}$) represents the jth level wavelet (scaling) coefficient for the discrete wavelet transform (DWT) at time t; $\widetilde{W}_{j,t}$ ($\tilde{V}_{j,t}$) represents the jth level wavelet (scaling) coefficient for the maximal overlap discrete wavelet transform (MODWT) at time t; $\widetilde{W}^a_{j,t}$ ($\tilde{V}^a_{j,t}$) represents the jth level wavelet (scaling) coefficient for the *à trous* algorithm (AT) at time t; h (g) is a wavelet (scaling) filter; N is the number of samples; and mod refers to the modulo operator. Note that $\tilde{g}_l \equiv g_l/\sqrt{2}$; $\tilde{h}_l \equiv h_l/\sqrt{2}$; and $g_l \equiv (-1)^{l+1} h_{L-1-t}$ where L is the length of the wavelet (scaling) filter (Maheswaran and Khosa, 2012a; Walden, 2001)

4.3.2. Use and Misuse of Wavelet and Scaling Coefficients in Hydrology and Water resources Wavelet-based Forecasting

Since the DWT wavelet ($\{W_1, W_2, \dots, W_J\}$) and scaling coefficients (V_J) are decimated at each scale, those adopting the DWT for wavelet-based forecasting must use multiresolution analysis (MRA) (Mallat, 1989) to convert the decimated wavelet and scaling coefficients in to detail ($\{D_1, D_2, \dots, D_J\}$) and approximation coefficients (S_J), respectively, that are of the same length as the original time series. Afterwards, the detail(s) and approximation coefficients are used as input to a given forecasting model (Kişi, 2011). We refer to the conversion of wavelet and scaling coefficients to detail and approximation coefficients, respectively, as DWT-based MRA or DWT-MRA (although it is often referred to in the literature as simply DWT (Sang, 2013) or MRA (Dixit et al., 2016)). Similar to the DWT-MRA, the same process can be followed for the MODWT to produce the MODWT-MRA detail(s) ($\widetilde{D}_1, \widetilde{D}_2, \dots, \widetilde{D}_J$) and approximation coefficients ($\tilde{S}_J$) (Percival and Walden, 2000).

An interesting characteristic of the DWT-MRA and MODWT-MRA is that similar to the AT, the detail(s) and approximation coefficients provide an additive reconstruction of the time series. Because of this property, the AT wavelet and scaling coefficients are often referred to as detail and approximation coefficients in the literature (Aussem et al., 1998; Maheswaran and Khosa, 2012a)). The additive reconstruction property of the DWT-MRA, MODWT-MRA, and AT is given as (Aussem et al., 1998; Percival and Walden, 2000):

$$X_t = \sum_{j=1}^{J} D_{j,t} + S_{J,t} = \sum_{j=1}^{J} \widetilde{D}_{j,t} + \tilde{S}_{J,t} = \sum_{j=1}^{J} \widetilde{W}_{j,t}^{a} + \tilde{V}_{J,t}^{a} \tag{4.10}$$

However, *unlike* the AT, the DWT-MRA and MODWT-MRA suffer from a serious drawback in terms of their usefulness in real-world forecasting applications: they require 'future data' in order to calculate detail and/or approximation coefficients at a given point in time. We give a detailed example showing why this occurs in the next sub-section. An inadvertent misunderstanding of this important drawback, has led to the detail(s) and approximation coefficients derived from both the DWT-MRA and MODWT-MRA being commonly used in hydrological and water resources wavelet-based forecasting models resulting in invalid forecasts (e.g., Alizadeh et al. (2017); Bachour et al. (2016); Barzegar et al. (2017a, 2017b); Ebrahimi and Rajaee (2017); Ghazvinei et al. (2017); He et al. (2015); Maslova et al. (2016); Patil and Deo (2017); Prasad et al. (2017); Seo et al. (2017); Shoaib et al. (2017); Yang et al. (2014)). While the wavelet (or detail) and

scaling (or approximation) coefficients derived by the AT have been used in a number of studies (Adamowski and Chan, 2011; Belayneh et al., 2016b; Bogner and Kalas, 2008; Djerbouai and Souag-Gamane, 2016; Khalil et al., 2015; Maheswaran and Khosa, 2012b; Rathinasamy et al., 2013), it is surprising to note (given that the MODWT-MRA detail(s) and approximation coefficients have been used in several studies) that the MODWT wavelet and scaling coefficients have yet to be used in any hydrological or water resources wavelet-based forecasting studies. Therefore, this study introduces the first use of the MODWT wavelet and scaling coefficients for wavelet-based forecasting of hydrological and water resources.

The two most common ways in which the detail(s) and approximations from the DWT-MRA, MODWT-MRA, and AT are used for wavelet-based forecasting are:

1. Decomposing only the explanatory variables into detail(s) and approximation coefficients and forecasting the target directly (Alizadeh and Kavianpour, 2015; Kişi, 2011; Samadianfard et al., 2018), also known as the direct approach (Nguyen and Nabney, 2010).
2. Decomposing both target and explanatory variables into detail(s) and approximation coefficients and forecasting each decomposed target series (detail or approximation) separately using the decomposed explanatory variables, and then aggregating the separate target series predictions using additive reconstruction (Eq. 4.10) (Barzegar et al., 2017; Shafaei and Kisi, 2016; Yadav and Eliza, 2017), also known as the multicomponent approach (Nguyen and Nabney, 2010). Another similar (but less popular) approach is to use another model (e.g., ANN) to reconstruct the separate target series predictions instead of Eq. 4.10 (Cannas et al., 2006; Kim and Valdés, 2003). This approach is unnecessary since it adds additional errors to the forecast due to the use of a model for reconstruction instead of the simple additive reconstruction offered by Eq. 4.10.

It is significant to note that both wavelet-based forecasting approaches (e.g., the direct and multicomponent methods) can lead to different levels of forecast accuracy and it is therefore important to consider both approaches in order to identify the best possible forecast for a given dataset (Nguyen and Nabney, 2010) a result that is confirmed later in this paper (section 4.6) and which is seldom considered in wavelet-based hydrological and water resources forecasting studies. We will also show later that both the multicomponent and direct approaches can be used by the AT for real-world wavelet-based forecasting, while only the direct approach is applicable to the MODWT.

We now discuss how the different BC-related issues (i.e., the 'future data' issue; inappropriately selecting decomposition levels and wavelet filters; and not carefully partitioning calibration and validation data) affect each of these wavelet-based forecasting approaches and propose our solutions to these problems.

4.3.3. Boundary Condition-related Issues and Solutions in Wavelet-based Forecasting

It is important to re-iterate that our intention is to circumvent the use of wavelet and scaling coefficients impacted by BC and to only use "boundary-corrected" wavelet and scaling coefficients (that do not suffer from *any* boundary condition) for wavelet-based forecasting. By meeting these conditions, "boundary-corrected" wavelet and scaling coefficients can be used correctly for real-world forecasting applications. We start by showing how the 'future data' issue is related to the DWT-MRA and MODWT-MRA and how it is avoided by the MODWT and AT.

4.3.3.1. The 'Future Data' Issue

The 'future data' issue occurs when a given WT (e.g., DWT-MRA and MODWT-MRA) requires time series observations existing after (ahead of) time t in order to perform wavelet decomposition on a time series observation at time t. In other words, data from the future of the time series ($> t$) is required to calculate a wavelet or scaling coefficient in the present (at t). It becomes obvious that if time series observations $> t$ are unavailable to the forecaster (which is the case in real-world scenarios), then they are unable to perform wavelet decomposition at time t. It is therefore of great importance that WT that do not use 'future data' (such as the MODWT and AT) are adopted in real-world wavelet-based forecasting models and those that require 'future data' (such as the DWT-MRA and MODWT-MRA) are avoided.

In Table 4.2, we give a very simple example showing how the 'future data' BC issue occurs in the DWT-MRA and MODWT-MRA and how it is avoided by the MODWT and AT using the formulae from Table 4.1. In this example, we consider a time series of length 32 ($N = 32$), a wavelet filter of length four ($L = 4$) (e.g., the Daubechies 4 filter– see section 4.A of the Supplementary Material [appendix]), and a decomposition level of one ($J = 1$). We calculate the first level detail coefficient at time index 16 ($t = 16$) for the DWT-MRA ($D_{1,16}$), MODWT-MRA ($\widetilde{D}_{1,16}$), and AT ($\widetilde{D}^{a}_{1,16}$) using Eqs. 4.7-4.9 (see also section 4.A of the Supplementary Material [appendix] for how the DWT-MRA and MODWT-MRA detail coefficients are calculated). Figure 4.1 provides a graphical depiction of the 'future data' issue showing which of the original time series observations are used for calculating the detail coefficients at $t = 16$ by the DWT-MRA, MODWT-MRA, and AT (note that the MODWT uses the same observations as the AT when calculating its wavelet and scaling coefficients)

Table 4.2. Example of wavelet decomposition using the different wavelet transforms (in Table 4.1)

Detail Coefficients	DWT - MRA	$D_{1,16}$	=	$\sum_{l=0}^{3} h_l\, W^{\uparrow}_{1,16+l \bmod 32} = h_0 W^{\uparrow}_{1,16} + h_1 W^{\uparrow}_{1,17} + h_2 W^{\uparrow}_{1,18} + h_3 W^{\uparrow}_{1,19} = h_0(0) + h_1 W_{1,8} + h_2(0) + h_3 W_{1,9}$
		$D_{1,16}$	=	$h_1\left(h_0 V_{0,17} + h_1 V_{0,16} + h_2 V_{0,15} + h_3 V_{0,14}\right) + h_3\left(h_0 V_{0,19} + h_1 V_{0,18} + h_2 V_{0,17} + h_3 V_{0,16}\right)$
	MODWT - MRA	$\widetilde{D}_{1,16}$	=	$\sum_{l=0}^{3} \tilde{h}_l\, \widetilde{W}_{1,16+l \bmod 32} = h_0 \widetilde{W}_{1,16} + h_1 \widetilde{W}_{1,17} + h_2 \widetilde{W}_{1,18} + h_3 \widetilde{W}_{1,19}$
		$\widetilde{D}_{1,16}$	=	$\tilde{h}_0\left(\tilde{h}_0 V_{0,16} + \tilde{h}_1 V_{0,15} +, \dots, + \tilde{h}_3 V_{0,13}\right) + \tilde{h}_1\left(\tilde{h}_0 V_{0,17} + \tilde{h}_1 V_{0,16} +, \dots, + \tilde{h}_3 V_{0,14}\right) +, \dots, + \tilde{h}_3\left(\tilde{h}_0 V_{0,19} + \tilde{h}_1 V_{0,18} +, \dots, + \tilde{h}_3 V_{0,16}\right)$
	AT	$\widetilde{D}^{a}_{1,16}$	=	$\tilde{V}^{a}_{0,16} - \tilde{V}^{a}_{1,16} = V_{0,16} - V_{1,16} = V_{0,16} - \sum_{l=0}^{3} \tilde{g}_l\, V_{0,t-l \bmod 32} = V_{0,16} - \left(\tilde{g}_0 V_{0,16} + \tilde{g}_1 V_{0,15} + \tilde{g}_2 V_{0,14} + \tilde{g}_3 V_{0,13}\right)$
Notes		$W^{\uparrow}_{1,t}$	≡	$\begin{cases} 0, & t = 0, 2, \dots, N-2 \\ W_{1,\frac{t-1}{2}}, & t = 1, 3, \dots, N-1 \end{cases}$
		$W_{1,t}$	=	$\sum_{l=0}^{3} h_l\, V_{0,2t+1-l \bmod 32} = h_0 V_{0,2t+1 \bmod 32} + h_1 V_{0,2t \bmod 32} + h_2 V_{0,2t-1 \bmod 32} + h_3 V_{0,2t-2 \bmod 32}$
		$\widetilde{W}_{1,t}$	=	$\sum_{l=0}^{3} \tilde{h}_l\, \tilde{V}_{0,t-l \bmod 32} = \sum_{l=0}^{3} \tilde{h}_l\, V_{0,t-l \bmod 32} = \tilde{h}_0 V_{0,t \bmod 32} + \tilde{h}_1 V_{0,t-1 \bmod 32} + \tilde{h}_2 V_{0,t-2 \bmod 32} + \tilde{h}_3 V_{0,t-3 \bmod 32}$
		$\tilde{V}^{a}_{1,t}$	=	$\sum_{l=0}^{3} \tilde{g}_l\, \tilde{V}^{a}_{0,t-l \bmod 32} = \sum_{l=0}^{3} \tilde{g}_l\, V_{0,t-l \bmod 32} = \tilde{g}_0 V_{0,t \bmod 32} + \tilde{g}_1 V_{0,t-1 \bmod 32} + \tilde{g}_2 V_{0,t-2 \bmod 32} + \tilde{g}_3 V_{0,t-3 \bmod 32}$

In this example, we calculate the detail coefficients (using Eqs. 1-9 Table 4.1) for the discrete wavelet transform (DWT) multiresolution analysis (MRA), i.e., the DWT-MRA; the maximal overlap DWT (MODWT) MRA, i.e., MODWT-MRA; and the *à trous* algorithm (AT) for a wavelet filter length of $L = 4$ at time index $t = 16$ where $N = 32$ samples and $J = 1$; t is a time index; $X \equiv V_{0,t} \equiv \tilde{V}_{0,t} \equiv \tilde{V}^{a}_{0,t}$ represents the original time series at time t; $W_{j,t}$ ($V_{j,t}$) represents the jth level wavelet (scaling) coefficient for the DWT at time t; $\widetilde{W}_{j,t}$ ($\tilde{V}_{j,t}$) represents the jth level wavelet (scaling) coefficient for the MODWT at time t; $\tilde{V}^{a}_{j,t}$ represents the jth level scaling coefficient for the AT at time t; h (g) is a wavelet (scaling) filter; N is the number of samples; and mod refers to the modulo operator. Note that $\tilde{g}_l \equiv g_l/\sqrt{2}$; $\tilde{h}_l \equiv h_l/\sqrt{2}$; and $g_l \equiv (-1)^{l+1} h_{L-1-t}$ where L is the length of the wavelet (scaling) filter.

Noting that $X_t \equiv V_{0,t}$, we can clearly see that both the DWT-MRA and MODWT-MRA require information from V_0 (X) as far into the future as $t = 19$ to calculate $D_{1,16}$ and $\widetilde{D}_{1,16}$, respectively. However, the AT does not require future information to exist as it requires only the present time index, or earlier, i.e., $t \leq 16$, to calculate $\widetilde{D}^a_{1,16}$ (which is equivalent to $\widetilde{W}^a_{1,16}$). Note also that the MODWT wavelet coefficient, $\widetilde{W}_{1,16}$, has the same requirement as the AT in this regard. To borrow a definition from linear filtering, the AT and MODWT are *causal* algorithms, as they do not require any future information to calculate their value at a given time index. Since this is not the case for the DWT-MRA and MODWT-MRA, they are *non-causal* algorithms (Bašta, 2014; Maheswaran and Khosa, 2012a).

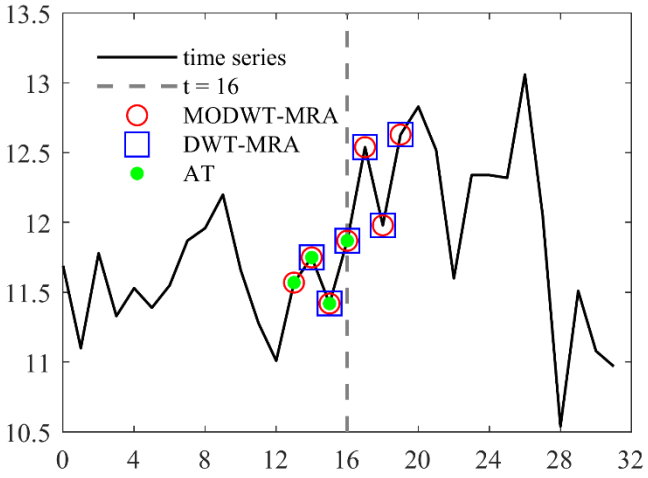

Figure 4.1. Depiction of the 'future data' issue boundary condition (see Table 4.1)

Another important point that we explore in more detail in section 4.3.3.3 is related to the calibration and validation set partitioning BC problem. Using the same example, if we consider that the first 17 time series observations as calibration records, and the remaining 15 as validation records, one can immediately realize that if we were to treat forecasts over the validation record in real-time (as in a real-world problem), one would not even have enough data to issue a forecast at time $t = 16$ for $t > 16$ using the DWT-MRA and MODWT-MRA. However, the AT and MODWT do not suffer from this issue as they do not require information in advance of $t = 16$ and could therefore be used to make a forecast for $t > 16$ at time $t = 16$. This very important issue is overlooked in the vast majority of studies considering wavelet-based forecasting models.

4.3.3.1.1. The 'Future Data' Solution

The solution to the 'future data' issue is simple. For any wavelet-based forecasting model, the *non-causal* DWT-MRA and MODWT-MRA (that use 'future data') should be avoided and the *causal* AT or MODWT algorithms should be used instead (since they *do not* use 'future data'). Henceforth, we limit our discussion of the DWT-MRA and MODWT-MRA (since they *cannot* be used correctly for real-world forecasting due to the 'future data' issue) and focus on the AT and MODWT (since they may be used for real-world forecasting) when discussing the remaining BC-related issues (e.g., selection of decomposition level(s) and wavelet filter(s) and the partitioning of calibration and validation sets).

A relevant question to ask is, "are there any 'catches' to using the AT and MODWT for real-world forecasting?". The answer is 'yes' and the 'catches' to using the AT and MODWT are as follows:

1. The MODWT cannot be used with the multicomponent approach, as it does not provide an additive reconstruction via Eq. 4.10;
2. The decomposition level and wavelet filter must be appropriately selected; and
3. A suitable number of calibration and validation records should be used.

We discuss the proper selection of the decomposition level and wavelet filter in the next sub-section and show how the second and third points are intimately tied to one another in section 4.3.3.3.

4.3.3.2. The Improper Selection of Decomposition Level(s) and Wavelet Filter(s) Issue

The BC due to decomposition level and wavelet filter selection introduces error into the wavelet and scaling coefficients at the beginning (and potentially end) of the time series (depending on the WT – see below) according to the following formula (Bašta, 2014; Maslova et al., 2016; Percival and Walden, 2000):

$$L_J = (2^J - 1)(L - 1) + 1 \tag{4.11}$$

where L_J represents the number of wavelet and scaling coefficients affected by the BC for decomposition level J and a wavelet filter of length L. For the AT and MODWT only the beginning of the time series, i.e., for $t = 0,1, \dots, L_J - 1$ is affected by the BC due to the selected decomposition level and wavelet filter, while the first and last L_J wavelet and scaling coefficients are affected (i.e., $t = 0,1, \dots, L_J - 1$ and $N - 1, N - 2, \dots, N - L_J + 1$) for the DWT-MRA and MODWT-MRA (Percival and Walden, 2000).)

The BC issue related to decomposition level and wavelet filter selection can be seen in Table 4.1 (e.g., refer to $t - 2^{j-1}l \bmod N$ for the MODWT and AT algorithms) where the modulo operator (mod) is used

to represent circular convolution (also referred to in the literature as the 'circularity boundary condition') and is necessary in order to calculate wavelet (detail) and scaling (approximation) coefficients that require observations for $t < 0$ ($t < 0$ and $t > N - 1$) (Bašta, 2014; Percival and Walden, 2000).

It is important to consider the significance of the BC-issue related to decomposition level and wavelet filter selection: when one selects a wavelet filter that is too wide (long) and chooses a decomposition level that is too high, it leaves very few wavelet and scaling coefficients (free from BC-related uncertainty) from which to calibrate a wavelet-based forecasting model (Aussem et al., 1998). This important item is often overlooked in numerous hydrological and water resources forecasting studies employing WT, where the common approach is to neglect the error introduced in those wavelet coefficients and to use the incorrect coefficients as input to a data-driven model (e.g., as in Barzegar et al. (2017b); Karbasi (2017); Nourani and Saeidifarzad (2017); and Shoaib et al. (2017)), when the correct approach is to not include those coefficients in the forecast and only use those wavelet and scaling coefficients that can be calculated with certainty (Aussem et al., 1998; Bašta, 2014).

The common practice of ignoring the error introduced due to improper selection of decomposition levels and wavelet filters in wavelet-based forecasting can be better understood through an example. A popular wavelet filter used in wavelet-based hydrological and water resources forecasting studies (Akrami et al., 2014; Barzegar et al., 2017; Ebrahimi and Rajaee, 2017; Mirbagheri et al., 2010; Rajaee, 2011; Ravansalar and Rajaee, 2015; Rezaie-balf et al., 2017; Sahay and Srivastava, 2014; Shoaib et al., 2016a), that very seldom should be used, except for cases considering very large time series records and low decomposition levels, is the Discrete Meyer (*dmey*) wavelet filter (with a filter length, L, of 102 coefficients). As an example, the *dmey* wavelet filter has been used to forecast a monthly precipitation time series that included only ~ 350 historical records, using the DWT-MRA at a decomposition level of three (Nourani et al., 2009a). Using the *dmey* with a three level decomposition produces 708 boundary coefficients (i.e., $L_3 = (2^3 - 1)(102 - 1) + 1 = 708$). This is more than double the amount of time series records that were available to the authors, which means that not only was the entire calibration dataset filled with wavelet and scaling coefficients that included errors due to the BC, but also the entire validation set. The model reported coefficient of determination scores of 0.935 and 0.890 on the calibration and validation partitions, respectively. However, these results are difficult to interpret in terms of model generalization since both datasets used wavelet and scaling coefficients that were incorrect, not only due to the 'future data' issue, but also due to the improper selection of decomposition level and wavelet filter. In addition to the references mentioned earlier in this paragraph, other very similar cases are present in

the literature (e.g., Krishna (2013); Moosavi et al. (2013); Shirmohammadi et al. (2013); Sehgal et al. (2014); Shoaib et al. (2015, 2016a)) where unrealistically wide wavelet filters are used with high decomposition levels causing most calibration and validation records to be tainted with error.

While the literature includes so-called 'boundary condition' correction methods to counteract the selection of high and wide decomposition level(s) and wavelet filter(s), respectively, such as those employed for the DWT-MRA and MODWT-MRA (Karthikeyan and Kumar, 2013; Maslova et al., 2016; Percival et al., 2011), the 'future data' issue still remains for these methods, which makes them not usable in real-world forecasting studies. We do not delve into BC correction methods as our goal is to avoid adding *any* error due to BC into our wavelet-based forecasting models (realistically, *any* BC correction-method adds some form of error since it is inherently a model in itself). Instead, we propose an alternative approach in the next sub-section and show how to properly address decomposition level and wavelet filter BC-related error for the AT and MODWT algorithms.

4.3.3.2.1. The Solution for Properly Selecting Decomposition Level(s) and Wavelet Filter(s)

The proper selection of decomposition level(s) and wavelet filter(s) is a three-step process:

1. Choose either the MODWT or AT for wavelet decomposition;
2. Select a decomposition level and wavelet filter; and
3. Remove the first L_J wavelet and scaling coefficients (as determined by Eq. 4.11), resulting in "boundary corrected" wavelet and scaling coefficients.

The final BC-related issue to address is related to the partitioning of data into calibration and validation sets. Since this item is tied to decomposition level and wavelet filter selection through the number of boundary-affected coefficients, as determined via Eq. 4.11 (recall our example of the *dmey* wavelet filter above), we will show below that it is imperative to select decomposition level(s) and wavelet filter(s) in conjunction with careful dataset partitioning.

4.3.3.3. The Dataset Partitioning Issue

A common practice in hydrological and water resources wavelet-based forecasting studies adopting the DWT-MRA and MODWT-MRA is to decompose both calibration and validation data separately to 'overcome' the 'future data' issue (Barzegar et al., 2017, 2018; Deo et al., 2017c; Prasad et al., 2017). However, it appears that this is a misunderstanding of the 'future data' issue (as described in this study), since we have shown that at *any* given time the DWT-MRA and MODWT-MRA require 'future data' to compute detail and approximation coefficients. In reality, if the calibration and validation data are

decomposed separately, this causes problems at the boundaries of each partition where: 1) the calibration data must 'wrap around' and use wavelet and scaling coefficients at the beginning of the time series as surrogates to calculate the boundary coefficients at the end of the calibration dataset and 2) the earlier boundary coefficients of the validation set use the wavelet and scaling coefficients at the end of the validation record as surrogate values for calculating their values (Du et al., 2017; Karthikeyan and Kumar, 2013; Zhang et al., 2015). We clearly show this to be the case by drawing on our example in Table 4.2. Assuming that the first 17 records are used for calibration and the remaining 15 for validation, if the DWT-MRA (or MODWT-MRA) is used and the calibration and validation data are first partitioned and then decomposed separately, in the calculation of $D_{1,16}$ (or $\widetilde{D}_{1,16}$), i.e., the final calibration record: $V_{0,19}$ would be swapped for $V_{0,2}$ (since $V_{0,19 \bmod 17} \equiv V_{0,2}$) as its surrogate, $V_{0,18}$ for $V_{0,1}$, $V_{0,17}$ for $V_{0,0}$, etc. Considering $D_{1,17}$ (or $\widetilde{D}_{1,17}$) (i.e., the first validation record), $V_{0,16}$ which is required to calculate $D_{1,17}$ (or $\widetilde{D}_{1,17}$), would no longer be taken as $V_{0,16}$, but would be replaced by $V_{0,32}$ as its surrogate. Similarly, $V_{0,15}$ would be replaced by $V_{0,31}$ as its surrogate and so on. It is very clear that the DWT-MRA and MODWT-MRA cause significant issues when attempting to calibrate and validate real-world wavelet-based forecasting models.

It is worth re-iterating that this issue with the DWT-MRA and MODWT-MRA can be circumvented by adopting the MODWT and AT algorithms for wavelet decomposition of the calibration and validation records and is discussed in more detail below.

4.3.3.3.1. The Solution for Correctly Partitioning a Dataset

Our solution for correctly partitioning data in wavelet-based forecasting models is founded on two key principles: 1) avoiding the case where there are more boundary coefficients than available time series records (as discussed in 4.3.3.2) and 2) to provide enough time series records from which to calibrate and validate a wavelet-based forecast. These two principles can be broken down in a set of steps:

1. Adopt either the MODWT or AT (since they are *causal* filters, that do not have any dependence on 'future data');
2. Select a maximum decomposition level (J_{max}) and a variety of wavelet filters of suitable length (see section 4.A of the Supplementary Material [appendix] for examples) - the value J_{max} and a series of wavelet filters can be selected by applying Eq. (4.11) to determine the number of boundary-affected coefficients and ensuring that $L_J \ll N$);
3. Remove the first L_J wavelet and scaling coefficients from the beginning of the wavelet and scaling coefficients (i.e., the boundary-effected coefficients), obtaining the "boundary-corrected"

wavelet and scaling coefficients, such that there are at least enough calibration records unaffected by the boundary condition (providing the wavelet-based forecasting model with enough records for calibrating the explanatory variables to the target variable);

4. After calibration, apply the MODWT or AT to the validation set one-record at a time and calculate the forecast for each validation record; and
5. Repeat steps 1 to 3 until one has identified suitable decomposition levels and wavelet filters for forecasting a given target process.

It may be useful to perform an initial exploratory analysis by selecting both a maximum decomposition level and wavelet filter length by cycling through a combination of each pair to identify suitable candidates. Suitable candidate decomposition levels and wavelet filters should be identified by the modeller's preference, which is usually a function of a trade-off between model accuracy on the validation records and computational efficiency (e.g., based on a tolerance that is appropriate to the modeller).

Another key point connected to data partitioning of the calibration and validation records is to ensure that any data scaling applied to the explanatory and target variables prior to their input to a data-driven model (a useful practice for ensuring that certain inputs are not favoured due to a larger range in comparison with other inputs, but instead due to their intrinsic relationships with the target process), is to apply the scaling properties (e.g., maximum and minimum values, mean and standard deviation, etc.) from the calibration data to both calibration and validation data, and to not perform this separately. Otherwise, the calibrated model parameters might not generalize properly to unseen data (Hastie et al., 2009). While this is a fundamental concept, we mention it here since to the best of our knowledge, we have not yet seen this important issue discussed in studies on wavelet-based hydrological and water resources forecasting.

4.4. Best (Correct) Practices and the Wavelet Data-Driven Forecasting Framework

The solutions to the BC-related issues discussed in the last section (e.g., the 'future data' issue, selection of decomposition levels and wavelet filters, and calibration and validation set partitioning) can now be formulated in to a general wavelet-based forecasting framework, Wavelet Data-Driven Forecasting Framework (WDDFF). We first summarize the best practices adopted within WDDFF and then discuss the 'building blocks' of WDDFF that can be varied based on the forecasters' preferences (e.g., type of WT (AT or MODWT), wavelet filter selection, data-driven model selection, etc.).

4.4.1. Best Practices Adopted by the Wavelet Data-Driven Forecasting Framework

The WDDFF is based on the following best practices:

1. Adopting either the AT or MODWT for wavelet decomposition.
2. Using the AT or MODWT to calculate "boundary-corrected" wavelet and scaling coefficients (see section 4.3.3.2 and 4.3.3.3.), which is carried out according to:
    a. The selection of a decomposition level and wavelet filter such that J_{max} and L results in $L_J \ll N$ (i.e., according to Eq. 4.11) and
    b. Removing the first L_J boundary-effected coefficients from the beginning of the wavelet and scaling coefficients such that after partitioning the dataset into calibration and validation partitions, there are at least enough calibration records unaffected by the boundary condition.
3. Using the "boundary-corrected" wavelet and scaling coefficients to calibrate the wavelet-based forecasting model.
4. Validating the wavelet-based forecasting model by applying the AT or MODWT to the validation set, one record at a time, and ensuring that the forecasting model's performance and computational efficiency meet the forecasters' requirements; else repeat steps 1-3 until a suitable forecasting model is obtained.

4.4.2. The 'Building Blocks' of the Wavelet Data-Driven Forecasting Framework

By assuming that the best practices described above are followed for wavelet-based forecasting, the main 'building blocks' of the WDDFF can be broken down into: 1) selecting a wavelet-based forecasting method (section 4.4.2.1.); 2) input variable selection for wavelet and scaling coefficients (section 4.4.2.2.); and 3) selection of data-driven model (section 4.4.2.3.). We discuss each of these items in detail in the sub-sections below and summarize these steps in a flowchart (Figure 4.2).

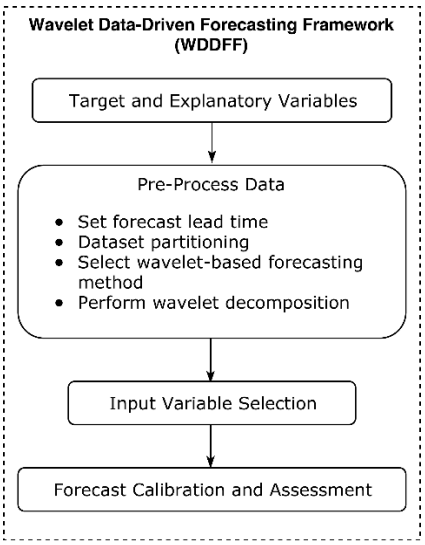

Figure 4.2. Wavelet Data-Driven Forecasting Framework flow chart

4.4.2.1. Pre-Processing Data

The first step in the WDDFF is to pre-process the data. To begin, the forecast lead time is set and the explanatory and target variables (i.e., input and output, respectively), are partitioned into calibration and validation sets. Depending on the wavelet-based forecasting method (see sub-section 4.4.2.2. and 4.4.2.1.1.), the explanatory variables are decomposed using the MODWT or AT (direct approach) or both explanatory and target variables are decomposed (multicomponent approach).

4.4.2.1.1. Different Wavelet-based Forecasting Methods

Since the AT can be used in either a direct or a multicomponent forecasting approach (i.e., both explanatory and target variables are decomposed and the target forecast is obtained through additive reconstruction (Eq. 4.10)) while the MODWT can only be used in a direct forecasting approach (i.e., only explanatory variables are decomposed and the target is forecasted directly) (see section 4.3.2.), one can develop a wide variety of wavelet-based forecasts under the WDDFF. Both direct and multicomponent approaches are important to consider as they can differ in forecast accuracy and it is thus significant to explore both approaches in order to obtain the best overall forecasting model for a given dataset. This also gives the modeller access to different models that may be assessed in terms of trade-offs between computational efficiency and forecast accuracy or for ensemble modelling (Rathinasamy et al., 2013). Using the direct and multicomponent approaches, we define six different ways in which the WDDFF may be used to develop wavelet-based forecasts via the MODWT and AT (although there potentially exists

other approaches, we focus on only these approaches since they are straightforward to apply and encompass the different direct and multicomponent wavelet-based forecasting approaches studied thus far in hydrology and water resources):

1. Single (Method 1), a direct approach - wavelet decomposed explanatory variables are used as input and the target variable as output for a given data-driven model.
2. Within (Method 2), a multicomponent approach - forecast each set of wavelet and scaling coefficients for the target variable using the wavelet and scaling coefficients from the explanatory variables at the same levels (i.e., the first level wavelet coefficients for the target are forecasted using the first level wavelet coefficients from the explanatory variables, the second level wavelet coefficients for the target are forecasted using the second level wavelet coefficients from the explanatory variables, and so on). We name this method Within, since only wavelet (and scaling) coefficients within the same level are used in the forecasts.
3. Across (Method 3), a multicomponent approach - forecast each set of wavelet and scaling coefficients for the target process using the wavelet and scaling coefficients from the explanatory variables for all levels (i.e., the first level wavelet coefficients for the target are forecasted using the first, second, third, etc. level wavelet coefficients (including the scaling coefficients) from the explanatory variables). We name this method Across, since wavelet (and scaling) coefficients from the explanatory variables across the different levels are used in the forecasts of the wavelet and scaling coefficients of the target process at each level.
4. Single-hybrid (Method 4), a direct approach – the same as Method 1, but the original un-decomposed explanatory variables are also used as input.
5. Within-hybrid (Method 5), a multicomponent approach – the same as Method 2, but the original un-decomposed explanatory variables are also used as input for each of the target variable's wavelet and scaling coefficients.
6. Across-hybrid (Method 6), a multicomponent approach – the same as Method 3, but the original un-decomposed explanatory variables are also used as input for each of the target variable's wavelet and scaling coefficients.

Eight different wavelet-based forecasting approaches can be used (six using the AT and two using the MODWT): both Within and Across-based multicomponent approaches (Method 2, 3, 5, and 6) can only be used with the AT, since it is the only WT (amongst MODWT and AT) that can be reconstructed additively (i.e., via Eq. 4.10). The reason for including the original un-decomposed explanatory variables along with

wavelet decomposed explanatory variables is because the original time series contain coarse, global details, while the wavelet decomposed time series contain specific, localized details – together they provide a more comprehensive structure (view) of the overall process that is being forecasted and this has been shown to increase forecast performance in earlier wavelet-based forecasting studies (Adamowski, 2008a, 2008b; Aggarwal et al., 2008; Nguyen and Nabney, 2010; Voronin and Partanen, 2013). We stress that Method 1-6 solely adopt "boundary-corrected" wavelet and scaling coefficients derived by MODWT and AT.

Methods 1-3 are generalizations of methods 4-6, with the difference being that non-wavelet-decomposed data is *not* considered (in methods 1-3) – since the original data is available, we feel that its predictive ability, in concert with that of the wavelet-decomposed data, should be explored in order to develop the best forecast model (Adamowski, 2008b; Nguyen and Nabney, 2010).

Method 1 has been used in numerous studies by the present authors (e.g., Adamowski and Sun (2010); Adamowski and Chan (2011); Adamowski et al., (2012); Campisi-Pinto et al. (2013); Belayneh et al. (2014, 2016a, 2016b); and Khalil et al. (2015)) and others (e.g., Benaouda et al. (2006); Maheswaran and Khosa (2012a); Li and Cheng (2014); Djerbouai and Souag-Gamane (2016)), but in these studies, solely the AT was adopted and the MODWT was not explored. Methods 4 and 5 were considered by Nguyen and Nabney (2010) (with methods 1 and 2 being specific cases of these methods), while a variant of method 3 (in addition to method 1) was considered by Murtagh et al. (2004), where the authors included lower level (higher frequency) wavelet coefficients as predictors for scales at the same level or above, while our approach includes the potential to use higher level (lower frequency) wavelet coefficients as predictors for lower level coefficients, and vice versa. We are not aware of any studies that have proposed method 6, or compared each of the proposed (eight different varieties of) wavelet-decomposition forecasting strategies (i.e., six different methods (methods 1-6) with eight different modes (methods 1-6 apply to AT and methods 1 and 4 apply to the MODWT).

4.4.2.2. Input variable selection for wavelet and scaling coefficients

After decomposing the input-output dataset (according to the wavelet-based forecasting method in the previous stage), input variable selection is used to identify only those wavelet and scaling coefficients pertinent to the forecasting model.

Input variable selection provides a systematic means for selecting which variables are useful in predicting a target process and can enable an automatic selection of relevant variables that consider relevancy and

redundancy amongst potential candidates (Quilty et al., 2016). Performing input variable selection using wavelet-decomposed data is not a new concept (Alsberg et al., 1998). Recent approaches have investigated using partial correlation input selection (PCIS) (Tran et al., 2016) , mutual information (MI) (Rana and Koprinska, 2016), and conditional (or partial) MI (CMI or PMI) (He et al., 2015; Li et al., 2016) for selecting the best wavelet-decomposed inputs for wavelet-based forecasting.

As part of WDDFF, we opt to select which "boundary-corrected" wavelet and scaling coefficients are to be used in forecasting the target process by adopting the input variable selection methods discussed in Quilty et al. (2016). We considered two input variable selection methods: the Edgeworth Approximations-based CMI (EA) approach and PCIS (to maintain brevity readers should consult Quilty et al. (2016) for details on the EA method). The EA method uses information theoretic concepts (e.g. multivariate differential entropy) to estimate CMI for multivariate datasets and is therefore well suited for capturing nonlinear dependence while the PCIS method is based on partial correlation using the typical linear Pearson correlation and is limited to identifying only linear relationships (Galelli et al., 2014). In general, other input variable selection algorithms can be used instead of EA or PCIS within the WDDFF (Andersen and Bro, 2010; Creaco et al., 2016; Fernando et al., 2009; Galelli and Castelletti, 2013; Kariwala et al., 2013; Taormina et al., 2016). We selected the EA method instead of other similar nonlinear methods, such as Partial Mutual Information Selection (Fernando et al., 2009), k-nearest neighbours-based conditional mutual information (Tsimpiris et al., 2012), Partial Information (Sharma et al., 2016), etc. since it has been shown to be computationally efficient and provides similar (if not, better) results for a number of benchmark input variable selection problems (Quilty et al., 2016). PCIS is adopted because it is a classic input variable selection method for linear (and nonlinear) regression problems. To the best of our knowledge, we are the first to apply a conditional mutual information approach for wavelet-based input variable selection using the MODWT and AT.

4.4.2.3. Data-Driven Model Selection

The wavelet-based inputs ("boundary-corrected" wavelet and scaling coefficients) selected in the last stage are used, together with the target variable (which may also be a set of wavelet and scaling coefficients, i.e., if the multicomponent wavelet-based forecasting approach is adopted – see section 4.4.2.1.1.), in nonlinear and linear regression frameworks to map inputs to target forecasts. We used the extreme learning machine (Huang et al., 2006; Stulp and Sigaud, 2015) and the second-order Volterra (SOV) model (Labat et al., 1999; Maheswaran and Khosa, 2012a) as the nonlinear regression methods and multiple linear regression (MLR) as the linear model. We adopted these three methods as they are easy

to program (they can be formulated as linear least-squares problems), are computationally efficient, and have performed very well in our experience for hydrological and water resources forecasting applications using wavelet- and non-wavelet-based approaches. We do not provide theoretical details for these models as they can be found in a number of our earlier works (Deo et al., 2017a; Quilty et al., 2016; Rathinasamy et al., 2014, 2013; Yaseen et al., 2016). As with the input variable selection methods, other regression models may easily be substituted for ELM, SOV, and MLR – popular methods such as kriging, fuzzy logic, support vector regression, model trees, multivariate adaptive regression splines, k-nearest neighbours, partial information, genetic programming, etc. (Altunkaynak et al., 2003; Deo et al., 2017b; Garg et al., 2014; Goyal et al., 2014; Kisi, 2016; Lee et al., 2017; Rahimikhoob, 2016; Sharma et al., 2016) may be used within WDDFF. EA is paired with the nonlinear regression models (ELM and SOV) and PCIS is paired with the MLR model.

4.4.3. Summary of the Wavelet Data-Driven Forecasting Framework

Our proposed WDDFF takes into account best practices for wavelet-based forecasting by overcoming boundary condition-related issues during wavelet decomposition (that if not addressed would render a wavelet-based forecast invalid for real-world application, such as in many existing wavelet-based forecasting frameworks in the literature): 1) avoiding the use of 'future data'; 2) selection of suitable decomposition levels and appropriate wavelet filters; and 3) proper dataset partitioning. Additionally, WDDFF incorporates different approaches for using wavelet-decomposed data (i.e., "boundary-corrected" wavelet and scaling coefficients) in wavelet-based forecasting models (i.e., via direct and multicomponent approaches, see section 4.4.2.1.1.), allowing one to identify the best wavelet-based forecast for a given input-output dataset, which is often overlooked but important to consider.

WDDFF can be summarized as follows:

1. WDDFF uses the MODWT and AT algorithms which do not consider 'future data' in calculating wavelet and scaling coefficients at a particular time;
2. Decomposition levels and wavelet filters are selected such that they provide enough data for calibrating the wavelet-based forecasting model, without including any boundary coefficients (i.e., after wavelet decomposition of the input-output dataset, we remove any boundary coefficients prior to calibrating our forecasts - we only use "boundary-corrected" wavelet and scaling coefficients in our forecasts) and such that they provide suitable forecasting performance on a validation set (based on forecasting accuracy and computational- efficiency, suited to the user's preference); and

3. Direct and multicomponent wavelet-based forecasting approaches can be adopted (Methods 1-6 listed in section 4.4.2.1.1.), allowing one to identify a forecasting approach that provides the best accuracy for a given dataset:
    a. Input variable selection is used to select which "boundary-corrected" wavelet and scaling coefficients should be used in the wavelet-based forecasting models and
    b. Data-driven methods are used to convert the selected "boundary-corrected" wavelet and scaling coefficients from the last step into target variable forecasts.

4.5. Experimental Setup

Here we first introduce the study area used in our experiments before discussing how the general methodology for the WDDFF is specifically adapted to our study area. Afterwards, we discuss the details of the different experiments that were designed for demonstrating the usefulness of WDDFF for real-world wavelet-based hydrological and water resources forecasting problems. To keep this work succinct, we refer readers to section 4.B of our Supplementary Material [appendix] for specific experiment details.

Briefly, our proposed WDDFF is tested on an urban water demand (UWD) forecasting case study in Montreal, Quebec where we forecasted average daily UWD, U, for lead times (1, 3, 5, 7, and 14 days) common to water utility functions (e.g., optimization of planning, design, management, and operations). Recent UWD forecast studies have focussed on forecasting Montreal's UWD (Adamowski et al., 2012; Tiwari and Adamowski, 2013). Adamowski et al. (2012) forecasted daily summer demands (i.e., May to August) at one day lead times using MLR, multiple nonlinear regression, autoregressive integrated moving average (ARIMA), ANN, and wavelet-ANN (WA-ANN) based on the AT (however, the boundary condition at the beginning of the time series was not taken into account in the WA-ANN model). The authors found that the WA-ANN provided the best one day-ahead forecast in the summer demand period between the years 2001-2009. Tiwari and Adamowski (2013) compared ARIMA, ARIMA with exogenous variables (ARIMAX), ANN, bootstrap ANN (BANN), WT-ANN, and wavelet BANN (WT-BANN) for forecasting average UWD at daily (1, 3, 5, 7, and 14) and monthly (1 and 2) lead times between February 27, 1999 and August 6, 2010. The authors used the DWT-MRA to develop their wavelet-based models and found that the WT-BANN models provided the best overall forecasts. In both studies, historical average daily UWD, total daily rainfall, and maximum daily air temperature were used as explanatory variables.

We incorporated the explanatory variables historical average daily UWD (U), total daily rainfall (R), maximum daily air temperature (T), and the daily antecedent precipitation index (A) to forecast the target variable (average daily UWD) at lead times of 1, 3, 5, 7, and 14 days ahead. We collected average daily

UWD (*U*) from the City of Montreal and daily meteorological variables (R, T, and A) from Environment Canada within the period February 27, 1999 to August 6, 2010 (4179 daily records).

4.5.1. Methodology for Wavelet Data-Driven Forecasting Framework

The general workflow (Figure 4.2) that was followed for developing the WDDFF forecasts for our UWD forecasting case study in Montreal Canada is summarized below (specific details may be found in section 4.C of our Supplementary Material [appendix]).

4.5.1.1. Quantitative Performance Comparisons across Input Variable Selection Algorithms

The target variable, U, was set to lead times of 1, 3, 5, 7, and 14 days ahead. Each explanatory variable (U, R, T, and A) was time-lagged up to 14 days in the past to allow historical information to be exploited through input variable selection (see Section 4.C of our Supplementary Material [appendix]). Calibration and validation sets were partitioned as follows: February 27, 1999 to December 31, 2007 for calibration (3230 records) and January 1, 2008 to August 6, 2010 for validation (949 records) – we found this data partitioning to provide good performance as it afforded us with enough records for calibration and a reasonable amount of data (over two and a half years) to validate the performance of the WDDFF out-of-sample.

Wavelet decomposition using the MODWT and AT algorithms was used to calculate "boundary-corrected" wavelet and scaling coefficients and was carried out using each of the six wavelet-based forecasting methods described in section 4.4.2.1.1 (i.e., Single (Method 1), Within (Method 2), Across (Method 3), Single-hybrid (Method 4), Within-hybrid (Method 5), and Across-hybrid (Method 6)). The maximum decomposition level was set to six ($J_{max} = 6$) and we considered wavelet filters up to length 14 ($L \leq 14$) (see section 4.C of our Supplementary Material [appendix] for further details on how J_{max} was determined and why these wavelet filters were selected).

The different wavelet families and their filters that were used in our study are as follows (their coefficients can be found in the Supplementary Material [appendix]): Haar (*haar*), Daubechies (*db2*, db3, *db4*, *db5*, *db6*, *db7*), Symlets (*sym4*, *sym6*,), Coiflets (*coif1*, *coif2*), Fejer-Korovkin (*fk4*, *fk6*, *fk8*, *fk14*), Least-Asymmetric (*la8*, *la12*, *la14*), and Best-Localized (*bl14*) (Crowley, 2007; Nielsen, 2001; Olhede and Walden, 2004; Percival and Walden, 2000; Rathinasamy et al., 2013; Zhang et al., 2016). Since the largest (widest) wavelet filter had 14 filter coefficients (*d7*, *sym7*, *fk14*, *la14*, and *bl14*) and we used a maximum decomposition level of six, there were $L_j = \left(2^j - 1\right)(L-1) + 1 = 820$ boundary-effected coefficients. Therefore, we removed the first 820 records from the beginning of our input-output datasets (creating

the “boundary-corrected” wavelet and scaling coefficients) prior to selecting input variables and before model calibration, to eliminate the effect of the boundary condition on our forecasts (see section 4.3.3.3.). Since explanatory variables were time lagged up to 14 days, forecast leads times of 1, 3, 5, 7, and 14 day were explored, and we removed the first 820 boundary coefficients from each dataset, the original calibration set of 3230 records was reduced to a total of: 2395, 2393, 2391, 2389, and 2382 calibration records for the 1, 3, 5, 7, and 14 day ahead forecasts, respectively. Therefore, 949 validation records were used for evaluating each lead time forecast out-of-sample (i.e., each lead time forecast had the exact same validation set for the target variable).

4.5.1.2. Input Variable Selection for Wavelet and Scaling Coefficients

The Edgeworth Approximation-based conditional mutual information and partial correlation input selection input variable selection methods (Quilty et al., 2016) were used for selecting wavelet and scaling coefficients as inputs for the different data-driven models within WDDFF (see section 4.C of our Supplementary Material [appendix] for further details).

4.5.1.3. Data-Driven Models

The extreme learning machine, second order Volterra, and multiple linear regression were selected as the data-driven models that were fed wavelet and scaling coefficient inputs according to the EA and PCIS input variable selection methods. The data-driven models using wavelet-decomposed inputs are appended with a ‘W’, i.e., WELM is a wavelet-based extreme learning machine, while WMLR is a wavelet-based multiple linear regression model. The EA method was used to select inputs for the nonlinear models (ELM and SOV) while PCIS was used to select inputs for the MLR. Non-wavelet based models were used as a benchmark for the wavelet-based models developed using WDDFF. A random walk (RW) (without wavelet-based inputs) was also considered as a benchmark model. Therefore, the following models were considered (input variable selection method-model): EA-WELM, EA-WSOV, PCIS-WMLR, EA-ELM, EA-SOV, PCIS-MLR, and RW.

4.5.1.4. Forecast Calibration and Evaluation

The ELM, SOV, and MLR-based models were calibrated using linear-least squares. Forecasts were evaluated over the validation set (an estimate of out-of-sample or generalization performance) using metrics commonly adopted in hydrological and water resources forecasting studies: Nash-Sutcliffe Efficiency Index, the root mean square error (RMSE), and mean absolute percentage error (MAPE) (Adamowski et al., 2012; Moriasi et al., 2007; Quilty et al., 2016) (see section 4.C of our Supplementary Material [appendix] for further details).

4.5.2. Experimental Details

In order to demonstrate the usefulness of WDDFF for wavelet-based hydrological and water resources forecasting we focussed on three main experiments, using the daily UWD dataset from Montreal, to highlight how the best practices adopted by WDDFF can be used for identifying the best performing wavelet-based forecast among a variety of alternatives based on different decomposition levels, wavelet filters, and forecasting strategies (direct and multicomponent) (i.e., optimal decomposition levels, wavelet filters, and forecasting approaches) and show its added value in comparison to a recent, and invalid (in terms of real-world applications), wavelet-based forecasting framework that adopts the MODWT-MRA (Bachour et al., 2016; Barzegar et al., 2018, 2017; Maslova et al., 2016; Prasad et al., 2017; Yang et al., 2014). The three different experiments are as follows:

1. Best performing models - for the different wavelet-based forecasts contained within the WDDFF (see section 4.4.3.1.1. and 4.5.1.), we identified which wavelet decomposition algorithm (MODWT or AT), wavelet-based forecasting approach (direct or multicomponent), and Method (1-6 - Single, Within, Across, Single-hybrid, Within-hybrid, or Across-hybrid) performed best.
2. Best decomposition levels and wavelet filters - of the best WDDFF-based forecasts, we showed that certain decomposition levels and wavelet filters were most useful in providing the best forecast accuracy. Some visible trends stood out.
3. WDDFF vs. MODWT-MRA – our proposed WDDFF, in comparison to a current wavelet-based hydrological and water resources forecasting approach, MODWT-MRA (Barzegar et al., 2018, 2017; He et al., 2015; Prasad et al., 2017; Yang et al., 2014), that we have demonstrated throughout this work to include 'future data' and that should be optimistically biased in terms of forecast accuracy (a sign of its incorrectness), provided more realistic forecast performance than MODWT-MRA; WDDFF's usefulness and realistic performance for real-world forecasting problems was clearly shown.

These experiments were chosen to demonstrate that:

1. WDDFF can offer widely different performances for different settings (e.g., direct or multicomponent approaches); but, for our case study, WDDFF always provided the best performance against benchmark models (e.g., MLR, ELM, SOV, and RW);
2. WDDFF had the tendency to provide the best performance for a given decomposition level and wavelet filter combination; and

3. WDDFF should be preferred over the MODWT-MRA-based (including the DWT-MRA-based) forecasts (currently adopted within many wavelet-based hydrological and water resources forecasting studies) as it provided realistic performance while MODWT-MRA provided incorrect performance for the dataset that we studied.

4.6. Results and Discussion

The results of our UWD forecasting experiment for the different daily lead times (1, 3, 5, 7, and 14) can be found in Table 4.3 (1 day lead time) while the remaining results (3, 5, 7, and 14 day lead times) are given in section 4.D of our Supplementary Material [appendix]. We included only the best forecasts for each model type, due to the large amount of models (14, 420 in total, see section 4.C of our Supplementary Material [appendix]), and reported their results for the validation partition (although the calibration performance is similar to the validation performance, affirming the stable generalization properties of the different forecasts). The best model for a particular lead time forecast is in bold-face within our tabulated results. Scatter plots for the best wavelet- and non-wavelet-based models are shown in Figure 4.3.

Table 4.3. 1 day lead time forecast results

1 Day Lead-Time Forecasts						
Strategy	IVS-Model	Level	Filter	RMSE (ML/D)	MAPE (%)	NASH
	RW			35.981	1.562	0.844
	PCIS-MLR			36.229	1.703	0.842
	EA-ELM			29.828	1.287	0.893
	EA-SOV			29.680	1.274	0.894
AT - single	PCIS-WMLR	5	*fk4*	26.986	1.162	0.912
	EA-WELM	1	*coif1*	30.091	1.289	0.891
	EA-WSOV	1	*coif2*	31.551	1.400	0.880
MODWT - single	PCIS-WMLR	3	*fk8*	26.660	1.148	0.914
	EA-WELM	2	*fk4*	32.581	1.406	0.872
	EA-WSOV	1	*d1*	33.888	1.476	0.862
AT - within	PCIS-WMLR	6	*fk8*	28.138	1.198	0.905
	EA-WELM	2	*fk4*	29.888	1.265	0.892
	EA-WSOV	2	*d1*	30.695	1.293	0.886
AT - across	PCIS-WMLR	5	*d2*	26.877	1.175	0.913
	EA-WELM	6	*d1*	28.266	1.212	0.904
	EA-WSOV	6	*d1*	28.779	1.243	0.900
AT - single - hybrid	PCIS-WMLR	3	*coif1*	27.040	1.187	0.912
	EA-WELM	2	*d1*	29.752	1.272	0.893
	EA-WSOV	1	*bl14*	29.680	1.274	0.894
MODWT - single - hybrid	**PCIS-WMLR**	**6**	***la14***	**26.358**	**1.133**	**0.916**
	EA-WELM	2	*d1*	29.752	1.272	0.893
	EA-WSOV	1	*bl14*	29.680	1.274	0.894
AT - within - hybrid	PCIS-WMLR	1	*d2*	29.372	1.304	0.896
	EA-WELM	1	*fk4*	28.372	1.226	0.903
	EA-WSOV	1	*fk4*	28.792	1.245	0.900
AT - across - hybrid	PCIS-WMLR	4	*coif1*	27.088	1.176	0.912
	EA-WELM	6	*d1*	28.332	1.215	0.903
	EA-WSOV	6	*d1*	28.528	1.229	0.902

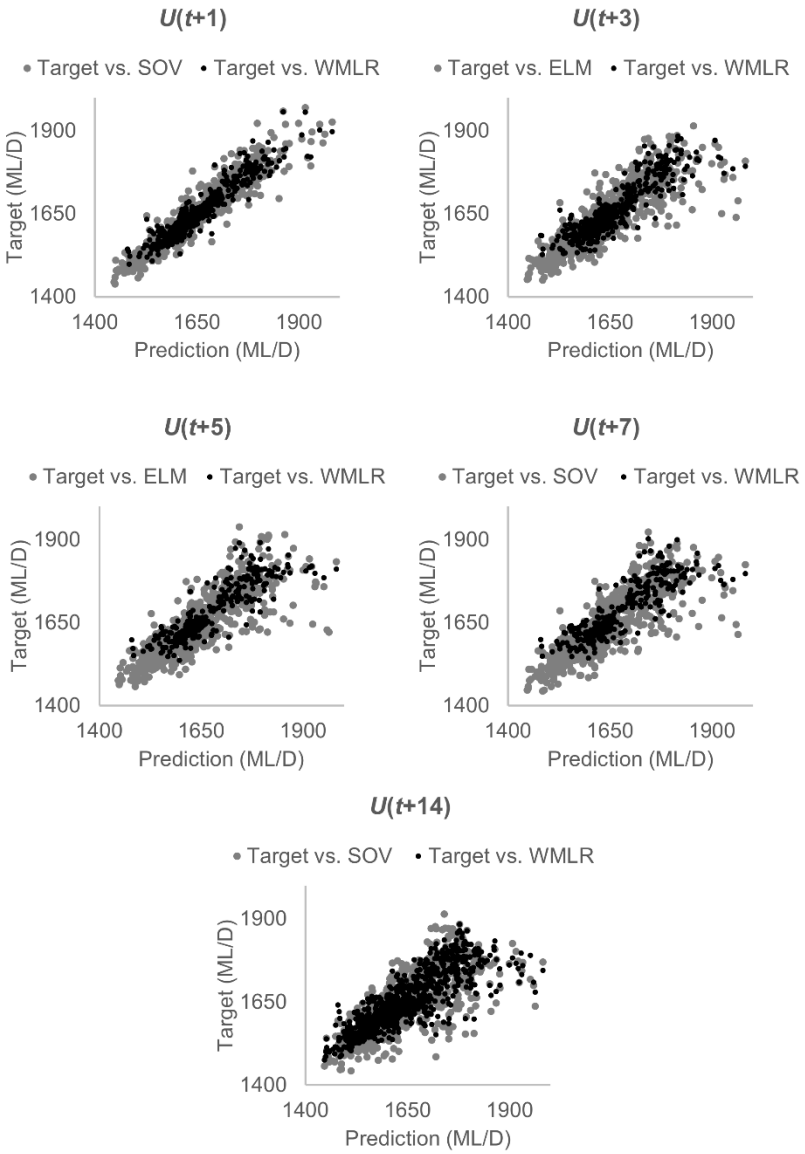

Figure 4.3. Scatter plots for the best wavelet-based and non-wavelet-based forecasts for each lead time (1, 3, 5, 7, and 14 days)

4.6.1 Best Performing Models

We may notice that the MODWT-single-hybrid class of WDDFF (wavelet-based) forecasts, particularly the WMLR model, provided the best performance across the explored lead times. Thus, for this dataset, it is the best of the eight different wavelet-decomposition forecasting strategies (see section 4.4.2.1.1. and 4.5.2.). When comparing the best wavelet-based forecasts against the non-wavelet-based forecasts (Table 4.3, Figure 4.3, and tables in section 4.D of our Supplementary Material [appendix]), we may notice that as the lead time increases, the difference in performance (generally) becomes more pronounced,

with the biggest gains in performance being realized for the 14 day lead time. This is a very promising quality of the WDDFF against the non-wavelet-based methods for the Montreal UWD dataset, as longer lead time forecasts can be used to constantly monitor sustainability, planning, and construction initiatives with reasonable foresight, such as projecting expected water use during droughts or the amount of supply required to a site or pressure zone, that would be beneficial in setting water use restrictions or allocating temporary services for water main maintenance or construction.

Since the MODWT-single-hybrid class of models provided the best forecasts for each lead time, it suggests, for this dataset, that including both wavelet-decomposed inputs and non-wavelet decomposed inputs can increase forecast performance considerably, in comparison to alternate cases where only wavelet-decomposed or non-wavelet decomposed inputs are solely used. This has also been the case in earlier wavelet-based forecasting studies (Adamowski, 2008b; Nguyen and Nabney, 2010; Voronin and Partanen, 2013). However, the use of the un-decomposed inputs alongside wavelet-decomposed inputs is very rarely explored within hydrology and water resources.

Furthermore, the superiority of the WMLR models shows that the linear wavelet-based models perform better than the nonlinear wavelet-based models for all lead times, while the nonlinear models tend to perform best when wavelet-decomposed data are *not* used as inputs. The reason behind this can be argued similarly to the discussion in Koutsoyiannis et al. (2010), where the authors provided evidence that while most hydrological and water resources processes are naturally nonlinear, when they are observed from the standpoint of their changes across different time scales, there tends to be a linear relationship between the original time series and the time series when viewed at its different scales of change.

We may also notice that in certain cases, the non-wavelet-based forecasts outperformed some of the WDDFF forecasts. For instance, at the 1 day lead time, the nonlinear non-wavelet-based forecasts outperform the nonlinear WDDFF models that use the AT-single, MODWT-single, and AT-within wavelet decomposition forecasting strategies; for 3, 5, 7, and 14 day lead times, the nonlinear non-wavelet-based forecasts outperform the nonlinear WDDFF models that use the MODWT-single wavelet decomposition forecasting strategy. Other similar cases can be inferred from Table 4.3 and section 4.D of the Supplementary Material [appendix]. These results highlight the importance of exploring a wide variety of different combinations wavelet decomposition strategies (MODWT and AT), decomposition levels and wavelet filters, and approaches for incorporating wavelet-decomposed data in forecast development (e.g., direct and multicomponent) – if this is not done it is possible to make generalized statements, such

as, wavelet decomposed inputs are not necessarily as useful as the original un-decomposed time series for forecasting purposes (Zhang et al., 2015).

4.6.2. Best Performing Decomposition Levels and Wavelet Filters

It can also be noted that a decomposition level of six and wavelet filters of length 14 provided the best performance for each lead time forecast (see Table 4.3 and section 4.D of the Supplementary Material [appendix]). Figure 4.4 shows the *U* time series decomposed by the MODWT and AT algorithms using the *la14* filter (the wavelet filter used in the best performing WDDFF forecasts at lead times 1, 3, 5, and 7). We can see that using a decomposition level of six with the *la14* filter results in smooth scaling coefficients that track inter-annual, annual, and intra-annual changes that are free of unnatural artefacts, such as blocks, sharks' fins, triangles, etc. (Percival and Walden, 2000, sec. 5.11).

It can also be seen that the AT algorithm exhibits much larger variation across the wavelet coefficients than the MODWT. The AT solely uses low-pass filtering, producing an additive decomposition, while the MODWT uses high-pass filtering to compute the wavelet coefficients, resulting in an energy-based decomposition (Aussem et al., 1998; Percival, 2008; Percival and Walden, 2000). The scaling coefficients (in Figure 4.4) re-affirms our earlier statement that the scaling coefficients produced by the MODWT and AT are the same.

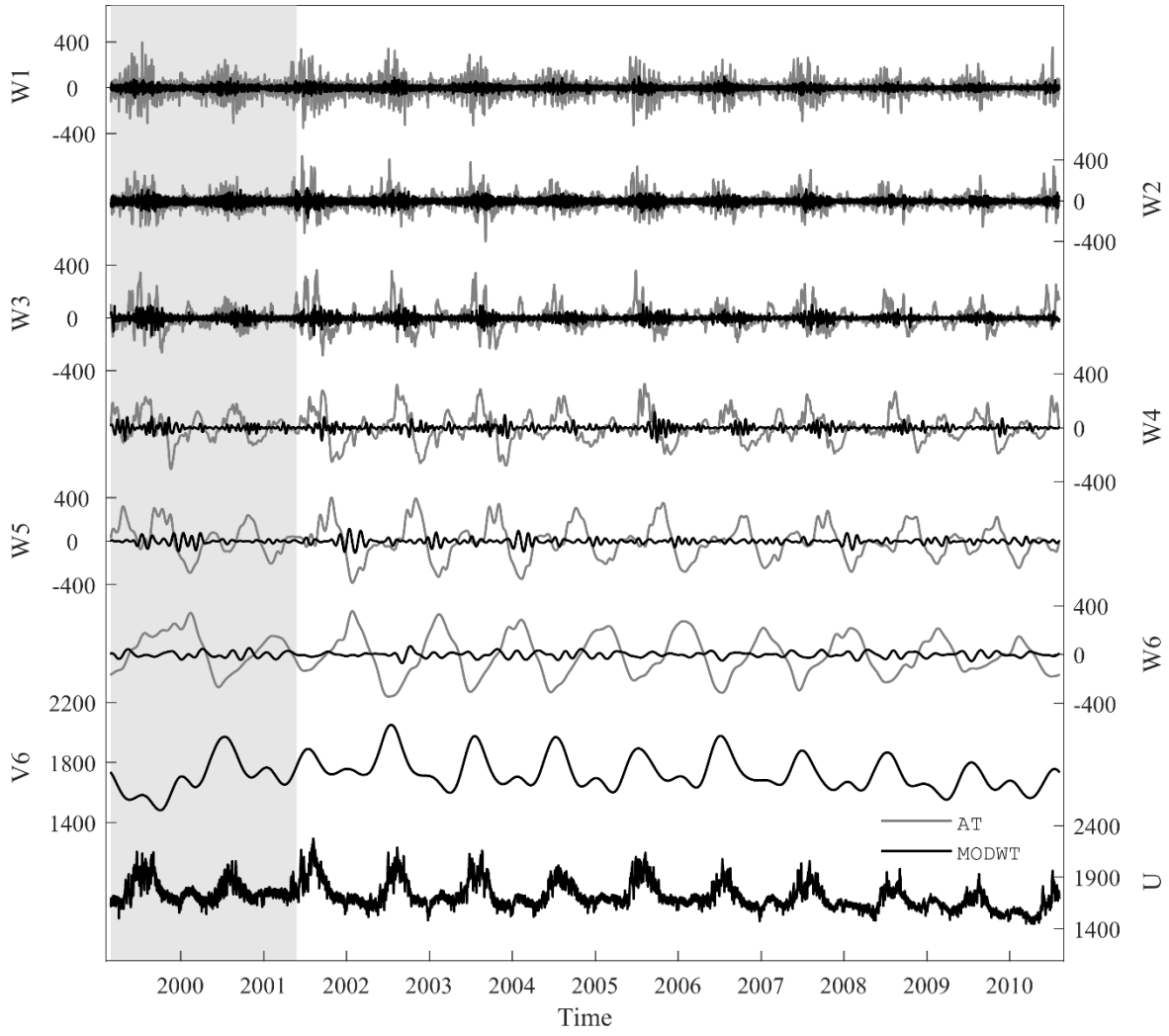

Figure 4.4. Average urban water demand time series 'U' (measured in megalitres per day) decomposed by the maximal overlap discrete wavelet transform (light colored line) and the *à trous* algorithm (dark colored line) using the la14 wavelet filter and a decomposition level of six; where each 'W' time series represents the wavelet coefficients at a particular scale and 'V' represents the scaling coefficients; the grayed out area represents the number of coefficients (820) that are affected by the boundary condition for the given wavelet filter and decomposition level

For a 1 day lead time, Figure 4.5 shows the performance of the best WDDFF and non-wavelet-based models for the last 366 time series records in the validation set. A pronounced weekly cycle is evident in the UWD time series. The weekly cycle is well captured by the *la14* and *bl14* wavelet filters within the

proposed WDDFF due to their larger support width, that is a multiple of a weekly periodicity for daily time series (when compared to shorter wavelet filters, such as the *haar*), which lend themselves well to capturing larger periods as opposed to smaller wavelet filters that are more suited to picking out transient events. In Figure 4.5 (b), one can notice the lower variation in the WMLR (WDDFF) residuals when compared to the (non-wavelet-based) SOV forecast giving further evidence to support the use of WDDFF over the benchmark models.

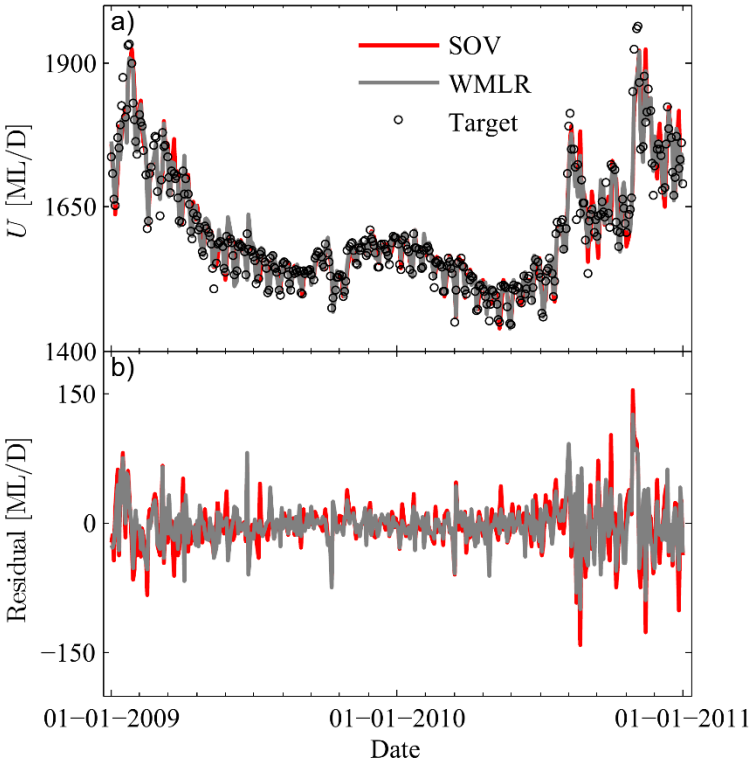

Figure 4.5. Comparison between the best wavelet-based (gray) and non-wavelet-based (red) 1 day lead time forecasts for the last 366 records in the validation set: a) represent the hydrograph while b) represents the residual, measured in megalitres per day (ML/D)

One may argue that certain wavelet-based forecasting approaches (e.g., Method 1-6, see section 4.4.2.1.1.) produce similar results and that they do not help to clearly identify a best model or a significant difference in performance between competitors. For example, let us consider the 1 day lead time (3, 5, 7, and 14 day lead times have a wider gap between the best performing model and its competitors in terms of RMSE, MAPE, and NASH scores): one can see that at this lead time, many similar forecast results

are achieved in terms of the NASH score (e.g., compare the AT-single, MODWT-single, AT-across, AT-single-hybrid, MODWT-single-hybrid, and AT-across-hybrid WMLR models in Table 4.3). However, one important point to note is that the optimal decomposition level and wavelet filter selected for each of these different wavelet-based forecasting approaches, that score similarly (e.g., in terms of NASH), tend to be different (e.g., WMLR AT-single-hybrid uses a decomposition level of three and the *coif1* wavelet filter while the MODWT-single-hybrid uses a decomposition level of six and the *la14* wavelet filter, both models' performance differ by a small margin, i.e., 0.002 in terms of NASH). This can be looked at positively by noting that methods that produced similar results but used different decomposition levels and wavelet filters could be assessed in terms of a trade-off between computational-efficiency and model accuracy. We argue that, without performing such a comprehensive exploration of these different wavelet-based forecasting approaches (Method 1-6, see section 4.4.2.1.1.), which has not yet been explored in such depth within the hydrology and water resources fields, one could *not* identify such cases. Therefore, our proposed WDDFF can be used as an exploratory analysis tool to identify optimal wavelet-based forecasting strategies (i.e., decomposition levels, wavelet filters, and forecasting approaches (Method 1-6)), trading-off between computational-efficiency and model accuracy. Using WDDFF as an exploratory analysis tool can be useful before selecting a single (or group of) model(s) for issuing forecasts using newly received data, such as for operational or planning tasks (e.g., opening of dam gates, identifying suitable crops to alleviate food shortages due to an impending drought, etc.).

We now compare WDDFF against a MODWT-MRA-based forecasting approach to show how its forecasts are realistic compared to the overly optimistic and pessimistic MODWT-MRA forecasts than *cannot* be used properly for real-world forecasting.

4.6.3. Comparing the Wavelet Data-Driven Forecasting Framework against a Forecasting Model using the Maximal Overlap Discrete Wavelet Transform-based Multiresolution Analysis

In Figure 4.6 we have shown how misinterpretation of the MODWT and instead applying the MODWT-MRA for forecasting (as in Kriechbaumer et al. (2014); Xiao et al. (2014); Yang et al. (2014); Zhu et al. (2014); Bachour et al. (2016); Barzegar et al. (2017a, 2017b); Prasad et al. (2017)), can lead to incorrect forecast performance. Figure 4.6 (a) shows the results of the best 1 day lead time forecast's residual on the validation set for the WDDFF which used the MODWT-single-hybrid approach and compares it against a MODWT-MRA-single-hybrid approach (Figure 4.6 (b)) - while we do not recommend this forecasting strategy, we simply adjusted the former by adopting the MODWT-MRA to decompose the target and explanatory variables instead of the MODWT. The same wavelet filter (*la14*), decomposition level (six),

and model type (WMLR) were used by both (MODWT- and MODWT-MRA-based) forecasts. Since the MODWT does not incorporate future information, the entire dataset can be decomposed at once without invoking uncertainty in the wavelet and scaling coefficients. However, the MODWT-MRA requires 'future data' when calculating detail and approximation coefficients at *any* given time. Since decomposing the calibration and validation set separately when using the MODWT-MRA causes 'wrap around' effects (see section 4.3.3.1.1.), we first decomposed the calibration data to obtain the calibration detail and approximation coefficients and then we decomposed the calibration and validation data together and extracted the detail and approximation coefficients associated with the validation indices to obtain the MODWT-MRA-based validation detail and approximation coefficients, in order to minimize the 'wrap around' effects (see section 4.C of our Supplementary Material [appendix] or refer to Eq. 4.2, 4.3, 4.5, and 4.6 in Table 4.1). Despite the fact that other studies commonly decompose both calibration and validation data separately (e.g., Barzegar et al. (2017a); Prasad et al. (2017)), which only exacerbates the BC-related uncertainty (see section 4.3.3.3.), our adjustment improved MODWT-MRA performance. A similar approach to our adjustment of the usual (and incorrect) manner in which the MODWT-MRA has been used in forecasting studies was also adopted in the DWT-MRA-based forecasting experiments in Du et al. (2017), where they demonstrated the incorrect forecasts produced by the DWT-MRA for a real-world monthly rainfall forecasting problem, which qualitatively resembles the results of the MODWT-MRA based forecasts in this study.

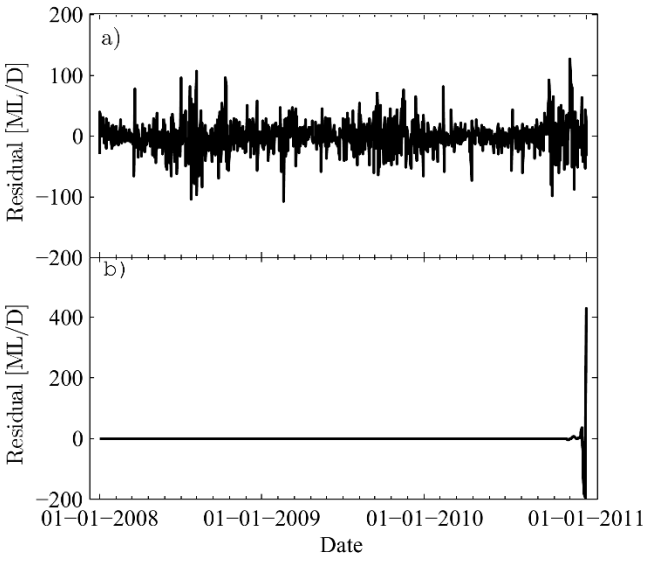

Figure 4.6. Plot showing the differences in 1 day lead time forecast residuals (in megalitres per day (ML/D)) for the validation partition where a) represents the maximal overlap discrete wavelet transform (valid for real-world forecasting problems) with a NASH score of 0.916 and b) the maximal overlap discrete wavelet transform-based multiresolution analysis (that should not be used for real-world forecasting problems due to its use of 'future data', see section 4.4.2.1., 4.4.2.3., and 4.6.1.) with a NASH score of 0.940

The MODWT-MRA-based forecast achieved a NASH score of 0.940 (which nears 1.0 when the poor quality forecasts at the end of the record are removed), significantly better than the MODWT (0.916). We may notice that the MODWT-MRA has negligible forecast error until it reaches the most recent time series observations where it utterly fails to provide a representative forecast, since it requires data from the future (which is replaced by surrogates, see section 4.3.3.3.) to accurately compute its detail and approximation components (which are obviously unavailable to us!), while the MODWT, which does not require 'future data' to calculate the wavelet and scaling coefficients at the same time positions, has stable performance with errors ranging (approximately) between $\pm$ 100 megalitres per day. Here we can clearly see the risks in applying the MODWT-MRA in forecasting studies – it truly cannot be used correctly for real-time forecasting.

4.7. Summary and Conclusions

Many wavelet-based water resources forecasting models, summarized in recent reviews (Afan et al., 2016; Dixit et al., 2016; Fahimi et al., 2017; Nourani et al., 2014; Sang, 2013; Yaseen et al., 2015), *cannot*

be used for real-world forecasting problems since they are *incorrectly* developed due to misunderstanding certain problems related to wavelet decomposition and the boundary condition: 1) the 'future data' issue; 2) inappropriately selecting decomposition levels and wavelet filters; and 3) not carefully partitioning calibration and validation data. By not addressing these boundary conditions during wavelet decomposition, the incorrectly developed wavelet-based forecasting models often result in much better performance than what is achievable in reality (Du et al., 2017) (for an example, see section 4.6.3.). To address these important and prevalent errors, we developed a set of best practices that were holistically combined in to a new wavelet-based forecasting framework, the Wavelet Data-Driven Forecasting Framework (WDDFF), that can be used to overcome these errors and that can be used for real-world hydrological and water resources forecasting problems (see section 4.3 and 4.4). Using a real-world daily urban water demand forecasting experiment in Montreal, Canada, we showed that WDDFF outperformed benchmark methods such as (non-wavelet-based) multiple linear regression, extreme learning machines, and a second-order Volterra model and provided more realistic and correct performance when compared against a recent wavelet-based forecasting strategy, that adopted the maximal overlap discrete wavelet transform-based multiresolution analysis, MODWT-MRA (Bachour et al., 2016; Barzegar et al., 2017; He et al., 2015; Prasad et al., 2017; Seo et al., 2017; Yang et al., 2014), and that showed unrealistic and incorrect performance during validation.

In addition to providing an approach that overcomes the errors mentioned above (thus allowing it to be used for real-world forecasting applications), a useful property of our proposed WDDFF is that it uses two different wavelet decomposition algorithms (maximal overlap discrete wavelet transform (MODWT) and *à trous* algorithm (AT)), two different wavelet-based forecasting approaches (direct and multicomponent), and has several methods (Single, Within, Across, Single-hybrid, Within-hybrid, or Across-hybrid - see section 4.4.2.1.1.) for adopting wavelet-decomposed data in the forecast design. Typically, in hydrological and water resources forecasting studies, only a single wavelet decomposition algorithm (e.g., MODWT-MRA), forecasting approach (e.g., direct), and method (e.g. Single) is used. For our experiments, the combination of different wavelet decomposition algorithms, wavelet-based forecasting approaches, and methods were demonstrated to provide significantly different performances when compared against one another, highlighting the usefulness, and supporting the practice, of exploring different strategies for incorporating wavelet-decomposed data in wavelet-based hydrological and water resources forecasting models. By addressing common pitfalls and errors of wavelet-based forecasts for real-world hydrological and water resources problems and providing a number of different strategies for using wavelet-decomposed data in the forecast design, WDDFF can be considered a promising new method for

hydrological and water resources forecasting and we expect its use to grow within hydrology and water resources (and potentially other domains).

Before discussing future research endeavours, we summarize below our main contributions and originality:

1. Since the wavelet-based forecasting literature in general (and specifically hydrology and water resources) does not contain any studies concerning best (and correct) practices for real-world wavelet-based forecasting models, this paper represents the first to comprehensively: a) compare correct (MODWT and AT) and incorrect (DWT-MRA and MODWT-MRA) wavelet decomposition methods for real-world forecasting (see section 4.2 and 4.3.2 of our revised manuscript); b) study the different boundary conditions that need to be properly addressed in order to correctly use a wavelet-based forecasting model in real-world problems (i.e., use of 'future data', improper selection of decomposition levels and wavelet filters, and improperly partitioning calibration and validation data - see section 4.2, 4.3.3, 4.3.4, and 4.3.5); c) propose (in addition to a set of best practices that addresses each boundary condition) a general wavelet-based data-driven forecasting framework (i.e., WDDFF) that can adopt any data–driven model and input variable selection routine, two wavelet decomposition algorithms (MODWT and AT), two wavelet-based forecasting approaches (direct and multicomponent), and several methods (Single, Within, Across, Single-hybrid, Within-hybrid, or Across-hybrid) for decomposing and including both target and explanatory data in the forecast design (see section 4.3.2., 4.4.2.1.1., and 4.4.3); and d) evaluate the forecasting performance of WDDFF on a real-world water resources forecasting problem using different combinations of the wavelet decomposition algorithms, wavelet-based forecasting approaches, and methods (see section 4.6).
2. This paper is the first to discuss and compare the differences between the MODWT and AT wavelet and scaling coefficients in terms of theory (see section 4.3.1) and forecasting performance (see section 4.6.2). This is also the first study to use the MODWT wavelet and scaling coefficients directly (i.e., instead of the MODWT-MRA detail and approximation coefficients) for hydrological or water resources forecasting (see section 4.3.2).
3. This is the first paper to apply a conditional mutual information-based input variable selection approach for identifying which MODWT and AT-based wavelet and scaling coefficients to use in a data-driven forecasting model (see section 4.4.2.2).

Future research will focus on converting WDDFF from a deterministic forecast, into a stochastic forecast - such a model has already been developed by the authors and will be published under the name Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF) – allowing for the forecasts from this method to be used for forecast systems subject to uncertainty and which may be useful for complimentary objectives such as risk assessment (Chen et al., 2013; Yung et al., 2011). One may expand WDDFF (and SWDDFF) by adopting singular spectrum analysis (SSA), to afford yet another method for taking into consideration the periodic and transient nature of hydrological and water resources processes – an interesting study would be to swap wavelets for SSA within our proposed WDDFF and compare performances between both methods. Finally, since WDDFF was only studied for a single water resources process (since this work served as an introduction to the new method), further research is required to explore its usefulness for forecasting other hydrological and water resources processes (e.g., drought, rainfall-runoff, water quality, evaporation, etc.).

4.8. Acknowledgements

The authors are grateful to the Associate Editor and anonymous reviewers for their constructive comments that have significantly improved the quality of this work. We thank the City of Montreal for providing the urban water demand data used in this study. Requests for urban water demand and meteorological data used in this study should be made through the City of Montreal or obtained from Environment Canada at climate.weather.gc.ca, respectively. Our appreciation is extended to Maheswaran Rathinasamy for his helpful discussions regarding the *à trous* algorithm and its relation to the boundary condition problem. Partial funding for this research was provided by an NSERC Discovery and Accelerator grants held by Jan Adamowski.

4.9. References

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407, 28–40. doi:10.1016/j.jhydrol.2011.06.013

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Adamowski, J., Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J. Hydrol. 390, 85–91.

doi:10.1016/j.jhydrol.2010.06.033

Adamowski, J.F., 2008a. River flow forecasting using wavelet and cross-wavelet transform models. Hydrol. Process. 22, 4877–4891. doi:10.1002/hyp.7107

Adamowski, J.F., 2008b. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. J. Hydrol. 353, 247–266. doi:10.1016/j.jhydrol.2008.02.013

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. doi:10.1016/j.jhydrol.2016.07.048

Aggarwal, S.K., Saini, L.M., Kumar, A., 2008. Price forecasting using wavelet transform and LSE based mixed model in Australian electricity market. Int. J. Energy Sect. Manag. 2, 521–546. doi:10.1108/17506220810919054

Akrami, S.A., El-Shafie, A., Naseri, M., Santos, C.A.G., 2014. Rainfall data analyzing using moving average (MA) model and wavelet multi-resolution intelligent model for noise evaluation to improve the forecasting accuracy. Neural Comput. Appl. 25, 1853–1861. doi:10.1007/s00521-014-1675-0

Alizadeh, M.J., Kavianpour, M.R., 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. Mar. Pollut. Bull. 98, 171–178. doi:10.1016/j.marpolbul.2015.06.052

Alizadeh, M.J., Kavianpour, M.R., Kisi, O., Nourani, V., 2017. A new approach for simulating and forecasting the rainfall-runoff process within the next two months. J. Hydrol. 548, 588–597. doi:10.1016/j.jhydrol.2017.03.032

Alsberg, B.K., Woodward, A.M., Winson, M.K., Rowland, J.J., Kell, D.B., 1998. Variable selection in wavelet regression models. Anal. Chim. Acta. doi:10.1016/S0003-2670(98)00194-9

Altunkaynak, a., Özger, M., Sen, Z., 2003. Triple diagram model of level fluctuations in Lake Van, Turkey. Hydrol. Earth Syst. Sci. 7, 235–244. doi:10.5194/hess-7-235-2003

Altunkaynak, A., Nigussie, T.A., 2017. Monthly Water Consumption Prediction Using Season Algorithm and Wavelet Transform–Based Models. J. Water Resour. Plan. Manag. 143, 04017011. doi:10.1061/(ASCE)WR.1943-5452.0000761

Andersen, C.M., Bro, R., 2010. Variable selection in regression-a tutorial. J. Chemom. 24, 728–737. doi:10.1002/cem.1360

Aussem, A., Campbell, J., Murtagh, F., 1998. Wavelet-based feature extraction and decomposition strategies for financial forecasting. J. Comput. Intell. Financ. 6, 5–12.

Bachour, R., Maslova, I., Ticlavilca, A.M., Walker, W.R., McKee, M., 2016. Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. Stoch. Environ. Res. Risk Assess. 30, 103–117. doi:10.1007/s00477-015-1039-z

Bakshi, B., 1999. Multiscale analysis and modeling using wavelets. J. Chemom. 434, 415–434.

Barzegar, R., Asghari Moghaddam, A., Adamowski, J., Ozga-Zielinski, B., 2018. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Environ. Res. Risk Assess. 32, 799–813. doi:10.1007/s00477-017-1394-z

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189

Bašta, M., 2014. Additive decomposition and boundary conditions in wavelet-based forecasting approaches. Acta Oeconomica Pragensia 2014, 48–70.

Baydaroğlu, Ö., Koçak, K., Duran, K., 2017. River flow prediction using hybrid models of support vector regression with the wavelet transform, singular spectrum analysis and chaotic approach. Meteorol. Atmos. Phys. doi:10.1007/s00703-017-0518-9

Belayneh, A., Adamowski, J., Khalil, B., 2016a. Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain. Water Resour. Manag. 2, 87–101. doi:10.1007/s40899-015-0040-5

Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models. J. Hydrol. 508, 418–429. doi:10.1016/j.jhydrol.2013.10.052

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016b. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172–173. doi:10.1016/j.atmosres.2015.12.017

Benaouda, D., Murtagh, F., Starck, J.L., Renaud, O., 2006. Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting. Neurocomputing. doi:10.1016/j.neucom.2006.04.005

Bogner, K., Kalas, M., 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. Atmos. Sci. Lett. doi:10.1002/asl.180

Campisi-Pinto, S., Adamowski, J., Oron, G., 2013. Forecasting Urban Water Demand Via Wavelet-Denoising and Neural Network Models. Case Study: City of Syracuse, Italy. Water Resour. Manag. 27, 319–321.

Cannas, B., Fanni, A., See, L., Sias, G., 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. Phys. Chem. Earth 31, 1164–1171. doi:10.1016/j.pce.2006.03.020

Chen, C., Huang, G.H., Li, Y.P., Zhou, Y., 2013. A robust risk analysis method for water resources allocation under uncertainty. Stoch. Environ. Res. Risk Assess. 27, 713–723. doi:10.1007/s00477-012-0634-5

Creaco, E., Berardi, L., Sun, S., Giustolisi, O., Savic, D., 2016. Selection of relevant input variables in storm water quality modeling by multiobjective evolutionary polynomial regression paradigm. Water Resour. Res. 52, 2403–2419. doi:10.1002/2015WR017971

Crowley, P.M., 2007. A guide to wavelets for economists. J. Econ. Surv. doi:10.1111/j.1467-6419.2006.00502.x

Deo, R.C., Downs, N., Parisi, A. V., Adamowski, J.F., Quilty, J.M., 2017a. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. Environ. Res. 155, 141–166. doi:10.1016/j.envres.2017.01.035

Deo, R.C., Kisi, O., Singh, V.P., 2017b. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos. Res. 184, 149–175. doi:10.1016/j.atmosres.2016.10.004

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2017c. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stoch. Environ. Res. Risk Assess. 31, 1211–1240. doi:10.1007/s00477-016-1265-z

Dixit, P., Londhe, S., Deo, M.C., 2016. Review of Applications of Neuro-Wavelet Techniques in Water Flows. Ina. Lett. 1, 99–104. doi:10.1007/s41403-016-0015-3

Djerbouai, S., Souag-Gamane, D., 2016. Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Stochastic Models: Case of the Algerois Basin in North Algeria. Water Resour. Manag. 30, 2445–2464. doi:10.1007/s11269-016-1298-6

Du, K., Zhao, Y., Lei, J., 2017. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. J. Hydrol. 552, 44–51. doi:10.1016/j.jhydrol.2017.06.019

Ebrahimi, H., Rajaee, T., 2017. Simulation of groundwater level variations using wavelet combined with neural network, linear regression and support vector machine. Glob. Planet. Change 148, 181–191. doi:10.1016/j.gloplacha.2016.11.014

Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor. Appl. Climatol. 128, 875–903. doi:10.1007/s00704-016-1735-8

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176. doi:10.1016/j.jhydrol.2008.10.019

Galelli, S., Castelletti, A., 2013. Tree-based iterative input variable selection for hydrological modeling. Water Resour. Res. 49, 4295–4310. doi:10.1002/wrcr.20339

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. doi:10.1016/j.envsoft.2014.08.015

Garg, A., Garg, A., Tai, K., 2014. A multi-gene genetic programming model for estimating stress-dependent soil water retention curves. Comput. Geosci. 18, 45–56. doi:10.1007/s10596-013-9381-z

Ghazvinei, P.T., Shamshirband, S., Motamedi, S., Hassanpour Darvishi, H., Salwana, E., 2017. Performance investigation of the dam intake physical hydraulic model using Support Vector Machine with a discrete wavelet transform algorithm. Comput. Electron. Agric. 140, 48–57. doi:10.1016/j.compag.2017.05.033

Goyal, M.K., Bharti, B., Quilty, J., Adamowski, J., Pandey, A., 2014. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst. Appl. 41.

doi:10.1016/j.eswa.2014.02.047

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed, Springer Series in Statistics. Springer New York, New York.

He, X., Guan, H., Qin, J., 2015. A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. J. Hydrol. 527, 88–100. doi:10.1016/j.jhydrol.2015.04.047

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. doi:10.1016/j.neucom.2005.12.126

Karbasi, M., 2017. Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet-Gaussian Process Regression Model. Water Resour. Manag. doi:10.1007/s11269-017-1853-9

Kariwala, V., Ye, L., Cao, Y., 2013. Branch and bound method for regression-based controlled variable selection. Comput. Chem. Eng. 54, 1–7. doi:10.1016/j.compchemeng.2013.03.006

Karthikeyan, L., Kumar, D.N., 2013. Predictability of nonstationary time series using wavelet and EMD based ARMA models. J. Hydrol. 502, 103–119. doi:10.1016/j.jhydrol.2013.08.030

Khalil, B., Broda, S., Adamowski, J., Ozga-Zielinski, B., Donohoe, A., 2015. Short-term forecasting of groundwater levels under conditions of mine-tailings recharge using wavelet ensemble neural network models. Hydrogeol. J. 23, 121–141. doi:10.1007/s10040-014-1204-3

Kim, T.-W., Valdés, J.B., 2003. Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks. J. Hydrol. Eng. 8, 319–328. doi:10.1061/(ASCE)1084-0699(2003)8:6(319)

Kisi, O., 2016. Modeling reference evapotranspiration using three different heuristic regression approaches. Agric. Water Manag. 169, 162–172. doi:10.1016/j.agwat.2016.02.026

Kişi, Ö., 2011. A combined generalized regression neural network wavelet model for monthly streamflow prediction. KSCE J. Civ. Eng. 15, 1469–1479. doi:10.1007/s12205-011-1004-4

Koutsoyiannis, D., Yao, H., Georgakakos, A., 2010. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrol. Sci. J. 53, 142–164. doi:10.1623/hysj.53.1.142

Kriechbaumer, T., Angus, A., Parsons, D., Rivas Casado, M., 2014. An improved wavelet-ARIMA approach for forecasting metal prices. Resour. Policy 39, 32–41. doi:10.1016/j.resourpol.2013.10.005

Krishna, B., 2013. Comparison of wavelet based ANN and Regression models for Reservoir Inflow Forecasting. J. Hydrol. Eng. 19, 1385–1400. doi:10.1061/(ASCE)HE.1943-5584.0000892

Labat, D., Ababou, R., Mangin, A., 1999. Linear and nonlinear input/output models for karstic springflow and flood prediction at different time scales. Stoch. Environ. Res. Risk Assess. 13, 337–364. doi:10.1007/s004770050055

Lee, T., Ouarda, T.B.M.J., Yoon, S., 2017. KNN-based local linear regression for the analysis and simulation of low flow extremes under climatic influence. Clim. Dyn. 49, 3493-3511. doi:10.1007/s00382-017-3525-0

Li, B.J., Cheng, C.T., 2014. Monthly discharge forecasting using wavelet neural networks with extreme learning machine. Sci. China Technol. Sci. 57, 2441–2452. doi:10.1007/s11431-014-5712-0

Li, S., Wang, P., Goel, L., 2016. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. IEEE Trans. Power Syst. 31, 1788–1798. doi:10.1109/TPWRS.2015.2438322

Maheswaran, R., Khosa, R., 2012a. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

Maheswaran, R., Khosa, R., 2012b. Wavelet-Volterra coupled model for monthly stream flow forecasting. J. Hydrol. 450–451, 320–335. doi:10.1016/j.jhydrol.2012.04.017

Mallat, S.G., 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693. doi:10.1109/34.192463

Maslova, I., Ticlavilca, A.M., Mckee, M., 2016. Adjusting wavelet-based multiresolution analysis boundary conditions for long-term streamflow forecasting. Hydrol. Process. doi:10.1002/hyp.10564

Mirbagheri, S.A., Nourani, V., Rajaee, T., Alikhani, A., 2010. Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. Hydrol. Sci. J. 55, 1175–1189. doi:10.1080/02626667.2010.508871

Moosavi, V., Vafakhah, M., Shirmohammadi, B., Behnia, N., 2013. A Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods. Water Resour. Manag. 27, 1301–

1321. doi:10.1007/s11269-012-0239-2

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. doi:10.13031/2013.23153

Murtagh, F., Starck, J.L., Renaud, O., 2004. On neuro-wavelet modeling. Decis. Support Syst. 37, 475–484. doi:10.1016/S0167-9236(03)00092-7

Nguyen, H.T., Nabney, I.T., 2010. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy 35, 3674–3685. doi:10.1016/j.energy.2010.05.013

Nielsen, M., 2001. On the Construction and Frequency Localization of Finite Orthogonal Quadrature Filters. J. Approx. Theory 108, 36–52.

Nourani, V., Alami, M.T., Aminfar, M.H., 2009a. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng. Appl. Artif. Intell. 22, 466–472. doi:10.1016/j.engappai.2008.09.003

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. J. Hydrol. 514, 358–377. doi:10.1016/j.jhydrol.2014.03.057

Nourani, V., Komasi, M., Mano, A., 2009b. A multivariate ANN-wavelet approach for rainfall-runoff modeling. Water Resour. Manag. 23, 2877–2894. doi:10.1007/s11269-009-9414-5

Nourani, V., Saeidifarzad, B., 2017. Detection of land use/cover change effect on watershed's response in generating runoff using computational intelligence approaches. Stoch. Environ. Res. Risk Assess. 31, 1341–1357. doi:10.1007/s00477-016-1220-z

Olhede, S., Walden, A.T., 2004. The Hilbert spectrum via wavelet projections. Proc. R. Soc. A Math. Phys. Eng. Sci. 460, 955–975. doi:10.1098/rspa.2003.1199

Patil, K., Deo, M.C., 2017. Prediction of daily sea surface temperature using efficient neural networks. Ocean Dyn. 67, 357–368. doi:10.1007/s10236-017-1032-9

Percival, D.B., 2008. Analysis of Geophysical Time Series Using Discrete Wavelet Transforms: An Overview, in: Donner, R. V, Barbosa, S.M. (Eds.), Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology, Geodynamics and Solar-Terrestrial Physics. Springer Berlin Heidelberg, Berlin,

Heidelberg, pp. 61–79. doi:10.1007/978-3-540-78938-3_4

Percival, D.B., Lennox, S.M., Wang, Y.G., Darnell, R.E., 2011. Wavelet-based multiresolution analysis of Wivenhoe Dam water temperatures. Water Resour. Res. doi:10.1029/2010WR009657

Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis, 1st ed, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ANN-based stream flow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos. Res. 197, 42–63. doi:10.1016/j.atmosres.2017.06.014

Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. Water Resour. Res. 52, 2299–2326. doi:10.1002/2015WR016959

Rahimikhoob, A., 2016. Comparison of M5 Model Tree and Artificial Neural Networks Methodologies in Modelling Daily Reference Evapotranspiration from NOAA Satellite Images. Water Resour. Manag. 1–13. doi:10.1007/s11269-016-1331-9

Rajaee, T., 2011. Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. Sci. Total Environ. 409, 2917–2928. doi:10.1016/j.scitotenv.2010.11.028

Rana, M., Koprinska, I., 2016. Forecasting electricity load with advanced wavelet neural networks . Neurocomputing 182, 118–132. doi:10.1016/j.neucom.2015.12.004

Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J. Hydrol. doi:10.1016/j.jhydrol.2013.09.025

Rathinasamy, M., Khosa, R., Adamowski, J., Ch, S., Partheepan, G., Anand, J., Narsimlu, B., 2014. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. Water Resour. Res. 50, 9721–9737. doi:10.1002/2013WR014650

Ravansalar, M., Rajaee, T., 2015. Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. Environ. Monit. Assess. 187, 366. doi:10.1007/s10661-015-4590-7

Rezaie-balf, M., Naganna, S.R., Ghaemi, A., Deka, P.C., 2017. Wavelet coupled MARS and M5 model tree approaches for groundwater level forecasting. J. Hydrol. 553, 356–373.

doi:10.1016/j.jhydrol.2017.08.006

Sahay, R.R., Srivastava, A., 2014. Predicting Monsoon Floods in Rivers Embedding Wavelet Transform, Genetic Algorithm and Neural Network. Water Resour. Manag. 28, 301–317. doi:10.1007/s11269-013-0446-5

Samadianfard, S., Asadi, E., Jarhan, S., Kazemi, H., Kheshtgar, S., Kisi, O., Sajjadi, S., Manaf, A.A., 2018. Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths. Soil Tillage Res. 175, 37–50. doi:10.1016/j.still.2017.08.012

Sang, Y.-F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. 122, 8–15. doi:10.1016/j.atmosres.2012.11.003

Sehgal, V., Tiwari, M.K., Chatterjee, C., Sehgal, V., Tiwari, M.K., Chatterjee, C., 2014. Wavelet Bootstrap Multiple Linear Regression Based Hybrid Modeling for Daily River Discharge Forecasting. Water Resour Manag. 28, 2793–2811. doi:10.1007/s11269-014-0638-7

Seo, Y., Choi, Y., Choi, J., 2017. River stage modeling by combining maximal overlap discrete wavelet transform, support vector machines and genetic algorithm. Water (Switzerland) 9. doi:10.3390/w9070525

Shafaei, M., Kisi, O., 2016. Lake Level Forecasting Using Wavelet-SVR, Wavelet-ANFIS and Wavelet-ARMA Conjunction Models. Water Resour. Manag. 30, 79–97. doi:10.1007/s11269-015-1147-z

Sharma, A., Mehrotra, R., Li, J., Jha, S., 2016. A programming tool for nonparametric system prediction using Partial Informational Correlation and Partial Weights. Environ. Model. Softw. 83, 271–275. doi:10.1016/j.envsoft.2016.05.021

Shi, B., Wang, P., Jiang, J., Liu, R., 2018. Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. Sci. Total Environ. 610–611, 1390–1399. doi:10.1016/j.scitotenv.2017.08.232

Shirmohammadi, B., Moradi, H., Moosavi, V., Semiromi, M.T., Zeinali, A., 2013. Forecasting of meteorological drought using Wavelet-ANFIS hybrid model for different time steps (case study: Southeastern part of east Azerbaijan province, Iran). Nat. Hazards 69, 389–402. doi:10.1007/s11069-013-0716-9

Shoaib, M., Shamseldin, A.Y., Khan, S., Khan, M.M., Khan, Z.M., Sultan, T., Melville, B.W., 2018. A

Comparative Study of Various Hybrid Wavelet Feedforward Neural Network Models for Runoff Forecasting. Water Resour. Manag. 32, 83–103. doi:10.1007/s11269-017-1796-1

Shoaib, M., Shamseldin, A.Y., Melville, B.W., Khan, M.M., 2016a. Hybrid Wavelet Neuro-Fuzzy Approach for Rainfall-Runoff Modeling. J. Comput. Civ. Eng. 30, 04014125. doi:10.1061/(ASCE)CP.1943-5487.0000457

Shoaib, M., Shamseldin, A.Y., Melville, B.W., Khan, M.M., 2016b. A comparison between wavelet based static and dynamic neural network approaches for runoff prediction. J. Hydrol. 535, 211–225. doi:10.1016/j.jhydrol.2016.01.076

Shoaib, M., Shamseldin, A.Y., Melville, B.W., Khan, M.M., 2015. Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. J. Hydrol. 527, 326–344. doi:10.1016/j.jhydrol.2015.04.072

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinformatics 10, 3–22. doi:10.2166/hydro.2008.015

Stulp, F., Sigaud, O., 2015. Many regression algorithms, one unified model: A review. Neural Networks 69, 60–79. doi:10.1016/j.neunet.2015.05.005

Taormina, R., Galelli, S., Karakaya, G., Ahipasaoglu, S.D., 2016. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models. J. Hydrol. 542, 18–34. doi:10.1016/j.jhydrol.2016.07.045

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Tran, H.D., Muttil, N., Perera, B.J.C., 2016. Enhancing accuracy of autoregressive time series forecasting with input selection and wavelet transformation. J. Hydroinformatics. doi:10.2166/hydro.2016.145

Tsimpiris, A., Vlachos, I., Kugiumtzis, D., 2012. Nearest neighbor estimate of conditional mutual information in feature selection. Expert Syst. Appl. 39, 12697–12708. doi:10.1016/j.eswa.2012.05.014

Voronin, S., Partanen, J., 2013. Price forecasting in the day-ahead energy market by an iterative method with separate normal price and price spike frameworks. Energies 6, 5897–5920.

doi:10.3390/en6115897

Walden, A.T., 2001. Wavelet Analysis of Discrete Time Series, in: Casacuberta, C., Miró-Roig, R.M., Verdera, J., Xambó-Descamps, S. (Eds.), European Congress of Mathematics: Barcelona, July 10--14, 2000 Volume II. Birkhäuser Basel, Basel, pp. 627–641. doi:10.1007/978-3-0348-8266-8_56

Xiao, Y., Xiao, J., Liu, J., Wang, S., 2014. A multiscale modeling approach incorporating ARIMA and anns for financial market volatility forecasting. J. Syst. Sci. Complex. 27, 225–236. doi:10.1007/s11424-014-3305-4

Yadav, B., Eliza, K., 2017. A hybrid wavelet-support vector machine model for prediction of Lake water level fluctuations using hydro-meteorological data. Meas. J. Int. Meas. Confed. 103, 294–301. doi:10.1016/j.measurement.2017.03.003

Yang, X., Zhang, H., Zhou, H., 2014. A Hybrid Methodology for Salinity Time Series Forecasting Based on Wavelet Transform and NARX Neural Networks. Arab. J. Sci. Eng. 39, 6895–6905. doi:10.1007/s13369-014-1243-z

Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. J. Hydrol. doi:10.1016/j.jhydrol.2015.10.038

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-Shafie, A., 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. J. Hydrol. 542, 603–614. doi:10.1016/j.jhydrol.2016.09.035

Yu, H., Wen, X., Feng, Q., Deo, R.C., Si, J., Wu, M., 2018. Comparative Study of Hybrid-Wavelet Artificial Intelligence Models for Monthly Groundwater Depth Forecasting in Extreme Arid Regions, Northwest China. Water Resour. Manag. 32, 301–323. doi:10.1007/s11269-017-1811-6

Yung, B.B., Tolson, B.A., Burn, D.H., 2011. Risk assessment of a water supply system under climate variability: a stochastic approach. Can. J. Civ. Eng. 38, 252–262. doi:10.1139/L10-132

Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J. Hydrol. 530, 137–152. doi:10.1016/j.jhydrol.2015.09.047

Zhang, Z., Telesford, Q.K., Giusti, C., Lim, K.O., Bassett, D.S., 2016. Choosing wavelet methods, filters, and lengths for functional brain network construction. PLoS One. doi:10.1371/journal.pone.0157243

Zhu, L., Wang, Y., Fan, Q., 2014. MODWT-ARMA model for time series prediction. Appl. Math. Model. doi:10.1016/j.apm.2013.10.002

Appendix

This appendix contains online supplementary material related to Chapter 4.

4. A.1. Details on Wavelet Decomposition Wavelet and Scaling Coefficients, and Wavelet and Scaling Filters

Wavelet decomposition through the discrete wavelet transform (DWT), maximal overlap DWT (MODWT), or *à trous* algorithm (AT) follows a recursive pyramid algorithm (Figure 4.A.1) and results in (DWT, MODWT, and AT) wavelet (W_j, $\widetilde{W}_j$, and $\widetilde{W}_{j,t}^a$) and scaling coefficients (V_j, $\tilde{V}_j$, and $\tilde{V}_{j,t}^a$), while the reconstruction follows a 'reverse' recursive pyramid algorithm (Figure 4.A.2 (a)) and converts the wavelet and scaling coefficients back into the original time series (see also Table 4.1 in the text).

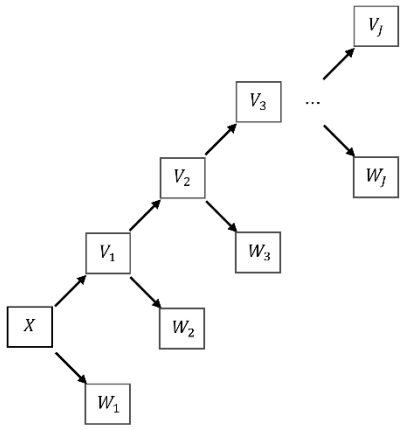

Figure 4.A.1. A block diagram illustrating the process of wavelet decomposition for a given time series (X) resulting in wavelet $\{W_1, \ldots, W_J\}$ and scaling coefficients $\{V_J\}$ using either the discrete wavelet transform (DWT), maximal overlap DWT (MODWT), or *à trous* algorithm (AT). Note that $X \equiv V_0$.

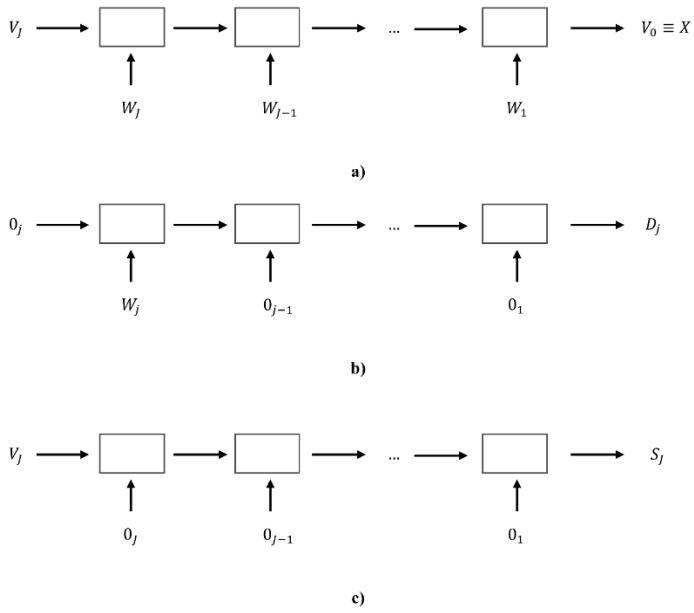

Figure 4.A.2. Using the block diagram notation in Percival and Walden (2000): a) The process for reconstructing a time series (X) from its wavelet $\{W_1, \dots, W_J\}$ and scaling coefficients $\{V_J\}$ using either the discrete wavelet transform, maximal overlap DWT, or *à trous* algorithm; b) the process of calculating the detail $\{D_1, \dots, D_J\}$ coefficients using multiresolution analysis (MRA) via the DWT-MRA and MODWT-MRA; and c) the process of calculating the approximation coefficients $\{S_J\}$ using MRA via the DWT-MRA and MODWT-MRA. Note that 0_k, for $k = 1, \dots, j$, represents a vector of zeros of size $N/2^k$ for the DWT-MRA and N for the MODWT-MRA. The regular reconstruction formula for the DWT and MODWT is used to recover the details and approximation coefficients. Also note that the AT is an additive deconstruction of the time series into wavelet and scaling coefficients, these coefficients can also be considered detail and approximation coefficients as they are equivalent (Aussem et al., 1998).

The DWT and MODWT wavelet coefficients, W_j and $\widetilde{W}_j$ (Eq. 4.1 and 4.4, respectively – see Table 4.1 in text), are associated with changes in averages over scale $\tau_j = 2^{j-1}$ ($j \in 1,2, \dots, J$) where J is the decomposition level, while the DWT, MODWT, and AT scaling coefficients, V_j, $\tilde{V}_j$, and $\tilde{V}_{j,t}^a$ (Eq. 4.2, 4.4, and 4.7, respectively) are linked to variations at scales $\lambda_j = 2^j$ and higher. The AT wavelet coefficients, $\widetilde{W}_{j,t}^a$, represent differences in variations between scales 2^{j-1} and 2^j(Aussem et al., 1998; Percival and Mofjeld, 1997). It follows that the scaling filters, g_l ($l \in 0, \dots, L-1$, where L is the filter length), are low-pass filters and the wavelet filters, h_l ($l \in 0, \dots, L-1$), are high-pass filters, fulfilling the quadrature mirror relationship, $g_l \equiv (-1)^{l+1} h_{L-1-l}$ (where $g_l \equiv \sqrt{2\tilde{g}_l}$ and $h_l \equiv \sqrt{2\tilde{h}_l}$). Therefore, the j-th level wavelet

coefficients are associated with frequencies in the pass-band $\left[\frac{1}{2^{j+1}}, \frac{1}{2^{j}}\right]$ while the j-th level scaling coefficients are associated with the interval of frequencies in $\left[0, \frac{1}{2^{j+1}}\right]$. Further properties of the wavelet and scaling filters can be found in Bašta (2014), Percival and Mofjeld (1997), Percival and Walden (2000), and Walden (2001). For illustration purposes, see Table 4.A.1 for the frequency bands corresponding to different decomposition levels at the daily scale (we use the daily scale since our case study involves daily time series forecasting, see section 4.5.2. and 4.6) and Table 4.A.2 for different scaling filters (up to length 14) (these filters are also used in our case study).

Another key difference between the DWT and the MODWT and AT is that the DWT involves decimation. Therefore, DWT wavelet and scaling coefficients at a given level j have 2^j fewer coefficients than the previous scale $(j-1)$. The decimation of the DWT causes several issues for wavelet-based forecasting applications, thus making it invalid for real-world forecasting problems:

1. It requires the time series to be an integer multiple of $2^J$to compute the wavelet and scaling coefficients (in cases where this is not met, time series "extension methods", such as zero-padding, are used to make the time series an integer multiple of 2^J (Percival and Walden, 2000));

2. It is shift-variant: circularly shifting the time series by some amount will not circularly shift the DWT wavelet and scaling coefficients by the same amount (Walden, 2001) – see Figure 1 in Maheswaran and Khosa (2012a) for a clear example of this issue; and

3. It is sensitive to adding new data points: including additional data points (such as acquiring new data from a sensor in real-time and updating a forecast model with this new data) and re-calculating the DWT creates errors, i.e., there is not a smooth continuity in the wavelet and scaling coefficients at time t and $t+1$ when the DWT is re-calculated using new data occurring at $t+1$ – this is a significant issue for real-world forecasting applications where forecast models must be updated regularly – see Figure 2 in Maheswaran and Khosa (2012a) for a clear example of this issue.

Due to the decimation of the DWT wavelet and scaling coefficients, one must adopt multiresolution analysis (MRA) to convert DWT wavelet and scaling coefficients into detail ($\{D_1, D_2, \dots, D_J\}$) and approximation coefficients (S_J) in order to use the DWT for wavelet-based forecasting (Kisi and Cimen, 2011). The detail and approximation coefficients for a given scale, e.g., D_j or A_J, are easily calculated by setting all other wavelet and/or scaling coefficients (i.e., except W_j or V_J) to 0 and sequentially applying

the DWT reconstruction algorithm (Eq. 4.7 in Table 4.1 in text; see also Figure 4.A.2 (b) and (c)). Sequential passes through the reconstruction algorithm (Eq. 4.7 in Table 4.1 in text) involves up-sampling (see the notes for the DWT in Table 4.1 text) and for this reason, the detail(s) and approximation coefficients retain the same number of observations as the original time series. The same process can be followed for the MODWT and an MODWT-based MRA (MODWT-MRA) can be produced via Eq. 4.8 (Table 4.1 in text), resulting in MODWT-MRA detail(s) ($\widetilde{D}_1, \widetilde{D}_2, \ldots, \widetilde{D}_J$) and approximation coefficients ($\tilde{S}_J$).

Table 4.A.1. Different decomposition levels and the corresponding frequency and time-scale ranges for daily time series (copied from Bašta (2011))

Decomposition Level	Corresponding Frequency Range	Changes Between Averages which are Calculated on the Time-Scale of:
1	$2^{-2} < f \leq 2^{-1}$	2^{1-1} day = 1 day
2	$2^{-3} < f \leq 2^{-2}$	2 days
3	$2^{-4} < f \leq 2^{-3}$	4 days
4	$2^{-5} < f \leq 2^{-4}$	8 days ≈ 1 week
5	$2^{-6} < f \leq 2^{-5}$	16 days
6	$2^{-7} < f \leq 2^{-6}$	32 days ≈ 1 month
7	$2^{-8} < f \leq 2^{-7}$	64 days
8	$2^{-9} < f \leq 2^{-8}$, contains the frequency year^{-1}	128 days
9	$2^{-10} < f \leq 2^{-9}$	256 days
10	$2^{-11} < f \leq 2^{-10}$	512 days

Table 4.A.2. Different scaling filters

Filter (g)	**Scaling Filter Coefficients**													
haar/d1	0.707	0.707												
d2	0.483	0.837	0.224	-0.129										
fk4	0.654	0.753	0.053	-0.046										
d3	0.333	0.807	0.460	-0.135	-0.085	0.035								
fk6	0.428	0.813	0.356	-0.146	-0.077	0.041								
coif1	-0.073	0.338	0.853	0.385	-0.073	-0.016								
d4	0.230	0.715	0.631	-0.028	-0.187	0.031	0.033	-0.011						
sym4	0.032	-0.013	-0.099	0.298	0.804	0.498	-0.030	-0.076						
fk8	0.349	0.783	0.475	-0.100	-0.160	0.043	0.043	-0.019						
la8	-0.076	-0.030	0.498	0.804	0.298	-0.099	-0.013	0.032						
d5	0.160	0.604	0.724	0.138	-0.242	-0.032	0.078	-0.006	-0.013	0.003				
la10	0.020	-0.021	-0.175	0.017	0.634	0.723	0.199	-0.039	0.030	0.027				
d6	0.112	0.495	0.751	0.315	-0.226	-0.130	0.098	0.028	-0.032	0.001	0.005	-0.001		
sym6	-0.008	0.002	0.045	-0.021	-0.073	0.338	0.788	0.491	-0.048	-0.118	0.003	0.015		
coif2	0.016	-0.041	-0.067	0.386	0.813	0.417	-0.076	-0.059	0.024	0.006	-0.002	-0.001		
la12	0.015	0.003	-0.118	-0.048	0.491	0.788	0.338	-0.073	-0.021	0.045	0.002	-0.008		
d7	0.078	0.397	0.729	0.470	-0.144	-0.224	0.071	0.081	-0.038	-0.017	0.013	0.000	-0.002	0.000
fk14	0.260	0.687	0.612	0.051	-0.246	-0.049	0.124	0.022	-0.064	-0.005	0.030	-0.003	-0.009	0.004
la14	0.010	0.004	-0.108	-0.140	0.289	0.768	0.536	0.017	-0.050	0.068	0.031	-0.013	-0.001	0.003
bl14	0.012	0.017	-0.065	-0.064	0.360	0.782	0.484	-0.057	-0.101	0.045	0.020	-0.018	-0.003	0.002

4. B.1. Study Area Details

Montreal, Quebec is the second largest city in Canada and provides safe drinking water to more than 1.9 million citizens. Montreal draws water from Riviere des Prairies, Lac Saint-Louis, and the St. Lawrence River where it is treated by six separate water treatment plants. The city has made improvements in reducing the production of their water supply (while still meeting demands) with a decrease in yearly production by about 110 million cubic meters between 2001 and 2013 and now resides at a daily production rate of about 1.75 million cubic meters (about 640 million cubic meters per annum). The City of Montreal transports water through their distribution system using nearly 5000 km of watermains (Tiwari and Adamowski, 2013; Ville de Montreal, 2014).

The City of Montreal made an agreement through the Montreal Community Sustainable Development Plan to reduce water production by 15% (from 2000 to 2015) through the following initiatives: 1) reducing water loss and illegal uses; 2) indexing water consumption by sector of activity; and 3) improving efficient use of potable water (Ville de Montreal, 2015). Since urban water demand forecasting (UWD) forecasting is very useful to apply for program tracking, the City of Montreal may use the proposed Wavelet Data-Driven Forecasting Framework (WDDFF) for UWD forecasting to track each of the above mentioned sustainability initiatives through time. This would provide water managers and/or city councillors access to UWD projections for a particular operational, planning and construction, or sustainability initiative that could allow more proactive measures to take place, such as accelerating certain system maintenance projects (e.g., to address water loss), planning temporary servicing requirements for large water main construction, repairs, or new site development, and/or implementing meaningful water use restrictions by sector of activity during times of need (e.g., severe droughts).

Montreal has a humid continental climate. The average yearly air temperature is 5.3 °C with the daily average air temperature peaking in July at 19.8 °C and at its lowest in January at -11.5 °C. Annually, on average, 166 days drop below freezing while 111 days exceed 20 °C. Average yearly rainfall is 836 mm with only 50 days exceeding 5 mm. Montreal's primary economy consists of manufacturing industries, retail and health sectors, professional, scientific, and technological services (Environment Canada, 2014; Learn Quebec, 2014).

We collected average daily UWD (*U*) from the City of Montreal and daily meteorological variables from Environment Canada within the period February 27, 1999 to August 6, 2010 (4179 daily records). Average daily UWD represents the dependent (target/response) variable which we forecasted at daily (1, 3, 5, 7, and 14) lead times. Based on previous UWD forecasting studies (Adamowski et al., 2012; Akuoko-Asibey

et al., 1993; Gato et al., 2007; Wong et al., 2010) and the data available to us through Environment Canada, the following explanatory (input) variables were considered as the independent variables used for forecasting UWD: historical average daily UWD, U, maximum daily air temperature (T), total daily rainfall (R), and a daily antecedent precipitation index (A, an indicator of soil moisture, an important variable to consider for outdoor water demand (Quilty et al., 2016)) with decay constant 0.95 (determined empirically via correlation analysis with UWD – typical values range between 0.85 – 0.99). antecedent precipitation index, A, was calculated from the daily rainfall time series using the relation (Fedora and Beschta, 1989):

$$A_t = k * A_{t-1} + R_t \tag{4.B.1}$$

where k is the decay constant.

While U, R, and A did not contain any missing observations, T had six missing records (missing at random) that were imputed using linear regression.

Descriptive statistics for the target and input variables are provided in Table 4.B.1 while the time series are plotted in Figure 4.B.1. Pearson correlation and information correlation (i.e., mutual information scaled between 0 and 1 (Sharma and Mehrotra, 2014), estimated using Edgeworth Approximations of differential entropy (Van Hulle, 2005) – the base method in our Edgeworth Approximations-based conditional mutual information input variable selection method, recently introduced in Quilty et al. (2016)) were calculated between each pair of variables and listed in Table 4.B.2.

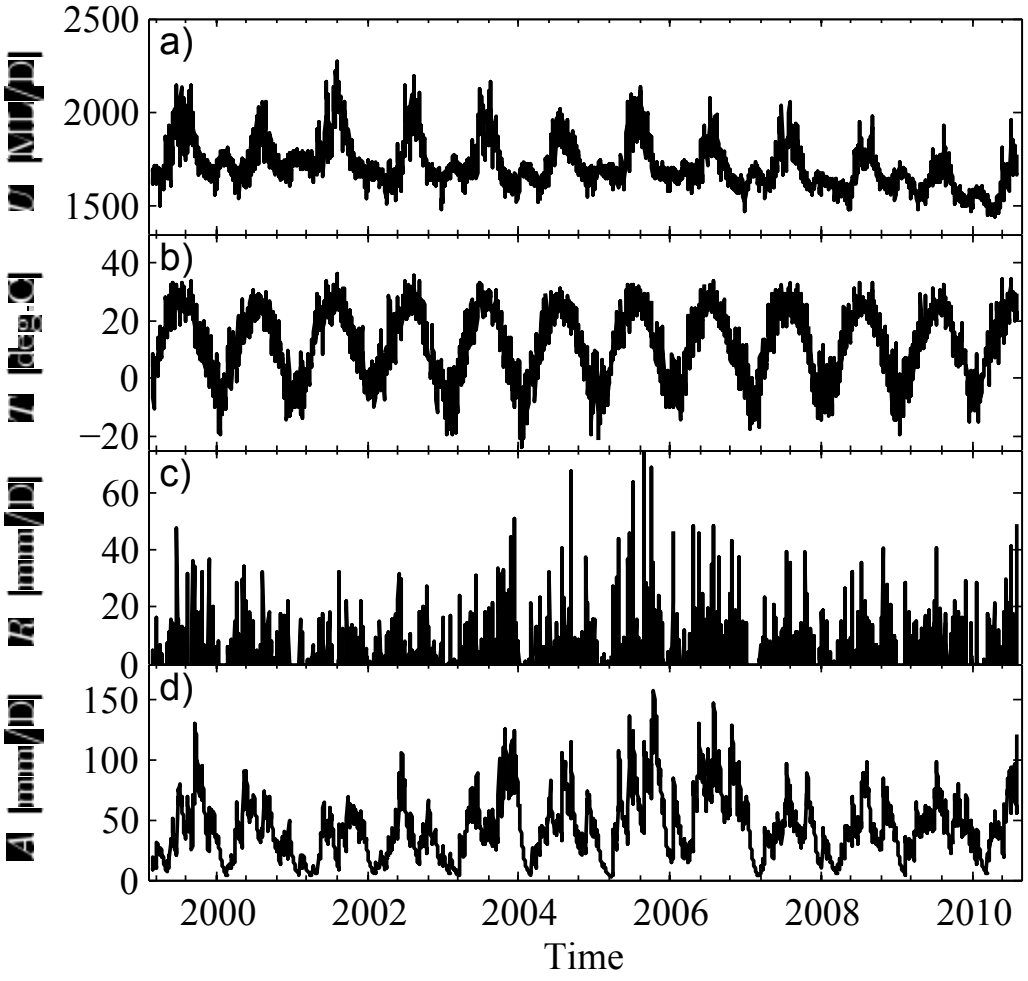

Figure 4.B.1. Time series plots for a) average daily urban water demand time series measured in megalitres per day (ML/D); b) daily maximum air temperature measured in degrees Celsius (deg. C); c) daily total rainfall measured in millimetres per day (mm/D); and d) antecedent precipitation index measured in millimetres per day (mm/D)

Table 4.B.1. Descriptive statistics for the time series in our experiment

Time Series	Units of Measure	Number of Records	Mean	Variance	Minimum	Maximum	Median
U	ML/D	4179	1728.275	15939.821	1446.212	2275.202	1705.833
T	°C		12.148	150.151	-23.900	36.200	13.100
R	mm/D		2.206	37.060	0.000	73.800	0.000
A	mm/D		43.613	704.581	2.177	158.297	40.103

Table 4.B.2. Pearson and information correlation for the time series in our experiment

Pearson Correlation				
Time Series	*U*	*T*	*R*	*A*
U	1.000	0.580	-0.009	0.121
T	0.580	1.000	0.122	0.385
R	-0.009	0.122	1.000	0.106
A	0.121	0.385	0.106	1.000
Information Correlation				
Time Series	*U*	*T*	*R*	*A*
U	1.000	0.692	0.090	0.124
T	0.692	1.000	0.285	0.403
R	0.090	0.285	1.000	0.269
A	0.124	0.403	0.269	1.000

4. C.1. Experimental Setup for the Wavelet Data-Driven Forecasting Framework

4. C.1.1. Target, Forecast Lead Times, and Explanatory Variables

The target time series, average UWD, U, was forecasted for lead times 1, 3, 5, 7, and 14 days ahead by using explanatory variables: previous observations of U, R, A, and, T; where each explanatory variable was time lagged up to 14 days. We chose this value as the UWD record displayed a strong weekly structure and we wanted to provide the developed models with enough temporal structure from which to identify suitable predictors while ensuring that we did not include too many variables that may potentially slow down the input variable selection process while not necessarily increasing the model performance (Galelli et al., 2014). We found that by increasing the lag time past this point, the model performance did not significantly change.

4. C.1.2. Wavelet Decomposition: Decomposition Level and Wavelet Filter Selection and Wavelet-based Forecasting Approaches

To obtain the "boundary-corrected" wavelet and scaling coefficients, i.e., those that avoid adding uncertainty into the wavelet and scaling coefficients due to the 'future data' problem, we adopted both the maximal overlap discrete wavelet transform (MODWT) and *à trous* (AT) algorithms for wavelet decomposition. Since both methods generate different sets of wavelet coefficients (see section 4.3.1. and 4.3.3.1.1.), it is important to understand which method is more useful in providing the best forecast accuracy for a given dataset.

Decomposition levels that allowed us to explore periodicities up to the inter- and intra-annual scale ($J \leq 9$ since we study a daily time series) were selected (i.e., $J_{max} = 9$, see section 4.3.3.3.1.) – see Table 4.A.1. We considered wavelet filters (see Table 4.A.2) of reasonable length ($L \leq 14$) since this filter length covered a two week period for the daily time series that we considered, which was useful since we selected explanatory variables with time lags of 14 days (time lags past this point did not improve our models significantly). The wavelet families we considered included: Haar (haar), Daubechies (db2, db3, db4, db5, db6, db7), Symlets (sym4, sym6,), Coiflets (coif1, coif2), Fejer-Korovkin (fk4, fk6, fk8, fk14), Least-Asymmetric (la8, la12, la14), and Best-Localized (bl14) (Crowley, 2007; Nielsen, 2001; Olhede and Walden, 2004; Percival and Walden, 2000; Rathinasamy et al., 2013; Zhang et al., 2016). A total of 20 wavelet filters were considered in our wavelet-based forecasts, a much larger number of wavelet filters than are commonly used in wavelet-based forecasting studies (Adamowski and Chan, 2011; Aussem et al., 1998; Barzegar et al., 2017; Belayneh et al., 2016; Murtagh et al., 2004; Nourani et al., 2015; Rathinasamy et al., 2013). The reader can find the different scaling filters for these wavelets in Table 4.A.2.

We used both direct and multicomponent wavelet-based forecasting approaches - each method (1-6 - Single, Within, Across, Single-hybrid, Within-hybrid, or Across-hybrid – see section 4.4.2.4.) were adopted in order to determine which forecasting approach and method performed best for our given dataset.

4. C.1.3. Dataset Partitioning

We selected our calibration and validation periods as follows: February 27, 1999 to December 31, 2007 for calibration (3230 records) and January 1, 2008 to August 6, 2010 for validation (949 records) – we found this data partitioning to provide good performance as it afforded us with enough records for calibration and a reasonable amount of data (over two and a half years) to validate forecast performance out-of-sample. Since the MODWT and AT avoid using 'future data' during wavelet decomposition this

allowed us to decompose the calibration and validation data in bulk (i.e., all calibration and validation records are decomposed at once) in order to obtain the "boundary-corrected" wavelet and scaling coefficients. Studies using the discrete wavelet transform multiresolution analysis (DWT-MRA) or the MODWT multiresolution analysis (MODWT-MRA) generally decompose both datasets separately by citing that this avoids incorporating 'future data' from the validation set in the calibration set (and vice versa) (e.g., (Barzegar et al., 2017; Deo et al., 2017)); however, as discussed in section 4.3.3.3., both the DWT-MRA and MODWT-MRA require 'future data' to be available at the decomposition stage for any given time t and that by decomposing both the calibration and validation records separately, this does not actually remedy this issue and instead incurs boundary artefacts in both the calibration and validation partitions (see section 4.3.3.3) which was verified mathematically in Table 4.2 (in text). In these cases where the DWT-MRA and MODWT-MRA have been used separately to decompose the calibration and validation records, it would have been more "appropriate" to first decompose the calibration records, then decompose the calibration and validation records together and "stitch" together the calibration records decomposed in the first step to the validation records decomposed in the second step (as this lessens the effect of the boundary condition on the validation records). In other words, the decomposed validation records (from the second step) should be placed ahead of the decomposed calibration records (from the first step) (see section 4.6.3.). Regardless, we do not recommend that one uses the DWT-MRA or MODWT-MRA and instead recommend the use of the MODWT and AT algorithms. In cases where one has access to a historical record and is calibrating then validating a wavelet-based forecast (as is the case in the literature) using the MODWT or AT, prior to adopting the model for real-world use, one does not have to decompose calibration and validation records separately since these methods do not incorporate 'future data'. However, if the model is to be subsequently used for forecasting when new data is received, then the newly received data needs to be decomposed after it is appended to the historical (combined calibration and validation) record.

It is important to note that although we originally explored decomposition levels from one to nine, we found that our forecasts were essentially identical at decomposition levels six and above. Thus, we reduced the maximum decomposition level to six and re-ran the models. Performance did not significantly change and the models were quicker to run since the wavelet decomposition stage had a lower number of decomposition levels to explore and there were less variables to consider during input variable selection. This is why we suggested (in section 4.3.3.3.1.) that a preliminary exploration of different decomposition levels and wavelet filters in conjunction with dataset partitioning should be

performed to identify reasonable decomposition levels and wavelet filters that suit the modellers preference in terms of model accuracy versus computation efficiency.

By setting the largest wavelet filter length at 14 filter coefficients (d7, sym7, fk14, la14, and bl14) and a maximum decomposition level of six, we had $L_j = (2^j - 1)(L - 1) + 1 = 820$ boundary-effected coefficients. We removed the first 820 records from the beginning of our input-output datasets (creating the "boundary-corrected" wavelet and scaling coefficients) prior to selecting input variables and before model calibration (eliminating the effect of the boundary condition on our forecasts (see section 4.3.3.3.1.)). This is a large number of records to remove (more than two years of daily observations), but it is a reasonable sacrifice to make if one does not want to introduce uncertainty into their wavelet-based forecasts due to the boundary condition - in our case, we found that the performance of the models was still reasonably high after removing these records. To ensure that each model used the same time datum for calibrating (and validating the forecasts) the same calibration and validation records were used for each decomposition level and wavelet filter combination (although, lower decomposition levels and narrower wavelet filters technically have less boundary-effected coefficients) - this was done to permit a fair comparison when evaluating the different wavelet-based forecasts that used different decomposition levels and wavelet filters.

Since we included explanatory variables with time lags up to 14 days, explored 1, 3, 5, 7, and 14 day ahead forecasts, and removed the first 820 boundary coefficients from each dataset there was a total of: 2395, 2393, 2391, 2389, and 2382 calibration records for the 1, 3, 5, 7, and 14 day ahead forecasts, respectively. This left us with 949 validation records for evaluating each lead time forecast out-of-sample (i.e., each lead time forecast had the exact same validation set for the target process). Each input was scaled such that it fell within the range of 0 to 1 for the calibration data, which required the calculation of the minimum and maximum values for each input variable (in a particular calibration dataset). The minimum and maximum values were then used to scale the validation dataset, a common practice for data-driven models that could suffer performance issues if inputs are on widely different scales, since certain model parameters may only focus on certain variables that contain much of the dataset variability within the original input variables (when compared to other inputs included in the set of explanatory variables) (Hastie et al., 2009).

4. C.1.4. Input Variable Selection, Forecast Calibration, and Assessment

As part of the WDDFF, we adopted the Edgeworth Approximation-based conditional mutual information (EA) and partial correlation input selection (PCIS) input variable selection methods (see section 4.4.2.2.)

to select inputs (i.e., "boundary-corrected" wavelet and scaling coefficients) for the different forecast models: multiple linear regression (MLR), extreme learning machine (ELM), and second order Volterra (SOV) (see section 4.5.1.). The PCIS, a linear method, was paired with MLR, and EA was paired with the nonlinear ELM and SOV methods.

We compared wavelet-based nonlinear models (EA paired with ELM and SOV) with wavelet-based linear models (PCIS paired with MLR) and a random walk (RW) benchmark. Standard models were also considered as comparison tools, i.e., models without wavelets were also considered. Models that used wavelets are appended with a 'W', i.e., WELM is a wavelet-based extreme learning machine, while WMLR is a wavelet-based multiple linear regression model. The following models were considered (input variable selection method-model): EA-WELM, EA-WSOV, PCIS-WMLR, EA-ELM, EA-SOV, PCIS-MLR, and RW. Each model based on wavelet-decomposed inputs was developed for each of the different wavelet-based forecasting approaches (direct and multicomponent) and method (1-6 - Single, Within, Across, Single-hybrid, Within-hybrid, or Across-hybrid – see section 4.4.2.1.1.) by considering each decomposition level (six in total) and wavelet filter (20 in total).

Since MLR, ELM, and SOV are traditionally formulated as linear least-squares problems, their parameters were calibrated such that they minimized the mean square error. The commonly adopted Nash-Sutcliffe Efficiency Index (NASH) (Moriasi et al., 2007) was used to select the best performing models by assessing performance on the validation set. We also adopted two criteria that we have used in other urban water demand forecasting studies (e.g., the root mean square error (RMSE) and mean absolute percentage error (MAPE) (Adamowski et al., 2012; Quilty et al., 2016; Tiwari and Adamowski, 2017)) to evaluate the different forecasts.

4. C.1.5. Summary of the Different Experiments

Since we explored: six decomposition levels, 20 different wavelet filters, eight different wavelet-based forecasts (Methods 1-6 - Single, Within, Across, Single-hybrid, Within-hybrid, or Across-hybrid (remember that the AT can be used with each method while the MODWT can only be used with methods 1 and 4, see section 4.3.3.1.1.)) combined with three different regression models (MLR, ELM, and SOV), and explored five forecast lead times, a total of 14, 400 different wavelet-based forecasts were developed and compared as part of our experiments. Additionally, we studied four different non-wavelet-based models (MLR, ELM, SOV, and a random walk model) for the different lead times. Therefore, 14, 420 different models were explored in this work.

4. D.1. Experiment Results

Table 4.D.1. 3 day lead time forecast results

3 Day Lead-Time Forecasts						
Strategy	IVS-Model	Level	Filter	RMSE (ML/D)	MAPE (%)	NASH
	RW			57.981	2.642	0.595
	PCIS-MLR			58.895	2.787	0.582
	EA-ELM			45.880	1.864	0.746
	EA-SOV			46.237	1.874	0.742
AT - single	PCIS-WMLR	3	*d5*	41.664	1.746	0.791
	EA-WELM	1	*la12*	43.246	1.799	0.774
	EA-WSOV	4	*d2*	46.178	1.836	0.743
MODWT - single	PCIS-WMLR	3	*fk6*	42.250	1.810	0.785
	EA-WELM	1	*fk6*	46.136	1.897	0.743
	EA-WSOV	1	*fk4*	46.946	1.909	0.734
AT - within	PCIS-WMLR	5	*fk4*	44.985	1.814	0.756
	EA-WELM	5	*fk4*	43.985	1.776	0.767
	EA-WSOV	6	*fk4*	44.680	1.792	0.759
AT - across	PCIS-WMLR	4	*d6*	41.156	1.743	0.796
	EA-WELM	5	*d3*	43.547	1.769	0.771
	EA-WSOV	4	*fk4*	44.694	1.863	0.759
AT - single - hybrid	PCIS-WMLR	2	*la10*	41.666	1.734	0.791
	EA-WELM	2	*d6*	44.141	1.845	0.765
	EA-WSOV	2	*d6*	44.246	1.880	0.764
MODWT - single - hybrid	**PCIS-WMLR**	**6**	***la14***	**39.048**	**1.611**	**0.816**
	EA-WELM	2	*d6*	44.141	1.845	0.765
	EA-WSOV	2	*d6*	44.246	1.880	0.764
AT - within - hybrid	PCIS-WMLR	1	*la8*	46.687	2.082	0.737
	EA-WELM	6	*fk4*	44.032	1.820	0.766
	EA-WSOV	2	*d1*	44.878	1.862	0.757
AT - across - hybrid	PCIS-WMLR	4	*d6*	41.276	1.747	0.795
	EA-WELM	4	*d2*	44.064	1.846	0.766
	EA-WSOV	4	*fk4*	44.364	1.836	0.763

Table 4.D.2. 5 day lead time forecast results

5 Day Lead-Time Forecasts						
Strategy	IVS-Model	Level	Filter	RMSE (ML/D)	MAPE (%)	NASH
	RW			61.096	2.676	0.550
	PCIS-MLR			62.023	2.908	0.536
	EA-WELM			49.785	2.039	0.701
	EA-WSOV			49.842	2.035	0.700
AT - single	PCIS-WMLR	3	*fk6*	47.912	2.005	0.723
	EA-WELM	4	*d6*	49.684	2.109	0.702
	EA-WSOV	1	*d6*	50.534	2.059	0.692
MODWT - single	PCIS-WMLR	2	*d2*	48.759	2.138	0.713
	EA-WELM	1	*d6*	50.019	2.079	0.698
	EA-WSOV	1	*d6*	50.534	2.059	0.692
AT - within	PCIS-WMLR	4	*d1*	49.837	2.011	0.701
	EA-WELM	4	*d1*	49.391	1.980	0.706
	EA-WSOV	3	*fk6*	49.212	1.937	0.708
AT - across	PCIS-WMLR	6	*coif1*	46.743	1.903	0.737
	EA-WELM	6	*d1*	47.604	1.911	0.727
	EA-WSOV	6	*d1*	48.630	2.003	0.715
AT - single - hybrid	PCIS-WMLR	3	*coif2*	49.025	2.058	0.710
	EA-WELM	3	*d2*	49.583	2.040	0.704
	EA-WSOV	3	*d2*	49.794	2.045	0.701
MODWT - single - hybrid	**PCIS-WMLR**	**6**	***la14***	**44.160**	**1.832**	**0.765**
	EA-WELM	1	*bl14*	49.779	2.039	0.701
	EA-WSOV	1	*bl14*	49.842	2.035	0.700
AT - within - hybrid	PCIS-WMLR	1	*sym4*	49.702	2.080	0.702
	EA-WELM	6	*d1*	48.929	2.115	0.711
	EA-WSOV	1	*fk4*	50.010	2.028	0.698
AT - across - hybrid	PCIS-WMLR	5	*d1*	47.202	1.964	0.731
	EA-WELM	6	*d1*	48.860	2.022	0.712
	EA-WSOV	1	*d7*	49.125	1.996	0.709

Table 4.D.3. 7 day lead time forecast results

7 Day Lead-Time Forecasts						
Strategy	IVS-Model	Level	Filter	RMSE (ML/D)	MAPE (%)	NASH
	RW			52.804	2.056	0.664
	PCIS-MLR			52.415	2.324	0.669
	EA-ELM			51.052	2.185	0.686
	EA-SOV			50.554	2.069	0.692
AT - single	PCIS-WMLR	1	*la14*	50.754	2.182	0.689
	EA-WELM	2	*bl14*	50.600	2.152	0.691
	EA-WSOV	1	*d1*	52.070	2.105	0.673
MODWT - single	PCIS-WMLR	3	*d1*	51.421	2.295	0.681
	EA-WELM	1	*d1*	51.863	2.242	0.676
	EA-WSOV	1	*d1*	52.070	2.105	0.673
AT - within	PCIS-WMLR	1	*d1*	52.416	2.300	0.669
	EA-WELM	4	*d1*	50.137	2.076	0.697
	EA-WSOV	6	*d1*	50.654	2.059	0.691
AT - across	PCIS-WMLR	6	*fk4*	49.913	2.171	0.700
	EA-WELM	6	*d1*	48.622	1.964	0.715
	EA-WSOV	1	*d7*	50.891	2.078	0.688
AT - single - hybrid	PCIS-WMLR	3	*fk8*	51.317	2.302	0.683
	EA-WELM	1	*sym6*	50.762	2.141	0.689
	EA-WSOV	2	*bl14*	50.554	2.069	0.692
MODWT - single - hybrid	**PCIS-WMLR**	**6**	***la14***	**46.840**	**1.974**	**0.735**
	EA-WELM	2	*bl14*	51.061	2.183	0.686
	EA-WSOV	2	*bl14*	50.554	2.069	0.692
AT - within - hybrid	PCIS-WMLR	1	*d1*	54.479	2.424	0.642
	EA-WELM	5	*d2*	50.045	2.112	0.698
	EA-WSOV	2	*d1*	50.877	2.100	0.688
AT - across - hybrid	PCIS-WMLR	4	*d1*	50.303	2.179	0.695
	EA-WELM	6	*d2*	49.993	2.119	0.699
	EA-WSOV	1	*d7*	51.032	2.098	0.686

Table 4.D.4. 14 day lead time forecast results

14 Day Lead-Time Forecasts						
Strategy	IVS-Model	Level	Filter	RMSE (ML/D)	MAPE (%)	NASH
	RW			64.914	2.628	0.492
	PCIS-MLR			64.205	3.037	0.503
	EA-ELM			62.759	2.959	0.525
	EA-SOV			61.218	2.680	0.548
AT - single	PCIS-WMLR	1	*d1*	64.751	3.081	0.494
	EA-WELM	2	*bl14*	61.944	2.877	0.537
	EA-WSOV	1	*fk4*	62.151	2.701	0.534
MODWT - single	PCIS-WMLR	1	*fk14*	63.691	3.017	0.511
	EA-WELM	1	*fk4*	63.017	2.979	0.521
	EA-WSOV	1	*fk4*	62.151	2.701	0.534
AT - within	PCIS-WMLR	3	*d1*	61.868	2.863	0.539
	EA-WELM	5	*fk6*	59.508	2.591	0.573
	EA-WSOV	5	*fk6*	60.406	2.531	0.560
AT - across	PCIS-WMLR	6	*fk8*	59.032	2.736	0.580
	EA-WELM	5	*coif1*	59.401	2.511	0.575
	EA-WSOV	1	*d2*	61.255	2.647	0.548
AT - single - hybrid	PCIS-WMLR	1	*d1*	64.205	3.037	0.503
	EA-WELM	6	*bl14*	52.396	2.283	0.669
	EA-WSOV	6	*bl14*	51.967	2.114	0.674
MODWT - single - hybrid	**PCIS-WMLR**	**6**	***bl14***	**51.048**	**2.234**	**0.686**
	EA-WELM	6	*bl14*	52.396	2.283	0.669
	EA-WSOV	6	*bl14*	51.967	2.114	0.674
AT - within - hybrid	PCIS-WMLR	1	*d1*	65.772	3.123	0.478
	EA-WELM	1	*sym6*	62.114	2.796	0.535
	EA-WSOV	2	*d1*	61.673	2.779	0.541
AT - across - hybrid	PCIS-WMLR	6	*coif1*	58.882	2.621	0.582
	EA-WELM	1	*sym6*	61.454	2.757	0.545
	EA-WSOV	3	*d2*	62.071	2.819	0.535

References

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407, 28–40. doi:10.1016/j.jhydrol.2011.06.013

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of

multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Akuoko-Asibey, A., Nkemdirim, L.C., Draper, D.L., 1993. the Impacts of Climatic Variables on Seasonal Water Consumption in Calgary, Alberta. Can. Water Resour. J. 18, 107–116. doi:10.4296/cwrj1802107

Aussem, A., Campbell, J., Murtagh, F., 1998. Wavelet-based feature extraction and decomposition strategies for financial forecasting. J. Comput. Intell. Financ. 6, 5–12.

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189

Bašta, M., 2014. Additive decomposition and boundary conditions in wavelet-based forecasting approaches. Acta Oeconomica Pragensia 2014, 48–70.

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172–173. doi:10.1016/j.atmosres.2015.12.017

Crowley, P.M., 2007. A guide to wavelets for economists. J. Econ. Surv. doi:10.1111/j.1467-6419.2006.00502.x

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2017. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stoch. Environ. Res. Risk Assess. 31, 1211–1240. doi:10.1007/s00477-016-1265-z

Environment Canada, 2014. Canadian Climate Normals 1981-2010 Station Data [Webpage]. URL: http://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=5616andlang=eanddCode=1andStationName=MONTREALandSearchType=Containsandprovince=ALLandprovBut=andmonth1=0andmonth2=12 (accessed 3.29.14).

Fedora, M.A., Beschta, R.L., 1989. Storm runoff simulation using an antecedent precipitation index (API) model. J. Hydrol. 112, 121–133. doi:10.1016/0022-1694(89)90184-4

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation

framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. doi:10.1016/j.envsoft.2014.08.015

Gato, S., Jayasuriya, N., Roberts, P., 2007. Temperature and rainfall thresholds for base use urban water demand modelling. J. Hydrol. 337, 364–376. doi:10.1016/j.jhydrol.2007.02.014

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed, Springer Series in Statistics. Springer New York, New York.

Kisi, O., Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J. Hydrol. 399, 132–140. doi:10.1016/j.jhydrol.2010.12.041

Learn Quebec, 2014. Metropolises - Montreal [Webpage]. URL: http://www.learnquebec.ca/en/content/curriculum/social_sciences/sec_geography/resources/metropolises/montreal.html (accessed 3.29.14).

Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. doi:10.13031/2013.23153

Murtagh, F., Starck, J.L., Renuad, O., 2004. On neuro-wavelet modeling. Decis. Support Syst. doi:10.1016/S0167-9236(03)00092-7

Nielsen, M., 2001. On the Construction and Frequency Localization of Finite Orthogonal Quadrature Filters. J. Approx. Theory 108, 36–52.

Nourani, V., Alami, M.T., Vousoughi, F.D., 2015. Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. J. Hydrol. 524, 255–269. doi: 10.1016/j.jhydrol.2015.02.048

Olhede, S., Walden, A.T., 2004. The Hilbert spectrum via wavelet projections. Proc. R. Soc. A Math. Phys. Eng. Sci. 460, 955–975. doi:10.1098/rspa.2003.1199

Percival, D.B., Mofjeld, H.O., 1997. Analysis of Subtidal Coastal Sea Level Fluctuations Using Wavelets. J. Am. Stat. Assoc. 92, 868. doi:10.2307/2965551

Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis, 1st ed, Cambridge Series in

Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. Water Resour. Res. 52, 2299–2326. doi:10.1002/2015WR016959

Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J. Hydrol. doi:10.1016/j.jhydrol.2013.09.025

Sharma, A., Mehrotra, R., 2014. An information theoretic alternative to model a natural system using observational information alone. Water Resour. Res. 50, 650–660. doi:10.1002/2013WR013845

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Tiwari, M.K., Adamowski, J.F., 2017. An ensemble wavelet bootstrap machine learning approach to water demand forecasting: a case study in the city of Calgary, Canada. Urban Water J. 14, 185–201. doi:10.1080/1573062X.2015.1084011

Van Hulle, M.M., 2005. Edgeworth Approximation of Multivariate Differential Entropy. Neural Comput. 17, 1903–1910. doi:10.1162/0899766054323026

Ville de Montreal, 2015. Montreal Community Sustainable Development Plan [Webpage]. URL: http://ville.montreal.qc.ca/portal/page?_pageid=7137,79233615and_dad=portaland_schema=PORTAL (accessed 3.14.15).

Ville de Montreal, 2014. Gestion durable de l'eau [Webpage]. URL: http://ville.montreal.qc.ca/portal/page?_pageid=6497,81365599and_dad=portaland_schema=PORTAL (accessed 3.29.14).

Walden, A.T., 2001. Wavelet Analysis of Discrete Time Series, in: Casacuberta, C., Miró-Roig, R.M., Verdera, J., Xambó-Descamps, S. (Eds.), European Congress of Mathematics: Barcelona, July 10--14, 2000 Volume II. Birkhäuser Basel, Basel, pp. 627–641. doi:10.1007/978-3-0348-8266-8_56

Wong, J.S., Zhang, Q., Chen, Y.D., 2010. Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. Water Resour. Res. 46. doi:10.1029/2009WR008147

Zhang, Z., Telesford, Q.K., Giusti, C., Lim, K.O., Bassett, D.S., 2016. Choosing wavelet methods, filters, and lengths for functional brain network construction. PLoS One. doi:10.1371/journal.pone.0157243

Connecting Text to Chapter 5

The Wavelet Data-Driven Forecasting Framework (WDDFF) developed in Chapter 4, which adopted the new computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods from Chapter 3, is extended in this chapter by including uncertainty assessment through a stochastic approach, resulting in the Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF). SWDDFF can be used for generating probability-based forecasts of water resources and is a very useful tool for accounting for the nonlinear, multiscale, and uncertain nature of water resources. SWDDFF can be used for decision-making tasks commonly encountered by water resources managers in the operation, planning, and management of water resources systems.

This chapter has been submitted to Journal of Hydrology. The format of the paper has been modified to ensure consistency with the style of this thesis. A list of references cited in this paper are available at the end of the chapter.

The author of the thesis was responsible for the development, testing, and application of the different methods and wrote the manuscript presented here. Prof. Adamowski, the supervisor of this thesis, provided valuable advice on all aspects of the research and contributed to the review and editing of the manuscript.

Chapter 5: A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes

5.1. Abstract

A recent blueprint was proposed for converting a deterministic process-based model into a stochastic one that accounts for input data, parameter, and model output uncertainty - yielding predictions of a target process in the form of a probability distribution. We extend this blueprint in three key directions, by: 1) focussing on data-driven (i.e., regression, machine learning, artificial intelligence, etc.) forecasting models (in contrast to process-based models), due to their flexibility and widespread use in hydrology and water resources; 2) modifying the original blueprint to include input variable selection uncertainty (in addition to input data, parameter and model output uncertainty) - a recent topic that has shown promise for improving forecasting performance; and 3) incorporating wavelet transformation of model inputs - a method that has been used in numerous studies to improve the ability of data-driven forecasting models to account for the multiscale nature of hydrological and water resources processes. The first two developments lead to a new forecasting framework named Stochastic Data-Driven Forecasting Framework (SDDFF). The third development is an extension of the SDDFF and our earlier work on wavelet-based data-driven forecasting which results in a second new framework, the Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF). Through a real-world urban water demand forecasting experiment in Montreal, Canada, we compare SDDFF and SWDDFF against benchmarks and demonstrate that both input variable selection uncertainty and wavelet transformation play key roles in improving forecasting accuracy and reliability according to deterministic and probabilistic forecast evaluation metrics commonly adopted in water resources. Many earlier wavelet-based (and non-wavelet-based) data-driven forecasting models are also shown to be special cases of the SDDFF and SWDDFF, highlighting the generality of the proposed frameworks.

Keywords: uncertainty, stochastic, data-driven forecasting, input variable selection, wavelets

5.2. Introduction

Accounting for the uncertainty in hydrological and water resources forecasts is recognized as a crucial task in the management, planning, and operation of hydrological and water resources systems (Krzysztofowicz, 2001). The literature abounds with studies seeking to estimate uncertainties related to input data,

parameters, model structure, model output, initial conditions, etc. and their impact on the resulting forecasts (see Gong et al. (2013); Beven (2015); and Nearing et al. (2016) and references therein for many different examples). Often the goal is to incorporate (one or many of) these different sources of uncertainty into a reliable forecasting framework that issues forecast probabilities (instead of single-valued deterministic forecasts) which can be used for decision-making purposes (Ramos et al., 2013). Therefore, much effort is spent on refining and innovating various probability-based forecasting methods to serve this end (Fortin et al., 2006; Pappenberger et al., 2015; Raftery, 2016; Sivillo et al., 1997; Thiboult et al., 2017; Wani et al., 2017). Our purpose is not to delve into the vast literature on probability-based hydrological and water resources forecasting approaches but rather to extend a recent stochastic modeling framework that is flexible and allows us to extend our earlier research on data-driven hydrological and water resources forecasting. In this work, the term data-driven is used to represent models based on regression, machine learning, artificial intelligence, etc. (Solomatine and Ostfeld, 2008).

Recently, Montanari and Koutsoyiannis (2012) introduced a general and simple 'blueprint' for converting a deterministic model into a stochastic one (see section 5.3.1.). The blueprint's generality lies in that it can be equally applied to any deterministic model, process-based (the focus of the original authors), data-driven (our focus), or anything in between, and result in a target forecast in the form of a probability distribution. The blueprint's simplicity stems from its incorporation of different uncertainty sources through use of stochastic perturbations to input data, parameters, and model error (among other sources), whose probability distributions may be defined with simple tools such as the bootstrap (Srivastav et al., 2007). The original authors focussed on process-based hydrological models (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) that incorporated input data, parameter, and model output uncertainty. Using these attractive features of the blueprint as our starting point, our objectives for extending the blueprint are three-fold:

1. We focus on data-driven models (i.e., multiple linear regression, neural networks, etc. (Solomatine and Ostfeld, 2008)) instead of process-based models – the blueprint authors mentioned this as a possible extension of their method (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015).

2. The blueprint is modified to include input variable selection uncertainty (this is in addition to input data, parameter, and model output uncertainty previously defined within the blueprint) (Quilty et al., 2016).

3. Wavelet transformation of the input data is adopted to account for the multiscale nature of hydrological and water resources processes (Kumar and Foufoula-Georgiou, 1997; Labat, 2005; Maheswaran and Khosa, 2012; Nourani et al., 2014) and improve forecasting performance.

Our justification for extending the blueprint in these three directions is as follows:

1. We chose to focus on data-driven models as they are more flexible (convenient) than process-based models and since they may be used, and have shown success, in modeling and forecasting complex problems where there is a lack of *apriori* knowledge, many potential input variables, and relatively few samples from which to build a model (Abrahart et al., 2012; Abrahart and See, 2007; Bowden et al., 2012, 2002; Dawson and Wilby, 2001).

2. In the case of adopting a data-driven forecasting framework and given that the necessary condition of data availability exists (i.e., that useful model inputs are available), the selection of model inputs is of great importance in guiding the development of accurate and useful models for real-world applications (Creaco et al., 2016; Fernando et al., 2009; Galelli et al., 2014; Sharma and Mehrotra, 2014; Sun and Bertrand-Krajewski, 2013). Although a relatively new topic, recent work has shown that by taking into account input variable selection uncertainty in data-driven models, hydrological and water resources model performance can be improved and may also act as a useful tool to gain further insight into the nature of the studied processes (Quilty et al., 2016; Taormina et al., 2016).

3. By transforming the input data in to the wavelet domain, we are able to localize events (such as periodicities, transients, trends, etc.) in both time and frequency (scales), which has proven useful for increasing data-driven forecasting performance of hydrological and water resources processes (Afan et al., 2016; Dixit et al., 2016; Fahimi et al., 2017). The likely reason why many data-driven models perform better when wavelet-decomposed inputs are used instead of the original inputs is because the variance of the original input data is dispersed amongst the different wavelet sub-series allowing for model parameters to be more easily fit to these sources of variation (a similar idea can be said for principal component analysis, although both methods approach this problem differently). This often results in a useful portrait of the original data that depicts its source of variation, sometimes in physically meaningful ways, such as seasonal periodicities, transients due to system failures (e.g., water main breaks), or level shifts due to anthropogenic factors (e.g., dam construction). Furthermore, the topic of input variable selection using wavelet-transformed data

is a new topic (He et al., 2015; Prasad et al., 2017; Quilty and Adamowski, 2018; Tran et al., 2016) and so far no studies have combined input variable selection uncertainty with wavelet-based data-driven forecasting models.

The first and second objectives (i.e., adjusting the blueprint for use with data-driven models and including input variable selection uncertainty, amongst input data, parameter, and model output uncertainty) leads us to the first of our proposed models, named the Stochastic Data-Driven Forecasting Framework (SDDFF). By altering the appellation of the blueprint to the Stochastic Data-Driven Forecasting Framework our intention is not to lay claim to the interesting framework developed by the earlier authors by a change of name, nor to suggest that data-driven models are superior or should be given more attention than process-based models, but rather to extend the blueprint by more prominently focussing on the case of data-driven forecasting since it offers a flexible alternative to process-based models that is becoming increasingly popular in hydrology and water resources (Remesan and Mathew, 2015). Our second proposed model combines wavelet transformation and the SDDFF; we name this model Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF).

The SDDFF is an extension of our earlier work on data-driven forecasting of hydrological and water resources processes (Adamowski et al., 2012; Belayneh et al., 2016; Ciupak et al., 2015; Goyal et al., 2014; Yaseen et al., 2016) and input variable selection uncertainty (Quilty et al., 2016) since SDDFF advances our forecasting strategies from evaluating and using a single best input variable selection set in a deterministic forecast to incorporating input variable selection uncertainty (many different input variable sets) in a stochastic forecasting framework (SDDFF). The SWDDFF is a direct extension of the Wavelet Data-Driven Forecasting Framework (WDDFF) recently proposed by Quilty and Adamowski (2018) and permits us to not only account for the multiscale nature of processes commonly encountered in hydrology and water resources - that often leads to poor accuracy when using data-driven forecasting methods (Dixit et al., 2016; Nourani et al., 2014; Yaseen et al., 2015) – but also allows us to quantify the uncertainty associated with input data, input variable selection, parameters, and model output.

A significant step in building and using the SDDFF and SWDDFF is the characterization of the probability distributions associated with input data, input variable selection, parameters, and model output uncertainty (see section 5.3.3., 5.4, and 5.5). To estimate these probability distributions we adopted the bootstrap resampling method (Efron and Tibshirani, 1993) due to its simplicity and ubiquity in empirically estimating probability distributions (Henderson, 2005; Polikar, 2007) (see section 5.5.1.). Since we used the bootstrap method for estimating the different sources of uncertainty and employed wavelet

transformation for improving our ability to forecast the multiscale nature of hydrological and water resources time series, we make note of earlier examples in the literature that share some form of resemblance to SWDDFF below.

Wavelet transformation and the bootstrap method have been combined with data-driven models (e.g., MLR, neural networks, and extreme learning machines) to account for parameter uncertainty in forecasting models used for various hydrological and water resources processes such as drought, groundwater, streamflow, reservoir inflows, urban water demand, and river discharge (Belayneh et al., 2016; Kasiviswanathan et al., 2016; Khalil et al., 2015; Kumar et al., 2015; Sehgal et al., 2014; Tiwari and Adamowski, 2013, 2015, 2017, Tiwari and Chatterjee, 2010a, 2011). The bootstrap has also been used to account for both parameter and model output uncertainty in wavelet-neural network-based data-driven forecasts of monthly water quality (Wang et al., 2013). Using a Relevance Vector Machine (RVM) (a Bayesian approach), earlier studies were able to estimate both parameter and model output uncertainty for daily evaporation and long-term streamflow forecasting (Bachour et al., 2016; Maslova et al., 2016). The wavelet-based data-driven forecasts that have adopted the bootstrap for parameter and/or model output uncertainty will be shown to be special cases of SWDDFF. It is significant to note that we are not aware of a single study that has combined wavelet transformation with data-driven models that accounts for input variable selection, parameters, and model output uncertainty (using the bootstrap or any other method). Therefore, the SWDDFF can be seen to improve upon the most advanced wavelet-based data-driven forecasting models available in the literature.

The main goal of this research is to move from the original blueprint (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) to the development of the SDDFF and SWDDFF and then to show, via a real-world water resources forecasting problem, how both input variable selection and wavelet transformation are key factors in improving forecasting accuracy and reliability. We note that for this study, we chose to focus on (empirically) estimating the input variable selection, parameter, and model error probability distributions. We leave the estimation of the input data uncertainty to future work, primarily because we are interested in:

1. Introducing how input variable selection uncertainty can be included in the blueprint (particularly in a data-driven context), resulting in the SDDFF;

2. How the SDDFF can be extended to include wavelet transformation (SWDDFF), making it amenable to improving the forecasting performance for multiscale processes commonly encountered in hydrology and water resources; and

3. Demonstrating, through a real-world water resources case study, the usefulness of SDDFF and SWDDFF for accurately and reliably forecasting hydrological and water resources processes against benchmarks, such as the earlier wavelet-based data-driven forecasting models that incorporated (solely parameter or both parameter and model output) uncertainty using the bootstrap (as mentioned two paragraphs above).

Notwithstanding, our exclusion of input data uncertainty in our case study, we still incorporate input data uncertainty in our theoretical development of SDDFF and SWDDFF.

Before outlining the rest of our study, we summarize our main contributions to the literature as follows:

1. We extend the blueprint for converting deterministic process-based models to stochastic ones (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) by focussing on the domain of data-driven forecasting methods.

2. The blueprint is modified to include input variable selection uncertainty, a relatively new topic, in addition to the input data, parameter, and model output uncertainty (that was explicitly accounted for in the original blueprint). The former contribution combined with this contribution results in our first proposed model SDDFF (Stochastic Data-Driven Forecasting Framework).

3. Our second contribution (SDDFF) is modified to include wavelet transformation of the model inputs, resulting in the SWDDFF (Stochastic Wavelet Data-Driven Forecasting Framework), a stochastic wavelet-based forecasting framework accounting for uncertainty in input data, input variable selection, parameter, and model output. SWDDFF innovates upon the most advanced wavelet-based forecasting models available in the literature.

The rest of our study is outlined as follows: section 5.3 contains the theory behind the blueprint, SDDFF, and SWDDFF; section 5.4 gives instructions for using SDDFF and SWDDFF in practice; section 5.5 provides experimental details concerning our real-world case study; section 5.6 discusses our experiment results; and section 5.7 concludes our study with a summary of our work and future research directions.

5.3. Theory

5.3.1. A Brief Overview of the Blueprint (Montanari and Koutsoyiannis, 2012)

Since we are interested in adapting the blueprint by focussing on its use for data-driven forecasting, including input variable selection, and the incorporating wavelet transformation of model inputs, we only give the main result of the blueprint and refer the interested reader to a more detailed treatment by the original authors (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). In what follows we attempt to keep as close as possible to the assumptions and notations of the original authors.

The blueprint begins with the simple deterministic formulation (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015):

$$Q = S(\boldsymbol{\Theta}, \boldsymbol{X}) \tag{5.1}$$

where Q is the model output (which through the deterministic formulation in Eq. 5.1 is assumed to be equal to the true value of the target variable to be forecasted); S represents the model structure (which in this case takes on the form of a data-driven model (e.g., multiple linear regression) and transform the input data and model parameters in to the model output using an empirical formulation), $\boldsymbol{X}$ represents the input data, and $\boldsymbol{\Theta}$ are the model parameters.

Omitting several detailed steps (captured by equations 2–6 in Montanari and Koutsoyiannis (2012)), the deterministic framework in Eq. 5.1 is converted to its stochastic counterpart resulting in (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015):

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{X}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X}) | \boldsymbol{\Theta}, \boldsymbol{X}) f_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{\Theta}, \boldsymbol{X}) d\boldsymbol{\Theta} d\boldsymbol{X} \tag{5.2}$$

where:

$e = Q - S(\boldsymbol{\Theta}, \boldsymbol{X})$ is the model error which incorporates all uncertainties not explicitly accounted for in Eq. 5.2, such as input and parameter uncertainty – it is important to note the emphasis placed upon the fact that, in the stochastic framework of Eq. 5.2, the blueprint relies on the assumption that Q is the true variable to be forecasted, which is an unknown quantity at the time of issuing the forecast, and therefore is to be treated as a random variable, see also Montanari and Koutsoyiannis (2014);

$f_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{\Theta},\boldsymbol{X})$ is the joint probability density function of the parameters and input data (which jointly quantifies parameter and input data uncertainty);

$f_e(Q - S(\boldsymbol{\Theta},\boldsymbol{X})|\boldsymbol{\Theta},\boldsymbol{X})$ or $f_e(e|\boldsymbol{\Theta},\boldsymbol{X})$ is the conditional probability density function of the model error (e), conditioned on the input data and parameters (which, as mentioned above, quantifies model uncertainty or uncertainties not explicitly accounted for in the model); and

$f_Q(Q)$ is the probability density function of the true value of the target variable to be forecasted (which quantifies uncertainty in the forecast of the true target variable).

One minor difference between Eq. 5.2 and that given in the original blueprint (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015), is that we choose to consider a joint probability density function of the parameters and input data instead of considering that parameter uncertainty is independent of data uncertainty. We use this dependence (between input data and parameters) as a convenience for introducing input variable selection uncertainty within our proposed Stochastic Data-Driven Forecasting Framework in the next sub-section.

In keeping with the original blueprint (Montanari and Koutsoyiannis, 2012) a number of assumptions are also maintained here: uncertainty is considered as randomness inherent in the hydrological or water resources process under study; the input data ($\boldsymbol{X}$) contains uncertainties mostly related to observation methods (e.g., in a water distribution system the measurement of system demands could be effected by meter inaccuracies); parameters ($\boldsymbol{\Theta}$) contain uncertainties due to the model structure, calibration schemes, and dataset consistency; while the model error (e), which is equal to the difference between the forecast and the observed target value, contains all uncertainties not captured in the input data and parameter uncertainties and is a result of the model's inability to replicate the real process (referred to as model structural error). The prediction (forecast) uncertainty is the uncertainty in the forecast of the true value of the given target process and is dependent on the input data uncertainty, parameter uncertainty, and model error. Normally, prediction (forecast) uncertainty is defined using prediction limits for the forecast which defines a range that the true value of the target process is contained within with probability equal to the nominal confidence level. These same assumptions are kept in both proposed SDDFF and SWDDFF methods.

As noted by the blueprint authors, a very useful aspect of this framework is related to its ability to explicitly incorporate other sources of uncertainty (e.g., input variable selection) in to Eq. 5.2, given that we have some knowledge pertaining to the related probability distribution for the source of uncertainty. Later

(see section 5.5.1.), we discuss how one may infer such probability distributions (for input data, input variable selection, parameters and model error). Next, we demonstrate how input variable selection uncertainty (a relatively new topic of interest in hydrological and water resources modeling) can be included within the blueprint.

5.3.2. From the Blueprint to the Stochastic Data-Driven Forecasting Framework

By slightly modifying Eq. 5.2, we can explicitly account for the uncertainty due to input variable selection (in addition to input data, parameters, and model error) within the blueprint as follows:

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{X}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{\Omega}}(\boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) d\boldsymbol{\Theta} d\boldsymbol{X} \tag{5.3}$$

where $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ is a binary vector spanning $\boldsymbol{X}$, identifying the variables in $\boldsymbol{X}$ which have been selected (i.e., $\boldsymbol{\omega} \in \{0,1\}^{\boldsymbol{D}}$) using an input variable selection routine (e.g., simple correlation ranking or more involved schemes such as iterative input selection (Galelli and Castelletti, 2013) or partial mutual information selection (Li et al., 2015)). The joint and conditional distributions take on similar meanings as in Eq. 5.2; specifically, the joint distribution $f_{\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{\Omega}}(\boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega})$ represents the uncertainty in the parameters, input data, and the selected input variables while the conditional distribution $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega})$ relates the model error to the parameters, input data, and selected input variables. Note in Eq. 5.3 that we use a summation operator instead of an integral for the selected input variables term, $\boldsymbol{\omega}$, since this variable takes on binary (discrete) values – this, however, does not hinder our ability to estimate the mixed joint probability distribution (between discrete and continuous variables) (Coelho et al., 2016; Moon et al., 2017; Ross, 2014), which we will show can be accomplished through bootstrap resampling (see section 5.5.1.).

Since it is well-accepted that the performance of a given data-driven model greatly depends upon the input variables used within the model (Fernando et al., 2009; Hejazi and Cai, 2009; Sharma and Mehrotra, 2014; Tran et al., 2016), conditioning the parameters of a data-driven model on selected input variables seems to be a natural and reasonable assumption. What appears less obvious, is how input data uncertainty plays a role in input variable selection (Galelli et al., 2014). While this latter point is outside the scope of this work (since we instead focus on introducing input variable selection uncertainty within the blueprint), it could also be a reasonable assumption to condition input variable selection on input data. In light of this idea, we modify Eq. 5.3, by using the chain rule (of probability), to take into account

the conditional dependence of parameters on input data and selected input variables and the conditional dependence of selected input variables on input data:

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{X}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta} | \boldsymbol{X}, \boldsymbol{\omega}) f_{\boldsymbol{\Omega}}(\boldsymbol{\omega} | \boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{\Theta} d\boldsymbol{X} \qquad (5.4)$$

where $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{X}, \boldsymbol{\omega})$ represents the conditional probability density function of the parameters given the input data and selected input variables and $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{X})$ is the conditional probability density function of the selected input variables conditioned on input data. Notably, we have converted Eq. 5.2 into a more general form Eq. 5.4) where model error is dependent on parameters, input data, and selected input variables; parameters are dependent on input data and selected input variables; and selected input variables are conditioned on input data. The process for estimating $f_Q(Q)$ using the different conditional probability density functions mentioned above is given in section 5.4.

We have moved from the blueprint used for process-based modeling by the original authors (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) to focus on data-driven applications, modifying the blueprint in terms of our earlier work on data-driven forecasting and input variable selection (Adamowski et al., 2012; Goyal et al., 2014; Quilty et al., 2016; Yaseen et al., 2016). In the next sub-section, using Eq. 5.4, we develop the Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF), an extension of both the Wavelet Data-Driven Forecasting Framework (WDDFF) proposed by Quilty and Adamowski (2018) and the SDDFF proposed above.

5.3.3. Stochastic Wavelet Data-Driven Forecasting Framework

The SWDDFF builds directly from the SDDFF, represented by Eq. 5.4, by applying wavelet transformation to the input data ($\boldsymbol{X}$) via the mapping $\boldsymbol{\mathcal{W}}$: $\boldsymbol{X} \rightarrow \boldsymbol{X_W}$; where $\boldsymbol{X_W}$ represents the wavelet-transformed inputs. The wavelet transformation of the input data maps an input matrix, $\boldsymbol{X}$ of dimension D to a matrix, $\boldsymbol{X_W}$ of dimension $D(J+1)$ (i.e., $\boldsymbol{\mathcal{W}}$: $\boldsymbol{X} \in \mathbb{R}^{N \mathrm{x} \mathrm{D}} \rightarrow \boldsymbol{X_W} \in \mathbb{R}^{N \mathrm{x} \mathrm{D}(J+1)}$); where N is the number of observations and $j \in J+1$ represents a scale of change for each input data vector. The wavelet-transformed inputs at scales $j = 1{:}J$, usually referred to as wavelet coefficients, represent changes in averages over a scale $\tau_j = 2^{j-1}$ while the wavelet-transformed inputs at scale $J+1$, usually referred to as scaling coefficients, are linked to variations at scales $\lambda_J = 2^J$ and higher (Percival and Walden, 2000, sec. 4.8). The major benefit of using wavelet-transformed inputs (instead of the original inputs) is that we can extract scales of change within the input data and use input variable selection to select only the relevant

scale-based information that is necessary to forecast the target process, which has been shown to be a key factor in improving model performance when forecasting multiscale processes in hydrology and water resources (Nourani et al., 2014; Rathinasamy et al., 2014; Sang, 2013). We adopted the best practices discussed in Quilty and Adamowski (2018) to perform wavelet transformation on the input data and refer the interested reader to their work for thorough theoretical background and a discussion of the key features of wavelet transformation and its use in forecasting applications.

The SWDDFF is realized by a simple adjustment to Eq. 5.4, i.e., modifying the SDDFF to include wavelet-transformed input data ($\boldsymbol{X_W}$):

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{X_W}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta} | \boldsymbol{X_W}, \boldsymbol{\omega}) f_{\boldsymbol{\Omega}}(\boldsymbol{\omega} | \boldsymbol{X_W}) f_{\boldsymbol{X_W}}(\boldsymbol{X_W}) d\boldsymbol{\Theta} d\boldsymbol{X_W} \quad (5.5)$$

We now move from the theoretical development of SDDFF (Eq. 5.4) and SWDDFF (Eq. 5.5) to show how they may be developed for practical applications.

5.4. Applying the Stochastic Wavelet Data-Driven Forecasting Framework

Here we give the workflow required for using the SDDFF and SWDDFF in practice. The workflow builds directly from that given by the blueprint authors (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) by considering the addition of input variable selection uncertainty and wavelet transformation (compare Eq. 5.2 with Eq. 5.4 and Eq. 5.5, respectively). Another source of difference between SDDFF, SWDDFF, and the original blueprint is the fact that we avoid assuming independence between the model error, parameters, (input variable selection) and input data. Due to the similarity between the workflow of SDDFF and SWDDFF, we only provide details for the latter. The workflow for the SWDDFF (as summarized in Figure 5.1) can be stated as follows:

1. A random sample is drawn from the probability density $f_{\boldsymbol{X}}(\boldsymbol{X})$.
2. Wavelet transformation is performed on the drawn sample from 1 (i.e., $\boldsymbol{\mathcal{W}}: \boldsymbol{X} \rightarrow \boldsymbol{X_W}$), obtaining an equivalent realization from $f_{\boldsymbol{X_W}}(\boldsymbol{X_W})$.
3. A random sample is drawn from the conditional probability density $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{X_W})$.
4. A random sample is drawn from the conditional probability density $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{X_W}, \boldsymbol{\omega})$.
5. Using the sampled information $(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$, a model prediction is computed via $S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$;

6. For the model prediction from 5, a random error is picked up from the conditional probability density $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})|\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$ and added to $S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$;
7. Steps 1 to 6 are repeated n times, giving n different predictions of Q.
8. The probability density $f_Q(Q)$ is realized by the n predictions of Q.

If one wished to forgo the wavelet transformation step and instead adopt the SDDFF, then one solely needs to remove step 2 above (substituting $\boldsymbol{X}$ for $\boldsymbol{X_W}$ in the remaining steps**)**. Furthermore, if one wanted to estimate uncertainty related to only specific components of SWDDFF (or SDDFF), e.g., parameter uncertainty, perhaps to compare the contribution of the different sources of uncertainty, then only those steps in the workflow need to be carried out (Montanari and Koutsoyiannis, 2012). We note that in our experiments we do not include the uncertainty due to the input data (i.e., step 1 in the SWDDFF workflow), primarily to focus on input variable selection uncertainty (see section 5.2); however, future work could consider input data uncertainty.

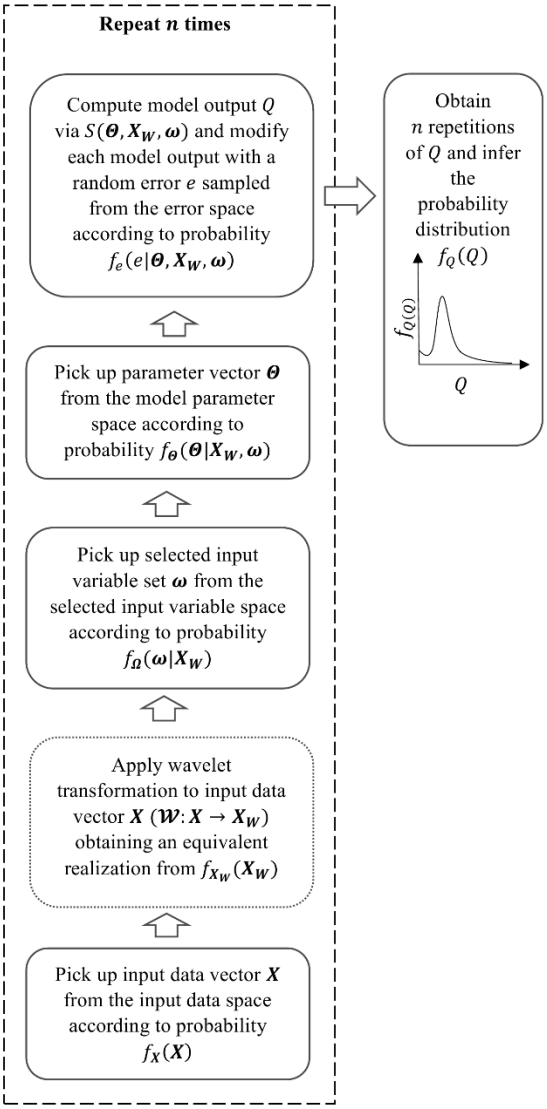

Figure 5.1. Workflow for the Stochastic Wavelet Data-Driven Forecasting Framework (modified from Montanari and Koutsoyiannis (2012) and Sikorska et al. (2015))

5.5. Experiment Settings, Case Study, and Forecast Evaluation

We discuss the details necessary to implement the SDDFF and SWDDFF by first demonstrating how one can both estimate and sample from the probability distributions in Eq. 5.4 and Eq. 5.5. Afterwards, we give details concerning our case study, and the evaluation metrics used to judge the quality of the forecasts produced by SDDFF and SWDDFF. In order to study the impact of input variable selection uncertainty, we simplify the application of the SWDDFF in our experiments (as discussed in the sub-section

above) by ignoring input data uncertainty, removing the need to estimate its related probability density function (step 1 in the SWDDFF workflow).

5.5.1. Estimation of Probability Density Functions

As noted in Montanari and Koutsoyiannis (2012), a key ingredient to using the blueprint (and therefore, SDDFF and SWDDFF), is the specification of the probability density functions. We used the bootstrap (Efron and Tibshirani, 1993; Henderson, 2005) to estimate the different probability density functions (e.g., input variable selection, parameters, and model error). We chose the bootstrap as it is a popular method that can be used for empirically estimating the probability density function of a random variable, leading to its use in a wide array of hydrological and water resources applications involving resampling and uncertainty estimation (Erkyihun et al., 2016; Faghih et al., 2017; Gupta, 2010; Hirsch et al., 2015; Lall and Sharma, 1996; Rustomji and Wilkinson, 2008; Sharma and Tiwari, 2009; Srinivas and Srinivasan, 2005). It was also suggested as a means to estimate parameter uncertainty in Montanari and Koutsoyiannis (2012) and was used in Sikorska et al. (2015) for estimating the model error uncertainty. The key benefits of the bootstrap is that it is non-parametric, likelihood-free, and simple to implement (as it solely relies on random sampling from a given dataset) (Srivastav et al., 2007; Wani et al., 2017). For those interested in non-bootstrap-based probability density function estimation methods we recommend that the reader reviews section 3, 4, and 6 in Montanari and Koutsoyiannis (2012) (and the references mentioned therein). Taormina et al. (2016) can be reviewed for an interesting information-theoretic approach to exploring input variable selection uncertainty.

We do not discuss theoretical details concerning the bootstrap as the above-mentioned references give excellent explanations about the method. For building the bootstrap-based probability distributions we rely on: Quilty et al. (2016) for estimating the input variable selection probability distribution; the paired bootstrap approach in Wan et al. (2014) for estimating the parameter probability density functions; and the k nearest-neighbours bootstrap approach for estimating the model error probability density function (Sikorska et al., 2015). Although it is not covered in this work, the bootstrap can also be used for estimating the input data probability density function (Barton et al., 2014; Freschi et al., 2017; Xie et al., 2016).

5.5.1.1. Input Variable Selection Uncertainty

The input variable selection probability distribution, $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{X_W})$, was estimated via the bootstrap by resampling the calibration data set pairs $(Y, \boldsymbol{X_W})$ several times (where Y represents past observations of Q, i.e., in the calibration dataset), evaluating a given input variable selection algorithm for each resample,

and storing its result (i.e., the selected input variables) in the binary variable $\boldsymbol{\omega} \in \boldsymbol{\Omega}$. The different $\boldsymbol{\omega}$ (for each resample) were used to infer the (empirical) input variable selection probability distribution. Note that the empirical input variable selection probability distribution could have multiple instances where the same input variables are selected for different bootstrap resamples. If the input variable selection uncertainty is minimal (or null), this could lead to the same input variables being selected for each bootstrap resample or, in the other extreme, each bootstrap resample may lead to completely different selected input variables, resulting in a flat (uniform) probability distribution. We note that for cases where each input variable selection set (i.e., for a given bootstrap resample) is unique, it is not to say that *each* selected input variable in that set is unique, when compared to the remaining selected input variable sets. In other words, even for an input variable selection probability distribution that is flat, there may be a single (or group of) input variable(s) that are selected in each input variable set.

5.5.1.2. Parameter Uncertainty

Since the parameter uncertainty is conditioned on the input variable selection uncertainty, for each unique input variable selection set, $\boldsymbol{\omega} \in \boldsymbol{\Omega}$, the parameter probability density function is estimated through the bootstrap by resampling the calibration data set pairs $(Y, \boldsymbol{X}_{\boldsymbol{W}}^{\boldsymbol{\omega}})$ several times (where $\boldsymbol{X}_{\boldsymbol{W}}^{\boldsymbol{\omega}}$ simply represents that only variables $\boldsymbol{\omega} \in \mathbf{1}$ have been selected in $\boldsymbol{X}_{\boldsymbol{W}}$**)** and optimizing the parameter vector for each resample using a given deterministic model, S. Note that $(Y, \boldsymbol{X}_{\boldsymbol{W}}^{\boldsymbol{\omega}})$ is a simplification of the triple $(Y, \boldsymbol{X}_{\boldsymbol{W}}, \boldsymbol{\omega})$, where $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ is held constant for each bootstrap resample when inferring the parameter uncertainty for that particular (unique) selected input variable set.

The different parameter sets for each $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ make up the conditional probability density function for the parameters, i.e., $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{X}_{\boldsymbol{W}}, \boldsymbol{\omega})$. Note that this conditional probability density function can be computationally intensive to estimate, as it requires a sufficient number of bootstrap resamples for *each* selected input variable set (which directly depends on the number of unique selected input variable sets in $\boldsymbol{\Omega}$ and the number of required bootstrap resamples).

5.5.1.3. Model Error Uncertainty

As the model error uncertainty is conditioned on both the parameter and input variable selection uncertainty, the conditional probability density function for the model error, $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X}_{\boldsymbol{W}}, \boldsymbol{\omega})|\boldsymbol{\Theta}, \boldsymbol{X}_{\boldsymbol{W}}, \boldsymbol{\omega})$, was obtained by estimating the error, $e = Q - S(\boldsymbol{\Theta}, \boldsymbol{X}_{\boldsymbol{W}}, \boldsymbol{\omega})$, on the validation set for the model's optimal parameter vector (which was determined by the Nash Sutcliffe Efficiency Index (NASH) (Moriasi et al., 2007)) associated with each (unique) input variable set (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). Therefore, a set (distribution) of model errors was associated

with each model attached to the different (unique) selected input variable sets. As noted in Montanari and Koutsoyiannis (2012), the optimal parameter vector could be exchanged for the parameter vector that provided average (or median) performance, or in the extreme case, one could use the model error for *each* parameter set. We did not follow these approaches, as the performance using our discussed method was high and reflective of performance that would generally satisfy most real-world applications. The caveat of estimating the model error distribution using this approach is that the model error in the validation set should be reflective of the model error which could be expected when running the model in simulation/forecast mode (which can easily be verified by a hold-out (test) set, as done in this study) (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). It is important to note the distinction between the use of Y for the calibration dataset and the use of Q in validation mode. In keeping with the assumptions of the blueprint, during validation mode Q is not yet observed at the time of issuing the forecast using $S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$ for some new input $\boldsymbol{X}^*$, while in calibration mode, past realizations of Q, i.e., Y, are available since it is on the basis of past realizations that input variables are selected and model parameters estimated.

Below, we briefly discuss how we sample from the input variable selection, parameter, and model error probability distributions in the SWDDFF workflow. For the SDDFF workflow, wavelet transformation is excluded and the remaining steps are the same.

5.5.2. Sampling from the Probability Density Functions

In order to use the SWDDFF in a forecasting application, i.e., to obtain $f_Q(Q)$ for a given input $\boldsymbol{X}^*$, which may be a realization from a hold-out (test) set or newly received information in a real-time setting, one needs to pass $\boldsymbol{X}^*$ through the SWDDFF workflow described in Figure 5.1 (section 5.4) by sampling from the probability distributions mentioned above (section 5.5.1.1., 5.5.1.2, and 5.5.1.3.).

Sampling from the probability distribution functions mentioned above (i.e., as part of the SWDDFF workflow) is straightforward. First, a selected input variable set $\boldsymbol{\omega}$ is picked up at random from $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{X_W})$; second, a parameter vector from $\boldsymbol{\Theta}$, associated with the selected input variable set ($\boldsymbol{\omega}$), is picked up at random from $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{X_W}, \boldsymbol{\omega})$; third, $S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$ is evaluated, resulting in a model output, Q; fourth, Q is then compared against the validation model outputs for the selected input variable set ($\boldsymbol{\omega}$) and using the k nearest-neighbour bootstrap (see Sikorska et al. (2015) for details), a model error is randomly sampled from $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})|\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$ and added to Q. This process is repeated a sufficient number of times and results in $f_Q(Q)$ for a given input $\boldsymbol{X}^*$.

5.5.3. Case Study

To demonstrate the potential usefulness of the SDDFF and SWDDFF for obtaining accurate and reliable hydrological and water resources forecasts we chose a real-world case study based on an urban water demand (UWD) dataset from Montreal, Canada that was recently studied in Quilty and Adamowski (2018).

The objective of this case study was to accurately and reliably forecast, via SDDFF and SWDDFF, average daily UWD (whose multiscale nature is shown in Figure 5.2) at lead times of 1, 7, and 14 day(s) ahead for the urban water supply system in Montreal, Canada using historical UWD, maximum air temperature, rainfall, and the antecedent precipitation index as model inputs.

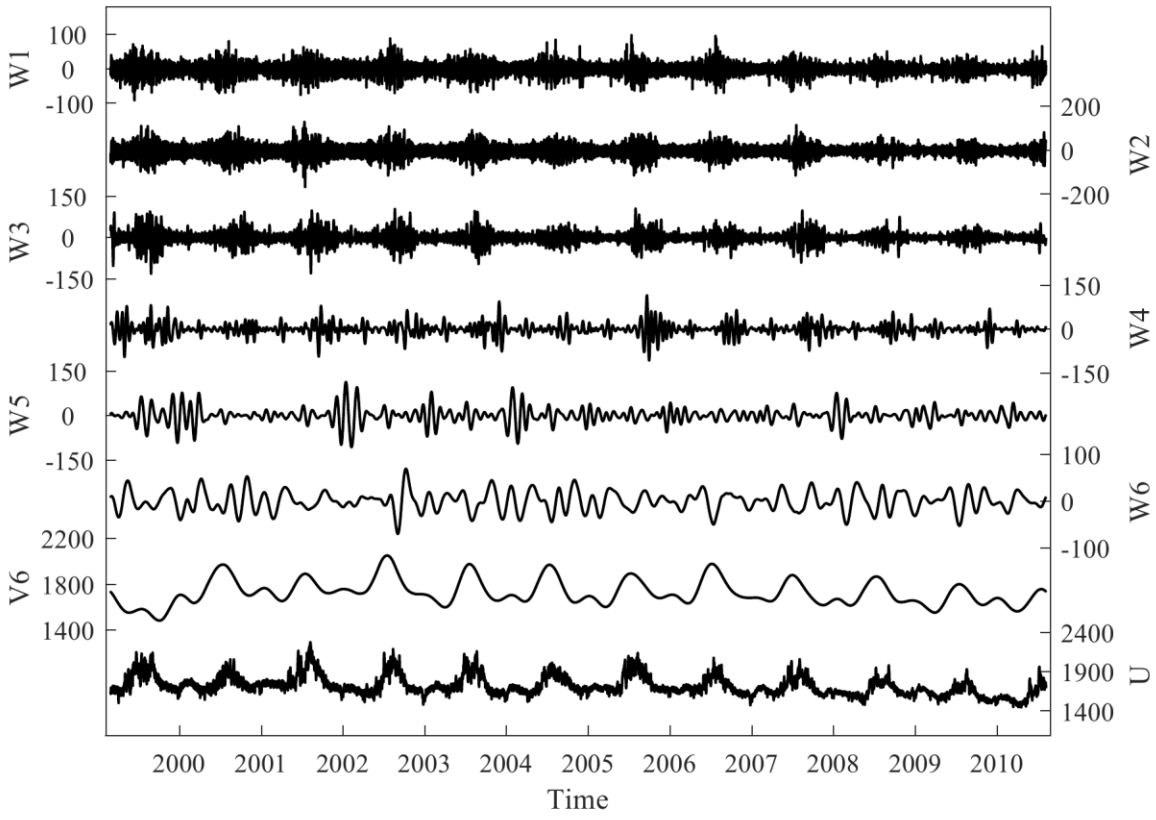

Figure 5.2. Average urban water demand time series, U, (measured in megalitres per day) decomposed by the maximal overlap discrete wavelet transform using the la14 wavelet filter and a decomposition level of six; where each 'W' time series represents the wavelet coefficients at a particular scale and 'V' represents the scaling coefficients; wavelet and scaling coefficients are also measured in megalitres per day

Since Montreal's urban water supply system is the second largest in Canada (supplying water to over 1.3 million persons), it is important that expected short-term UWD be known in advance for optimizing system operations (e.g., pump scheduling, reservoir operations, etc.), planning (e.g., water main maintenance, hydrant flow testing, etc.) and construction (e.g., chlorination, temporary servicing, etc.) for which 1, 7, and 14 day ahead forecasts of average daily UWD are useful.

The authors in Quilty and Adamowski (2018) developed a number of wavelet (WDDFF) and non-wavelet data-driven forecasts for this dataset for the purpose of demonstrating: 1) best practices for wavelet-based forecasting and 2) the improvement in accuracy that can be achieved by wavelet-based forecasts in comparison to their non-wavelet-based counterparts. We used the WDDFF (and non-wavelet-based) models (which we term Data-Driven Forecasting Framework (DDFF)) from this earlier study as benchmarks for the SWDDFF and SDDFF models developed in this paper. The SWDDFF and SDDFF models were developed by modifying the respective (WDDFF and DDFF) models from Quilty and Adamowski (2018) through application of the workflow described in section 5.4 and given in Figure 5.1.

In Table 5.1, we list the different models (and their properties) that were used in Quilty and Adamowski (2018) along with their SDDFF and SWDDFF counterparts. For details on the original models we refer the reader to Quilty and Adamowski (2018). Briefly, the models shown in Table 5.1 can be described by the following information: Model, i.e., model name (DDFF, SDDFF, WDDFF, and SWDDFF); Method, i.e., data-driven method used in the model (second order Volterra (SOV) and multiple linear regression (MLR)); input variable selection (IVS), i.e., input variable selection algorithm used to select input variables for the model (Edgeworth Approximations–based conditional mutual information (EA) and partial correlation input selection (PCIS)); Wavelet, i.e., wavelet filter used in the WDDFF models (Least-Asymmetric with 14 coefficients (la14) and Best-Localized with 14 coefficients (bl14)); and Decomposition Level, i.e., the number of scales, J, used in the wavelet transformation of the inputs for the WDDFF models.

Table 5.1. Description of models

Model	Method	IVS	Wavelet	Decomposition Level
1 Day Lead Time				
DDFF, SDDFF	SOV	EA		
WDDFF, SWDDFF	MLR	PCIS	la14	6
7 Day Lead Time				
DDFF, SDDFF	SOV	EA		
WDDFF, SWDDFF	MLR	PCIS	la14	6
14 Day Lead Time				
DDFF, SDDFF	SOV	EA		
WDDFF, SWDDFF	MLR	PCIS	bl14	6

The dataset (i.e., consisting of average daily historical UWD, daily maximum air temperature, daily rainfall, and daily antecedent precipitation index) extends over the period February 1999 to December 2010. For the 1 (7 and 14) day lead time(s), i.e., the target variable, there are a total of 2395 (2389 and 2382) calibration, 583 validation, and 366 hold-out (test) records (i.e., the same validation and test records are used for each lead time forecast). The non-wavelet-based models (DDFF and SDDFF) had a total of 56 inputs that were considered during input variable selection, which included 14 days of time lagged time series for each explanatory variable (historical UWD, daily maximum air temperature, daily rainfall, and daily antecedent precipitation index). The wavelet-based models (WDDFF and SWDDFF) had a total of 448 inputs that were considered during input variable selection, comprising of the same inputs as the non-wavelet-based models in addition to their wavelet-transformed counterparts at a decomposition level of six ($J = 6$) (56 + (6 +1) * 56 = 448). We refer the reader to Quilty and Adamowski (2018) for further details on the process followed for creating the input-output dataset (the target and input variables) discussed above.

5.5.3.1. Model Settings

It is important to re-iterate that the calibration records were used for building the input variable selection and parameter probability distributions, thus characterizing their respective sources of uncertainty, while the validation set was used for building the model error probability distribution. The hold-out (test) set was used to assess the accuracy and reliability of the forecasts since it was not involved in the estimation of the previously mentioned probability distributions, allowing one to infer the generalization abilities of the SDDFF and SWDDFF models and their potential for real-world problems. This approach is consistent

with what was followed by the blueprint authors (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015).

In our experiments we used a total of 500 bootstrap resamples to develop the different probability distributions (input variable selection, parameter, and model error) and drew from the respective probability distributions 500 times ($n = 500$) to generate the probability distribution of the model output $f_Q(Q)$. Ten ($k = 10$) nearest-neighbours were considered in the k nearest-neighbour bootstrap for drawing from the model error probability density function. We found that these settings provided generally good performance that would be suitable for real-world applications.

To demonstrate the role and importance of the different sources of uncertainty (i.e., input variable selection, parameters, and model output) on the forecast results, we explored three different settings for the SDDFF and SWDDFF models. The three different settings that were explored explicitly considered the following sources of uncertainty: 1) parameter (SDDFF_1 and SWDDFF_1); 2) parameter and model output (SDDFF_2 and SWDDFF_2); and 3) input variable selection, parameter, and model output (SDDFF_3 and SWDDFF_3). We chose these three settings to compare the performance achieved by the SDDFF and SWDDFF models when different sources of uncertainty were explicitly accounted for, allowing us to discern the relative importance of each source of uncertainty, with a focus on the role of input variable selection uncertainty and its impact on forecast accuracy and reliability. We note that these sources of uncertainty should not be considered additive (Montanari and Koutsoyiannis, 2012).

It is relevant to note earlier studies in hydrology and water resources whose models are special cases of the different SDDFF and SWDDFF models described above:

- SDDFF_1 is comparable to the bootstrap data-driven forecasting models that focussed on estimating parameter uncertainty (Jia and Culver, 2006; Tiwari and Chatterjee, 2010b; Zaier et al., 2010);
- SDDFF_2 is similar to the bootstrap neural network forecasting models in Wang et al., (2013) that considered parameter and model output uncertainty;
- SWDDFF_1 is akin to the bootstrap wavelet-data-driven forecasting models that considered parameter uncertainty (Belayneh et al., 2016; Kasiviswanathan et al., 2016; Tiwari and Adamowski, 2017).

- SWDDFF_2 is most alike the bootstrap wavelet-neural network forecasting model in Wang et al., (2013), however the authors in this study used a wavelet-based activation function for the neural network nodes instead of applying wavelet transformation directly to the input data; and
- Finally, SWDDFF_3 is the first wavelet-based forecasting method that considered input variable selection, parameter, and model output uncertainty.

In the next sub-section we briefly describe the evaluation metrics used to judge the accuracy and reliability of the SDDFF and SWDDFF forecasts for our case study.

5.5.4. Forecast Evaluation Metrics

To assess the accuracy and reliability of the SDDFF and SWDDFF models, we applied both deterministic and probabilistic performance evaluation metrics commonly adopted in hydrology and water resources studies. To measure the accuracy of the different forecasts we used the NASH, root mean square error (RMSE), and mean absolute error (MAE) (Hauduc et al., 2015; Legates and McCabe Jr., 1999; Moriasi et al., 2007). We considered the mean continuous ranked probability score (CRPS), prediction interval coverage probability (PICP), average prediction interval width (AW), and interval score (IS) (Boucher et al., 2011; Del Giudice et al., 2015; Kasiviswanathan et al., 2013; Solomatine and Shrestha, 2009) to assess the SDDFF and SWDDFF forecasts from a probabilistic perspective. To supplement these performance evaluation metrics, we also used scatter plots and time series plots for a graphical assessment of the different models.

The NASH, RMSE, and MAE measure the deviation of the forecast from the observations – both NASH and RMSE are based on the square of errors and are more sensitive to large errors while the MAE treats small and large errors equally. The mean CRPS - a useful measure reflecting both forecast sharpness and reliability - compares the distribution of the forecast with that of the observations and reduces to the mean absolute error for point forecasts, allowing for the direct comparison between point and probabilistic forecasts (Fortin et al., 2006; Gneiting and Katzfuss, 2014; Hersbach, 2000; Pappenberger et al., 2015). The PICP measures the coverage of the observations within the forecasts' prediction intervals, which should ideally match that of the prediction interval nominal confidence of $100(1 - \alpha)$ % (we set α at 0.05 in our experiments), and is related to the reliability of the forecast (Wan et al., 2014). The AW simply measures the average width of the prediction intervals and is related to the forecasts' sharpness (Gneiting et al., 2007). The IS combines both sharpness and reliability into a single metric by considering the width of the uncertainty bound in combination with the position of the observation relative to the uncertainty bound (Bourgin et al., 2015; Gneiting and Raftery, 2007).

For the SDDFF and SWDDFF models, the deterministic measures (NASH, RMSE, and MAE) were calculated by using the mean of the n different realizations defining $f_Q(Q)$, enabling comparisons with the DDFF and WDDFF models. The probabilistic measures (CRPS, PICP, AW, and IS) use the entire distribution (i.e., all n different realizations from) $f_Q(Q)$ for evaluating their functions.

Since the SDDFF_1 and SWDDFF_1 models only account for parameter uncertainty, their n forecasts provide an estimate of the variance (or uncertainty) of only the parameters and not the true value of the target process. However, SDDFF_2, SWDDFF_2, SDDFF_3, and SWDDFF_3 through their n forecasts include the variance of not only the parameters but also of the future (forecasted) target process. Therefore, the former results in confidence intervals while the latter results in prediction intervals (which are always wider and encompass the confidence intervals) (Faraway, 2014, sec. 4.1). Since prediction intervals are required to estimate PICP, AW, and IS (Gneiting and Raftery, 2007), we assess only the SDDFF_2, SWDDFF_2, SDDFF_3, and SWDDFF_3 using these measures. However, the CRPS can be used more generally to assess an ensemble of forecasts (e.g., whether generated by drawing samples from a stochastic model (such as in this study), making perturbations to initial conditions, tweaking hyper-parameters, or combining forecasts from different models (Boucher et al., 2015; Zamo and Naveau, 2017)) and is therefore used for assessing how well the n realizations from the SDDFF and SWDDFF models forecast the target process. Therefore, the CRPS acts as a useful measure for assessing the impact that the different sources of uncertainty (input variable selection, parameter, and model output) have on SDDFF and SWDDFF accuracy and reliability. At the same time, since CRPS reduces to the MAE for deterministic forecasts, it is a useful tool for comparing the SDDFF and SWDDFF against the DDFF and WDDFF forecasts.

In terms of the probabilistic assessment of the WDDFF and SWDDFF models, we prefer reliable forecasts to sharp forecasts since users of probability-based forecasts are often interested in the uncertainty of the next outcome and unreliable forecasts are usually of little use regardless of their sharpness (Boucher et al., 2010). Therefore, we are most interested in the PICP score then the CRPS, IS, and AW scores. In other words, if a forecasting model has the best PICP score and relatively close CRPS, IS, AW, and deterministic (NASH, RMSE, and MAE) scores when compared to its competitors, the model with the best PICP score is deemed superior for the task of probability-based forecasting.

5.6. Results and Discussion

We now turn to the evaluation of the proposed SDDFF and SWDDFF models and their benchmarks (DDFF and WDDFF). Table 5.2 records the deterministic (NASH, RMSE, and MAE) and probabilistic (CRPS, PICP,

AW, and IS) forecast evaluation metrics for our UWD forecasting case study in Montreal, Canada. Our focus is on comparing the SDDFF and SWDDFF forecasts against their competitors. The results in Table 5.2 supports our thesis (at least for this case study) that input variable selection uncertainty and wavelet transformation are two important factors in improving forecast accuracy and reliability in comparison to forecasts that do not consider either (or only one) of these tools (compare SDDFF_3 and SWDDFF_3 forecasts against the others in Table 5.2). To support this claim we summarize our main findings below. (Remember that only the test set performances are discussed in this sub-section, as we are solely concerned with how the forecasting models performed out-of-sample.)

1. For each lead time, the SWDDFF_3 model exhibited the most reliable forecasts in terms of the PICP score (ideally the PICP should match the prediction interval nominal confidence of 95%). Therefore, in a probabilistic sense, the inclusion of input variable selection uncertainty and wavelet transformation were important factors that resulted in the most reliable forecasts for our case study.
2. For each lead time, the SWDDFF_3 model provided either the best (lowest) interval score or CRPS score (in conjunction with the best PICP score). Since the CRPS reduces to the MAE for a point forecast, SWDDFF and SDDFF models demonstrated better forecast performance by accounting for forecast probabilities rather than using the methods deterministically, since in all cases the CRPS scores were lower than the MAE scores.
3. For each lead time and most performance measures, wavelet transformation improved forecast performance over the non-wavelet-based models; however, there were exceptions for the 7 day lead time:
    a. The CRPS for the SWDDFF_1 model was inferior to that of the SDDFF_1. However, the SWDDFF_3, which accounted for input variable selection uncertainty, had the lowest CRPS score at the 7 day lead time.
    b. The PICP for the SWDDFF_2 model indicated that it was less reliable than the SDDFF_2 model although it had better CRPS, PICP, AW, and IS scores. Regardless, the SWDDFF_3 exhibited the best reliability (PICP).
4. Input variable selection uncertainty appears to become more important in improving model performance, both in terms of deterministic and probabilistic measures, at the 7 and 14 day lead times in contrast to the 1 day lead time.

Table 5.2. Test set performance for different models using the deterministic (NASH, RMSE, and MAE) and probabilistic (CRPS, PICP, AW, and IS) forecast evaluation metrics

Model	NASH	RMSE (ML/D)	MAE (ML/D)	CRPS (ML/D)	PICP (%)	AW (ML/D)	IS (ML/D)
1 Day Lead Time							
DDFF	0.909	30.631	20.908				
WDDFF	0.929	27.026	18.535				
SDDFF_1	0.909	30.694	20.907	18.289			
SWDDFF_1	0.929	27.019	18.521	16.668			
SDDFF_2	0.903	31.659	21.901	16.714	92.077	117.009	191.467
SWDDFF_2	0.924	27.960	19.418	14.880	92.077	106.163	160.089
SDDFF_3	0.904	31.467	21.800	16.552	92.623	116.377	185.815
SWDDFF_3	0.926	27.689	19.178	14.640	93.169	111.449	160.598
7 Day Lead Time							
DDFF	0.662	59.101	38.857				
WDDFF	0.707	55.027	38.421				
SDDFF_1	0.659	59.394	39.172	35.658			
SWDDFF_1	0.707	55.019	38.405	36.893			
SDDFF_2	0.640	60.967	42.091	31.078	87.158	169.206	406.591
SWDDFF_2	0.694	56.248	39.110	29.621	86.066	141.776	400.841
SDDFF_3	0.642	60.832	42.223	31.111	87.978	170.834	404.286
SWDDFF_3	0.726	53.207	35.923	27.197	91.803	155.963	340.225
14 Day Lead Time							
DDFF	0.504	71.613	51.136				
WDDFF	0.665	58.834	42.919				
SDDFF_1	0.503	71.629	51.048	47.653			
SWDDFF_1	0.665	58.838	42.908	41.281			
SDDFF_2	0.503	71.651	52.947	37.885	83.880	201.832	505.484
SWDDFF_2	0.651	60.022	43.120	32.094	85.246	157.778	416.632
SDDFF_3	0.505	71.498	52.996	37.601	85.246	207.422	468.415
SWDDFF_3	0.674	58.049	42.490	30.770	89.071	168.042	352.173

One may observe that the DDFF and WDDFF models often provided “better” deterministic performance than their respective SDDFF and SWDDFF counterparts (e.g., see the 1 day lead time results in Table 5.2). This is easily explained by the fact that these models (DDFF and WDDFF) did not take into account any form of uncertainty assessment and therefore were biased, with their forecasts relying on only a single set of parameters and selected input variables. Likewise, there exist cases where the SDDFF_1 and SWDDFF_1 models performed “better” than, or as nearly good as, in a deterministic sense, their

respective SDDFF_2 (and SDDFF_3) and SWDDFF_2 (and SWDDFF_3) counterparts (e.g., refer again to the 1 day lead time results in Table 5.2). This is met with similar reasoning as above – the SDDFF_1 and SWDDFF_1 forecasts do not explain the variance in their predictions, only in their parameters and therefore the mean of their n forecasts are biased, as evidenced by the CRPS score. Furthermore, a comparison between the SDDFF_2 and SWDDFF_2 with their respective SDDFF_3 and SWDDFF_3 counterparts reveals that the forecasts of the former are also biased since they do not consider the source of error stemming from forecasts produced using different input variable sets, which resulted in lower reliability for all cases (for the SDDFF_2 and SWDDFF_2 models).

The last point is worth re-iterating: in each case that input variable selection uncertainty was considered (whether wavelets were used or not), forecast reliability improved as well as either or both the CRPS and IS scores. This performance was further increased by incorporating wavelet transformation. We believe that this is evidence supporting the importance of including input variable selection uncertainty and wavelet transformation in stochastic data-driven forecasting of multiscale processes, such as those commonly encountered in hydrology and water resources. Furthermore, by including input variable selection uncertainty alongside parameter and model output uncertainty, SWDDFF (SWDDFF_3 in particular) improved upon the most advanced wavelet-based forecasting models (e.g., those comparable to SWDDFF_1 and SWDDF_2, see the end of section 5.5.3.1.) in the hydrological and water resources literature that to date have either included an assessment of only parameter uncertainty (Belayneh et al., 2016; Kasiviswanathan et al., 2016; Sehgal et al., 2014; Tiwari and Adamowski, 2017) or parameter and output uncertainty (Bachour et al., 2016; Maslova et al., 2016; Wang et al., 2013). While there have been some wavelet-based forecasting models that have focussed on combining multiple wavelet-based forecasts to reduce forecasting uncertainty (Barzegar et al., 2018, 2017; Rathinasamy et al., 2013), they were not concerned with providing probability-based forecasts but rather an average (pooled) ensemble forecast.

Given that the SWDDFF_3 models provided the best overall performance, we now compare its performance against SDDFF_3 using different graphical tools (note that in these figures SWDDFF and SDDFF are equivalent to SWDDFF_3 and SDDFF_3, respectively):

1. Figure 5.3 - A plot of the CRPS score versus the number of resamples (n) included in $f_Q(Q)$;
2. Figure 5.4 and 5.5 - Scatter plots for the observed versus the mean forecast of $f_Q(Q)$ and each of the n forecasts in $f_Q(Q)$, respectively); and

3. Figure 5.6 - Time series plots for the mean of the $f_Q(Q)$ forecasts and for the 0.025 and 0.975 quantiles of the $f_Q(Q)$ forecasts versus the observations.

We notice in Figure 5.3 that the CRPS steadily decreases as the number of resamples (n) is increased. There is a noticeable drop in CRPS around n=50 but the CRPS continues to descend after this point. This is an indication that n=500 is more than sufficient for stabilizing performance of the SDDFF and SWDDFF models for this case study.

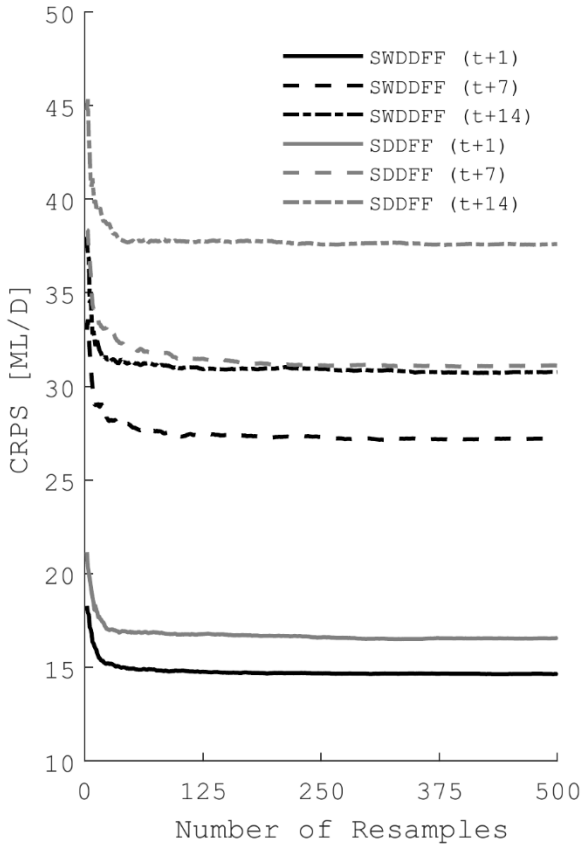

Figure 5.3. CRPS versus the number of resamples ($\boldsymbol{n}$) for different lead times

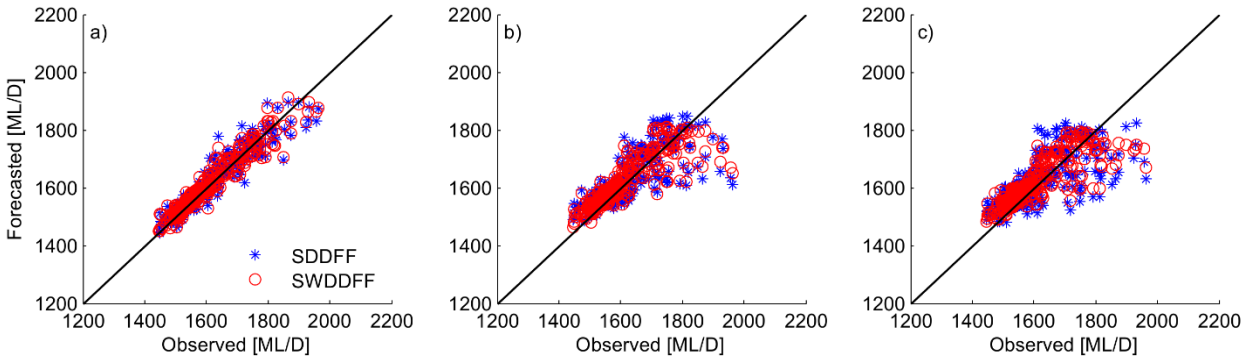

Figure 5.4. Scatter plots for the mean of the $\boldsymbol{n}$ different forecasts defining $\boldsymbol{f_Q(Q)}$ for different lead times a) 1 day ahead, b) 7 day ahead, and c) 14 day ahead

In Figure 5.4 and 5.5, the SWDDFF forecasts are more tightly centered on the bisector line indicating a better fit than the SDDFF forecasts. The SDDFF forecasts tend to have larger outliers for medium and higher flows than the SWDDFF.

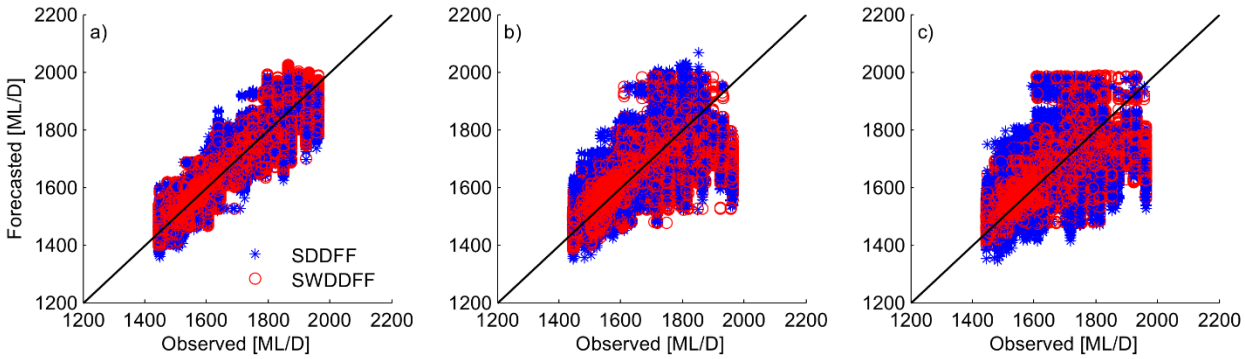

Figure 5.5. Scatter plots for the $\boldsymbol{n}$ different forecasts defining $\boldsymbol{f_Q(Q)}$ for different lead times a) 1 day ahead, b) 7 day ahead, and c) 14 day ahead

Figure 5.6 compliments Figure 5.4 and 5.5, demonstrating that the SDDFF and SWDDFF forecasts are not only accurate but successful at matching the different scales of change in the UWD time series through time. For each lead time, the SDDFF and SWDDFF captured the weekly cycle during lower flows (time indices around 60 – 280) reasonably well in terms of its mean forecast and its prediction intervals (i.e., 0.025 and 0.0975 quantiles). During the seasonal drop in UWD (time indices around 0 – 60), the SDDFF and SWDDFF dropped in accuracy but were still reliable across all lead times. When the seasonal demand picked up (time indices around 280 onwards), the 1 day lead time forecasts were both accurate and

reliable. However, the accuracy and reliability of the mean forecast and its prediction intervals were noticeably worse for 7 and 14 day lead times. In general, the prediction intervals for the SWDDFF can be deemed superior to those for the SDDFF (visually and statistically) as there is not only an improvement in sharpness (see the AW scores) but also in reliability (see PICP, CPRS, and IS). However, while both forecasting models tended to exhibit good reliability, their upper prediction interval tended to overestimate lower flows, especially at longer lead times (e.g., 7 and 14 days ahead).

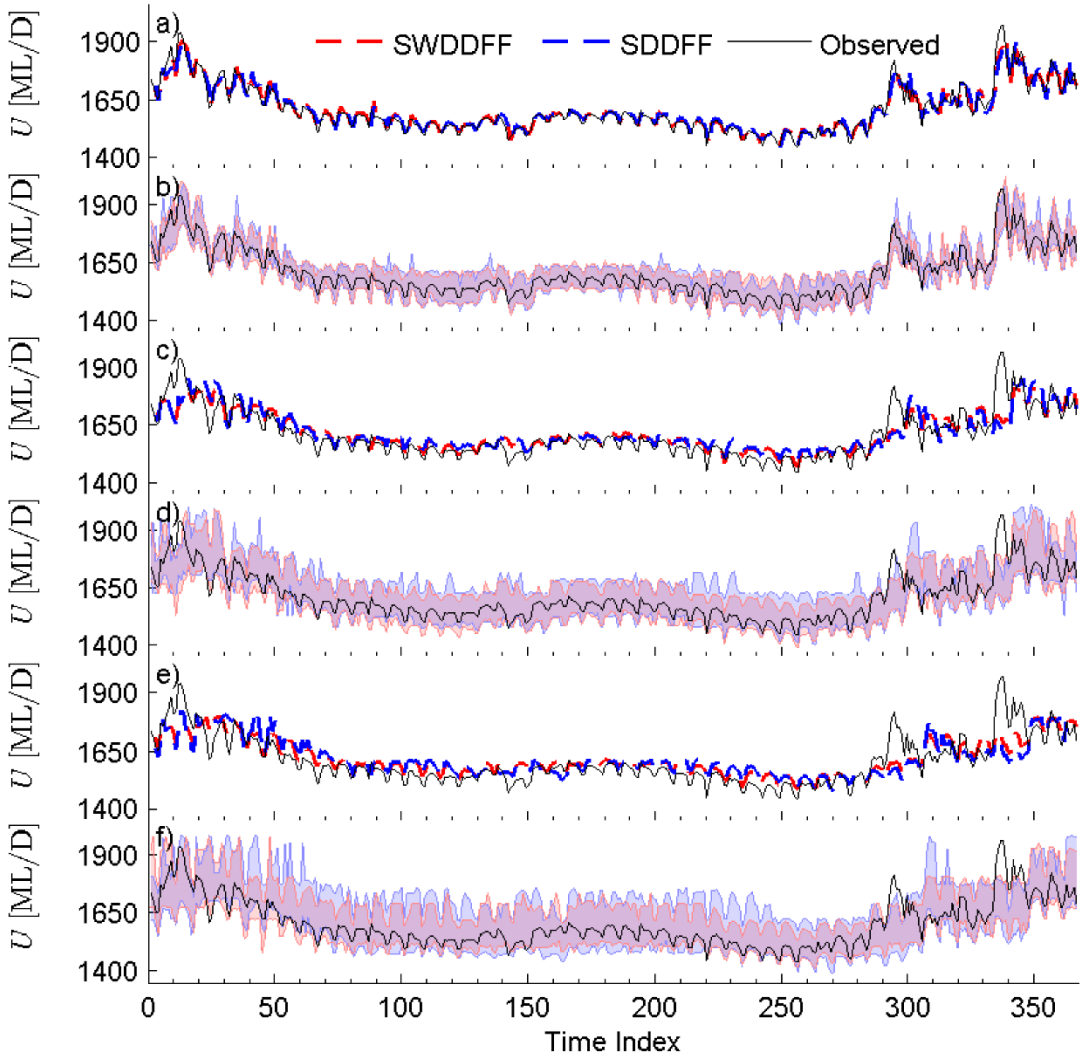

Figure 5.6. Time series plots for the mean of the $\boldsymbol{f_Q(Q)}$ forecasts (at a 1 day lead time (a), 7 day lead time (c), and 14 day lead time (e)) and for the 0.025 and 0.975 quantiles of the $\boldsymbol{f_Q(Q)}$ forecasts (at a 1 day lead time (b), 7 day lead time (d), and 14 day lead time (f)) versus the observations

5.7. Summary and Conclusions

This study introduced two new models: a stochastic data-driven forecasting framework (Stochastic Data-Driven Forecasting Framework (SDDFF)) and a wavelet-based stochastic data-driven forecasting framework (Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF)) for forecasting of

multiscale hydrological and water resources processes. Our proposed methods contribute to the literature by:

1. Extending the recent stochastic process-based modeling blueprint (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) by:
    a. Adopting a more general data-driven framework (instead of process-based models);
    b. Explicitly accounting for input variable selection uncertainty, alongside input data, parameter, and model output uncertainty (resulting in the SDDFF); and
    c. Incorporating wavelet transformation of model inputs to address the multiscale nature of hydrological and water resources processes (resulting in the SWDDFF).
2. By accounting for input variable selection, input data, parameter, and model output uncertainty we improve upon the most advanced wavelet-based data driven forecasting models that until now have either only considered parameter uncertainty (Belayneh et al., 2016; Kasiviswanathan et al., 2016; Khalil et al., 2015; Sang et al., 2013; Tiwari and Adamowski, 2013, 2017; Tiwari and Chatterjee, 2011) or parameter and model output uncertainty (Bachour et al., 2016; Maslova et al., 2016; Wang et al., 2013) in their forecasts.

We demonstrated through a real-world daily urban water demand forecasting study in Montreal, Canada that both input variable selection uncertainty and wavelet transformation were key factors in improving both forecast accuracy and reliability as measured by deterministic and probabilistic forecasting metrics commonly adopted in hydrology and water resources. Since input variable selection uncertainty is a new topic in this domain (Quilty et al., 2016; Taormina et al., 2016) and since input variable selection has just recently began to be studied in conjunction with wavelet transformation (He et al., 2015; Prasad et al., 2017; Tran et al., 2016) (but not yet in terms of stochastic forecasting), this study introduces a simple and effective manner for building stochastic wavelet-based forecasts incorporating input variable selection, parameter, and model output uncertainty. We hope that SWDDFF serves as a useful reference for those interested in building stochastic wavelet-based forecasting models.

We note that our study was limited in a number of ways: 1) we did not consider a wide variety of wavelet filters and decomposition levels; 2) only a small number of data-driven models and input variable selection methods were studied; 3) we did not consider input data uncertainty (in order to focus on input variable selection uncertainty); and 4) we only used a single case study. We believe that these limitations are reasonable for the present study, as our main purpose was to introduce SDDFF and SWDDFF and test its

efficacy on a case study familiar to the authors. However, different research avenues can be explored in future studies such as:

1. Testing the suitability of different methods (i.e., in addition to the bootstrap) for generating the various probability density functions (e.g., input data, input variable selection, parameters, and model output);
2. The evaluation of a number of different wavelet families and decomposition levels;
3. Exploring a wider range of data-driven models and input variable selection methods;
4. Considering an ensemble-SWDDFF based on different wavelet families, decomposition levels, data-driven models, and input variable selection methods;
5. Including input data uncertainty alongside input variable selection, parameter, and model uncertainty;
6. Comparing the performance of SDDFF and SWDDFF against process-based models, such as those used in the original blueprint; and
7. Testing SDDFF and SWDDFF on a larger number of hydrological and water resources processes as well as time series stemming from different domains.

5.8. Acknowledgements

We thank the City of Montreal for providing the urban water demand data used in this study. Requests for urban water demand and meteorological data used in this study should be made through the City of Montreal or obtained from Environment Canada at climate.weather.gc.ca, respectively. Partial funding for this research was provided by NSERC Discovery and Accelerator grants held by Jan Adamowski.

5.9. References

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L., 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog. Phys. Geogr. 36, 480–513. doi:10.1177/0309133312444943

Abrahart, R.J., See, L.M., 2007. Neural network modelling of non-linear hydrological relationships. Hydrol. Earth Syst. Sci. 11, 1563–1579. doi:10.5194/hess-11-1563-2007

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in

Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. doi:10.1016/j.jhydrol.2016.07.048

Bachour, R., Maslova, I., Ticlavilca, A.M., Walker, W.R., McKee, M., 2016. Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. Stoch. Environ. Res. Risk Assess. 30, 103–117. doi:10.1007/s00477-015-1039-z

Barton, R.R., Nelson, B.L., Xie, W., 2014. Quantifying input uncertainty via simulation confidence intervals. INFORMS J. Comput. 26, 74–87. doi:10.1287/ijoc.2013.0548

Barzegar, R., Asghari Moghaddam, A., Adamowski, J., Ozga-Zielinski, B., 2018. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Environ. Res. Risk Assess. 32, 799–813. doi:10.1007/s00477-017-1394-z

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172–173. doi:10.1016/j.atmosres.2015.12.017

Beven, K., 2015. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol. Sci. J. 6667, 150527103244004. doi:10.1080/02626667.2015.1031761

Boucher, M.-A., Laliberté, J.-P., Anctil, F., 2010. An experiment on the evolution of an ensemble of neural networks for streamflow forecasting. Hydrol. Earth Syst. Sci. 14, 603–612. doi:10.5194/hess-14-603-2010

Boucher, M.A., Anctil, F., Perreault, L., Tremblay, D., 2011. A comparison between ensemble and deterministic hydrological forecasts in an operational context. Adv. Geosci. 29, 85–94. doi:10.5194/adgeo-29-85-2011

Boucher, M.A., Perreault, L., Anctil, F., Favre, A.C., 2015. Exploratory analysis of statistical post-processing

methods for hydrological ensemble forecasts. Hydrol. Process. 29, 1141–1155. doi:10.1002/hyp.10234

Bourgin, F., Andréassian, V., Perrin, C., Oudin, L., 2015. Transferring global uncertainty estimates from gauged to ungauged catchments. Hydrol. Earth Syst. Sci. 19, 2535–2546. doi:10.5194/hess-19-2535-2015

Bowden, G.J., Maier, H.R., Dandy, G.C., 2012. Real-time deployment of artificial neural network forecasting models: Understanding the range of applicability. Water Resour. Res. 48, 1–16. doi:10.1029/2012WR011984

Bowden, G.J., Maier, H.R., Dandy, G.C., 2002. Optimal division of data for neural network models in water resources applications. Water Resour. Res. 38, 2-1-2–11. doi:10.1029/2001WR000266

Ciupak, M., Ozga-Zielinski, B., Adamowski, J., Quilty, J., Khalil, B., 2015. The application of Dynamic Linear Bayesian Models in hydrological forecasting: Varying Coefficient Regression and Discount Weighted Regression. J. Hydrol. 530, 762–784. doi:10.1016/j.jhydrol.2015.10.023

Coelho, F., Braga, A.P., Verleysen, M., 2016. A Mutual Information estimator for continuous and discrete variables applied to Feature Selection and Classification problems. Int. J. Comput. Intell. Syst. 9, 726–733. doi:10.1080/18756891.2016.1204120

Creaco, E., Berardi, L., Sun, S., Giustolisi, O., Savic, D., 2016. Selection of relevant input variables in storm water quality modeling by multiobjective evolutionary polynomial regression paradigm. Water Resour. Res. 52, 2403–2419. doi:10.1002/2015WR017971

Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Prog. Phys. Geogr. 25, 80–108. doi:10.1191/030913301674775671

Del Giudice, D., Löwe, R., Madsen, H., Mikkelsen, P.S., Rieckermann, J., 2015. Comparison of two stochastic techniques for reliable urban runoff prediction by modeling systematic errors. Water Resour. Res. 51, 5004–5022. doi:10.1002/2014WR016678

Dixit, P., Londhe, S., Deo, M.C., 2016. Review of Applications of Neuro-Wavelet Techniques in Water Flows. Ina. Lett. 1, 99–104. doi:10.1007/s41403-016-0015-3

Efron, B., Tibshirani, R.J.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, London, U.K.

Erkyihun, S.T., Rajagopalan, B., Zagona, E., Lall, U., Nowak, K., 2016. Wavelet-based time series bootstrap

model for multidecadal streamflow simulation using climate indicators. Water Resour. Res. 52, 4061–4077. doi:10.1002/2016WR018696

Faghih, M., Mirzaei, M., Adamowski, J., Lee, J., El-Shafie, A., 2017. Uncertainty Estimation in Flood Inundation Mapping: An Application of Non-parametric Bootstrapping. River Res. Appl. 33, 611–619. doi:10.1002/rra.3108

Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor. Appl. Climatol. 128, 875–903. doi:10.1007/s00704-016-1735-8

Faraway, J.J., 2014. Linear models with R. CRC press.

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176. doi:10.1016/j.jhydrol.2008.10.019

Fortin, V., Favre, A.-C., Saïd, M., 2006. Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. Q. J. R. Meteorol. Soc. 132, 1349–1369. doi:10.1256/qj.05.167

Freschi, V., Delpriori, S., Lattanzi, E., Bogliolo, A., 2017. Bootstrap Based Uncertainty Propagation for Data Quality Estimation in Crowdsensing Systems. IEEE Access 5, 1146–1155.

Galelli, S., Castelletti, A., 2013. Tree-based iterative input variable selection for hydrological modeling. Water Resour. Res. 49, 4295–4310. doi:10.1002/wrcr.20339

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. doi:10.1016/j.envsoft.2014.08.015

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat. Methodol. 69, 243–268. doi:10.1111/j.1467-9868.2007.00587.x

Gneiting, T., Katzfuss, M., 2014. Probabilistic Forecasting. Annu. Rev. Stat. Its Appl. 1, 125–151. doi:10.1146/annurev-statistics-062713-085831

Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. J. Am. Stat. Assoc. 102, 359–378. doi:10.1198/016214506000001437

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., Hero, A.O., 2013. Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. Water Resour. Res. 49, 2253–2273. doi:10.1002/wrcr.20161

Goyal, M.K., Bharti, B., Quilty, J., Adamowski, J., Pandey, A., 2014. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst. Appl. 41. doi:10.1016/j.eswa.2014.02.047

Gupta, H. V, 2010. Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling. Water Resour. Res. 46, W07515.

Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., Vanrolleghem, P.A., 2015. Efficiency criteria for environmental model quality assessment: Areview and its application to wastewater treatment. Environ. Model. Softw. 68, 196–204. doi:10.1016/j.envsoft.2015.02.004

He, X., Guan, H., Qin, J., 2015. A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. J. Hydrol. 527, 88–100. doi:10.1016/j.jhydrol.2015.04.047

Hejazi, M.I., Cai, X., 2009. Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm. Adv. Water Resour. 32, 582–593. doi:10.1016/j.advwatres.2009.01.009

Henderson, A.R., 2005. The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. Clin. Chim. Acta 359, 1–26. doi:10.1016/j.cccn.2005.04.002

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather Forecast. 15, 559–570. doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Hirsch, R.M., Archfield, S.A., De Cicco, L.A., 2015. A bootstrap method for estimating uncertainty of water quality trends. Environ. Model. Softw. 73, 148–166. doi:10.1016/j.envsoft.2015.07.017

Jia, Y., Culver, T.B., 2006. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. J. Hydrol. 331, 580–590. doi:10.1016/j.jhydrol.2006.06.005

Kasiviswanathan, K.S., Cibin, R., Sudheer, K.P., Chaubey, I., 2013. Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. J. Hydrol. 499, 275–

288. doi:10.1016/j.jhydrol.2013.06.043

Kasiviswanathan, K.S., He, J., Sudheer, K.P., Tay, J.-H., 2016. Potential application of wavelet neural network ensemble to forecast streamflow for flood management. J. Hydrol. 536, 161–173. doi:10.1016/j.jhydrol.2016.02.044

Khalil, B., Broda, S., Adamowski, J., Ozga-Zielinski, B., Donohoe, A., 2015. Short-term forecasting of groundwater levels under conditions of mine-tailings recharge using wavelet ensemble neural network models. Hydrogeol. J. 23, 121–141. doi:10.1007/s10040-014-1204-3

Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology. J. Hydrol. 249, 2–9. doi:10.1016/S0022-1694(01)00420-6

Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet analysis for geophysical applications. Rev. Geophys. 35, 385–412. doi:10.1029/97RG00427

Kumar, S., Tiwari, M.K., Chatterjee, C., Mishra, A., 2015. Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method. Water Resour. Manag. 29, 4863–4883. doi:10.1007/s11269-015-1095-7

Labat, D., 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. J. Hydrol. 314, 275–288. doi:10.1016/j.jhydrol.2005.04.003

Lall, U., Sharma, A., 1996. A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series. Water Resour. Res. 32, 679–693. doi:10.1029/95WR02966

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the Use of "Goodness of Fit" Measures in Hydrologic and Hydroclimatic Model Validation. Water Resour. Res. 35, 233–241. doi:10.1029/1998WR900018

Li, X., Maier, H.R., Zecchin, A.C., 2015. Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. Environ. Model. Softw. 65, 15–29. doi:10.1016/j.envsoft.2014.11.028

Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

Maslova, I., Ticlavilca, A.M., Mckee, M., 2016. Adjusting wavelet-based multiresolution analysis boundary conditions for long-term streamflow forecasting. Hydrol. Process. doi:10.1002/hyp.10564

Montanari, A., Koutsoyiannis, D., 2014. Reply to comment by Grey Nearing on "A blueprint for process-based modeling of uncertain hydrological systems." Water Resour. Res. 50, 6264–6268. doi:10.1002/2013WR014987

Montanari, A., Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. Water Resour. Res. 48. doi:10.1029/2011WR011412

Moon, K.R., Sricharan, K., Hero III, A.O., 2017. Ensemble Estimation of Mutual Information, in: 2017 IEEE International Symposium on Information Theory (ISIT). Aachen, Germany, pp. 3030–3034. doi:10.1109/ISIT.2017.8007086

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. doi:10.13031/2013.23153

Nearing, G.S., Tian, Y., Gupta, H. V., Clark, M.P., Harrison, K.W., Weijs, S. V., 2016. A Philosophical Basis for Hydrologic Uncertainty. Hydrol. Sci. J. 6667, 1–20. doi:10.1080/02626667.2016.1183009

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. J. Hydrol. 514, 358–377. doi:10.1016/j.jhydrol.2014.03.057

Pappenberger, F., Ramos, M.H., Cloke, H.L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., Salamon, P., 2015. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. J. Hydrol. 522, 697–713. doi:10.1016/j.jhydrol.2015.01.024

Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis, 1st ed, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Polikar, R., 2007. Bootstrap-inspired techniques in computation intelligence. IEEE Signal Process. Mag. 24, 59–72. doi:10.1109/MSP.2007.4286565

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ANN-based stream flow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos. Res. 197, 42–63. doi:10.1016/j.atmosres.2017.06.014

Quilty, J., Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting

framework. J. Hydrol. 563, 336 – 353. doi:10.1016/j.jhydrol.2018.05.003

Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. Water Resour. Res. 52, 2299–2326. doi:10.1002/2015WR016959

Raftery, A.E., 2016. Use and communication of probabilistic forecasts. Stat. Anal. Data Min. ASA Data Sci. J. 9, 397–410. doi:10.1002/sam.11302

Ramos, M.H., Van Andel, S.J., Pappenberger, F., 2013. Do probabilistic forecasts lead to better decisions? Hydrol. Earth Syst. Sci. 17, 2219–2232. doi:10.5194/hess-17-2219-2013

Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J. Hydrol. doi:10.1016/j.jhydrol.2013.09.025

Rathinasamy, M., Khosa, R., Adamowski, J., Ch, S., Partheepan, G., Anand, J., Narsimlu, B., 2014. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. Water Resour. Res. 50, 9721–9737. doi:10.1002/2013WR014650

Remesan, R., Mathew, J., 2015. Hydrological data driven modelling: A case study approach, Hydrological Data Driven Modelling: A Case Study Approach. Springer International Publishing.

Ross, B.C., 2014. Mutual information between discrete and continuous data sets. PLoS One 9. doi:10.1371/journal.pone.0087357

Rustomji, P., Wilkinson, S.N., 2008. Applying bootstrap resampling to quantify uncertainty in fluvial suspended sediment loads estimated using rating curves. Water Resour. Res. 44, 1–12. doi:10.1029/2007WR006088

Sang, Y.-F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. 122, 8–15. doi:10.1016/j.atmosres.2012.11.003

Sang, Y.F., Shang, L.Y., Wang, Z.G., Liu, C.M., Yang, M.G., 2013. Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting. Chinese Sci. Bull. 58, 3796–3805. doi:10.1007/s11434-013-5903-2

Sehgal, V., Tiwari, M.K., Chatterjee, C., Sehgal, V., Tiwari, M.K., Chatterjee, C., 2014. Wavelet Bootstrap Multiple Linear Regression Based Hybrid Modeling for Daily River Discharge Forecasting. Water

Resour Manag. 28, 2793–2811. doi:10.1007/s11269-014-0638-7

Sharma, A., Mehrotra, R., 2014. An information theoretic alternative to model a natural system using observational information alone. Water Resour. Res. 50, 650–660. doi:10.1002/2013WR013845

Sharma, S.K., Tiwari, K.N., 2009. Bootstrap based artificial neural network (BANN) analysis for hierarchical prediction of monthly runoff in Upper Damodar Valley Catchment. J. Hydrol. 374, 209–222. doi:10.1016/j.jhydrol.2009.06.003

Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2015. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. J. Hydrol. Eng. 20.

Sivillo, J.K., Ahlquist, J.E., Toth, Z., 1997. An Ensemble Forecasting Primer. Weather Forecast. 12, 809–818. doi:10.1175/1520-0434(1997)012<0809:AEFP>2.0.CO;2

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinformatics 10, 3–22. doi:10.2166/hydro.2008.015

Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resour. Res. 45, 1–16. doi:10.1029/2008WR006839

Srinivas, V. V., Srinivasan, K., 2005. Matched block bootstrap for resampling multiseason hydrologic time series. Hydrol. Process. 19, 3659–3682. doi:10.1002/hyp.5849

Srivastav, R.K., Sudheer, K.P., Chaubey, I., 2007. A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. Water Resour. Res. 43, 1–12. doi:10.1029/2006WR005352

Sun, S., Bertrand-Krajewski, J.-L., 2013. Input variable selection and calibration data selection for storm water quality regression models. Water Sci. Technol. 68, 50. doi:10.2166/wst.2013.222

Taormina, R., Galelli, S., Karakaya, G., Ahipasaoglu, S.D., 2016. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models. J. Hydrol. 542, 18–34. doi:10.1016/j.jhydrol.2016.07.045

Thiboult, A., Anctil, F., Ramos, M.H., 2017. How does the quantification of uncertainties affect the quality and value of flood early warning systems? J. Hydrol. 551, 365–373. doi:10.1016/j.jhydrol.2017.05.014

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Tiwari, M.K., Adamowski, J.F., 2017. An ensemble wavelet bootstrap machine learning approach to water demand forecasting: a case study in the city of Calgary, Canada. Urban Water J. 14, 185–201. doi:10.1080/1573062X.2015.1084011

Tiwari, M.K., Adamowski, J.F., 2015. Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet–Bootstrap Machine-Learning Approach. J. Water Resour. Plan. Manag. 141. doi:10.1061/(ASCE)WR.1943-5452.0000454

Tiwari, M.K., Chatterjee, C., 2011. A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. J. Hydroinformatics 13, 500. doi:10.2166/hydro.2010.142

Tiwari, M.K., Chatterjee, C., 2010a. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. J. Hydrol. 394, 458–470. doi:10.1016/j.jhydrol.2010.10.001

Tiwari, M.K., Chatterjee, C., 2010b. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). J. Hydrol. 382, 20–33. doi:10.1016/j.jhydrol.2009.12.013

Tran, H.D., Muttil, N., Perera, B.J.C., 2016. Enhancing accuracy of autoregressive time series forecasting with input selection and wavelet transformation. J. Hydroinformatics. doi:10.2166/hydro.2016.145

Wan, C., Xu, Z., Pinson, P., Dong, Z.Y., Wong, K.P., 2014. Probabilistic forecasting of wind power generation using extreme learning machine. IEEE Trans. Power Syst. 29, 1033–1044. doi:10.1109/TPWRS.2013.2287871

Wang, Y., Zheng, T., Zhao, Y., Jiang, J., Wang, Y., Guo, L., Wang, P., 2013. Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. Environ. Sci. Pollut. Res. 20, 8909–8923. doi:10.1007/s11356-013-1874-8

Wani, O., Beckers, J.V.L., Weerts, A.H., Solomatine, D.P., 2017. Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting. Hydrol. Earth Syst. Sci. 21, 4021–4036. doi:10.5194/hess-21-4021-2017

Xie, W., Nelson, B.L., Barton, R.R., 2016. Multivariate Input Uncertainty in Output Analysis for Stochastic Simulation. ACM Trans. Model. Comput. Simul. 27, 5:1--5:22. doi:10.1145/2990190

Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. J. Hydrol. doi:10.1016/j.jhydrol.2015.10.038

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-Shafie, A., 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. J. Hydrol. 542, 603–614. doi:10.1016/j.jhydrol.2016.09.035

Zaier, I., Shu, C., Ouarda, T.B.M.J., Seidou, O., Chebana, F., 2010. Estimation of ice thickness on lakes using artificial neural network ensembles. J. Hydrol. 383, 330–340. doi:10.1016/j.jhydrol.2010.01.006

Zamo, M., Naveau, P., 2017. Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. Math. Geosci. 1–26. doi:10.1007/s11004-017-9709-7

Connecting Text to Chapter 6

The Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF) from Chapter 5 accounts for the multiscale nature of water resources by relying on a single wavelet basis function (wavelet filter). This hinders the ability of SWDDFF to comprehensively capture the complex behaviour of water resources which may differ across scales. Therefore, in this chapter, an ensemble multi-wavelet version of SWDDFF, Ensemble Wavelet – Stochastic Data-Driven Forecasting Framework (EW-SDDFF) is proposed to incorporate the strengths of multiple wavelet filters through an ensemble approach to better capture the complex multiscale nature of water resources and to generate probability-based forecasts of the target process. EW-SDDFF uses multiple forecasts previously developed by the Wavelet Data-Driven Forecasting Framework (WDDFF) from Chapter 4 and identifies and weights the importance of each forecast using input variable selection and data-driven methods, respectively. EW-SDDFF is a useful probability-based forecasting tool for capturing the nonlinear, multiscale, and uncertain nature of water resources (that can vary across scales) improving upon its single-wavelet counterpart from Chapter 5. EW-SDDFF can be used by water resources managers for operating, planning, and managing water resources systems under uncertainty and is particularly useful in the decision-making stages due to its probability-based interpretation.

This chapter has been submitted to Water Resources Research. The format of the paper has been modified to ensure consistency with the style of this thesis. A list of references cited in this paper are available at the end of the chapter.

The author of the thesis was responsible for the development, testing, and application of the different methods and wrote the manuscript presented here. Prof. Adamowski, the supervisor of this thesis, provided valuable advice on all aspects of the research and contributed to the review and editing of the manuscript. Prof. Marie-Amélie Boucher at the Civil Engineering Department, Université de Sherbrooke, provided valuable advice on ensemble and probabilistic forecasting model evaluation and contributed to the review and editing of the manuscript.

Chapter 6: A Stochastic Data-Driven Ensemble Forecasting Framework for Water Resources: A Case Study using Ensemble Members Derived from a Database of Deterministic Wavelet-based Models

6.1. Abstract

The computational burden in operational water resources forecasting can be very high. It can involve combining a large number of forecasts via post-hoc ensemble model selection and weighting strategies. Since this is prohibitive for project resources and timelines, we propose and test a stochastic data-driven ensemble forecasting framework that uses archived deterministic forecasts as input and results in water resources forecasts in the form of a probability distribution. In addition to input data and (ensemble) model output uncertainty, the proposed approach integrates both model selection and weighting uncertainties, using input variable selection and data-driven methods, respectively (first contribution). Therefore, it does not require one to perform model selection and weighting separately. We applied the proposed forecasting framework to a previous real-world case study in Montreal, Canada to forecast daily urban water demand (UWD) at multiple lead times. Using wavelet-based forecasts as input data, we develop the Ensemble Wavelet - Stochastic Data-Driven Forecasting Framework (EW-SDDFF), the first multi-wavelet ensemble stochastic forecasting framework that produces probability-based forecasts (second contribution). For the considered case study, several variants of EW-SDDFF, produced using different input variable selection methods (partial correlation input selection and Edgeworth Approximations-based conditional mutual information) and data-driven models (multiple linear regression, extreme learning machines, and second order Volterra series models), are shown to outperform wavelet- and non-wavelet-based benchmarks, especially during a heatwave (first time studied in the UWD forecasting literature). Future work will consider using the stochastic data-driven ensemble forecasting framework for combining forecasts produced by physically-based and data-driven models for a variety of water resources processes.

Keywords: ensemble forecasting, stochastics, input variable selection, data-driven, wavelets

6.2. Introduction

Research in to the development and application of ensemble forecasting systems is of key interest in the hydrological and water resources domains due to their importance in operations, decision-making, and communicating risks to the public (Cloke and Pappenberger, 2009; Silvestro et al., 2017; Thiboult et al.,

2017). Understandably, the topic of probability-based forecasts (i.e., a forecast in the form of a probability density function) draws much attention in ensemble forecasting since probability-based forecasts easily translate into estimations of risk and provides a range of potential outcomes (forecasts) instead of only a single (deterministic) forecast (Farmer and Vogel, 2016; Han and Coulibaly, 2017; Krzysztofowicz, 2001; Yung et al., 2011). Generally, the purpose behind developing an ensemble forecast is to capture the individual strengths of candidate models (which themselves might be deterministic or probability-based (Hemri et al., 2015)) through model selection (Adhikari et al., 2015; Brochero et al., 2011a; Doycheva et al., 2017; Herger et al., 2018; Tapiador and Gallardo, 2006) and to combine them through model weighting (Li et al., 2017; Weijs and van de Giesen, 2013; Zeng et al., 2016). The ensemble forecasting framework then provides a better overall forecast, usually judged by its accuracy and reliability, when compared to a suitable benchmark (such as the single 'best' candidate forecasting model) (e.g., Brochero et al. (2011b); Pappenberger et al. (2015a); Thiboult et al. (2017)). However, *what is one to do when selecting an optimal set of models to include in the ensemble and determining their individual weights or identifying a suitable benchmark (i.e., the single 'best' candidate forecasting model) for which to compare a particular ensemble forecasting model against, is computationally prohibitive for a given budget or set of project resources*? This set of questions (or rather constraints) motivates the present paper. In the remainder of this section, we: 1) give a brief history of how we arrived at these constraints during previous research; 2) introduce our attempt to mitigate this problem, resulting in a general ensemble stochastic data-driven forecasting framework; and 3) discuss the benefits of the proposed approach, including our contributions to the literature.

6.2.1. A Blueprint for Converting Deterministic to Stochastic Forecasts

Recently, Montanari and Koutsoyiannis (2012) proposed a versatile 'blueprint' that converts any deterministic model into a stochastic one and that results in a probability-based forecast (i.e., a forecast in the form of a probability density function). Two main facets of their blueprint are extremely useful for forecasting problems (particularly those in hydrology and water resources): 1) it can be used with both process-based and data-driven models and 2) it can explicitly account for many different sources of uncertainty (e.g., input data, parameters, model output, model structure, initial conditions, etc.) by estimating their probability density functions, using simple tools such as the bootstrap (Efron and Tibshirani, 1993). To estimate the probability distribution of the target forecast, stochastic perturbations are made to the input data, model parameters, and model output via random draws from the respective probability distributions. The authors showed how their blueprint could also be used to create a ensemble forecast by weighting the output probability distributions of different candidate models (Montanari and

Koutsoyiannis, 2012). The only potential drawback of the blueprint is the computational time required to estimate the various probability functions and to draw from them during forecasting. We discuss this drawback in more detail in section 6.3.2, 6.3.3 and 6.4.1. Notwithstanding, the proposed blueprint still serves as a powerful tool for hydrological and water resources modelling practitioners.

6.2.2. From a Blueprint for Stochastic Process-based Models to a Stochastic Data-Driven Forecasting Framework Involving Wavelets

Montanari and Koutsoyiannis (2012) focussed on process-based models that include input data, parameter, and model output uncertainty; however, the authors alluded to the fact that their method could be adapted for data-driven modelling (Sikorska et al., 2015). In that vein, Quilty and Adamowski (2018a) extended the original blueprint in three main directions by: 1) focussing on the general case of data-driven models (e.g., multiple linear regression (MLR), neural networks, etc.) (Solomatine and Ostfeld, 2008); 2) amending the original blueprint to also include input variable selection uncertainty; and 3) incorporating wavelet transformation of model inputs to extract (potentially complex) time-frequency information into different sub-series. The first two developments resulted in the Stochastic Data-Driven Forecasting Framework (SDDFF). By using wavelet transformed model inputs within the SDDFF, the authors developed the Stochastic Wavelet Data-Driven Forecasting Framework (SWDDFF). The SWDDFF is a combination of the SDDFF and the Wavelet Data-Driven Forecasting Framework (WDDFF), which was developed in Quilty and Adamowski (2018b) according to best practices for real-world wavelet-based forecasting. Quilty and Adamowski (2018a) converted the deterministic WDDFF forecasts from Quilty and Adamowski (2018b) into their stochastic counterparts (SWDDFF) by considering input variable selection, parameter, and model output uncertainties. The authors demonstrated that SWDDFF was able to outperform its non-wavelet-based version (SDDFF) and its non-stochastic version (WDDFF) on a real-world urban water demand forecasting experiment in Montreal, Canada. The main finding of their experiment was that input variable selection uncertainty and wavelet transformation of the model inputs were key factors in improving forecast accuracy and reliability.

6.2.3. The Case for a Multi-Wavelet Stochastic Data-Driven Ensemble Forecasting Framework

However, our previous study (Quilty and Adamowski, 2018a) was limited in that it only explored the single 'best' (deterministic) WDDFF model and converted it into its stochastic counterpart. Quilty and Adamowski (2018b) clearly demonstrated that different wavelet decomposition algorithms, decomposition levels, and wavelet filters lead to substantially different performance. The SWDDFF only considered a single wavelet transform algorithm, decomposition level, and wavelet filter. Although they

performed better than their deterministic WDDFF counterparts and their (non-wavelet-based) SDDFF benchmarks, it remains that other combinations of wavelet transform algorithms, decomposition levels, and wavelet filters could have been used within SWDDFF to provide improved performance. In fact, many recent studies have shown that different combinations of wavelet decomposition levels and wavelet filters perform better at capturing different features of hydrological and water resources processes than others, such as at low-flows or flood regimes (Maheswaran and Khosa, 2012; Rathinasamy et al., 2014, 2013). Furthermore, the decomposition of model inputs through different wavelet decomposition levels and wavelet filters often results in physically interpretable information such as periodicities, trends, transients, and level-shifts, leading to increased forecasting accuracy when wavelet decomposed inputs are used as input to data-driven models (Dixit et al., 2016; Fahimi et al., 2017; Nourani et al., 2014). In connection with these finding, three very recent topics in the wavelet-based forecasting community have been converging: 1) accounting for uncertainty in wavelet-based forecasting models (Belayneh et al., 2016; Bogner and Pappenberger, 2011; Khalil et al., 2015; Liu et al., 2015; Tiwari and Adamowski, 2013); 2) using input variable selection to select the most important wavelet sub-series for forecasting the target process (He et al., 2015; Prasad et al., 2017; Quilty and Adamowski, 2018b; Tran et al., 2016); and 3) combining different wavelet-based forecasts that have been developed using different decomposition levels and wavelet filters in an ensemble forecasting system (as in Alizadeh et al. (2017); Barzegar et al. (2018, 2017); Rathinasamy et al. (2013)).

Therefore, in designing this study, we chose to tackle the problem of creating a multi-wavelet ensemble version of SWDDFF. The goal to develop a multi-wavelet ensemble SWDDFF aligns with the theme of generating probability-based forecasts using ensembles, prominent in hydrology and water resources, along with advancing the wavelet-based forecasting literature (in the same domains). Furthermore, our proposed approach accounts for both model selection and weighting uncertainties (a difficult task in ensemble forecasting (Doycheva et al., 2017; Han and Coulibaly, 2017; Herger et al., 2018)). We are not aware of any studies that have developed a multi-wavelet ensemble stochastic data-driven forecasting model that encompasses each of the features mentioned in this sub-section.

Our intention is to use variations of different wavelet transformation algorithms, decomposition levels, and wavelet filters to create a multi-wavelet ensemble SWDDFF that outperforms the single 'best' SWDDFF model, taking advantage of the strengths of multiple SWDDFF models built on these different variations. Since SWDDFF is the most advanced wavelet-based forecasting method proposed to date, as it is the first to consider input variable selection, parameter, and model output uncertainty (Quilty and

Adamowski, 2018a), extending it to a multi-wavelet ensemble framework that generates probability-based forecasts further improves upon earlier multi-wavelet ensemble approaches that have only sought to develop an aggregated multi-wavelet forecast without taking into account forecast probabilities nor studying the reliability of the forecasts (such as in Alizadeh et al. (2017); Barzegar et al. (2018a, 2017)).

6.2.4. How to Make a Computationally Efficient Ensemble Multi-Wavelet Stochastic Data-Driven Forecasting Framework

In order to show that a multi-wavelet ensemble SWDDFF model can outperform the 'single' best SWDDFF model, we would have to convert each of the 4320 WDDFF models from Quilty and Adamowski (2018b) into their stochastic counterparts (SWDDFF (Quilty and Adamowski, 2018a)), select which SWDDFF models to include in the ensemble and weight them. Then the performance between each individual SWDDFF and the multi-wavelet ensemble SWDDFF could be compared to verify that the multi-wavelet ensemble SWDDFF performed better than the 'single' best SWDDFF. When attempting to strike a balance between the accuracy of the model and computational time required in estimating the various probability density functions in the SWDDFF, we approximated that it would take about 60 days of computation time on a desktop PC (with specifications as follows: Intel(R) Xeon (R) CPU @ 2.40 GHz with 32.0 GB RAM) to convert all WDDFF models into SWDDFF models before selecting then weighting the individual models in the multi-wavelet ensemble SWDDFF. This computational requirement would breach most project timelines and resources and would be impractical for most organizations to consider. Thus, we re-explored how we could still use the WDDFF models in a stochastic ensemble forecasting system that generated probability-based forecasts.

The above computational constraint lead us to use the already produced forecasts from each WDDFF model as input data to a single SDDFF model. Therefore, instead of having to compute thousands of models, perform model selection, and then weight each individual model in the ensemble, this can all be accomplished in one single model. We named our model the Ensemble Wavelet - SDDFF (EW-SDDFF) since it uses an ensemble of deterministic WDDFF forecasts as input data to a SDDFF (inherently performing ensemble member selection and weighting).

6.2.5. Contributions and Benefits of the Proposed Framework

Before moving into the details of the various methods used to develop the EW-SDDFF, we briefly discuss how the SDDFF can be used as a general stochastic ensemble forecasting system, including some of its benefits in this regard; we end this sub-section with a list of our contributions and an outline of the remainder of this paper.

Using the SDDFF model as an ensemble forecasting system by adopting deterministic forecasts as input data reduces the burden of (separately) adopting model selection and weighting strategies within the ensemble since it uses input variable selection to determine useful model inputs and data-driven methods to estimate model weights - inherently handling both model selection and weighting tasks at once. In other words, since the SDDFF incorporates input variable selection uncertainty it selects, amongst the different individual models used as input data, the models whose forecasts are similar to the target process, which can be seen likened to estimating (ensemble) model selection uncertainty. The modeling component (i.e., mapping inputs to the target via a set of parameters in a given model), can be understood from the viewpoint of estimating model weight (parameter) uncertainty based on the selected models in the ensemble. We are not aware of any studies that have used an ensemble stochastic data-driven forecasting framework in terms of probability-based forecasting that inherently considered both ensemble model selection and ensemble weight uncertainties.

The other benefit of using the SDDFF as an ensemble forecasting system is that it is flexible in what can be used as input data, owing to its data-driven nature. The ensemble SDDFF could take as input: different deterministic model outputs from process-based models and/or data-driven models (e.g., WDDFF), large-scale climate indicators, numerical weather predictions, etc. Earlier studies have used data-driven methods to create ensemble forecasting systems, combining process-based and/or data-driven model outputs (Barzegar et al., 2018b; Fernando et al., 2012; Humphrey et al., 2016; Noori and Kalin, 2016; Phanida et al., 2016; Shamseldin et al., 1997; Shoaib et al., 2018; Sun and Trevor, 2018, 2017; Young et al., 2015). Yet none of these approaches inherently incorporated ensemble member selection uncertainty nor produced probability-based forecasts, with the exception of Humphrey et al. (2016). Their approach used GR4J (Perrin et al., 2003) model outputs as input to a Bayesian neural network and produced probabilistic monthly stream flow forecasts near South Australia. Another notable approach, although quite different in the manner of combining model outputs, is that of Tyralis and Koutsoyiannis (2017). The authors combined historical observations, Hurst-Kolmogorov processes, and the output of different general circulation models (GCMs) through the Bayesian processor of forecasts (BPF) in order to probabilistically forecast mean annual air temperature and annual precipitation across the USA for the time period 2016-2100. Both approaches differ from ours since they did not implicitly consider both ensemble model selection and weighting.

The implicit consideration of both ensemble model selection (via input variable selection) and model weighting (via data-driven models) tasks at once is the first contribution of this research to the literature.

The proposed EW-SDDFF is also the first multi-wavelet ensemble stochastic forecasting framework that produces probability-based forecasts accounting for ensemble model weighting and model selection uncertainties, alongside input data and (ensemble) model output uncertainty. This is our second main contribution to the scientific literature. A further contribution of this study is the first application and comparison of deterministic and stochastic forecasts of urban water demand during a heatwave period (see section 6.5.3).

To demonstrate the development and use of the EW-SDDFF on a real-world case study, the rest of this paper is organized as follows: in section 6.3 we discuss the different methods used to develop EW-SDDFF; section 6.4 describes our experiments involving EW-SDDFF and its benchmarks; section 6.5 discusses our main results and their significance; and section 6.6 with a summary and some recommendations for future work.

6.3. Methods

Here we briefly describe the main methods that influenced the development of the proposed EW-SDDFF.

6.3.1. Wavelet Data-Driven Forecasting Framework

In Quilty and Adamowski (2018b) a set of best practices were proposed for the correct development of wavelet-based forecasting models for use in real-world forecasting problems, culminating in the Wavelet Data-Driven Forecasting Framework. In their study, the authors proposed two different wavelet decomposition algorithms, the MODWT (Olhede and Walden, 2004) and *à trous* algorithm (AT) (Aussem et al., 1998), that do not use future information (i.e., data from the present and past is solely required) during wavelet decomposition and therefore may be used correctly for real-world forecasting problems.

Briefly, the MODWT and AT are used to carry out wavelet decomposition (transformation) on given input data ($\boldsymbol{X} \in \mathbb{R}^{N\mathrm{xD}}$, where N is the number of samples and D is the number of inputs) via the mapping $\boldsymbol{\mathcal{W}}$: $\boldsymbol{X} \rightarrow \boldsymbol{X_W}$; where $\boldsymbol{X_W}$ represents the wavelet-decomposed (transformed) inputs. The wavelet transformation of the input data maps an input matrix, $\boldsymbol{X}$ of dimension D to a matrix, $\boldsymbol{X_W}$ of dimension $D(J+1)$ (i.e., $\boldsymbol{\mathcal{W}}$: $\boldsymbol{X} \in \mathbb{R}^{N\mathrm{xD}} \rightarrow \boldsymbol{X_W} \in \mathbb{R}^{N\mathrm{xD}(J+1)}$); where each $j \in J+1$ represents a scale of change for each input data vector Quilty and Adamowski (2018a). We refer to the wavelet-transformed inputs at scales $j = 1{:}J$ as wavelet coefficients (which represent changes in averages over a scale $\tau_j = 2^{j-1}$); the wavelet-transformed inputs at scale $J+1$ are termed scaling coefficients (and represent variations in the data at scales $\lambda_J = 2^J$ and higher) (Percival and Walden, 2000, sec. 4.8).

The WDDFF can be used to forecast a given hydrological or water resources process through the following steps (Quilty and Adamowski, 2018b):

1. The MODWT or AT are used to decompose input data;
2. The decomposed sub-series are fed to an input variable selection method that determines which wavelet-decomposed sub-series (wavelet and scaling coefficients) to use in the forecasting model;
3. A data-driven method (e.g., multiple linear regression (MLR), neural networks, etc.) is then used to map the selected wavelet-decomposed sub-series into a target forecast, $Q_{W_{i,j}}$, where $W_{i,j}$ represents that the forecast was produced using a wavelet decomposition algorithm (MODWT or AT), decomposition level i, and wavelet filter j for the target (Q); and
4. Common forecast evaluation methods (e.g., Nash-Sutcliffe Efficiency Index, root mean square error, etc.) are used to select the best forecasting model(s) – which can be built using different variations of the wavelet decomposition algorithm, decomposition levels, wavelet filters, input variable selection methods, data-driven methods, and so on.

6.3.2. Stochastic Data-Driven Forecasting Framework

The Stochastic Data-Driven Forecasting Framework takes input data ($\boldsymbol{X}$), a given data-driven model (S), any input variable selection method (e.g., partial correlation input selection (PCIS) (May et al., 2008; Quilty et al., 2016; Tran et al., 2016)), and estimates the probability density function of the true target variable to be forecasted $f_Q(Q)$ via (Quilty and Adamowski, 2018a):

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{X}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta} | \boldsymbol{X}, \boldsymbol{\omega}) f_{\boldsymbol{\Omega}}(\boldsymbol{\omega} | \boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{\Theta} d\boldsymbol{X} \tag{6.1}$$

where:

$f_{\boldsymbol{X}}(\boldsymbol{X})$ is the marginal probability density of the input data;

$\boldsymbol{\omega} \in \boldsymbol{\Omega}$ is a binary vector spanning $\boldsymbol{X}$, identifying the variables in $\boldsymbol{X}$ which have been selected (i.e., $\boldsymbol{\omega} \in \{0,1\}^{\boldsymbol{D}}$) using an input variable selection routine;

$f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{X})$ is the conditional probability density function of the selected input variables conditioned on input data;

$f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\omega})$ represents the conditional probability density function of the parameters given the input data and selected input variables;

$S(\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{\omega})$ is the deterministic output of the model S for parameters $\boldsymbol{\Theta}$, input data $\boldsymbol{X}$, and selected input variables $\boldsymbol{\omega}$;

$f_e(Q - S(\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{\omega})|\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{\omega})$ is the conditional probability density function of the model error ($e = Q - S(\boldsymbol{\Theta},\boldsymbol{X})$) conditioned on the parameters, input data, and selected input variables (which quantifies model uncertainty or uncertainties not explicitly accounted for in the model, such as model structural uncertainty, uncertainty due to initial conditions, etc.).

The original blueprint (Montanari and Koutsoyiannis, 2012), and therefore SDDFF (Quilty and Adamowski, 2018a), relies on the assumption that Q is the true variable to be forecasted, which is an unknown quantity at the time of issuing the forecast, and therefore is to be treated as a random variable, see also Montanari and Koutsoyiannis (2014).

Essentially, the SDDFF is a tool to convert any deterministic forecasting model (process-based, data-driven, etc.) into a stochastic one, with the ability to account for various sources of uncertainty explicitly (e.g., input data, input variable selection, parameter, and model output) or implicitly (e.g., model structure, initial conditions, etc.). For further details on the theoretical basis of the SDDFF, readers are referred to the papers on the original blueprint (i.e., Montanari and Koutsoyiannis (2012); Sikorska et al. (2015)) as well as Quilty and Adamowski (2018a).

Perhaps one of the most convenient and flexible features of SDDFF (and the original blueprint) is that it is likelihood-free and that the various sources of uncertainty can each be formulated under different hypotheses. For example, in Montanari and Koutsoyiannis (2012), parameter uncertainty was estimated using the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt and Robinson, 2007); input variable selection and parameter uncertainty were estimated using the paired bootstrap in Quilty and Adamowski (2018a); the meta-Gaussian (Montanari and Koutsoyiannis, 2012) and the k nearest-neighbour (Quilty and Adamowski, 2018a; Sikorska et al., 2015) approaches were used for model error uncertainty in other studies. Furthermore, akin to the blueprint, one may easily 'turn-off' various sources of uncertainty within the SDDFF. One could choose to solely estimate parameter uncertainty, or combined sources of uncertainty (such as input variable selection, parameter, and model output, as in Quilty and Adamowski (2018a)). Alternatively, one could choose to aggregate all sources of uncertainty into the model output uncertainty, similar to Wani et al. (2017).

We now discuss how the SDDFF and WDDFF can be converted into the SWDDFF.

6.3.3. Stochastic Wavelet Data-Driven Forecasting Framework

Merging together the WDDFF and SDDFF, Quilty and Adamowski (2018a) developed the SWDDFF. The main difference between SDDFF and its wavelet-based counterpart, is that SWDDFF includes wavelet-transformation of the input data. The main equation for the SWDDFF is given as:

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{X_W}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta} | \boldsymbol{X_W}, \boldsymbol{\omega}) f_{\boldsymbol{\Omega}}(\boldsymbol{\omega} | \boldsymbol{X_W}) f_{\boldsymbol{X_W}}(\boldsymbol{X_W}) d\boldsymbol{\Theta} d\boldsymbol{X_W} \tag{6.2}$$

where the meanings remain the same for the marginal and conditional probability distributions in Eq. 6.1, with the exception that $\boldsymbol{X}$ in SDDFF is swapped for the wavelet-transformed inputs $\boldsymbol{X_W}$.

The workflow for applying the SWDDFF (and SDDFF) follows directly from the original blueprint (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015) and is summarized below and in Figure 6.1 (which is an adapted version of the original blueprint (Montanari and Koutsoyiannis, 2012) taken from Quilty and Adamowski (2018a)).

1. A random sample is drawn from the probability density function $f_{\boldsymbol{X}}(\boldsymbol{X})$.
2. Wavelet transformation is performed on the drawn sample from 1 (i.e., $\boldsymbol{\mathcal{W}}: \boldsymbol{X} \rightarrow \boldsymbol{X_W}$), obtaining an equivalent realization from $f_{\boldsymbol{X_W}}(\boldsymbol{X_W})$.
3. A random sample is drawn from the conditional probability density function $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega} | \boldsymbol{X_W})$.
4. A random sample is drawn from the conditional probability density function $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta} | \boldsymbol{X_W}, \boldsymbol{\omega})$.
5. Using the sampled information $(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$, a forecast is computed via $S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$;
6. For the forecast from 5, a random error is picked up from the conditional probability density function $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$ and added to $S(\boldsymbol{\Theta}, \boldsymbol{X_W}, \boldsymbol{\omega})$;
7. Steps 1 to 6 are repeated n times, giving n different forecasts of Q.
8. The probability density function $f_Q(Q)$ is realized by the n forecasts of Q.

By skipping step 2 (wavelet transformation), the SWDDFF reverts to the SDDFF. It is worth recalling that one can "turn-off" various sources of uncertainty – this is accomplished by skipping a particular step in the above workflow. For example, if one did not want to include input data uncertainty (which we do not incorporate in our case study given later), one can skip step 1.

An important item to mention is that the different hypotheses employed for estimating the various marginal and conditional probability density functions within SDDFF and SWDDFF can take a significant amount of time to compute, especially when estimating conditional probability density functions. Both Montanari and Koutsoyiannis (2012) and Sikorska et al. (2015) relaxed assumptions on the conditional dependencies and instead considered independence between input data, parameters, and model error and found that their blueprint resulted in sufficient performance for their case studies. However, Quilty and Adamowski (2018a) explicitly accounted for input variable selection, parameter, and model output uncertainty, which, as given in Eq. 6.1 and Eq. 6.2, requires the estimation of the conditional probability density function, i.e., parameter and model output uncertainties and the input variable selection marginal probability density function (since input data uncertainty was not considered).

Finally, the development of the SDDFF and SWDDFF tend to increase in computation time when there is high input variable selection uncertainty. This is because parameter uncertainty is conditioned on input variable selection uncertainty and model error uncertainty is conditioned on both parameter uncertainty and input variable selection uncertainty. The reader is referred to section 6.4 below as well as Quilty and Adamowski (2018a) for further details. We now discuss our proposed EW-SDDFF.

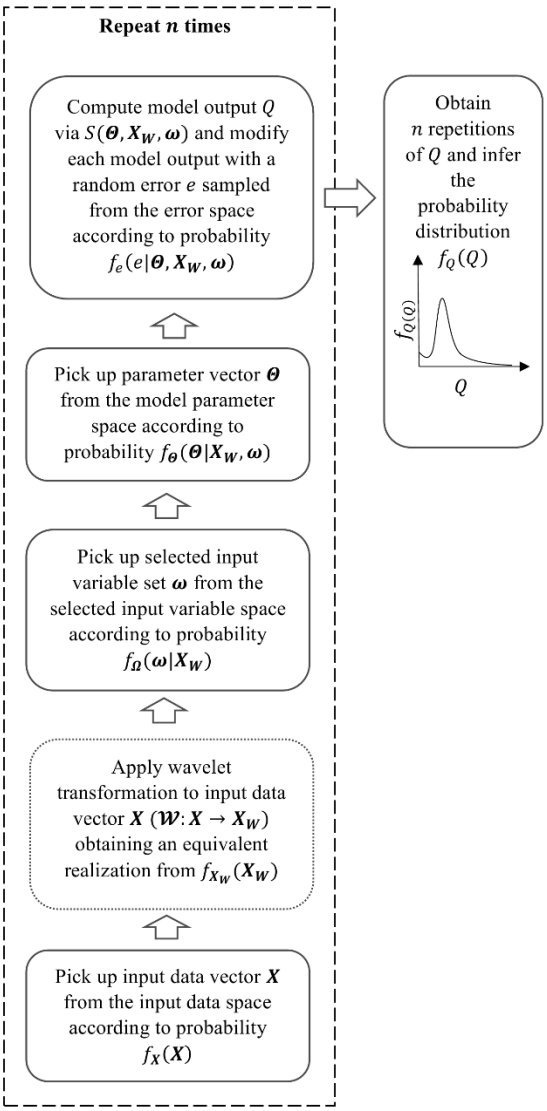

Figure 6.1. Workflow for the Stochastic Wavelet Data-Driven Forecasting Framework (according to Quilty and Adamowski (2018a) and originally modified from Montanari and Koutsoyiannis (2012) and Sikorska et al. (2015))

6.3.4. Ensemble Stochastic Data-Driven Forecasting Framework using Wavelet-based Forecasts as Input Data

The original blueprint authors discussed how their stochastic modelling framework could be extended to a multi-model (ensemble) stochastic modelling framework by weighting the predictive distribution of each *separate* model output according to (Montanari and Koutsoyiannis, 2012):

$$f_Q(Q) = \sum_{k=1}^{M} w_k f_Q^{(k)}(Q) \tag{6.3}$$

where M is the number of models in the ensemble, w_k is the weight associated to each ensemble member (with positive weights that sum to unity), and $f_Q^{(k)}(Q)$ is the probability distribution for each k-th model as given in Eq. 6.1. However, equation Eq. 6.3 can be extremely computationally demanding when one is working with a large number of candidate models. This is particularly true when there is no clear indicator *a priori* to judge which models should *not* be included in the multi-model ensemble. We elaborate more on this topic below giving an example from our experience during this study.

In light of the above, our goal was to develop an ensemble SWDDFF that could outperform the best single SWDDFF. For our case study (described in more detail in section 6.4.2), to develop an SWDDFF for each candidate model, we estimated that it would have taken nearly 60 days on a desktop PC (Intel(R) Xeon (R) CPU @ 2.40 GHz with 32.0 GB RAM) to produce all single SWDDFF models prior to determining which models to include in the ensemble as well as their respective weights (i.e., in the multi-model scheme given in Eq. 6.3). We instead decided to use the SDDFF as the ensemble model itself, by using the various deterministic WDDFF forecasts as input data (to SDDFF), leading to a single probability distribution of the target to be forecasted, rather than M separate probability distributions that would then need to be combined via ensemble member weighting as per Eq. 6.3. In other words, if we were to develop an ensemble SWDDFF it would have taken nearly 60 days to obtain all ensemble members, which would then have to be selected individually for consideration in the ensemble and only afterwards, would the weights (w_k for $k = 1, \dots, M$) for each model in the ensemble be calculated (in three separate steps); instead, EW-SDDFF performs each of the steps required by the ensemble SWDDF at once (i.e., via a single model, $M = 1$) using the previously developed WDDFF as input data to an SDDFF, which uses input variable selection for model selection and data-driven modelling for model weighting. This reduces computing time to within an hour or less (for our particular case study discussed section 6.4.2).

Furthermore, due its data-driven nature, the SDDFF does not require any strictly defined relationships between input data and the target variable, a significant difference when compared with its blueprint counterpart that is process-based (Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). Therefore, it benefits from flexibility in what can be used as input data. For this study, we used the different WDDFF forecasts (i.e., $Q_{W_{i,j}}$) from Quilty and Adamowski (2018b) as input data to forecast the target process,

resulting in our proposed EW-SDDFF. The EW-SDDFF allows us to avoid the excessive computational requirements of a multi-wavelet ensemble of SWDDFF models built according to Eq. 6.3 while still benefitting from the strengths of using different wavelet decomposition levels and wavelet filters in a multi-wavelet ensemble forecasting system.

The EW-SDDFF may be written as follows:

$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{Q_W}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega}) | \boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta} | \boldsymbol{Q_W}, \boldsymbol{\omega}) f_{\boldsymbol{\Omega}}(\boldsymbol{\omega} | \boldsymbol{Q_W}) f_{\boldsymbol{Q_W}}(\boldsymbol{Q_W}) d\boldsymbol{\Theta} d\boldsymbol{Q_W} \quad (6.4)$$

where $\boldsymbol{Q_W}$ is used to represent the collection of the different $Q_{W_{i,j}}$ forecasts from Quilty and Adamowski (2018b). Notice that the only difference between Eq. 6.1 and Eq. 6.4 is that $\boldsymbol{X_W}$ has been swapped for $\boldsymbol{Q_W}$ and the only difference between the SWDDFF workflow and the workflow for EW-SDDFF is that wavelet transformation is no longer required since $\boldsymbol{Q_W}$ embodies the outputs of different forecasts that used wavelet-transformed data as model inputs (i.e., the WDDFF models in Quilty and Adamowski (2018b)).

We now discuss the different experiment settings along with our case study adopted for exploring the use of the EW-SDDFF for a real-world water resources forecasting task.

6.4. Experiment Details

6.4.1. Estimation and Sampling from the Different Probability Density Functions

The most important components of the EW-SDDFF (SDDFF and SWDDFF) is the estimation of and sampling from the different probability density functions. For the estimation and sampling from the different probability density functions in EW-SDDFF (and SWDDFF), we follow our earlier study (Quilty and Adamowski (2018a)):

1. We focus on the estimation of ensemble model selection, model weighting, and model output uncertainty. We do not consider input data uncertainty in order to maintain consistency with our earlier study.
2. The bootstrap resampling method (Efron and Tibshirani, 1993) is used to estimate the input variable selection (ensemble model), parameter (weights), and model output uncertainties:

a. The input variable selection probability distribution $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{Q_W})$ is estimated using the paired bootstrap approach (Wan et al., 2014) by relying on the calibration dataset pairs $(Y, \boldsymbol{Q_W})$ where Y represents past observations of Q, i.e., in the calibration dataset. In other words, Q is a random variable to be observed In the future while Y are previously observed realizations of Q;
b. The conditional probability density function for the parameters, i.e., $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{Q_W}, \boldsymbol{\omega})$ was also estimated using the paired bootstrap approach (Wan et al., 2014) by relying on the calibration data and selected input variables $(Y, \boldsymbol{Q_W}, \boldsymbol{\omega})$. This requires a parameter distribution to be estimated for each unique selected input variable set (i.e., the different selected input variable sets in $\boldsymbol{\Omega}$); and
c. The conditional probability density function for the model error, $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega})|\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega})$, is obtained by estimating the error, $e = Q - S(\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega})$, on the validation set for each model's 'optimal' parameter vector (judged by the Nash Sutcliffe Efficiency Index (Krause and Boyle, 2005; Moriasi et al., 2007)) associated with each unique input variable set through the k nearest-neighbour bootstrap (Sikorska et al., 2015).

The EW-SDDFF can be used to generate the probability density function of the target forecast $f_Q(Q)$ for a given input $\boldsymbol{Q_W}^*$. This input can be the test (hold-out) set data or newly received data, such as from a sensor in a monitoring network. To generate $f_Q(Q)$, sampling from the various probability density functions is required and is accomplished by following the workflow below. It is equivalent to the steps shown in Figure 6.1 but is modified according to section 6.3.4. See also Quilty and Adamowski (2018a)):

1. First, a selected input variable set $\boldsymbol{\omega}$ is picked up at random from $f_{\boldsymbol{\Omega}}(\boldsymbol{\omega}|\boldsymbol{Q_W})$;
2. Second, a parameter vector from $\boldsymbol{\Theta}$, associated with the selected input variable set ($\boldsymbol{\omega}$), is picked up at random from $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\boldsymbol{Q_W}, \boldsymbol{\omega})$;
3. Third, $S(\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega})$ is evaluated, resulting in a model output, Q; and
4. Finally, Q is compared against the validation model outputs for the selected input variable set ($\boldsymbol{\omega}$) and using the k nearest-neighbour bootstrap (see Sikorska et al. (2015) for details). A model error is randomly sampled from $f_e(Q - S(\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega})|\boldsymbol{\Theta}, \boldsymbol{Q_W}, \boldsymbol{\omega})$ and added to Q. This process is repeated a sufficient number of times (n) and results in $f_Q(Q)$ for a given input $\boldsymbol{Q_W}^*$.

We used 100 bootstrap resamples ($n = 100$) to estimate each of the input variable selection, parameter, and model error probability distributions. For the k nearest-neighbour bootstrap, we used ten nearest-

neighbours for picking up a random error from the validation set predictions used to infer the conditional probability density function for the model error. Similarly, 100 bootstrap resamples were used for generating $f_Q(Q)$ by randomly sampling from the different probability distributions. In Quilty and Adamowski (2018a), we found that the reliability of the SDDFF and SWDDFF for generating $f_Q(Q)$ was sufficient for 50 bootstrap resamples and greater; in the same study we also found that 10 nearest-neighbours provided good performance.

We now discuss the details of our case study, which includes a discussion of the forecasts produced by the WDDFF (Quilty and Adamowski, 2018b), which are used as input data for the EW-SDDFF.

6.4.2. Case Study

Our objective is to use the real-world water resources dataset discussed in section 6.4.2.1 in a case study to demonstrate the usefulness of the proposed EW-SDDFF (section 6.3.4 and 6.4.2.2) by evaluating its forecasts through deterministic and probabilistic evaluation metrics commonly adopted in water resources (section 6.4.2.3) and comparing its performance against suitable benchmarks (section 6.4.2.4).

6.4.2.1. Study Site Overview

In earlier studies, Quilty and Adamowski (2018a, 2018b) focussed on a real-world daily urban water demand (UWD) forecasting experiment in Montreal, Canada. Daily UWD forecasting in Montreal is very important as the municipality's water supply system provides drinking water to over 1.3 million people and is subject to changing demands due to seasonal usage. The period where UWD in Montreal is most variable occurs during the summer when water demand is heavily dependent on weather (e.g., lawn watering, outdoor recreational activities, construction, etc. (Toth et al., 2018)). The available dataset consists of average daily historical UWD, daily maximum air temperature, daily rainfall, and daily antecedent precipitation index collected from both the City of Montreal (UWD) and Environment and Canada (meteorological data) and covers the period from February 1999 to December 2010. The lead times explored in this study are 1, 7, and 14 day(s) ahead. These lead times were selected since they are important for operations, planning, and construction activities associated with the water supply system (Quilty and Adamowski (2018a, 2018b)). For the 1 (7 and 14) day lead time(s) there were: 2395 (2389 and 2382) calibration, 583 validation, and 366 hold-out (test) records. The same validation and test records are used for each lead time forecast. The dataset partitioning follows our previous study (Quilty and Adamowski, 2018a). More information on this UWD dataset can also be found in earlier work (Adamowski et al., 2013, 2012; Mouatadid and Adamowski, 2017; Tiwari and Adamowski, 2013).

6.4.2.2. Input Data for the Ensemble Wavelet – Stochastic Data-Driven Forecasting Framework

In this study, we consider the forecasts produced by the WDDFF models from Quilty and Adamowski (2018b) that forecasted the target variable directly (i.e., the target variable was not decomposed into separate sub-series). Those forecasts are used as input data to our proposed EW-SDDFF. In this regard, we consider WDDFF models that: (1) used the MODWT and AT for wavelet decomposition; (2) considered decomposition levels between one and six and 20 different wavelet filters; (3) adopted either the partial correlation input selection (PCIS) or Edgeworth Approximations-based conditional mutual information (EA) input variable selection methods; and (4) used either multiple linear regression (MLR), extreme learning machines (ELM), or second order Volterra series models (SOV) for the data-driven method. For each WDDFF we considered two cases: (1) using only wavelet-decomposed input data or (2) using both wavelet-decomposed and the original (non-decomposed) input data as potential model inputs during input variable selection. In total, for each lead time, there were 1440 different WDDFF forecasts (scenarios). In addition to the WDDFF forecasts, we also include their non-wavelet-based counterparts, resulting in a total of 1443 forecasts that were considered as input data for the EW-SDDFF.

Following earlier studies on ensemble forecasting we removed poorly performing forecasts for inclusion in the ensemble forecast model (Lee et al., 2012; Rathinasamy et al., 2013). In order to remove poorly performing WDDFF and non-wavelet-based forecasts from the input dataset for each EW-SDDFF model, we took a practical approach. We removed any forecast from the input dataset that could not beat a random walk (RW) model on the calibration dataset. We think that this is justified as persistence-based forecast measures (such as the RW) have been used in earlier UWD forecasting studies to rationalize the adoption of more 'complicated' models (Fullerton et al., 2016; Fullerton and Cárdenas, 2016; Fullerton and Molina, 2010). We also removed any WDDFF forecasts that produced identical forecasts. For example, input variable selection was used in WDDFF to select which wavelet-decomposed input data to use in the forecast, which also included the original (non-decomposed) input data. In a number of cases, only the original (non-decomposed) input data was selected during input variable selection. Therefore, when the same (deterministic) data-driven model was used in such instances, the same calibrated parameters were estimated for the different models, resulting in the same forecast. In other words, in certain cases input variable selection within WDDFF did not select wavelet-decomposed data as useful inputs. In these cases, we kept only those forecasts that were unique and removed any other forecast that could not beat the RW benchmark on the calibration dataset.

We now discuss the different deterministic and probabilistic forecast evaluation metrics used to judge the performance of our proposed EW-SDDFF against benchmark methods.

6.4.2.3. Forecast Evaluation Metrics

We adopted the same deterministic and probabilistic performance measures as Quilty and Adamowski (2018a). The interested reader should consult our earlier work and references therein for further details on each method. The deterministic forecast evaluation metrics are: the Nash-Sutcliffe Efficiency Index (NASH); root mean square error (RMSE); and mean absolute error (MAE). The probabilistic forecast evaluation metrics: mean continuous ranked probability score (CRPS), prediction interval coverage probability (PICP), average prediction interval width (AW), and interval score (IS).

The PICP is a measure of forecast reliability while the AW in a measure of forecast sharpness. A reliable forecasting model does not always indicate useful forecasts since it can also lack sharpness and provide poor forecasts, while an unreliable model will always indicate poor forecasts (in the probabilistic sense) (Humphrey et al., 2016). Both the CRPS and IS simultaneously measure forecast sharpness and reliability (Gneiting et al., 2007; Gneiting and Raftery, 2007). However, it is not straightforward to identify the level of sharpness and/or reliability solely by these measures, which is why we also incorporated the PICP and AW. In order to calculate the PICP, AW, and IS one needs to specify the confidence level ($1 - \alpha$ %) at which they are to be evaluated. In this study, we selected a 95 % confidence level (i.e., $\alpha = 0.05$).

Additionally, we include the Coverage Probability Plots (CPP) (Laio and Tamea, 2007) to visually inspect the quality of the forecast quantiles against the theoretical quantiles. The CPP has been used for the same task in a number of other recent studies for assessing forecast quality (Humphrey et al., 2016; Montanari and Koutsoyiannis, 2012; Sikorska et al., 2015). Essentially, the CPP produces a scatter plot between theoretical and forecast quantiles and in the ideal case each point in the scatter plot lies on the bisector. Using the ideal case, one can estimate the mean square error (MSE) between the bisector line and the plotted points in the CPP. Using this concept, we adopt what we call the 'CPP_mse' as a means to summarize the CPP in terms of a MSE-based scoring metric, similar to the MSE calculated for reliability diagrams (Brochero et al., 2011b).

6.4.2.4. Developing our Proposed Method and its Comparison with Benchmarks

Our proposed EW-SDDFF was built using WDDFF forecasts (Quilty and Adamowski, 2018b) as input data. Based on our discussion in section 6.4.2.2, the WDDFF forecasts that beat the RW benchmark on the calibration dataset were considered as input data when developing the EW-SDDFF for each lead time (1,

7, and 14 days ahead). To evaluate whether each individual WDDFF was better than the RW benchmark, we used the MAE as an indicator, since it is related to the CRPS (discussed further below). We considered two different input variable selection methods in the design of each EW-SDDFF; the linear PCIS and the nonlinear EA. We then matched each input variable selection method with linear (MLR) and nonlinear (ELM and SOV) data-driven models. Recall that input variable selection within the EW-SDDFF acts as a method for ensemble model selection while the data-driven model acts as a tool for ensemble model weighting, both of which account for uncertainty in each process. In addition to ensemble model selection (input variable selection) and weighting (parameter) uncertainties, EW-SDDFF also considers model output uncertainty.

As a type of EW-SDDFF-based benchmark, we also considered the case where input variable selection *was not* performed (i.e., ensemble model selection was not considered). In that case, we used the MLR as the data-driven model (i.e., simple linear ensemble weighting). As mentioned before, other options were discarded because of computation issues. Therefore, there were four variants of the EW-SDDFF used in our case study (input variable selection method – data-driven model): 1) None-MLR; 2) PCIS-MLR; 3) EA-ELM; and 4) EA-SOV.

To evaluate the proposed EW-SDDFF we compared it against both deterministic and stochastic forecasting benchmarks. We selected the SWDDFF that was built using the WDDFF that performed best on the calibration dataset in terms of the MAE. This 'best' SWDDFF served as the main stochastic forecasting method benchmark for EW-SDDFF. The deterministic benchmarks included the RW, the WDDFF that performed best on the calibration dataset (i.e., the non-stochastic version of the 'best' SWDDFF), and the non-stochastic versions of the different EW-SDDFF approaches, which we named EW-DDFF (there were four in total). The EW-DDFF did not include input variable selection (ensemble model selection), parameter (ensemble model weighting), or model output uncertainty. Overall, we compared five different methods (EW-SDDFF, EW-DDFF, SWDDFF, WDDFF, and RW). Since EW-SDDFF and its non-stochastic counterpart (EW-DDFF) were considered using four separate variations of input variable selection methods and data-driven models (None – MLR, PCIS-MLR, EA-ELM; and EA-SOV), there was a total of ten different model combinations explored in our case study for each lead time. Therefore, there is a total of 30 models for all lead times: 1, 7, and 14 days ahead.

In order to compare deterministic and stochastic forecasts, we adopted the CRPS since it reduces to the MAE in the case of deterministic models (Hersbach, 2000). Therefore, a stochastic forecasting method

was deemed preferable to a deterministic method when the CRPS of the stochastic forecast was lower than the MAE of the deterministic forecast.

In the following section we present and discuss the results of our experiments.

6.5. Results and Discussion

We begin with a discussion on the use of the RW benchmark to select input data for the EW-SDDFF and give details on the 'best' WDDFF models that were subsequently converted to their SWDDFF versions. Following this, we compare the performance of EW-SDDFF against its stochastic (SWDDFF) and deterministic benchmarks (EW-DDFF, WDDFF, and RW). Finally, we explore the quality of the EW-SDDFF and SWDDFF models' probability-based forecasts during a heatwave that occurred in Montreal during July 2010. All results mentioned in this section are for the test set and therefore represent the generalization capabilities of the different forecasting models (EW-SDDFF, EW-DDFF, SWDDFF, WDDFF, and RW).

6.5.1. Selecting Input Data for the Ensemble Wavelet-Stochastic Data-Driven Forecasting Framework and Deterministic Benchmark Results

In Table 6.1, we have noted for each lead time: the number of WDDFF (1440 for each lead time) and non-wavelet-based forecasting models (three for each lead time) including how many unique forecasts were produced as well as the number of models that had a MAE lower than that of the RW benchmark (see section 6.4.2.2 and 6.4.2.4). The performance of the RW benchmarks in terms of MAE are also provided for both the calibration and test sets. Therefore, 441 (1), 416 (1), and 604 (1) different WDDFF and non-wavelet-based-forecasts were considered as input data for the different EW-SDDFF and EW-DDFF models.

Table 6.1. Random walk (RW) benchmark results used for identifying input data for the EW-SDDFF

Lead Time (Days)	No. Total Models	No. Unique Models	No. Models Better than RW	Calibration Set RW MAE [ML/D]	Test Set RW MAE [ML/D]
1	1443	874	442	35.135	26.363
7	1443	893	417	44.648	38.708
14	1443	938	605	56.573	50.945

Table 6.2 summarizes information regarding the best WDDFF forecasting models including their test set performance. The best WDDFF forecasting models were selected according to their calibration set MAE.

Each of the selected models incorporated a combination of both wavelet- and non-wavelet-decomposed input data. The best WDDFF models, according to the calibration set MAE, had high decomposition levels (six, five, and six) and wide wavelet filters. The wavelet filters contained fourteen, eight, and twelve wavelet coefficients for the different (1, 7, and 14 day) lead times. Since high decomposition levels and wide wavelet filters provided the best performance for the UWD time series, it is likely an indication of a process that changes across both high- and low-frequency scales and that exhibits a polynomial behaviour (Maheswaran and Khosa, 2012; Rathinasamy et al., 2014). This is a reasonable hypothesis for this time series since it has prominent weekly, bi-weekly, seasonal, and annual cycles and corresponds with seasonal cycles of air temperature and rainfall (Adamowski et al., 2013).

Table 6.2. Information and results for the best WDDFF models

	WDDFF Benchmark Model Details					Test Set Performance		
Lead Time (Days)	**Wavelet Decomposition Algorithm**	**Wavelet Decomposition Level**	**Wavelet Filter**	**IVS**	**Data-Driven Model**	**NASH**	**RMSE [ML/D]**	**MAE [ML/D]**
1	MODWT	6	la14	PCIS	MLR	0.929	27.026	18.535
7	MODWT	5	sym4	PCIS	MLR	0.724	53.418	34.679
14	AT	6	la12	EA	SOV	0.640	60.981	39.389

When comparing the RW to the WDDFF benchmarks, the biggest gain in terms of forecasting accuracy (MAE) appears at the 14 day lead time. Even at the 14 day lead time, the WDDFF provides good forecast performance as its NASH score is above 0.5 (Moriasi et al., 2007). The RW had test set NASH scores of 0.864, 0.635, and 0.439 for the 1, 7, and 14 day lead times, respectively.

6.5.2. Comparing the Proposed Model against its Benchmarks

A summary of our experiments comparing EW-SDDFF against competing methods (SWDDFF, EW-DDFF, and WDDFF) is provided in Table 6.3. The best EW-SDDFF and EW-DDFF models for a particular metric are bolded while the SWDDFF and WDDFF models that outperformed their ensemble counterparts for a particular metric are highlighted in grey. The deterministic forecast evaluation metrics for the stochastic methods (EW-SDDFF and SWDDFF) were calculated by taking the mean of $f_Q(Q)$. The forecasts for each method are compared against the observed (target) time series in Figure 6.2. Our experiments indicated that the EW-SDDFF was at least as computationally efficient as a single SWDDFF and incorporated the same information that would be needed be needed to generate an SWDDFF for each WDDFF model (to maintain brevity these results are not shown).

The most encouraging result of this experiment is that the EW-SDDFF provided the best forecast performance in terms of CRPS across all lead times when compared against its competitors. Remembering that the CRPS reduces to the MAE for deterministic forecasts, this enables us to compare the CRPS for a stochastic forecast against the MAE for a deterministic forecast (allowing us to identify if the stochastic forecast should be preferred over the deterministic forecast).

However, no single variation of EW-SDDFF (None – MLR, PCIS-MLR, EA-ELM, and EA-SOV) performed best across all lead times or forecast evaluation metrics. This indicates the importance of considering different methods for input variable selection and data-driven modelling within the EW-SDDFF. In general, it can be said that by considering each EW-SDDFF variant, the EW-SDDFF provided improved performance when compared against the SWDDFF, which is significant given that both methods are nearly equal in computational requirements (at least for this study).

Table 6.3. Test set performance for the EW-SDDFF, EW-DDFF, SWDDFF, and WDDFF models (Note: the results for best EW-SDDFF and EW-DDFF models for a particular forecast evaluation metric are bolded while the results for the SWDDFF and WDDFF models that outperformed their EW-SDDFF and EW-DDFF counterparts are highlighted in grey)

Lead Time (Days)	Forecasting Framework	IVS	Data-Driven Model	NASH	RMSE [ML/D]	MAE [ML/D]	CRPS [ML/D]	PICP [%]	AW [ML/D]	IS [ML/D]	CPP_mse
1	**WDDFF**	PCIS	MLR	0.929	27.026	18.535					
	EW-DDFF	None	MLR	**0.929**	**27.087**	19.116					
		PCIS	MLR	0.923	28.191	20.134					
		EA	ELM	0.929	27.120	**18.575**					
		EA	SOV	0.924	27.991	20.320					
	SWDDFF	PCIS	MLR	0.925	27.886	19.409	14.790	92.350	109.715	166.143	0.006
	EW-SDDFF	None	MLR	0.926	27.706	19.645	15.421	**94.809**	126.919	**160.580**	0.004
		PCIS	MLR	0.921	28.524	20.170	15.503	86.885	**96.685**	193.962	0.004
		EA	ELM	0.926	27.625	19.098	**14.651**	91.803	108.985	161.693	**0.002**
		EA	SOV	**0.928**	**27.356**	**18.858**	14.803	93.716	121.532	165.711	0.003
7	**WDDFF**	PCIS	MLR	0.724	53.418	34.679					
	EW-DDFF	None	MLR	0.684	57.183	43.314					
		PCIS	MLR	**0.713**	**54.494**	**41.312**					
		EA	ELM	0.478	73.434	65.321					
		EA	SOV	0.616	63.007	46.354					
	SWDDFF	PCIS	MLR	0.691	56.509	39.624	29.756	92.350	192.773	347.416	0.021
	EW-SDDFF	None	MLR	0.728	52.966	36.334	27.553	**92.077**	186.646	321.638	**0.004**
		PCIS	MLR	0.736	52.241	37.455	27.810	91.257	**152.883**	313.316	0.042
		EA	ELM	0.718	53.938	40.179	29.203	89.891	163.760	**308.462**	0.047
		EA	SOV	**0.736**	**52.197**	**36.069**	**27.028**	89.617	162.892	327.956	0.025

14	**WDDFF**	EA	SOV	0.640	60.981	39.389					
	EW-DDFF	None	MLR	0.043	99.451	80.338					
		PCIS	MLR	0.388	79.500	71.299					
		EA	ELM	0.294	85.437	75.095					
		EA	SOV	**0.479**	**73.394**	**59.426**					
	SWDDFF	EA	SOV	0.641	60.930	43.324	31.949	87.432	186.574	367.335	0.021
	EW-SDDFF	None	MLR	0.181	92.000	73.475	52.777	79.508	264.444	552.345	**0.011**
		PCIS	MLR	0.689	56.688	45.524	32.248	88.525	**175.975**	282.629	0.063
		EA	ELM	0.679	57.580	46.507	32.718	90.437	182.539	271.016	0.060
		EA	SOV	**0.704**	**55.291**	**43.813**	**31.101**	**91.530**	178.876	**267.488**	0.052

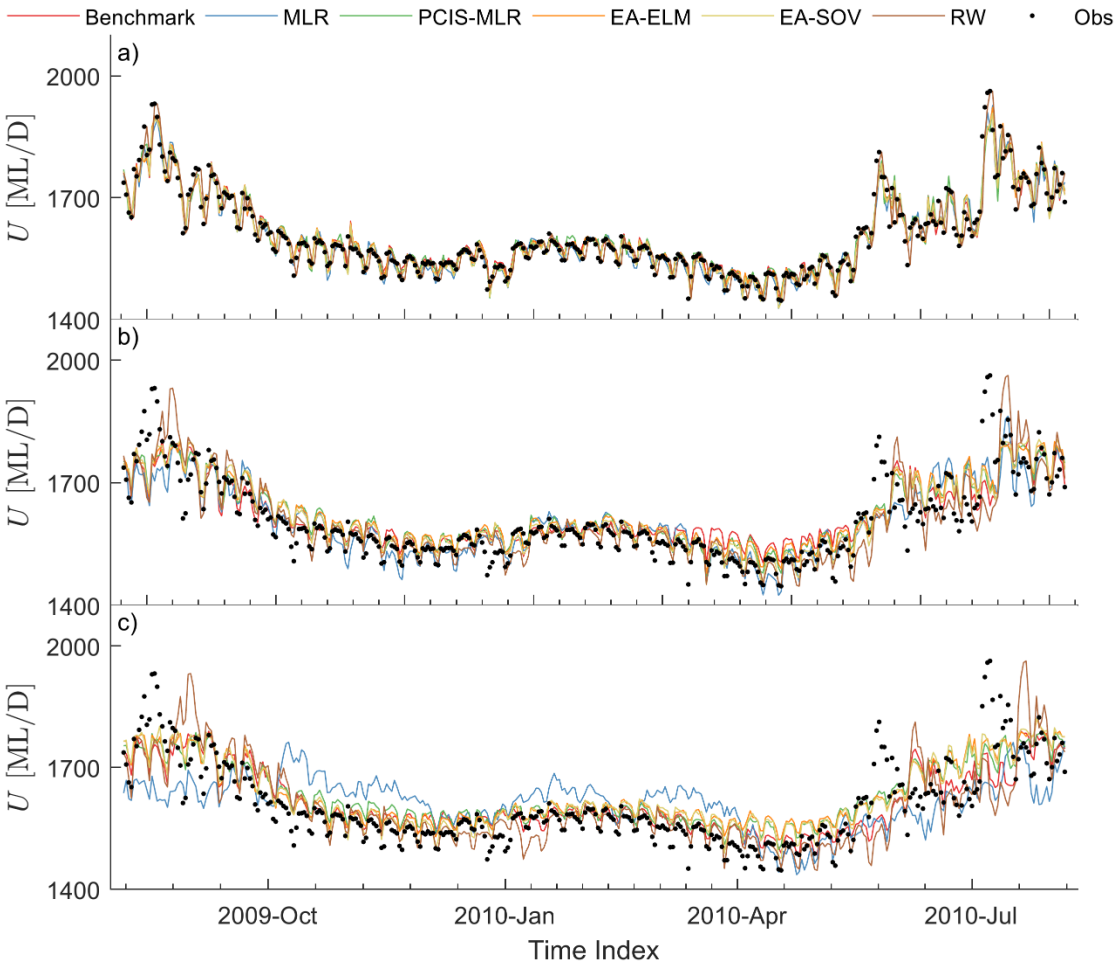

Figure 6.2. Time series plots for the benchmark (SWDDFF), EW-SDDFF variants (MLR (without input variable selection), PCIS-MLR, EA-ELM, and EA-SOV), and RW against the observed time series for a) 1, b) 7, and c) day lead time(s); the variable $\boldsymbol{U}$ denotes urban water demand measured in megalitres per day [ML/D]

6.5.2.1. Comparing the Proposed Method against its Deterministic Version

The EW-SDDFF outperformed its deterministic counterpart (EW-DDFF) when comparing the CRPS versus the MAE for each lead time. It also substantially outperformed the EW-DDFF in terms of deterministic forecast evaluation metrics at 7 and 14 day lead times, while providing similar performance for the 1 day lead time across the same metrics. These results indicate that it is very important to consider the uncertainties related to ensemble model selection and weighting in the EW-SDDFF as the lead time increases, at least for the dataset under study. For example, for the 7 and 14 day lead times the EW-

SDDFF reduced the MAE between 9 and 38 % when compared to the EW-DDFF. The improvement in performance shown by the EW-SDDFF when compared to the EW-DDFF at these longer lead times is not surprising and could be expected for a process such as UWD since it is intimately tied to meteorological conditions that can change quite suddenly over the course of 7 and 14 days. Therefore, by taking into account the variability in the different WDDFF forecasts via ensemble model selection and weighting uncertainties, the EW-SDDFF combines and appropriately weights their individual strengths. Such advantages (of the EW-SDDFF) are not present in the EW-DDFF nor the SWDDFF, since the former does not include ensemble model selection and weighing uncertainty and the latter only includes stochastic perturbations stemming from a single (wavelet-based) forecasting model (i.e., that incorporates only a single wavelet decomposition algorithm, decomposition level, wavelet filter, input variable selection method, and data-driven model). Therefore, both EW-DDFF and SWDDFF lack the variability in their input data that is required to improve UWD forecasting performance at longer lead times. However, since the EW-SDDFF contains such information as input data, it is able to provide better longer-lead time forecasts than its benchmarks.

6.5.2.2. Comparing the Proposed Method against its Stochastic Benchmark

When comparing the EW-SDDFF against its stochastic benchmark (SWDDFF), we notice that, with the exception of the PICP and the MAE at the 7 and 14 day lead times, respectively, at least one of the EW-SDDFF variants (None-MLR, PCIS-MLR, EA-ELM, and EA-SOV) outperformed the SWDDFF across the remaining forecast evaluation metrics. In most cases, the EW-SDDFF provided sharper forecasts (see the AW metric) and was nearly as, or more, reliable than the SWDDFF for a number of EW-SDDFF variants. This can be seen for instance by comparing the EA-SOV variant of the EW-SDDFF against the SWDDFF at the 14 day lead time. This is a positive indication that including multiple wavelet-based forecasts in an ensemble framework can lead to more accurate and reliable probability-based forecasts than wavelet-based forecasting models that include only a single wavelet decomposition algorithm, decomposition level, wavelet filter, input variable selection method, and data-driven model.

6.5.2.2.1. Using Coverage Probability Plots to Supplement Probabilistic Forecast Evaluation Metrics

To supplement Table 6.3, we explore Coverage Probability Plots (Figure 6.3) to evaluate the quality of the forecast quantiles produced by the different EW-SDDFF variants and the SWDDFF. We notice through the CPP (Figure 6.3) that each model performed fairly well by considering the relative closeness of each model's forecasted quantiles versus theoretical quantiles to the bisector, summarized by the CPP_mse in Table 6.3. The EA-ELM variant of EW-SDDFF, deemed the best model in terms of CRPS, also provided the

best CPP_mse at the 1 day lead time. However, like the remaining models, it overestimated the UWD for the upper quantiles. The 'S' shape of the MLR variant of the EW-SDDFF (which did not consider input variable selection) is characteristic of an over-dispersed forecast (see Figure 2 and supporting text in Laio and Tamea (2007)). This was probably the result of some ensemble members that were biased towards large predictions that were not included in the remaining EW-SDDFF variants since these members were not selected during input variable selection. Interestingly, the same model provided the best PICP and IS, while the CPP_mse was lower than that of the SWDDFF. The PCIS-MLR variant of EW-SDDFF provided the sharpest forecasts (based on the AW) but also provided the highest (worst) CRPS and IS as well as the lowest PICP. However, it still had a better CPP_mse score than the benchmark (SWDDFF).

Inspecting the CPP at the 7 day lead time we see that the MLR variant of EW-SDDFF had the best CPP_mse and the sharpest forecasts according to the AW. It also maintained the second best coverage probability next to the benchmark (SWDDFF). While the remaining EW-SDDFF variants exhibited better (lower) CRPS than the SWDDFF, their CPP displayed inferior coverage probabilities at most quantiles with the exception of the EA-SOV variant of the EW-SDDFF.

Moving to the 14 day lead time CPP we notice a very interesting result for the MLR variant of the EW-SDDFF. It over-predicted the lower forecast quantiles, but was nearly perfect for the upper quantiles and provided the best CPP_mse. However, it gave considerably worse performance for the other probabilistic performance measures (CRPS, PICP, AW and IS). The remaining EW-SDDFF variants over-predicted each quantile along with the SWDDFF. Although, the different EW-SDDFF provided better or similar CRPS, sharper forecasts (AW), and substantially better interval scores.

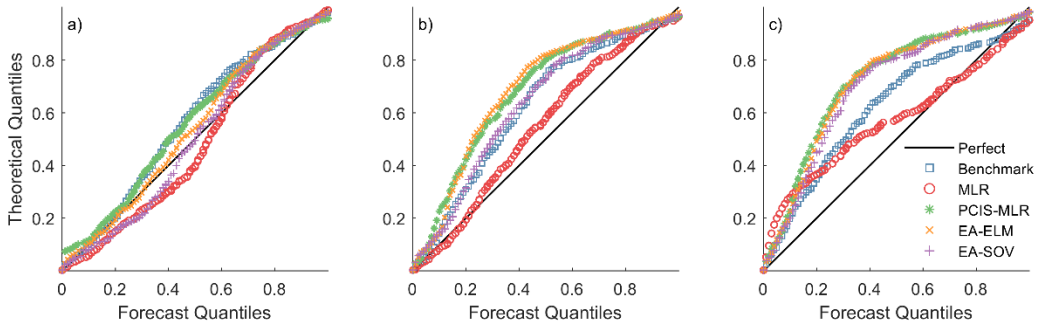

Figure 6.3. Coverage Probability Plot for the benchmark (SWDDFF) and EW-SDDFF variants (MLR (without input variable selection), PCIS-MLR, EA-ELM, and EA-SOV) for a) 1, b) 7, and c) 14 day lead time(s)

The CPP can be seen to provide a useful basis for expanding upon the 'summarized' probabilistic metrics in Table 6.3. It also highlights how each of the different EW-SDDFF variants have their own strengths and weaknesses, again supporting the practice of exploring different input variable selection and data-driven models in the EW-SDDFF.

6.5.2.2.2. The Effect of Ensemble Size on Performance

The various probability density functions required in the development of the EW-SDDFF and SWDDFF were estimated using 100 bootstrap resamples. In order to assess the effect of ensemble size (n) on generating sharp and reliable forecast probability densities $f_Q(Q)$, we explored the performance of different probabilistic forecast evaluation metrics (i.e., CRPS, AW, IS, and CPP_mse) as a function of ensemble size (i.e., number of stochastic resamples (n) used to infer $f_Q(Q)$). We note that the MLR variant for the EW-SDDFF at the 14 day lead time is not included in most plots (i.e., CRPS, AW, and IS) due to its substantially inferior performance. The interested reader can find a study on the effect of ensemble size on the CRPS in Ferro et al. (2008).

Figure 6.4 shows the CRPS as a function of ensemble size for 1 (a), 7 (b), and 14 (c) day lead times. In general, we notice that for each lead time an 'elbow' appears in each plot around 10 members, with decreasing CRPS that tends to stabilize around $n = 80$. Similar to the CRPS, an elbow in the AW versus ensemble size plot (Figure 6.5) appears around 15-20 members for the different lead times. The PICP displayed very similar characteristics to the AW and for that reason it is not included. Figure 6.6 shows the IS versus ensemble size. Note that for this plot we only consider ensemble sizes from five to 100 ($n = 5{:}100$) to improve interpretability. Studying the IS as a function of ensemble size, we notice that each method requires a relatively high ensemble size before stable performance is achieved (at least $n = 50$ for most of the methods and lead times). Similar to the IS, the CPP_mse versus ensemble size plots (Figure 6.7) also indicate that relatively high ensemble sizes are needed to stabilize performance. Since the CPP_mse requires the estimation of empirical quantiles (similar to the IS), a sufficiently large ensemble size is needed to get accurate estimates of the empirical quantiles. Figure 6.6 and 6.7 suggest that ensemble sizes of around 50 members is required for most lead times and methods in order to achieve stable IS and CPP_mse scores. Furthermore, both the IS (Figure 6.6) and CPP_mse (Figure 6.7) versus ensemble size plots provide strong evidence that at most ensemble sizes at least one of the EW-SDDFF variants outperform the benchmark (SWDDFF).

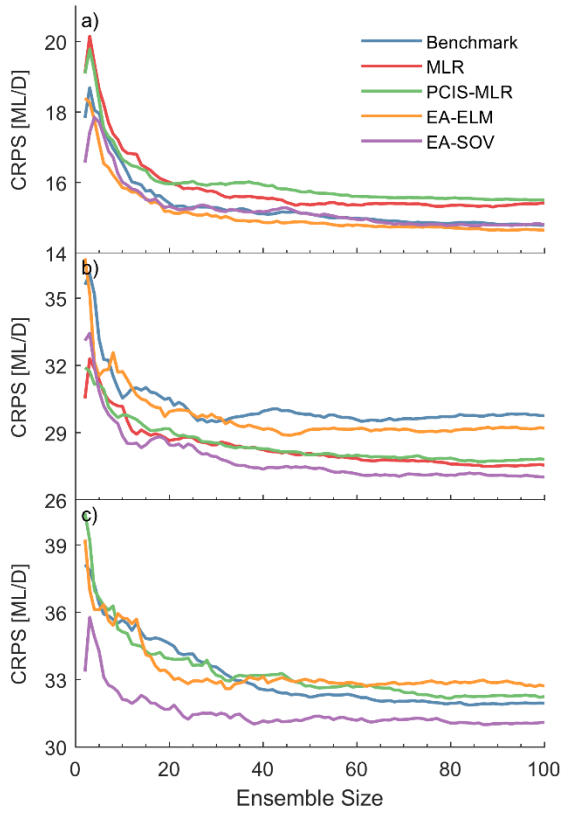

Figure 6.4. CRPS versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead time(s)

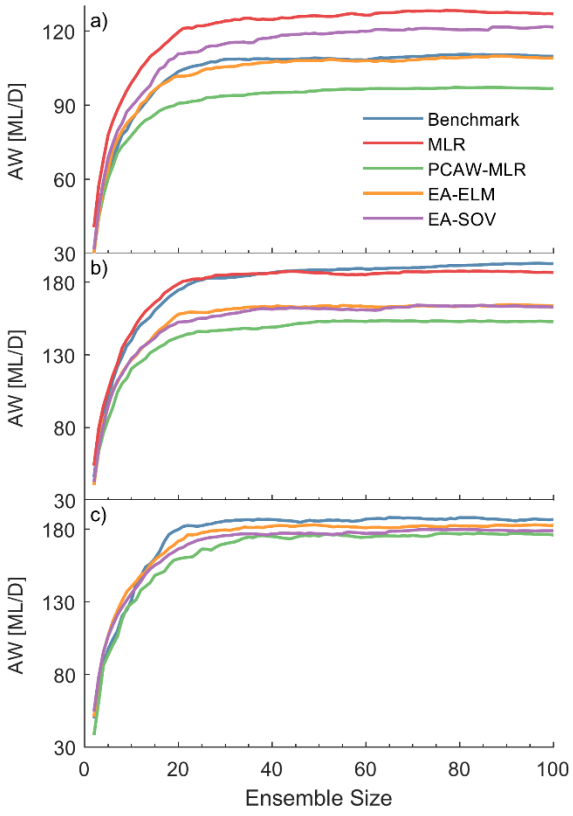

Figure 6.5. AW versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead time(s)

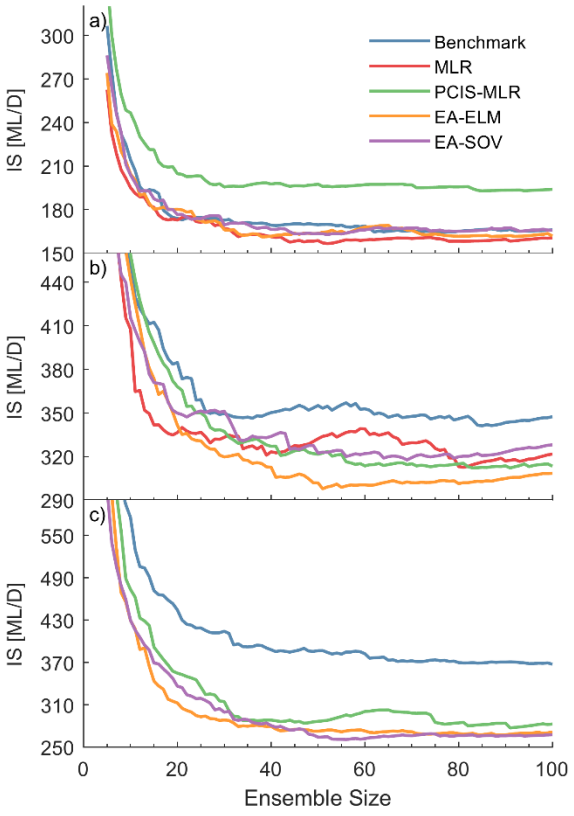

Figure 6.6. IS versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead time(s)

By reviewing the different probabilistic forecast evaluation metrics as a function of ensemble size we cannot identify an optimal ensemble size according to all metrics. However, we can establish that for each metric separately, with the exception of CPP_mse at the 1 day lead time for the benchmark (SWDDFF), performance has stabilized before $n = 100$, justifying the selection of this ensemble size in our experiments. It is very important to note that we did not vary the number of bootstrap samples used to estimate the different probability densities and therefore the characteristics of the plots shown here may differ for a lesser or greater number of bootstrap resamples. It is outside the scope of this work to study the effect of bootstrap size on estimating the different probability density functions; however, this could be an interesting study in the future. We also note that for each performance metric, the best performing model was the EW-SDDFF and the best EW-SDDFF tended to outperform the benchmark (SWDDFF) across each ensemble size with the exception of the CPP_mse for very low ensemble sizes ($n < 10$). It could be argued that performance across the different models had yet to stabilize at this point.

Finally, corroborating our results in Table 6.3, there was not a single EW-SDDFF variant that performed best across each metric or lead time, demonstrating the importance of exploring different input variable selection (ensemble member selection) and data-driven models (ensemble weighting methods) within our proposed approach.

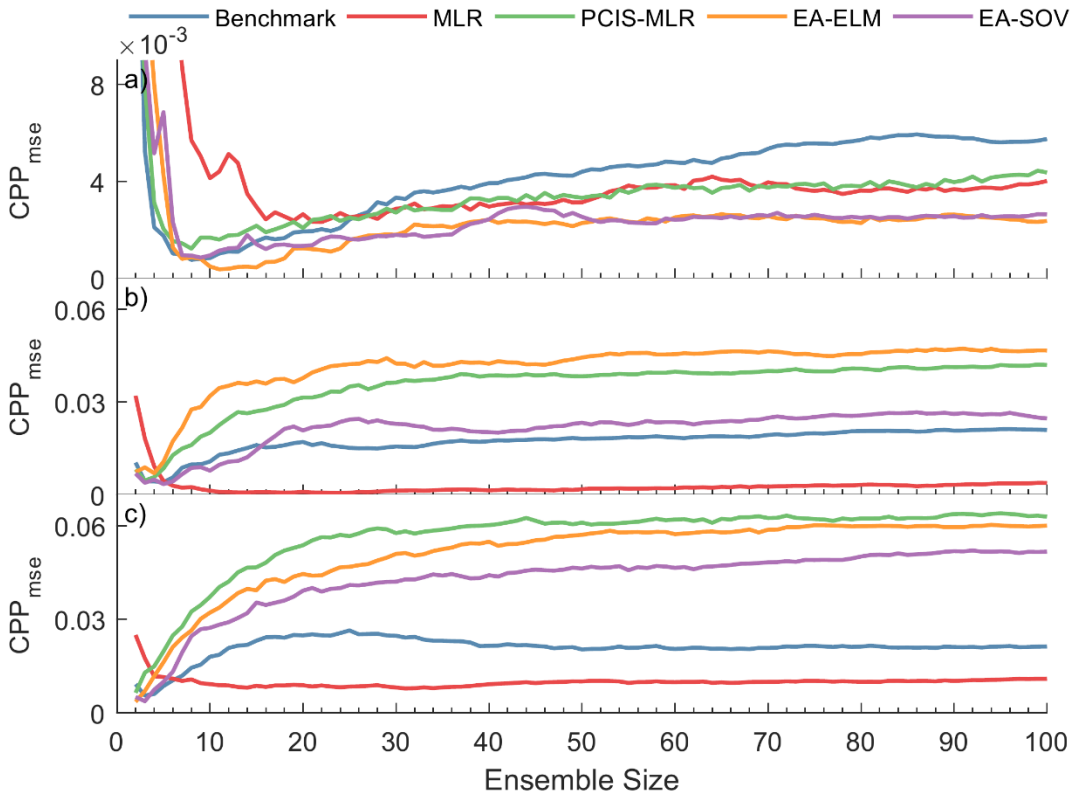

Figure 6.7. CPP_mse versus Ensemble Size ($\boldsymbol{n}$) for a) 1, b) 7, and c) 14 day lead times

We now move on to an evaluation of the different methods concerning the forecasting of UWD during the July 2010 heatwave in Montreal (Bustinza et al., 2013; Price et al., 2013). The reaction of UWD to a heatwave is a very useful test case for any water supply system since this type of scenario could be expected to substantially increase demand on the system. Interestingly, the study of the impacts of heatwaves on UWD is a research topic that has received very little attention (Hatvani-Kovacs et al., 2016a, 2016b). This study is the first to test different deterministic and stochastic forecasting methods for this task.

6.5.3. Stochastic Urban Water Demand Forecasting During a Heatwave

Starting on July 5, 2010 and lasting five days, Montreal (Canada) experienced a heat wave where the maximum temperature reached 33.7 °C three days into the heatwave with the minimum temperature staying above 25.4 °C during the same period (Bustinza et al., 2013; Price et al., 2013). We now study the performance of the EW-SDDFF by comparing it against the SWDDFF during the period leading up to, during, and after the heatwave (July 2 – 18, 2010) in order to further demonstrate the usefulness of our proposed approach. In Figure 6.2, one can see that during this period the different deterministic models provide adequate performance for the 1 day lead time. However, at 7 and 14 day lead times they are quite poor. In Figure 6.8 we compare the mean forecast and its 95 % prediction intervals for the SWDDFF and the EW-SDDFF model that provided the best CRPS score over the entire test set for the 1, 7, and 14 day lead times (EA-ELM, EA-SOV, and EA-SOV, respectively). Figure 6.9 is the same as Figure 6.8, with the exception that at the 1 and 7 day lead times the best EW-SDDFF (MLR in both cases) is different when considering the models with the best CRPS during the heatwave period. Our Supplementary Material [appendix] contains the same figures but for each EW-SDDFF variant plotted separately. Table 6.4 includes the same forecast evaluation metrics as Table 6.3, but instead considers the heat wave period.

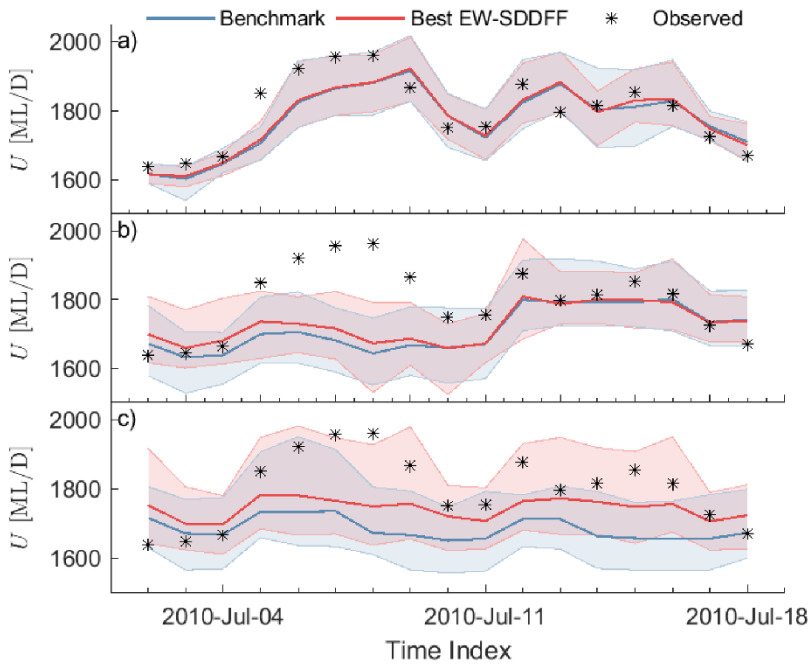

Figure 6.8. Mean forecasts and their 95 % prediction intervals for the SWDDFF and the EW-SDDFF for the a) 1 (EA-ELM), b) 7 (EA-SOV), and c) 14 (EA-SOV) day lead time(s)

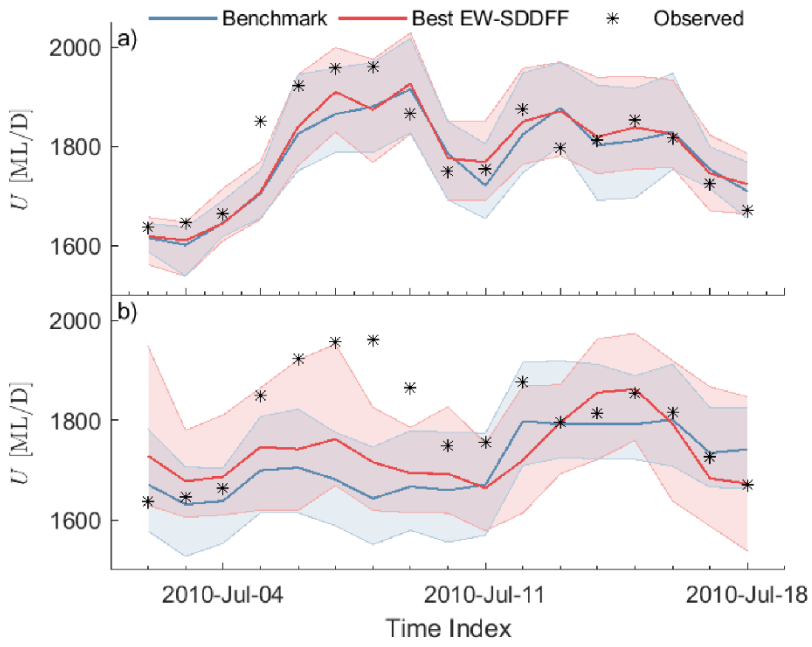

Figure 6.9. Mean forecasts and their 95 % prediction intervals for the SWDDFF and the EW-SDDFF for the a) 1 (MLR) and b) 7 (MLR) day lead time(s)

For each lead time, the EW-SDDFF provided better performance than the SWDDFF across each forecast evaluation metric with the exception of the PICP at the 7 day lead time. In that case, both methods had the same performance. At the 1 day lead time, both methods provided satisfactory performance in terms of deterministic and probabilistic forecast evaluation metrics. The deterministic performance severely degraded for the 7 and 14 day lead times, which is not surprising given the rapidity of the onset of the heatwave and the lack of such information as input to the different forecasting models. However, in terms of probabilistic performance, especially in terms of the PICP, the EW-SDDFF performed quite well even at 7 and 14 day lead times. They provided substantially better performance than the SWDDFF. It is interesting to note that the EW-SDDFF at the 14 day lead time had better performance than at the 7 day lead time (with the exception of the CPP_mse). This could be due to the magnitude of the prediction errors over the validation set, which is used within the stochastic framework to draw from the model error conditional probability density function. As is evident from Figure 6.2, the UWD in Montreal has a prominent weekly cycle and therefore there is a high correlation between UWD at time t and $t + 7$. The magnitude of errors at the 7 day lead time are smaller than at the 14 day lead time. Thus, at the 14 day lead time there is a higher probability of drawing a larger value from the model error conditional

probability density function than at the 7 day lead time. This could be the reason for the improved coverage probability at the 14 day lead time. However, at an aggregated scale, when comparing the best EW-SDDFF, the 7 day lead time had lower overall errors between the forecasted quantiles and the theoretical quantiles as indicated by the CPP_mse.

We believe that the reason for the EW-SDDFF providing superior performance than the SWDDFF during the heatwave period is that the EW-SDDFF combines forecasts that have quite different characteristics. Those forecasts were produced using different wavelet decomposition algorithms, decomposition levels, wavelet filters, input variable selection methods, and data-driven models. This variety of characteristics increased the diversity in the generated forecast probability density function compared to the SWDDFF (which relies only on a single wavelet decomposition algorithm, decomposition level, wavelet filter, input variable selection method, and data-driven model).

Even though the deterministic forecasting performance of the EW-SDDFF was quite low for 7 and 14 day lead times, the EW-SDDFF still has skill when considering forecast in terms of probabilities (e.g., compare the MAE against the CRPS). This supports the potential usefulness of stochastic forecasting models even when their deterministic forecasting performance is low. Besides, during a heatwave, it is logical to deduce that an urban water supply system will be more closely monitored, and with such strong performance as that provided by the EW-SDDFF at one day lead times, suitable adjustments to the operation of pumps and reservoirs can be made in sufficient time (i.e., to ensure adequate demand is available at a suitable pressure). However, reliable 7 and 14 day lead times could be valuable tools for planning the purchase of additional supplies in water scarce areas, exploring water demand reduction programs, or making adjustments to construction and maintenance projects planned prior to the heatwave.

Table 6.4. Test set performance for the EW-SDDFF, EW-DDFF, SWDDFF, and WDDFF models leading up to, during, and after the July 2010 heatwave (July 3 – 17, 2010) (Note: the results for best forecast models for a particular forecast evaluation metric are bolded)

Lead Time (Days)	Forecasting Framework	IVS	Data-Driven Model	NASH	RMSE [ML/D]	MAE [ML/D]	CRPS [ML/D]	PICP [%]	AW [ML/D]	IS [ML/D]	CPP_mse
1	**SWDDFF**	PCIS	MLR	0.622	62.786	52.048	36.763	82.353	151.913	413.889	0.033
	EW-SDDFF	None	MLR	**0.699**	**56.022**	**43.287**	**30.994**	**94.118**	160.304	351.695	**0.004**
		PCIS	MLR	0.620	62.969	52.330	39.664	64.706	**127.725**	368.327	0.033
		EA	ELM	0.659	59.640	48.935	34.885	88.235	136.381	**349.202**	0.024
		EA	SOV	0.670	58.647	48.396	34.181	88.235	148.340	363.879	0.019
7	**SWDDFF**	PCIS	MLR	-0.810	137.443	98.374	82.634	**70.588**	189.820	1669.746	0.076
	EW-SDDFF	None	MLR	**-0.251**	**114.253**	**86.345**	**69.325**	64.706	238.104	**786.056**	**0.035**
		PCIS	MLR	-0.520	125.941	95.040	77.685	58.824	**169.263**	1657.527	0.067
		EA	ELM	-0.369	119.551	91.125	73.686	**70.588**	188.520	1356.765	0.061
		EA	SOV	-0.454	123.209	90.034	74.064	58.824	184.466	1474.118	0.062
14	**SWDDFF**	EA	SOV	-1.129	149.083	126.882	106.945	52.941	**214.924**	1483.119	0.164
	EW-SDDFF	None	MLR	-1.616	165.241	137.326	109.792	64.706	293.696	1571.699	0.113
		PCIS	MLR	-0.295	116.250	96.025	75.335	64.706	232.779	586.373	0.061
		EA	ELM	-0.063	105.331	86.842	66.288	**88.235**	260.675	402.240	**0.041**
		EA	SOV	**0.021**	**101.092**	**84.253**	**62.946**	82.353	244.063	**360.424**	**0.041**

Through the results and discussion presented in this section, it is evident that the EW-SDDFF provided the best overall forecasting performance when compared against deterministic (EW-DDFF, WDDFF, and RW) and stochastic (SWDDFF) benchmarks. Perhaps the most useful feature of the EW-SDDFF is that it can incorporate multiple wavelet-based deterministic forecasts as input data. It also has the ability to include other information, such as large-scale climate indicators or numerical weather predictions and inherently accounts for the uncertainty in selecting and weighting different forecasts. Combined with model output uncertainty, this results in probability-based forecasts of the target process. This advantage of the proposed approach is different than many other ensemble forecasting methods (wavelet-based and otherwise) in hydrology and water resources that apply ensemble model selection and weighting separately (as in Rathinasamy et al. (2013); Alizadeh et al. (2017a); Barzegar et al. (2018b, 2017), Sun and Trevor (2018, 2017)).

6.6. Summary and Conclusions

Extending earlier work (Quilty and Adamowski (2018a) on a stochastic data-driven forecasting framework (SDDFF) inspired by Montanari and Koutsoyiannis (2012)), this study introduced a new *ensemble* SDDFF that uses the output of different deterministic models as input data to generate a forecast in the form of a probability density function. The different deterministic models can be process-based, data-driven, etc. Although not considered in this study, additional inputs could eventually include numerical weather predictions, large-scale climate indicators, etc. The proposed framework can explicitly account for various sources of uncertainty: input data (forecast data), input variable selection (ensemble model selection), parameter (ensemble model weighting), and model output. Additionally, other sources of uncertainty (e.g., initial conditions) could be explicitly defined within the framework in future studies. To the best of our knowledge, this is the first study to introduce an ensemble stochastic data-driven forecasting framework that inherently accounts for ensemble model selection and weighting uncertainties. The former can be estimated using a variety of different input variable selection methods and the latter by data-driven models.

Data-driven wavelet-based forecasting is a growing field in hydrology and water resources. However, despite the popularity of wavelet-based data-driven forecasting in hydrology and water resources (Afan et al., 2016; Dixit et al., 2016; Nourani et al., 2014; Sang, 2013), only a small number of studies have considered multiple wavelet-based forecasts (i.e., derived using different wavelet decomposition algorithms, decomposition levels, wavelet filters, etc.) in an ensemble forecasting model (Alizadeh et al., 2017a; Barzegar et al., 2018a, 2017; Rathinasamy et al., 2013). Furthermore, none exist that consider an

ensemble stochastic data-driven forecasting framework based on wavelets. To address these gaps and demonstrate the usefulness of our proposed framework, we used the large number of deterministic forecasts derived from our previous study (Quilty and Adamowski, 2018b) as input data to the method proposed herein. This results in the Ensemble Wavelet-SDDFF (EW-SDDFF). For the same case study as in Quilty and Adamowski (2018a, 2018b), we used EW-SDDFF to forecast urban water demand (UWD) in Montreal, Canada at several lead times. This forecasting system was then compared against several benchmarks, including its non-stochastic version (EW-DDFF), the WDDFF, and its stochastic version SWDDFF (Quilty and Adamowski, 2018a). We also used EW-SDDFF to incorporate the same information (i.e., the forecasts produced by the different WDDFF models) more efficiently (than an ensemble SWDDFF). Another important benefit of adopting EW-SDDFF instead of an ensemble of SWDDFF models is that it inherently accounts for both ensemble model selection and weighting uncertainties that would have to be done separately for an ensemble of SWDDFF models.

By exploring different input variable selection methods and data-driven models for ensemble model selection and weighting, respectively, within EW-SDDFF, we were able to demonstrate across a wide number of deterministic and probabilistic forecast evaluation metrics that EW-SDDFF led to more accurate and reliable forecasts than the 'best' single SWDDFF. Furthermore, for a historical heatwave that occurred during July 2010 and caused very high UWD in Montreal, our EW-SDDFF provided strong probabilistic forecast performance, even as far into the future as 14 days ahead. It also significantly outperformed its SWDDFF counterpart. During this period, different variants of EW-SDDFF (MLR, PCIS-MLR, EA-ELM, and EA-SOV) proved to be instrumental in achieving a high level of performance (as some performed better at different lead times). This finding highlights the importance of considering different input variable selection methods and data-driven models when accounting for ensemble model selection and weighting within EW-SDDFF for our case study.

Our results indicate that EW-SDDFF and the ensemble SDDFF, in general, are promising forecasting frameworks for hydrology and water resources. However, our study was limited in several ways: 1) only a small number of input variable selection methods (PCIS and EA) and data-driven models (MLR, ELM, and SOV) were included; 2) we did not consider input data uncertainty; and 3) we only used a single case study. These limitations are reasonable for this study as our focus was on introducing the ensemble SDDFF and demonstrating its use for the relatively under-explored topic of multi-wavelet ensemble forecasting, and the unexplored topic of ensemble wavelet-based stochastic data-driven forecasting. However, there is

much more to be explored by considering the ensemble SDDFF as a general method in ensemble hydrological and water resources forecasting. Some future research topics include:

1. Using the outputs of process-based and data-driven models alongside additional process-related information (e.g., large-scale climate indicators and numerical weather predictions) for forecasting different hydrological and water resources processes such as streamflow, drought, evaporation, etc.
2. Exploring the impact of bootstrap resample size for both estimating and sampling from the different probability density functions in terms of forecasting performance (in this study we only considered the impact of bootstrap resampling size on forecasting performance);
3. Testing the suitability of different methods (i.e., in addition to the bootstrap) for generating the various probability density functions (e.g., input data, input variable selection, parameters, and model output); and
4. Considering a wider range of input variable selection methods and data-driven models for ensemble model selection and weighting, respectively.

6.7. Acknowledgements

We thank the City of Montreal for providing the urban water demand data used in this study. Requests for urban water demand and meteorological data used in this study should be made through the City of Montreal or obtained from Environment Canada at climate.weather.gc.ca, respectively. Partial funding for this research was provided by NSERC Discovery and Accelerator grants held by Jan Adamowski.

6.8. References

Adamowski, J., Adamowski, K., Prokoph, A., 2013. A Spectral Analysis Based Methodology to Detect Climatological Influences on Daily Urban Water Demand. Math. Geosci. doi:10.1007/s11004-012-9427-0

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 48, W01528. doi:10.1029/2010WR009945

Adhikari, R., Verma, G., Khandelwal, I., 2015. A model ranking based selective ensemble approach for time series forecasting. Procedia Comput. Sci. 48, 14–21. doi:10.1016/j.procs.2015.04.104

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. doi:10.1016/j.jhydrol.2016.07.048

Alizadeh, M.J., Jafari Nodoushan, E., Kalarestaghi, N., Chau, K.W., 2017a. Toward multi-day-ahead forecasting of suspended sediment concentration using ensemble models. Environ. Sci. Pollut. Res. 28017–28025. doi:10.1007/s11356-017-0405-4

Alizadeh, M.J., Kavianpour, M.R., Kisi, O., Nourani, V., 2017b. A new approach for simulating and forecasting the rainfall-runoff process within the next two months. J. Hydrol. 548, 588–597. doi:10.1016/j.jhydrol.2017.03.032

Aussem, A., Campbell, J., Murtagh, F., 1998. Wavelet-based feature extraction and decomposition strategies for financial forecasting. J. Comput. Intell. Financ. 6, 5–12.

Barzegar, R., Asghari Moghaddam, A., Adamowski, J., Ozga-Zielinski, B., 2018a. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Environ. Res. Risk Assess. 32, 799–813. doi:10.1007/s00477-017-1394-z

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189

Barzegar, R., Moghaddam, A.A., Deo, R., Fijani, E., Tziritis, E., 2018b. Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. Sci. Total Environ. 621, 697–712. doi:10.1016/j.scitotenv.2017.11.185

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172–173. doi:10.1016/j.atmosres.2015.12.017

Bogner, K., Pappenberger, F., 2011. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. Water Resour. Res. doi:10.1029/2010WR009137

Brochero, D., Anctil, F., Gagné, C., 2011a. Simplifying a hydrological ensemble prediction system with a backward greedy selection of members – Part 1: Optimization criteria. Hydrol. Earth Syst. Sci. 15, 3307–3325. doi:10.5194/hess-15-3307-2011

Brochero, D., Anctil, F., Gagné, C., 2011b. Simplifying a hydrological ensemble prediction system with a backward greedy selection of members - Part 2: Generalization in time and space. Hydrol. Earth Syst. Sci. 15, 3327–3341. doi:10.5194/hess-15-3327-2011

Bustinza, R., Lebel, G., Gosselin, P., Bélanger, D., Chebana, F., 2013. Health impacts of the July 2010 heat wave in Québec, Canada. BMC Public Health 13, 56. doi:10.1186/1471-2458-13-56

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. J. Hydrol. 375, 613–626. doi:10.1016/j.jhydrol.2009.06.005

Dixit, P., Londhe, S., Deo, M.C., 2016. Review of Applications of Neuro-Wavelet Techniques in Water Flows. Ina. Lett. 1, 99–104. doi:10.1007/s41403-016-0015-3

Doycheva, K., Horn, G., Koch, C., Schumann, A., König, M., 2017. Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning. Adv. Eng. Informatics 33, 427–439. doi:10.1016/j.aei.2016.11.001

Efron, B., Tibshirani, R.J.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, London, U.K.

Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor. Appl. Climatol. 128, 875–903. doi:10.1007/s00704-016-1735-8

Farmer, W.H., Vogel, R.M., 2016. On the deterministic and stochastic use of hydrologic models. Water Resour. Res. 52, 5619–5633. doi:10.1002/2016WR019129

Fernando, A.K., Shamseldin, A.Y., Abrahart, R.J., 2012. Use of Gene Expression Programming for Multimodel Combination of Rainfall-Runoff Models. J. Hydrol. Eng. 17, 975–985. doi:10.1061/(ASCE)HE.1943-5584.0000533

Ferro, C.A.T., Richardson, D.S., Weigel, A.P., 2008. On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorol. Appl. 15, 19–24. doi:10.1002/met.45

Fullerton, T.M., Cárdenas, J.P., 2016. Forecasting water demand in Phoenix, Ariz. J. Am. Water Works Assoc. 108, E533–E545. doi:10.5942/jawwa.2016.108.0156

Fullerton, T.M., Molina, A.L., 2010. Municipal water consumption forecast accuracy. Water Resour. Res. 46, 1–8. doi:10.1029/2009WR008450

Fullerton, T.M., Walke, A.G., Asefa, T., 2016. Short-run water demand forecast accuracy for the tampa bay area. J. Am. Water Works Assoc. 108, 126–136. doi:10.5942/jawwa.2016.108.0011

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat. Methodol. 69, 243–268. doi:10.1111/j.1467-9868.2007.00587.x

Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. J. Am. Stat. Assoc. 102, 359–378. doi:10.1198/016214506000001437

Han, S., Coulibaly, P., 2017. Bayesian flood forecasting methods: A review. J. Hydrol. 551, 340–351. doi:10.1016/j.jhydrol.2017.06.004

Hatvani-Kovacs, G., Belusko, M., Pockett, J., Boland, J., 2016a. Assessment of Heatwave Impacts. Procedia Eng. 169, 316–323. doi:10.1016/j.proeng.2016.10.039

Hatvani-Kovacs, G., Belusko, M., Skinner, N., Pockett, J., Boland, J., 2016b. Heat stress risk and resilience in the urban environment. Sustain. Cities Soc. 26, 278–288. doi:10.1016/j.scs.2016.06.019

He, X., Guan, H., Qin, J., 2015. A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. J. Hydrol. 527, 88–100. doi:10.1016/j.jhydrol.2015.04.047

Hemri, S., Lisniak, D., Klein, B., 2015. Multivariate postprocessing techniques for probabilistic hydrological forecasting. Water Resour. Res. 51, 7436–7451. doi:10.1002/2014WR016473

Herger, N., Abramowitz, G., Knutti, R., Angélil, O., Lehmann, K., Sanderson, B.M., 2018. Selecting a climate model subset to optimise key ensemble properties. Earth Syst. Dyn. 9, 135–151. doi:10.5194/esd-9-135-2018

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather Forecast. 15, 559–570. doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Humphrey, G.B., Gibbs, M.S., Dandy, G.C., Maier, H.R., 2016. A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. J. Hydrol. 540, 623–640. doi:10.1016/j.jhydrol.2016.06.026

Khalil, B., Broda, S., Adamowski, J., Ozga-Zielinski, B., Donohoe, A., 2015. Short-term forecasting of groundwater levels under conditions of mine-tailings recharge using wavelet ensemble neural

network models. Hydrogeol. J. 23, 121–141. doi:10.1007/s10040-014-1204-3

Krause, P., Boyle, D.P., 2005. Advances in Geosciences Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97. doi:10.5194/adgeo-5-89-2005

Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology. J. Hydrol. 249, 2–9. doi:10.1016/S0022-1694(01)00420-6

Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrol. Earth Syst. Sci. 11, 1267–1277. doi:10.5194/hess-11-1267-2007

Lee, J.A., Kolczynski, W.C., McCandless, T.C., Haupt, S.E., 2012. An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Wind Prediction. Mon. Weather Rev. 140, 2270–2286. doi:10.1175/MWR-D-11-00065.1

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z., 2017. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. Wiley Interdiscip. Rev. Water 4, e1246. doi:10.1002/wat2.1246

Liu, Z., Zhou, P., Zhang, Y., 2015. A Probabilistic Wavelet–Support Vector Regression Model for Streamflow Forecasting with Rainfall and Climate Information Input*. J. Hydrometeorol. 16, 2209–2229. doi:10.1175/JHM-D-14-0210.1

Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

May, R.J., Maier, H.R., Dandy, G.C., Fernando, T.M.K.G., 2008. Non-linear variable selection for artificial neural networks using partial mutual information. Environ. Model. Softw. 23, 1312–1326. doi:10.1016/j.envsoft.2008.03.007

Montanari, A., Koutsoyiannis, D., 2014. Reply to comment by Grey Nearing on “A blueprint for process-based modeling of uncertain hydrological systems.” Water Resour. Res. 50, 6264–6268. doi:10.1002/2013WR014987

Montanari, A., Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. Water Resour. Res. 48. doi:10.1029/2011WR011412

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50,

885–900. doi:10.13031/2013.23153

Mouatadid, S., Adamowski, J., 2017. Using extreme learning machines for short-term urban water demand forecasting. Urban Water J. 14, 630–638. doi:10.1080/1573062X.2016.1236133

Noori, N., Kalin, L., 2016. Coupling SWAT and ANN models for enhanced daily streamflow prediction. J. Hydrol. 533, 141–151. doi:10.1016/j.jhydrol.2015.11.050

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. J. Hydrol. 514, 358–377. doi:10.1016/j.jhydrol.2014.03.057

Olhede, S., Walden, A.T., 2004. The Hilbert spectrum via wavelet projections. Proc. R. Soc. A Math. Phys. Eng. Sci. 460, 955–975. doi:10.1098/rspa.2003.1199

Pappenberger, F., Ramos, M.H., Cloke, H.L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., Salamon, P., 2015. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. J. Hydrol. 522, 697–713. doi:10.1016/j.jhydrol.2015.01.024

Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis, 1st ed, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289. doi:10.1016/S0022-1694(03)00225-7

Phanida, P., Y., S.A., Keith, A., 2016. Multimodel Approach Using Neural Networks and Symbolic Regression to Combine the Estimated Discharges of Rainfall-Runoff Models. J. Hydrol. Eng. 21, 4016022. doi:10.1061/(ASCE)HE.1943-5584.0001332

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ANN-based stream flow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos. Res. 197, 42–63. doi:10.1016/j.atmosres.2017.06.014

Price, K., Perron, S., King, N., 2013. Implementation of the Montreal Heat Response Plan During the 2010 Heat Wave. Can J Public Heal. 104, 96--100.

Quilty, J., Adamowski, J., 2018a. A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes. Submitted to J. Hydrol.

Quilty, J., Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J. Hydrol. 563, 336 – 353. doi:10.1016/j.jhydrol.2018.05.003

Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. Water Resour. Res. 52, 2299–2326. doi:10.1002/2015WR016959

Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. J. Hydrol. doi:10.1016/j.jhydrol.2013.09.025

Rathinasamy, M., Khosa, R., Adamowski, J., Ch, S., Partheepan, G., Anand, J., Narsimlu, B., 2014. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. Water Resour. Res. 50, 9721–9737. doi:10.1002/2013WR014650

Sang, Y.-F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. 122, 8–15. doi:10.1016/j.atmosres.2012.11.003

Shamseldin, A.Y., O'Connor, K.M., Liang, G.C., 1997. Methods for combining the outputs of different rainfall-runoff models. J. Hydrol. 197, 203–229. doi:10.1016/S0022-1694(96)03259-3

Shoaib, M., Shamseldin, A.Y., Khan, S., Khan, M.M., Khan, Z.M., Melville, B.W., 2018. A wavelet based approach for combining the outputs of different rainfall–runoff models. Stoch. Environ. Res. Risk Assess. 32, 155–168. doi:10.1007/s00477-016-1364-x

Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2015. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. J. Hydrol. Eng. 20.

Silvestro, F., Rebora, N., Cummings, G., Ferraris, L., 2017. Experiences of dealing with flash floods using an ensemble hydrological nowcasting chain: implications of communication, accessibility and distribution of the results. J. Flood Risk Manag. 10, 446–462. doi:10.1111/jfr3.12161

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinformatics 10, 3–22. doi:10.2166/hydro.2008.015

Sun, W., Trevor, B., 2018. Multiple model combination methods for annual maximum water level prediction during river ice breakup. Hydrol. Process. 32, 421–435. doi:10.1002/hyp.11429

Sun, W., Trevor, B., 2017. Combining k-nearest-neighbor models for annual peak breakup flow forecasting. Cold Reg. Sci. Technol. 143, 59–69. doi:10.1016/j.coldregions.2017.08.009

Tapiador, F.J., Gallardo, C., 2006. Entropy-based member selection in a GCM ensemble forecasting. Geophys. Res. Lett. 33, 2–5. doi:10.1029/2005GL024888

Thiboult, A., Anctil, F., Ramos, M.H., 2017. How does the quantification of uncertainties affect the quality and value of flood early warning systems? J. Hydrol. 551, 365–373. doi:10.1016/j.jhydrol.2017.05.014

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. doi:10.1002/wrcr.20517

Toth, E., Bragalli, C., Neri, M., 2018. Assessing the significance of tourism and climate on residential water demand: Panel-data analysis and non-linear modelling of monthly water consumptions. Environ. Model. Softw. 103, 52–61. doi:10.1016/j.envsoft.2018.01.011

Tran, H.D., Muttil, N., Perera, B.J.C., 2016. Enhancing accuracy of autoregressive time series forecasting with input selection and wavelet transformation. J. Hydroinformatics. doi:10.2166/hydro.2016.145

Tyralis, H., Koutsoyiannis, D., 2017. On the prediction of persistent processes using the output of deterministic models. Hydrol. Sci. J. 62, 2083–2102. doi:10.1080/02626667.2017.1361535

Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. Proc. Natl. Acad. Sci. 104, 708–711. doi:10.1073/pnas.0610471104

Wan, C., Xu, Z., Pinson, P., Dong, Z.Y., Wong, K.P., 2014. Probabilistic forecasting of wind power generation using extreme learning machine. IEEE Trans. Power Syst. 29, 1033–1044. doi:10.1109/TPWRS.2013.2287871

Wani, O., Beckers, J.V.L., Weerts, A.H., Solomatine, D.P., 2017. Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting. Hydrol. Earth Syst. Sci. 21, 4021–4036. doi:10.5194/hess-21-4021-2017

Weijs, S. V., van de Giesen, N., 2013. An information-theoretical perspective on weighted ensemble forecasts. J. Hydrol. 498, 177–190. doi:10.1016/j.jhydrol.2013.06.033

Young, C.-C., Liu, W.-C., Chung, C.-E., 2015. Genetic algorithm and fuzzy neural networks combined with

the hydrological modeling system for forecasting watershed runoff discharge. Neural Comput. Appl. 26, 1631–1643. doi:10.1007/s00521-015-1832-0

Yung, B.B., Tolson, B.A., Burn, D.H., 2011. Risk assessment of a water supply system under climate variability: a stochastic approach. Can. J. Civ. Eng. 38, 252–262. doi:10.1139/L10-132

Zeng, X., Wu, J., Wang, D., Zhu, X., Long, Y., 2016. Assessing Bayesian model averaging uncertainty of groundwater modeling based on information entropy method. J. Hydrol. 538, 689–704. doi:10.1016/j.jhydrol.2016.04.038

Appendix

This appendix contains online supplementary material related to Chapter 6.

6. A.1. Different Variants of Figure 6.8 in Text

This supporting information provides different versions of Figure 6.8 (in-text) for different variants of the Ensemble Wavelet-Stochastic Data-Driven Forecasting Framework (EW-SDDFF).

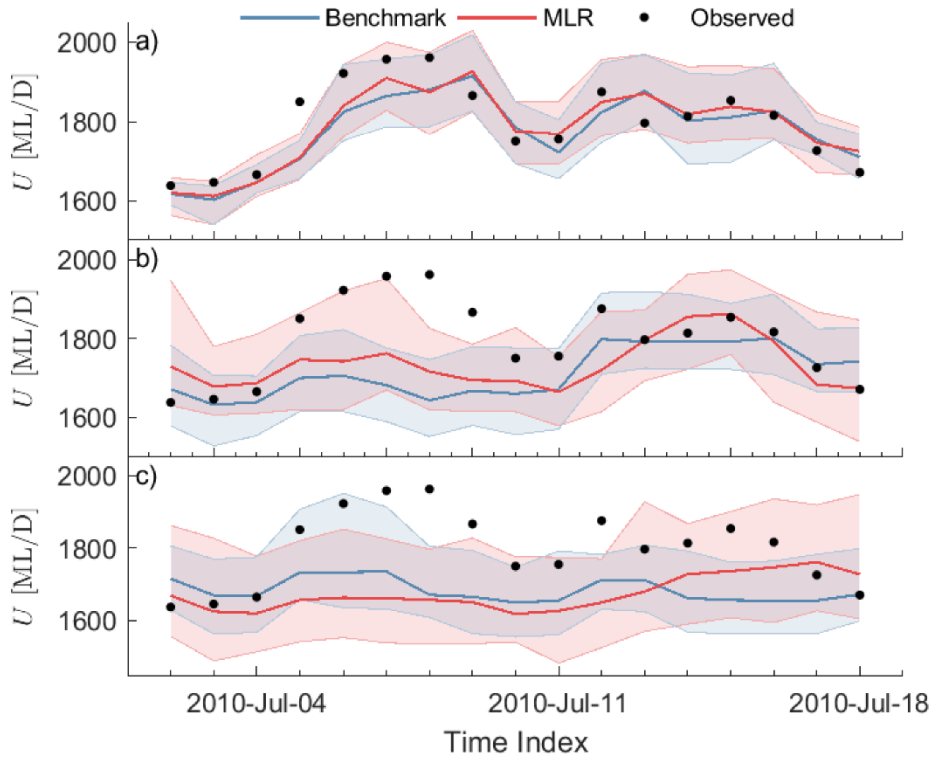

Figure 6.A.1. Mean forecasts and their 95 % prediction intervals for the SWDDFF (benchmark) and the MLR variant of EW-SDDFF for the a) 1, b) 7, and c) 14 day lead time(s) (see Figure 6.8 in text)

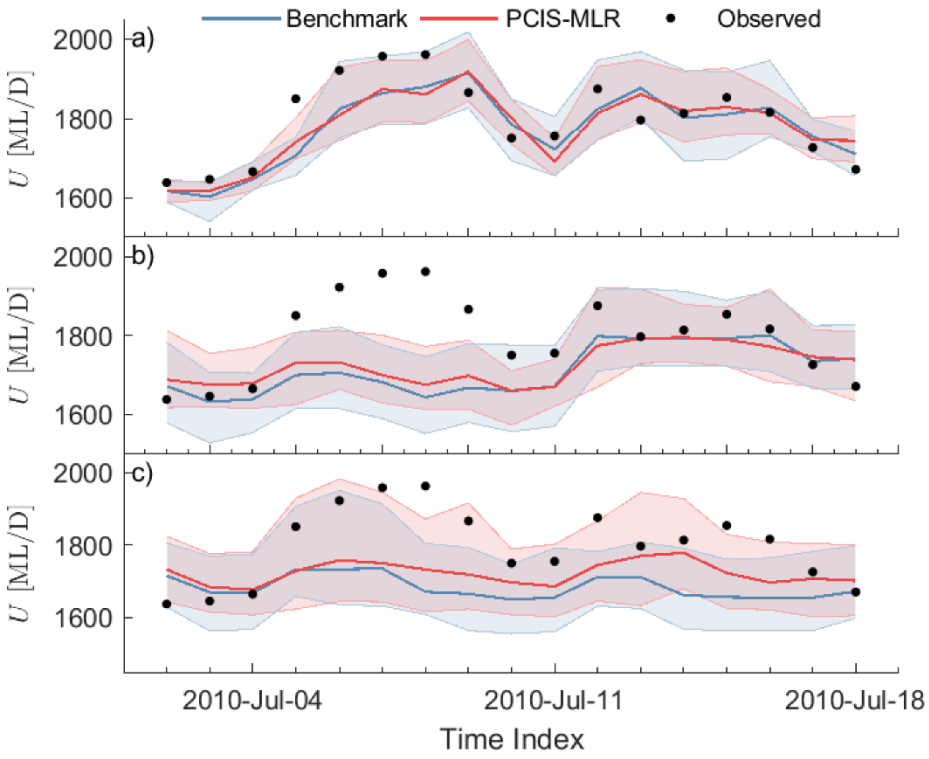

Figure 6.A.2. Mean forecasts and their 95 % prediction intervals for the SWDDFF (benchmark) and the PCIS-MLR variant of EW-SDDFF for the a) 1, b) 7, and c) 14 day lead time(s) (see Figure 6.8 in text)

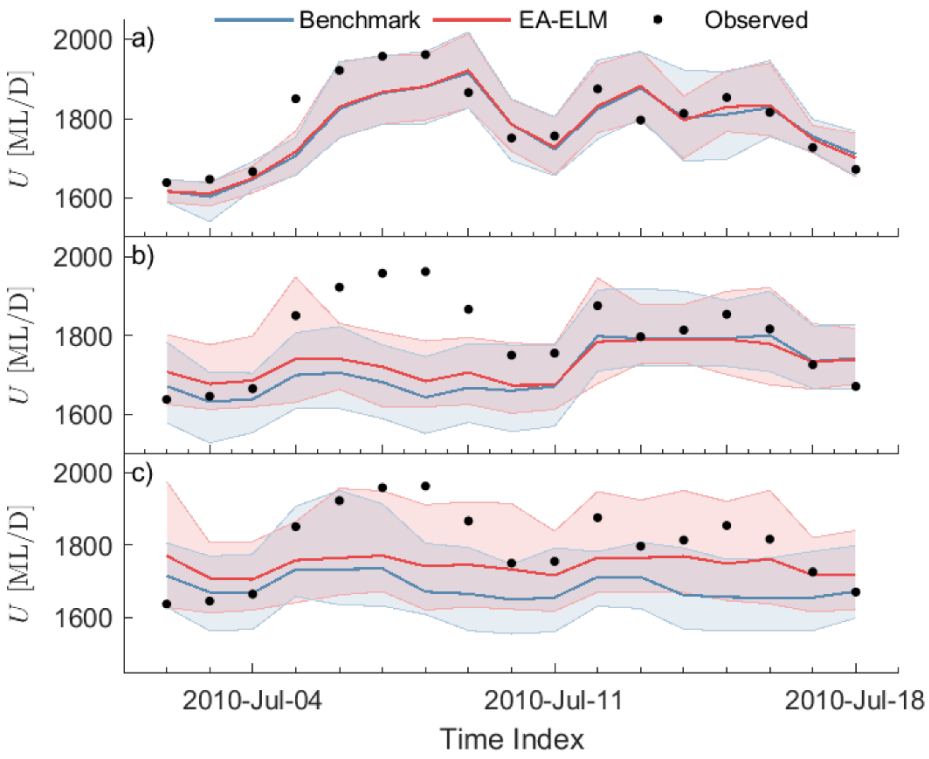

Figure 6.A.3. Mean forecasts and their 95 % prediction intervals for the SWDDFF (benchmark) and the EA-ELM variant of EW-SDDFF for the a) 1, b) 7, and c) 14 day lead time(s) (see Figure 6.8 in text)

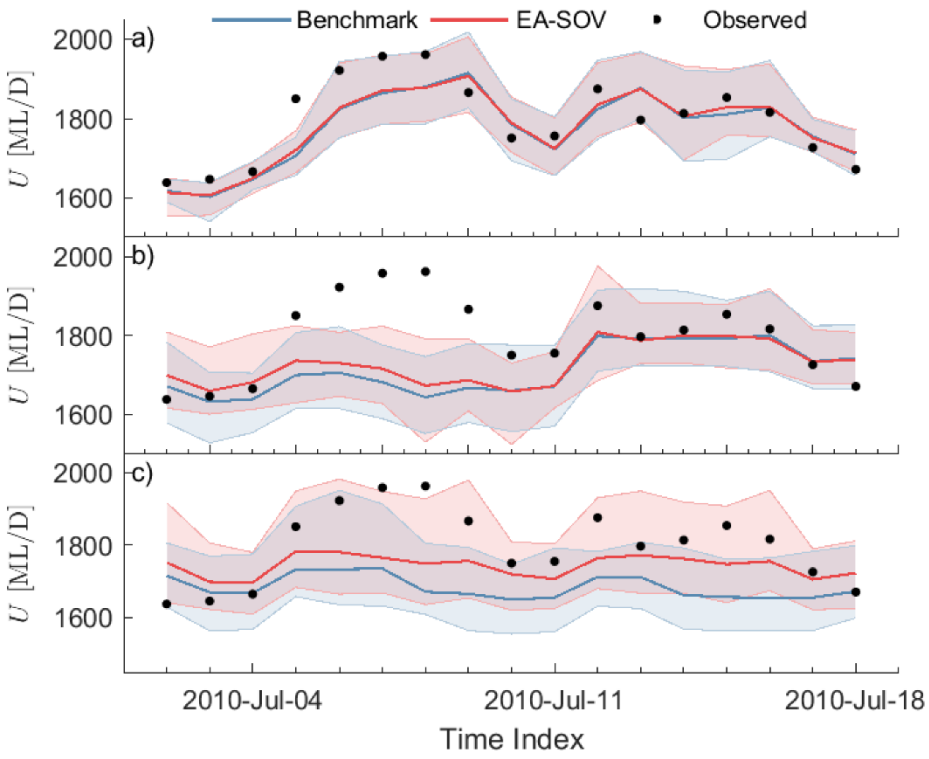

Figure 6.A.4. Mean forecasts and their 95 % prediction intervals for the SWDDFF (benchmark) and the EA-SOV variant of EW-SDDFF for the a) 1, b) 7, and c) 14 day lead time(s) (see Figure 6.8 in text)

Chapter 7: Summary and Conclusions

The main objective of this research was to develop, test, and apply a new ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) for real-world water resources forecasting applications. The main goal of this thesis was to demonstrate that EW-SDDFF can be used to generate accurate and reliable probability-based forecasts, which are very useful for supporting operational, management, and planning tasks commonly encountered by water resources managers. To demonstrate the usefulness of the method, EW-SDDFF was applied for the task of daily urban water demand forecasting in the City of Montreal, Quebec and was shown to significantly outperform benchmark methods in terms of accuracy and reliability, especially during a heatwave that occurred in July, 2010.

EW-SDDFF was primarily motivated by the nonlinear, multiscale, and uncertain nature of water resources that creates significant challenges for data-driven forecasting methods and often hampers their accuracy and reliability. A second motivating factor behind the development of EW-SDDFF was related to the incorrect development of the majority of wavelet-based forecasting models in hydrology and water resources that has led to erroneous forecasts that cannot be used properly in real-world forecasting applications. EW-SDDFF provided solutions to each of these obstacles. In particular, EW-SDDFF addressed:

1. Nonlinearity, through the use of new computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods, developed in this research, and fed selected inputs to nonlinear data-driven forecasting methods.
2. Multiscale change, by using wavelet transforms (a useful tool for time-frequency localization) that were adopted within a new wavelet-based data-driven forecasting framework (WDDFF) incorporating a set of best (correct) practices that addressed the incorrect development of wavelet-based forecasting models (which is prevalent within water resources forecasting). Additionally, WDDFF is able to be used with any input variable selection and data-driven forecasting method.
3. Uncertainty, by using stochastics to:
    a. Transform the WDDFF into a new wavelet-based stochastic data-driven forecasting framework (SWDDFF) that accounted for different sources of uncertainty (input data, input variable selection, parameter, and model output) and resulted in probability-based forecasts.

b. To take advantage of the strengths of different wavelet transforms, an ensemble multi-wavelet version of the single-wavelet framework (SWDDFF) was created (EW-SDDFF), accounting for ensemble model selection and weighting uncertainties (along with input data and ensemble model output uncertainties), and was shown to provide significantly more accurate and reliable (probability-based) forecasts than SWDDFF.

This research was divided into four main parts (each resulting in peer-reviewed journal manuscripts) based on the four developments mentioned above. In the next four sub-sections, a summary and set of conclusions is given for each of the four components of this research.

7.1. Computationally Efficient, Non-parametric, Nonlinear Information-Theoretic Input Variable Selection Methods

Two new computationally efficient, non-parametric, nonlinear information-theoretic input variable selection methods were developed to overcome the drawbacks of existing nonlinear information-theoretic methods that require careful parametric tuning that is often computationally prohibitive. The first method is the Edgeworth Approximations-based conditional mutual information (EA) approach. By coupling input variable selection uncertainty (through bootstrap resampling) with the EA method, the bootstrap rank-ordering conditional mutual information (broCMI) approach was developed. The EA and broCMI approaches were compared against existing methods on several synthetic, partially-synthetic, and a real-world urban water demand forecasting experiment. The main findings of this study are:

1. The EA method provided similar or better performance than existing nonlinear information-theoretic input variable selection methods but had the benefit of being much more computationally efficient since it did not require the tuning of parameters.
2. By including bootstrap rank-ordering, broCMI was able to provide substantially better and more robust input variable selection accuracy and forecasting model performance when compared against its EA counterpart (and other parametric nonlinear methods).
3. Bootstrap rank-ordering is a very simple procedure for incorporating uncertainty assessment in the input variable selection procedure and can be applied to any input variable selection method.

7.2. Best (Correct) Practices for Wavelet-based Forecasting and the Wavelet Data-Driven Forecasting Framework

In order to address an important gap in the literature, the first set of best (correct) practices for wavelet-based forecasting models for real-world applications was developed. This set of best practices addressed

deficiencies with many incorrect wavelet-based forecasting models reported in the literature and provided a step-by-step guide that can be followed in order to develop correct wavelet-based forecasting models that can be applied for real-world water resources forecasting scenarios. The best practices were then formed into a general wavelet-based data-driven forecasting framework (WDDFF) that can be used with both the maximal overlap discrete wavelet transform (MODWT) and *à trous* algorithm (AT), any input variable selection method, and any data-driven model for forecasting a target (water resources) process. The WDDFF was used with different: wavelet transforms (MODWT and AT), decomposition levels and wavelet filters, input variable selection methods, and data-driven models for a daily urban water demand forecasting experiment in Montreal, Quebec where multiple forecast lead times were considered. For the same case study, the WDDFF was also compared with a commonly adopted and incorrect wavelet-based forecasting method (that incorporated the MODWT multiresolution analysis (MODWT-MRA)). The main conclusions of this study are:

1. The WDDFF, which adopted best (correct) practices for wavelet-based forecasting, provided realistic and correct forecast performance while the MODWT-MRA (a method based on the incorrect usage of the wavelet transform for forecasting purposes) provided unrealistic forecast performance that demonstrated its incorrectness.
2. WDDFF provided significantly different forecasting performance depending on the wavelet transform (MODWT or AT), decomposition level, and wavelet filter that was used in the forecasting model. This is useful because different decomposition levels and wavelet filters are better suited to capturing different phenomena and therefore multiple WDDFF forecasts could be combined in an ensemble framework to improve forecast performance.
3. WDDFF provided significantly better performance than benchmark methods that did not include the use of wavelets.

7.3. Stochastic Wavelet Data-Driven Forecasting Framework

To provide a holistic means for uncertainty assessment in the WDDFF, a stochastic framework was adopted. The stochastic framework allowed for the estimation of uncertainties related to input data, input variable selection, parameter, and model output to be incorporated in a stochastic version of WDDFF (i.e., SWDDFF) and resulted in probability-based forecasts. The SWDDFF was implemented using three different settings that considered varying sources of uncertainty: 1) parameter; 2) parameter and model output; and 3) input variable selection, parameter, and model output. SWDDFF was compared against a) its non-wavelet-based version (i.e., SDDFF) which also considered the same varying levels of

uncertainty and b) deterministic benchmarks. Similar to the WDDFF, each of these methods were tested on the daily urban water demand dataset in Montreal. The main results of this study indicated that:

1. In all cases, including input variable selection, parameter, and model output uncertainty in SWDDFF and SDDFF provided the most accurate and reliable forecasts when compared against those forecasts that only included parameter or parameter and model output uncertainties.
2. In all cases, wavelet transformation was shown to improve forecasting accuracy and reliability.
3. Benchmark methods that did not include uncertainty assessment or only various sources of uncertainty provided biased performance.

7.4. Ensemble Wavelet – Data-Driven Forecasting Framework

An ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) was developed to extend its single-wavelet counterpart (SWDDFF) in order to improve forecast accuracy and reliability. This was done by using multiple forecasts generated by the WDDFF as input data. The WDDFF forecasts vary according to the different wavelet transforms (MODWT and AT), decomposition levels, wavelet filters, input variable selection methods, and data-driven models used in their production. Therefore, each of the different WDDFF forecasts displayed similar or complementary strengths in capturing different phenomena in the studied process. The EW-SDDFF was able to take advantage of the varying strengths of the different WDDFF models by using input variable selection and data-driven methods to perform ensemble model selection and weighting, respectively. For the same urban water demand forecasting case study in Montreal (used to demonstrate the useful of WDDFF and SWDDFF), the EW-SDDFF was compared against its single-wavelet counterpart (SWDDFF), it's non-stochastic counterpart (EW-DDFF), WDDFF, and a random walk (RW) benchmark. Different variants of EW-SDDFF and EW-DDFF were considered based on varying the input variable selection and data-driven methods used for ensemble model selection and weighting. EW-SDDFF considered ensemble model selection, weighting, and model output uncertainties (while EW-DDFF did not consider uncertainty), SWDDFF considered input variable selection, parameter, and model output uncertainty (while WDDFF did not consider uncertainty). Based on the comparison of EW-SDDFF variants and its benchmark methods (SWDDFF, EW-DDFF, WDDFF, and RW), the conclusions drawn from this study are as follows:

1. The EW-SDDFF provided the best performance (in terms of forecast accuracy and reliability) when compared against the different benchmarks, including its single-wavelet counterpart, SWDDFF, highlighting the usefulness of the multi-wavelet ensemble framework.

2. EW-SDDFF provided exceptionally better performance than SWDDFF during the July, 2010 heatwave in Montreal. Therefore, EW-SDDFF could be a very useful tool for water resources managers for planning the purchase of additional water supplies, implementing water restrictions, or making adjustments to construction and/or maintenance projects leading up to, during, and after heatwave periods.
3. Different variants of EW-SDDFF provided different levels of forecast accuracy and reliability and no single method performed best across each forecast lead time. This highlights the usefulness of exploring different input variable selection methods and data-driven models for ensemble model selection and weighting, respectively.

The EW-SDDFF is a very useful tool that can be applied towards the operation, management, and planning of water resources systems due to its ability to address the nonlinear, multiscale, and uncertain nature of water resources in a holistic manner, resulting in probability-based forecasts that can provide valuable information to water resources managers, especially during decision-making stages.

Chapter 8: Contributions to Knowledge, Limitations, and Recommendations for Further Research

A new ensemble multi-wavelet stochastic data-driven forecasting framework (EW-SDDFF) was developed, tested, and applied for the purpose of generating probability-based forecasts of water resources. Firstly, contributions to knowledge stemming from this research are given. Secondly, limitations concerning the application of the main methods developed in this work are discussed. Lastly, avenues for future research are noted. These research ideas may be useful in addressing some of the noted limitations and may serve as motivation for the study of new topics related to the methods discussed herein.

8.1. Contributions to Knowledge

1. Two new computationally efficient, non-parametric, nonlinear information-theoretic approaches for input variable selection (i.e., Edgeworth Approximations-based conditional mutual information (EA) and bootstrap rank-ordered conditional mutual information (broCMI)) have been developed. The EA and broCMI (a version of the EA method that accounts for input variable selection uncertainty) provide similar or better performance than existing nonlinear information-theoretic input variable selection methods that are computationally expensive (due to the need to optimize parameters and their settings).
2. A new procedure for assessing input variable selection uncertainty (bootstrap rank-ordering) was developed and shown to improve input variable selection accuracy and robustness. The new procedure can be used to assess input variable selection uncertainty for any input variable selection method.
3. The majority of wavelet-based forecasting models developed for water resources forecasting are incorrect and cannot be used properly in real-world applications. To address this significant issue, a set of best (correct) practices for using wavelet-based forecasting models for real-world applications were developed, addressing earlier methodological problems that led to incorrect wavelet-based forecasts.
4. The best practices for wavelet-based forecasting were adopted in a new wavelet-based data-driven forecasting framework (WDDFF) that can be applied for real-world forecasting applications. WDDFF can be used with different wavelet transforms (maximal overlap discrete wavelet transform (MODWT) and *à trous* algorithm (AT)), any input variable selection method (e.g., EA and broCMI), and any data-driven forecasting method (e.g., multiple linear regression,

artificial neural networks, etc.). This method is useful for addressing nonlinearity and multiscale changes in water resources forecasting applications.

5. A new stochastic wavelet-based data-driven forecasting framework (SWDDFF) was developed using stochastics (that makes use of bootstrap resampling) to quantify uncertainty in the WDDFF. SWDDFF considers input data, input variable selection, parameter, and model output uncertainties and results in forecasts in the form of a probability density function. This framework is useful for holistically addressing nonlinearity, multiscale change, and uncertainty in water resources forecasting applications.
6. A new ensemble multi-wavelet data-driven forecasting framework (EW-SDDFF) was proposed to improve forecast accuracy and reliability in relation to its single-wavelet counterpart (SWDDFF). EW-SDDFF takes advantage of the strengths of multiple wavelet-based forecasts (that may vary according to wavelet transform (MODWT or AT), input variable selection method, and data-driven model) by using different WDDFF forecasts as input data. EW-SDDFF considers ensemble model selection and weighting uncertainties, in addition to input data and ensemble model output uncertainties. EW-SDDFF is very useful as it has nearly the same computational requirements as its single-wavelet counterpart (EW-SDDFF), yet provides improved forecast accuracy and reliability. Similar to SWDDFF, EW-SDDFF results in probability-based forecasts and holistically addresses nonlinearity, multiscale change, and uncertainty in water resources forecasting applications.

8.2. Limitations

The present research has several limitations:

1. While the new input variable selection algorithms (EA and broCMI) were tested on a number of synthetic, partially-synthetic, and real-world problems, WDDFF, SWDDFF, and EW-SDDFF were only tested using a single study site (i.e., for forecasting daily urban water demand within the City of Montreal's urban water supply system). This was deemed acceptable for this research as the intention was to develop, test, and apply these new methods (WDDFF, SWDDFF, and EW-SDDFF) to a case study familiar to the author (see section 8.3 for an idea to address this limitation). Therefore, it should be noted that while the WDDFF, SWDDFF, and EW-SDDFF provided accurate and reliable forecasts for the study site explored in this research, its performance at other sites or for other water resources processes may differ significantly. Notwithstanding, it is believed that due to the generality of the proposed frameworks (WDDFF, SWDDFF, and EW-SDDFF) they

have the potential to provide accurate and reliable forecasts for different water resources processes exhibiting nonlinear, multiscale, and uncertain properties.

2. Both the SWDDFF and EW-SDDFF rely on the estimation of different probability density functions relating to input data, input variable selection, parameter, and model output uncertainties (see Chapter 5 and 6). In this research, the bootstrap method was used for this since it is a very simple, intuitive, and widely-used approach in water resources forecasting for uncertainty quantification. However, different uncertainty estimation methods (e.g., Bayesian or evolutionary approaches) may be adopted instead and may lead to different levels of forecast accuracy and reliability.
3. Multiscale change can be accounted for using methods other than the wavelet transform (e.g., singular spectrum analysis) and such methods may lead to different levels of forecast accuracy and reliability. The wavelet transform was adopted in this research since it is growing in popularity within water resources forecasting applications and has been used incorrectly in the vast majority of cases.
4. This research only adopted two nonlinear data-driven forecasting methods (extreme learning machine and second order Volterra series models). These methods were adopted since they are simple to use, are computationally efficient, and have performed well in other studies familiar to the author. However, different nonlinear data-driven methods may provide different levels of forecast accuracy and reliability.

8.3. Recommendations for Further Research

1. Forecasting different water resources (e.g., streamflow, rainfall, evaporation, groundwater, etc.) using the WDDFF, SWDDFF, and EW-SDDFF should be explored in order to understand the usefulness of these approaches for forecasting a wide range of processes important to water resources managers.
2. To highlight the importance of the best (correct) practices for wavelet-based forecasting introduced in this thesis and to explore their usefulness in real-world forecasting applications: the WDDFF, SWDDFF, and EW-SDDFF should be compared against incorrect wavelet-based forecasting approaches proposed in the literature for a number of different case studies involving operational (e.g., reservoir operation) and decision-making (e.g., whether to impose water use restrictions during a drought) tasks in order to demonstrate how incorrect wavelet-based forecasts can cause unrealistically high forecast performance that can lead to non-optimal water resources system operations and poor decisions while WDDFF, SWDDFF, and EW-SDDFF can

provide correct forecasts that can improve operational efficiencies and lead to beneficial decisions in water resources system management. This is an important area to study as the vast majority of wavelet-based forecasting approaches proposed in the literature are incorrect and (outside of this research) little work has been done to highlight and correct these deficiencies. Furthermore, no work has been done in terms of identifying the risks posed by adopting incorrect wavelet-based forecasts in terms of water resources system operations and decision-making scenarios.

3. Different input variable selection methods and data-driven models should be considered within WDDFF, SWDDFF, and EW-SDDFF in order to gauge whether improvements can be made in forecasting accuracy and/or reliability. For example, input variable selection based on random forests, which are very useful at identifying all relevant input variables in very large datasets, can be considered to explore the use of randomly selected features, in contrast to the filter-based approaches studied in this thesis (that identified the best inputs one input at a time and sometimes missed out on identifying all relevant variables due to an overly simplistic input variable selection stopping criteria). New data-driven methods that can rapidly calibrate model parameters such as the 'No-prop' class of neural networks, deep-learning approaches such as convolutional neural networks (that can be used to forecast time-synchronous datasets when a large amount of data is available) or deep representation extreme learning machines (that are useful when forecasting shorter datasets), have not been explored in any detail within water resources forecasting and should therefore be explored in order to assess potential performance improvements (within WDDFF, SWDDFF, and EW-SDDFF).
4. Testing different approaches (i.e., other than the bootstrap) for estimating the sources uncertainty in SWDDFF and EW-SDDFF (i.e., input data, input variable selection, parameter, and model output uncertainties) would be very useful and could form the basis of identifying the best methods for estimating a particular source of uncertainty. For example, the restricted Boltzmann machine might be used for estimating input data uncertainty; the block-bootstrap or maximum entropy bootstrap may be used for estimating input variable selection uncertainty; parameter uncertainty could be estimated using the Differential Evolution Adaptive Metropolis algorithm; and model output uncertainty can be estimated using Distribution Element Trees.
5. In the EW-SDDFF approach, multiple WDDFF-based forecasts were used as input data. Since data-driven methods are flexible in what can be used as input data, it would be interesting to include the outputs of different types of (physically-based, conceptual, or data-driven) forecast models

and other information sources (such as numerical weather predictions or large-scale climate indicators, etc.) to assess whether forecasting accuracy and reliability can be improved within EW-SDDFF.

6. The WDDFF, SWDDFF, and EW-SDDFF can be modified by using singular spectrum analysis instead of wavelets for the purpose of capturing multiscale change(s). It would be interesting to assess the differences in forecast accuracy and reliability that could be achieved when using singular spectrum analysis instead of wavelets.