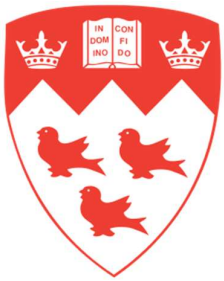
Identification of multiple-input, linear and nonlinear, time-varying systems and binary response systems for biomedical applications.

Kyriaki Kostoglou



Department of Electrical & Computer Engineering
McGill University
Montreal, Canada
August 2017

A thesis submitted to the Department of Electrical and Computer Engineering in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Acknowledgements

My deep and sincere gratitude goes first to my supervisor, Professor Georgios Mitsis, for having faith and confidence in me. It has been a great privilege and honor being a PhD student in his group. His motivation and his immense knowledge have guided me in all the time of research and writing of this thesis. I greatly appreciate all his contributions to make my PhD experience more productive and I am thankful for the excellent example he has provided as a researcher, a mentor and a man of high character and integrity.

I want to thank my beloved family and especially my parents, Kostantinos and Eugenia, for supporting and encouraging me during my PhD studies. I feel indebted to them for the tremenduous sacrifices they made to ensure that I achieve my goals. I dedicate this dissertation to my younger brother and sister, Nikolaos and Sofia, and I hope we fill up, one day, our bookshelves back home with our own small achievements.

My heartful thanks to my wonderful labmates, Alba, Charalampos, Foivia, Gregory, Manolis, Maria, Michalis, Michael, Prokopis and Sara for the sleepless nights working and supporting each other. I couldn't have imagined working in a better environment.

I am grateful to my doctoral committee members, Professor Robert Kearney and Professor Yannis Psaromiligkos for their invaluable feedback and their guidance through this process. I have benefited greatly from their advice.

Finally, I would like to express my sincere appreciation to all the gifted coauthors and researchers that contributed in the succesful completion of my PhD work.

Abstract

System identification is the process of building dynamical models from measured data in order to determine and quantify the underlying relationships between them. Ideally, the resulting mathematical models ought to imitate precisely the observed behavior of the system under examination. In this doctoral dissertation, we focus on developing effective methodologies for quantifying dynamic interrelationships in physiological systems using parametric, nonparametric and connectionist approaches. Due to the complex nature of physiological functions, standard system identification methods, which usually assume linear and time-invariant interrelationships, fail. Thus, this work describes fast and reliable modeling schemes that are capable of dealing with (a) multiple input systems (b) nonlinear dynamics (c) nonstationarities in system dynamics and (d) binary responses. These schemes were applied in combination with Laguerre-Volterra (LV) models, which can capture a wide range of nonlinear dynamic input-output causal interrelationships and Multivariate Autoregressive models (MVAR), which are used to detect couplings and causality between time series. The performance of the abovementioned methodologies was assessed using both simulations and experimental data. Specifically, we examined,

- The time-varying (TV) characteristics of Cerebral Autoregulation (CA) in patients suffering from Vasovagal Syncope (VVS) during Head-Up Tilt (HUT) testing.
- Exercise-induced cardiovascular and cerebrovascular changes in healthy subjects and stroke survivors.
- Neuronal responses to subthalamic nucleus (STN) Local Field Potentials (LFP) in Parkinson's Disease (PD) patients undergoing Deep Brain Stimulation (DBS).

Résumé

L'identification des systèmes est le processus de construction de modèles dynamiques à partir des données mesurées afin de déterminer et de quantifier les relations sous-jacentes entre eux. Idéalement, les modèles mathématiques qui en résultent doivent imiter précisément le comportement observé du système examiné. Dans cette thèse de doctorat, nous nous efforçons d'élaborer des méthodologies efficaces pour quantifier les interrelations dynamiques dans les systèmes physiologiques à l'aide d'approches paramétriques, non paramétriques et connexionnistes. En raison de la nature complexe des fonctions physiologiques, les méthodes standard d'identification du système, qui supposent habituellement des interrelations linéaires et temporelles, échouent. Ainsi, ce travail décrit des schémas de modélisation rapides et fiables qui sont capables de traiter (a) des systèmes d'entrée multiples (b) des dynamiques non linéaires (c) des systèmes non stationnaires et (d) des réponses binaires. Ces schémas ont été appliqués en combinaison avec les modèles Laguerre-Volterra (LV), qui peuvent capturer une large gamme d'interrelations causales d'entrée-sortie dynamiques non linéaires et de modèles autogressifs multivariés (MVAR), qui sont utilisés pour détecter les couplages et la causalité entre les séries temporelles. La performance des méthodes susmentionnées a été évaluée en utilisant des simulations et des données expérimentales. Plus précisément, nous avons examiné,

- Les caractéristiques de l'Autoregulation Cérébral (AC) variant dans le temps (TV) dans les patients souffrant du Syncope Vasovagale (VVS) pendant le test de Head-Up Tilt (HUT).
- Changements cardiovasculaires et cérébrovasculaires induits par l'exercice chez des sujets sains et des survivants d'accident vasculaire cérébral.
- Réponses neuronales aux potentiels de terrain locaux (LFP) du noyau subthalamique (NST) chez les patients atteints de la maladie de Parkinson (PD) subissant une chirurgie de Stimulation Cérébrale Profonde (SCP).

Preface and Contribution of the Authors

This dissertation is a partial fulfillment for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering of McGill University, Montreal, Canada. The research described herein was carried out under the supervision of Professor Georgios Mitsis from the Department of Bioengineering, McGill University, between September 2014 and July 2017. This is a manuscript-based thesis consisting of the following three articles,

- **"Modeling of multiple-input, time-varying nonlinear systems with recursively estimated basis expansions,"** Kostoglou K., Schondorf R. and Mitsis G.D. (*Submitted to Elsevier Signal Processing*).

  This manuscript is my original work. The proposed modeling framework is an extension of a previously published paper (J1) and has been presented in several well-known conferences (C2, C8, C10). Professor Mitsis and his expertise has been invaluable in the successful completion of this research topic as he provided excellent feedback and expert advice related with the methodological aspects of this work, as well as data analysis. Furthermore, he contributed to manuscript editing and evaluation. Dr. Schondorf, from the McGill Department of Neurology, provided the data for the experimental applications and helped with the physiological interpretation of my findings.

- **"Modeling time-varying couplings between time series for biomedical applications,"** Kostoglou K., Robertson A., Macintosh B.J and Mitsis G.D. (*Submitted to IEEE Transactions on Biomedical Engineering*).

  This is my original work. Professor Mitsis supervised the research, guided me with his expertise and contributed to manuscript editing and evaluation. Dr. A. Robertson and Professor B.J. Macintosh from the Sunnybrook Research Institute, Toronto, Canada, provided the experimental data, assisted with the physiological interpretation of the results and participated in editing of the manuscript. Part of this work has been presented in biomedical conferences (C1 and C3).

- **"Prediction of the spiking activity in the Parkinsonian subthalamic nucleus using Local Field Potentials and Laguerre Volterra Networks,"** Kostoglou K., Michmizos K.P., Stathis P., Sakas D., Nikita K.S. and Mitsis G.D. *(To be submitted to Journal of Neuroscience)*

  This paper is my original work. Professor Mitsis supervised the research, monitored my progress and contributed to manuscript editing and evaluation. Professor K.P Michmizos from the Department of Computer Science, Rutgers University, New Jersey, USA, offered expert advice, helped with the physiological interpretation of the results and contributed to manuscript editing and evaluation. Neurosurgeons P. Stathis and D. Sakas from the Department of Neurosurgery, National and Kapodistrian University of Athens, Greece, and Professor K.S. Nikita from the School of Electrical and Computer Engineering, National Technical University of Athens, Greece provided the data for analysis, as well as technical advice regarding the surgical procedure of Deep Brain Stimulation. The data analysed in this work have been previously used in publication J3 to answer other related scientific questions.

A full list of my journal and conference publications during my PhD studies is provided below,

Peer reviewed journal publications

J1. **"Nonstationary multivariate modelling of cerebral autoregulation during hypercapnia,"** Kostoglou K., Debert C.T., Poulin M.J. and Mitsis G.D., *Medical Engineering and Physics*, 36.5: 592-600, 2014, (http://dx.doi.org/10.1016/j.medengphy.2013.10.011).

J2. **"Between-centre variability in transfer function analysis, a widely used method for linear quantification of the dynamic pressure-flow relation: the CARNet study,"** A.S.S.M. van den Abeelen, A.S.S.M, Simpson D., Zhang, R., Tarumi, T., Rickards, C.A., Payne, S.J., Mitsis G.D., Kostoglou K., Marmarelis V.Z., Shin D., Tzeng Y-C., Gommer E., Müller M., Caicedo Dorado A., Yelicich B., Puppo C., Liu X., Czosnyka M., Wang N., Novak, V., Panerai R.B. and J.A.H.R Claassen, *Medical Engineering and Physics*, 36.5: 620-627, 2014 (http://dx.doi.org/10.1016/j.medengphy.2014.02.002).

J3. **"Classification and prediction of clinical improvement in Deep Brain Stimulation from intraoperative microelectrode recordings,"** Kostoglou K., Michmizos K.P., Stathis P., Sakas D., Nikita K.S. and Mitsis G.D., *IEEE Transactions on Biomedical Engineering*, 64.5: 1123-1130, 2017 (https://doi.org/10.1109/TBME.2016.2591827).

J4. **"Modelling of multiple-input, time-varying nonlinear systems with recursively estimated basis expansions,"** Kostoglou K., Schondorf R. and Mitsis G.D. (*Submitted to Elsevier Signal Processing*).

J5. **"Modeling time-varying couplings between time series for biomedical applications,"** Kostoglou K., Robertson A., Macintosh B.J. and Mitsis G.D. *(Submitted to IEEE Transactions on Biomedical Engineering*).

J6. **"Aerobic exercise as a cerebral artery stress test to assess the regulation of pulsatile hemodynamics in stroke and small vessel disease,"** Robertson A., Atwi S., Kostoglou K., Verhoeff P., Mitsis G.D. and Macintosh B.J. (*Submitted to Journal of Applied Physiology*).

Peer reviewed journal publications in preparation

JP1. **"Prediction of the spiking activity in the Parkinsonian subthalamic nucleus using Local Field Potentials and Laguerre Volterra Networks,"** Kostoglou K., Michmizos K.P., Stathis P., Sakas D., Nikita K.S. and Mitsis G.D. (*To be submitted to Journal of Neuroscience*).

JP2. **"Nonlinear Autoregressive models with exogenous inputs using Laguerre Volterra Networks,"** Kostoglou K. and Mitsis G.D. (*To be submitted to IEEE Transactions on Neural Networks*).

Peer reviewed conference abstracts and publications

C1. **"Exercise-induced hemodynamic changes in healthy subjects and patients with stroke assessed with time-varying models,"** Kostoglou K., Robertson A., Macintosh B.J. and Mitsis G.D, 7th International CARNET meeting in association with Brain, April 2017 (abstract).

C2. **"Cerebral autoregulation during syncopal events in patients with vasovagal syncope,"** Kostoglou K., Schondorf R., Benoit J., Balegh S. and Mitsis G.D., 7th International CARNET meeting in association with Brain, April 2017 (abstract).

C3. **"Cortico-cortical and corticomuscular coherence using time-varying multivariate autoregressive models",** Porxas A. X., Larivière S., Kostoglou K., Kassinopoulos M., Niso G., Boudrias M.H. and Mitsis G., Organization of Human Brain Mapping (*OHBM)*, June 2017 (abstract).

C4. **"Prediction of the time to syncope occurrence in patients diagnosed with vasovagal syncope,"** Kostoglou K., Schondorf R., Benoit J., Balegh S. and Mitsis G.D., 16th International Symposium on Intracranial Pressure and Neuromonitoring (ICP), ICP 2016 Proceedings, Acta Neurochirurgica Supplements (7 page paper in press).

C5. **"Dynamic cerebral autoregulation in young athletes following concussion,"** Kostoglou, K., Wright, A. D., Smirl, J. D., Bryk, K., van Donkelaar, P. and Mitsis, G. D., 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2016), pp.696-699, Orlando, Florida, USA, August 2016.

C6. **"Prediction of the outcome of Subthalamic Nucleus Deep Brain Stimulation in patients with Parkinson's Disease,"** Kostoglou K. and Mitsis G.D., 2015 Biomedical Engineering Society Annual Meeting (BMES 2015), Tampa, Florida, USA, October 2015.

C7. **"GPU technology as a platform for accelerating physiological systems modeling based on Laguerre-Volterra networks,"** Papadopoulos A, Kostoglou K., Mitsis G.D. and Theocharides T., 37th International Conference of IEEE Engineering in Medicine and Biology Society (EMBC 2015), pp. 3283-3286, Milano, Italy, January 2015.

C8. **"Multivariate nonstationary modeling of cerebral hemodynamics,"** Kostoglou K., Debert C.T., Poulin M.J. and Mitsis G.D., 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2014), pp. 6028-6031, Chicago, USA, August 2014.

C9. **"Accurate spike time prediction from local field potentials in monkey visual cortex: A nonlinear system identification approach,"** Kostoglou K., Hadjipapas A., Lowet E., Roberts M.J., De Weerd P., and Mitsis G.D., 9th Forum of the Federation of European Neuroscience Societies (FENS 2014), Milan, Italy, July 2014.

C10. **"Nonstationary multivariate modelling of cerebral autoregulation during free-breathing and hypercapnic conditions,"** Kostoglou K., Debert C.T., Poulin M.J. and Mitsis G.D., 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO 2014), pp. 193-194, Trento, Italy, May 2014.

C11. **"Prediction of the Parkinsonian subthalamic nucleus spike activity from Local Field Potentials using nonlinear dynamic models,"** Kostoglou K., Michmizos K.P., Stathis P., Sakas D., Nikita K.S. and Mitsis G.D., 12th IEEE International Biomedical Informatics and Bioengineering Conference (BIBE 2012), pp. 298-302, Larnaca, Cyprus, November 2012.

Contents

List of Acronyms

AIC	Akaike Information Criterion
AP	Action Potential
AR	Autoregressive
ARX	Autoregressive with exogenous input
BIC	Bayesian Information Criterion
BUA	Background Unit Activity
CA	Cerebral Autoregulation
CBFV	Cerebral Blood Flow Velocity
CFC	Cross Frequency Coupling
COH	Coherence
CO_2	Carbon Dioxide
D	Delta band (1–4Hz)
DBP	Diastolic Blood Pressure
DBS	Deep Brain Stimulation
DC	Directed Coherence
dCA	Dynamic Cerebral Autoregulation
DLF	Discrete-time Laguerre Function
EEG	Electroencephalogram
FF	Forgetting Factor
FNR	False Negative Rate
FPR	False Positive Rate

G	Gamma band (45-100Hz)
GA	Genetic Algorithm
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GLS	Generalized Least Squares
GWN	Gaussian White Noise
HB	High Beta band (30-45Hz)
HF	High Frequency band (0.15-0.3Hz)
HG	High Gamma band (100-200Hz)
HR	Heart Rate
HUT	Head-up Tilt
ISI	Interspike Interval
KF	Kalman Filter
KFA	Adaptive Kalman Filter
KFM	Kalman Filter for Multiple inputs
KFMA	Adaptive Kalman Filter for Multiple inputs
LASSO	Least Absolute Shrinkage and Selection Operator
LB	Low Beta band (12-30Hz)
LET	Laguerre Expansion Technique
LF	Low Frequency band (0.04-0.15Hz)
LFB	Laguerre Filter Bank
LFP	Local Field Potential
LPV	Linear Parameter-Varying
LS	Least-Squares
LTI	Linear time-invariant
LV	Laguerre Volterra
LVN	Laguerre Volterra Network

MA	Moving Average
MABP	Mean Arterial Blood Pressure
MCC	Matthews Correlation Coefficient
MER	Microelectrode Recording
MI	Multiple Input
MIn	Mutual Information
MSE	Mean Squared Error
MU	Multi-unit
MUA	Multi-unit activity
MVAR	Multivariate Autoregressive
MVARX	Multivariate Autoregressive with exogenous inputs
NARX	Nonlinear Autoregressive with exogenous inputs
NMSE	Normalized Mean Squared Error
nVAF	Negative Variance Accounted For
OLS	Ordinary Least-Squares
PCOH	Partial Coherencec
PD	Parkinson's Disease
PDC	Partial Directed Coherence
PDM	Principal Dynamic Mode
$PETCO_2$	End-tidal Partial Pressure of Carbon Dioxide
PI	Pulsatility Index
PL	Phaselocking
PP	Pulse Pressure
RCT	Regularized Constant Trace
RLS	Recursive Least Squares
RLSA	Recursive Least Squares with Adaptive forgetting factor

RLSC	Recursive Least Squares with Constant forgetting factor
RLSM	Recursive Least Squares with Multiple constant forgetting factors
RLSMA	Recursive Least Squares with Multiple Adaptive forgetting factors
RMS	Root Mean Square
SBP	Systolic Blood Pressure
SI	Single Input
SME	Sinusoidally Modulated Exponential
SNR	Signal to Noise Ratio
STN	Subthalamic Nucleus
SU	Single-Unit
SUA	Single-Unit Activity
SV	Stroke Volume
SVa	Scheduling Variable
T	Theta band(4-12Hz)
TNR	True Negative Rate
TPR	True Positive Rate
TV	Time-Varying
VAF	Variance Accounted For
VHF	Very High Frequency band (0.3-0.4Hz)
VLF	Very Low Frequency band (0.01-0.04Hz)
VRD	Van Rossum Distance
VVS	Vasovagal Syncope
WLS	Weighted Least Squares

Chapter 1 Introduction

1.1 Motivation

System identification is the science of deriving mathematical models to describe the relationship between measured input and output data of a dynamical system. Four main modeling approaches exist in the literature; the parametric, the nonparametric, the modular and the connectionist approach [1]. In the parametric approach, the input-output relationship is represented by using algebraic or differential/difference equations. In the nonparametric approach (known also as black-box models), no particular prior knowledge of the system is assumed (i.e. the model parameters are not necessarily known or fixed). The structure of the system is determined solely from the data. The modular approach is a hybrid parametric and nonparametric approach which is based on block structures that represent different system components, whereas in connectionist models input-output mappings are created using network-type models (e.g. artificial neural networks). Ideally, the extracted models should be able to replicate in an accurate and parsimonious manner the observed functional behavior of the system under consideration.

The past decade, physiological system identification [1], [2] has gained significant interest due to the emergence of a new generation of signal/imaging recording techniques. These techniques provide means of measuring data from a plethora of physiological processes at both micro and macroscale. Therefore, understanding and quantifying the underlying mechanisms that give rise to the obtained signals is the natural next step. Due to the complex nature of physiological functions however, standard system identification methods, which usually

assume linear and time-invariant relationships, are often not adequate. Proper study of real physiological systems requires reliable modeling methods that are capable of dealing with,

A. **Multiple variables of interest:** Physiological mechanisms are usually subject to multiple influences and thus, multiple inputs (MI) should be taken into account when building mathematical models [3]–[5]. For example, in hemodynamics, blood flow to the brain is affected by both arterial blood pressure and carbon dioxide [4], [6]–[8]. Ignoring one of these variables may lead to biased representations of the blood flow regulation mechanisms and induce variability on the obtained models. An important aspect of MI modeling is that it provides the opportunity and the flexibility to study dynamical systems on a multidimensional scale. For instance, MI models can be applied to investigate neuronal network interactions in the brain [9] or study the multijoint properties of a limb [10] among others.
B. **Nonlinear dynamics:** Nonlinearities are observed in most living systems [6], [7], [9], [11]–[16] and they manifest either as a nonproportional relationship between input and output or switching between a set of linear systems. Static nonlinearities are usually easier to model, however dynamic nonlinearities (i.e., the present value of the output depends nonlinearly on the present and past values of the input) pose significant challenges.
C. **Nonstationarities in system dynamics:** Nonstationarities may reflect the intrinsic time-varying (TV) properties of a system, as in the case of cerebrovascular/cardiovascular [17]–[20] and neuromuscular systems [21], but also the indirect effects of unobserved or unmodeled factors. In the vast majority of studies, nonstationarities are usually not considered for reasons of simplification. This, however, may result into substantial errors when describing system characteristics. On the other hand, studies that apply TV analysis assume constant rate of variations. Yet, in a nonstationary environment, system parameters may exhibit intervals of slow, fast or even abrupt changes. In the case of TV MI systems, the corresponding model parameters may vary with different rates too since different inputs can give rise to different TV dynamics.
D. **Binary responses:** Biomedical researchers usually analyse continuous variables, however, there are applications where event-related dynamics are of main interest e.g. epileptic seizure prediction based on electroencephalogram (EEG) signals [22], prediction of the neuronal spiking activity of a brain area based on the recorded Local Field Potentials (LFP) [9], [23], [24], and assessment of heartbeat dynamics [25] to name a few examples. In

practice, when modeling such systems, the output is assumed to be a binary signal where '1' and '0' indicate whether or not an event occurs at a specific time point. Traditional system identification approaches are designed primarily for continuous-valued data and therefore adapting these approaches to systems with binary outputs is vital for their proper analysis.

Herein, we focus on developing effective methodologies for quantifying dynamic interrelationships among signals of biomedical relevance using parametric, nonparametric and connectionist approaches. Our main interest relies on accurately identifying TV, linear and nonlinear physiological systems, as well as systems that produce binary outputs using computationally efficient and fast algorithms. Aside of the new methodological schemes presented in this work, we provide model optimization techniques that reduce significantly computation time and allow further exploration of the model capabilities under various conditions. Our developed methodologies are applied to answer important research questions regarding various physiological functions under healthy and pathological states. Specifically,

- We investigate the link between Cerebral Autoregulation (CA) and syncope occurrence in patients suffering from Vasovagal Syncope (VVS). The main question that we address is whether loss of consciousness during Head-Up Tilt (HUT) testing is an aftereffect of impaired CA function.
- We explore the hemodynamic changes that occur during and after exercise in healthy subjects (old and young), as well as stroke survivors in order to understand the interplay between different cardiovascular and cerebrovascular mechanisms. Our main aim is to extrapolate conclusions regarding the beneficial effects of a single bout of exercise on cardiovascular regulation.
- Finally, we examine the response of different neuronal populations to Subthalamic Nucleus (STN) Local Field Potentials (LFP) in Parkinson's Disease (PD) patients undergoing Deep Brain Stimulation (DBS). The main purpose of this study is to explain the observed variability in predicting accurately neuronal spikes from LFPs and explore the potential use of spike timing predictability as a predictive biomarker of DBS surgery response.

1.2 Thesis Organization and Contributions

This is a manuscript-based dissertation interpolating material from three papers. Chapter 2 is an introductory chapter presenting the methodological background for the work described in later chapters. Chapter 3, 4 and 5 are based on the following references,

Chapter 3 **- "Modeling of multiple-input, time-varying nonlinear systems with recursively estimated basis expansions,"** Kostoglou K., Schondorf R. and Mitsis G.D. (*Submitted to Elsevier Signal Processing*).

Chapter 4 **- "Modeling time-varying couplings between time series for biomedical applications,"** Kostoglou K., Robertson A., Macintosh B.J and Mitsis G.D. (*Submitted to IEEE Transactions on Biomedical Engineering*).

Chapter 5 **- "Prediction of the spiking activity in the Parkinsonian subthalamic nucleus using Local Field Potentials and Laguerre Volterra Networks,"** Kostoglou K., Michmizos K.P., Stathis P., Sakas D., Nikita K.S. and Mitsis G.D. *(To be submitted to Journal of Neuroscience).*

In more detail,

Chapter 2 introduces the time-invariant models used as a basis for Chapters 3,4 and 5. Three type of models are examined; the univariate/multivariate Laguerre-Volterra (LV) models which can be seen as nonparametric models, the Laguerre Volterra Network (LVN) models that belong to the category of connectionist models and the univariate/multivariate autoregressive models with or without exogenous inputs (AR/ARX, MVAR/MVARX) which can be described as parametric infinite memory models. Chapter 2 also introduces the Recursive Least Squares (RLS) and the Kalman Filter (KF) technique for TV model parameter estimation and outlines some of the model order selection methodologies developed for TV systems. Last but not least, the Genetic Algorithm (GA), proposed in subsequent chapters as a model optimization technique, is explored in detail.

Chapters 3 and 4 concentrate on the development of algorithmic schemes for TV system identification (A, B, and C of Section 1.1), whereas Chapter 5 tackles with the estimation of systems with binary outputs (D from Section 1.1).

Chapter 3 describes novel recursive schemes for estimating single- (SI) and multiple-input (MI) TV systems, based on a LV model formulation. The major contributions of this chapter are,

- We developed computational schemes that are applicable to a wide class of TV systems (linear/nonlinear, SI/MI systems with finite memory) by extending and modifying the conventional RLS and KF techniques.
- The proposed schemes achieved superior performance, compared to conventional RLS and KF, under different nonstationarity types (slow, fast/abrupt variations).

- Model order selection and tuning of the estimator hyperparameters were implemented using Genetic Algorithms (GA), significantly improving performance and reducing computation time.
- The link between hyperparameters, model complexity and TV system characteristics was explored in detail.
- We examined the behavior of different model order selection criteria under TV environments.
- Our methodology was used to track changes in CA of patients suffering from VVS during a HUT test protocol.

Chapter 4 extends the TV methodology described in Chapter 3 to MVAR models. The major contributions of this chapter are presented below,

- The methodology of Chapter 3 was adapted to the case of TV-MVAR model estimation.
- Based on the extracted TV-MVAR models we computed TV measures of coupling and causality between time series.
- We corrected biases on the extracted TV-MVAR connectivity measures due to heteroskedasticity in the error terms observed usually in experimental protocols that consist of phase transitions or event-related changes.
- We investigated the TV characteristics of cerebrovascular/cardiovascular regulation in healthy subjects and stroke survivors during exercise.

Chapter 5 describes a modified LVN model that produces binary output. The main contributions of this chapter are summarized below,

- Time-invariant LVN models were modified to produce probabilities of event occurrence in their output. The optimal threshold for transforming these probabilities to '0' and '1' was chosen based on the Matthews Correlation Coefficient (MCC).
- Network parameters were trained by applying a hybrid optimization scheme (GA and interior nonlinear methods) in order to overcome the gradient-based local minima and convergence problems. Instead of the conventional Mean Squared Error (MSE), we minimized the cross-entropy loss function.
- The modified model was extended to account for autoregressive (AR) terms (LVN-ARX).
- The proposed framework was applied to neuronal data (i.e. LFPs and spikes) acquired by PD patients undergoing STN DBS.

Chapter 6 presents the conclusions of this study and potential future avenues.

Chapter 2 Background

2.1 Time-Invariant models

System identification methods are divided into four main groups; the nonparametric, the parametric, the modular and the connectionist approaches. In the following subsections, we briefly present some representative nonparametric and parametric models, in order to facilitate the understanding of the reader. As explained in Chapter 1, nonparametric modeling is useful when the primary interest is in fitting the data without particular knowledge of the underlying structure of the system. Parametric models, on the other hand, assume a predefined structure. Both methods exhibit advantages and disadvantages and the choice of approach depends on the amount of *a priori* information about the identified system.

2.1.1 Nonparametric models

2.1.1.1 Linear

The task of black-box linear models is to describe basic dynamic properties of the system such as the system's impulse or frequency response. The impulse response describes the reaction of a system to input changes and more specifically it is the output of the system when presented with an impulse. The unit impulse function is defined as,

$$\delta(n) = \begin{cases} 0 \; for \; all \; n \neq 0, \\ \\ 1 \; for \; n = 0, \end{cases} \tag{2.1}$$

Let the impulsive input $\delta(n)$ produce the response $h(n)$. For a linear time-invariant (LTI) system, a time shifted impulse $\delta(n-k)$ will result in an output signal $h(n-k)$. In addition, due to linearity, $x(n)\delta(n-k)$ will produce $x(n)h(n-k)$. By taking the infinite sum of both input and output it follows,

$$\sum_{k=-\infty}^{+\infty} x(n)\delta(n-k) = x(n) \tag{2.2}$$

due to the properties of the impulse function and,

$$\sum_{k=-\infty}^{+\infty} x(n)h(n-k) = y(n) \tag{2.3}$$

Equation (2.3) is known as the *convolutional model* and it describes the relationship between input and output. Consequently, if the impulse response of the system $h(n)$ is known then the output $y(n)$ can be easily computed for any input $x(n)$. Impulse response estimation relies on methods such as impulse response and correlation analysis [26], [27].

The frequency response of a system describes the relationship between input and output in the frequency domain. If we choose a complex exponential as input to the LTI system, i.e. $x(t) = Ae^{j(\omega n+\theta)}$, then Equation (2.3) becomes,

$$y(n) = \sum_{k=-\infty}^{+\infty} x(n)h(n-k) = \sum_{k=-\infty}^{+\infty} h(n)x(n-k) = \sum_{k=-\infty}^{+\infty} h(n)Ae^{j[\omega(n-k)+\theta]} \Rightarrow$$

$$y(n) = Ae^{j(\omega n+\theta)} \sum_{k=-\infty}^{+\infty} h(k)e^{-j\omega k} = Ae^{j(\omega n+\theta)}H(j\omega) \tag{2.4}$$

Equation (2.4) states that the output is of the same form as the input, differing only by a scaling factor. The frequency response of the system is defined as $H(j\omega) = \sum_{k=-\infty}^{+\infty} h(k)e^{-j\omega k} = |H(j\omega)|e^{j\sphericalangle H(j\omega)}$. The magnitude $|H(j\omega)|$ and the phase $\sphericalangle H(j\omega)$ are commonly referred to as the gain and the phase shift of the system, respectively. Generally, a frequency domain formulation of the system can be extracted using sinusoidal, frequency response or correlation analysis [26]–[28].

2.1.1.2 Nonlinear

Nonlinear black-box modeling techniques include the Volterra and Wiener series models. The Volterra model assumes causal linear or nonlinear relationships where the input drives the

output. The transformation between input-output is generally dynamic in the sense that the present values of the output depend on present and past values of the inputs. The discrete-time single-input single-output Volterra model can be expressed as [1],

$$y(n) = k0 + \sum_{m} k_1(m)x(n-m) + \sum_{m_1,m_2} k_2(m_1,m_2)x(n-m_1)x(n-m_2) + \cdots + \sum_{m_1..m_Q} k_Q(m_1,\dots,m_Q)x(n-m_1)\dots x(n-m_Q) + \varepsilon(n) \quad (2.5)$$

where $x(n)$ is the input, $y(n)$ is the output, $\varepsilon(n)$ is zero-mean white noise and k_Q are the Q-th order Volterra kernels of the system. Volterra kernels can be interpreted as weighting functions that describe the effect of past input values (k_1; linear kernels), as well as the effect of the Q-th order products between past values of the input (k_Q; Q-th order nonlinear kernel) in order to generate the output signal (Figure 2.1). The zeroth-order Volterra kernel k_0 is the output of the system when all inputs are absent. The Wiener model can be characterized as an orthogonal version of the Volterra model designed specifically for gaussian white noise (GWN) input.

An efficient way to estimate the Volterra kernels is by expanding them to a basis of discrete Laguerre functions (DLF) reducing this way considerably the number of free parameters especially in the case of nonlinear systems. This procedure is known as the Laguerre Expansion Technique (LET) [1], [29] and the resulting models are called *Laguerre Volterra (LV) models*. An equivalent LV representation is the *Laguerre Volterra Network* (LVN) [14], [30] that combines LET with artificial neural networks. The LVN belongs to the connectionist framework and its main advantage is that it can reduce further the total number of required model parameters.

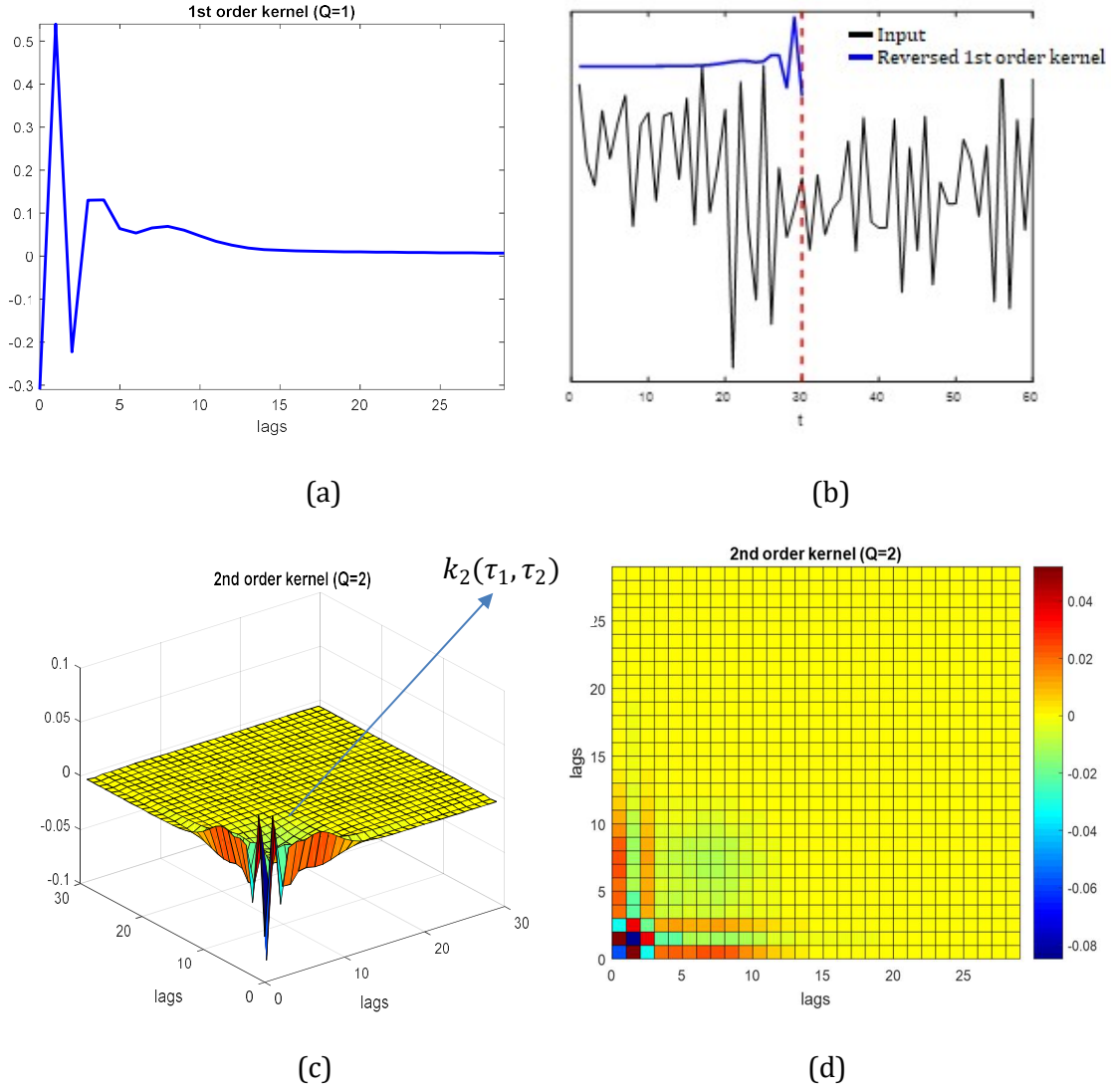


Figure 2.1 (a) A representative 1st order Volterra kernel (b) The 1st-order kernel, akin to the "impulse response function" of linear system theory, can be described as a weighting function that describes the effect of present and past input (signal depicted in black) values in order to generate the output. At time point t=30 (red dashed line) the output is equal to the sum of present and past input values (signal depicted in black before t=30) weighted by the 1st-order kernel (c,d) A representative 2nd-order Volterra kernel that describes the effect of the 2nd-order products between two input values (present or past) in order to generate the output signal.

2.1.2 Parametric models

2.1.2.1 Linear

One example of a linear parametric model is the autoregressive model with exogenous input (ARX). In the ARX model, the output depends on its history, as well as on present and past input values. In discrete-time, a SI ARX [26] of order (n_a, n_b) is defined as,

$$y(n) = \sum_{i=1}^{n_a} a_i y(n-i) + \sum_{0}^{n_b} b_j u(n-j) + \varepsilon(n) \tag{2.6}$$

where $\boldsymbol{y}$ and $\boldsymbol{u}$ are the output and the input of the system respectively, $\boldsymbol{\varepsilon}$ is assumed to be zero-mean white noise. Equation (2.6) can be extended for multiple time series in the form of the multivariate ARX (MVARX) model.

2.1.2.2 Nonlinear

The ARX of Equation (2.6) can be transformed into a nonlinear model if we assume that the output is a nonlinear function of past input and output values. The nonlinear ARX model (NARX) describes the data as follows [31], [32],

$$y(t) = F[y(t-1), \dots, y(t-n_a), u(t), \dots, u(t-n_b)] + \varepsilon(t) \tag{2.7}$$

where $F[.]$ denotes a nonlinear function of lagged input and output signals. NARX models are more difficult to estimate compared to the standard ARX models. Nonetheless, they constitute powerful tools for nonlinear system identification.

2.2 Time-Invariant Basis models

Herein, we describe in detail the three basic models used in this work to study various physiological mechanisms. The LV model is a nonparametric model ideal for identifying finite memory linear and nonlinear systems. The LVN is functionally similar to the LV model, however structurally it resembles an artificial neural network and thus it belongs to the connectionist approaches. On the other hand, the MVAR/MVARX is a parametric technique applied usually to study directed interactions between time series. In the following subsections, the LV, LVN and MVAR/MVARX models are presented in their conventional form, deterministic and time-invariant (i.e. the model parameters remain constant through time). Possible stochastic variations of the system characteristics, measurement noise or modelling error are described

by the stochastic error term of the model which is treated as a zero-mean stationary random process, statistically independent from the input and the noise-free output signals.

2.2.1 Laguerre-Volterra (LV) model

2.2.1.1 Methodology

As described in Section 2.1.1.2, the LV model is an optimal expansion of the Volterra model using Laguerre functions. The Laguerre functions are infinite response (IIR) filters and they are defined in the z-domain as follows [33],

$$B_j(z) = \sqrt{1-a^2}\,\frac{(a-z^{-1})^j}{(1-az^{-1})^{j+1}} \tag{2.8}$$

where $B_j(z)$ is the transfer function of the *j*-th order discrete Laguerre function (DLF) and a is a real pole, known as Laguerre parameter, that defines the decay rate of the DLFs, satisfying $|a| < 1$ (The constraint on the pole $|a| < 1$ leads to causal and stable functions). If the pole is chosen closer to the unit circle, then the decay rate of the DLFs will become slower accordingly. The DLFs form an orthonormal set that spans the function space $L_2(0, +\infty)$, i.e. the space of square integrable functions on the time interval $(0, +\infty)$. Subsequently, any function $f \in L_2$ can be represented as a Laguerre series,

$$f(m) = \sum_{j=0}^{+\infty} c_j\, b_j(m) \tag{2.9}$$

where $b_j(m)$ is the *j*-th order DLF in the time domain,

$$b_j(m) = \alpha^{(m-j)/2}(1-a)^{1/2} \sum_{k=0}^{j} (-1)^k \binom{m}{k}\binom{j}{k} a^{j-k}(1-a)^k \tag{2.10}$$

and the Laguerre coefficients are given by,

$$c_j = < f, b_j > = \sum_{m=0}^{+\infty} f(m)\, b_j(m) \tag{2.11}$$

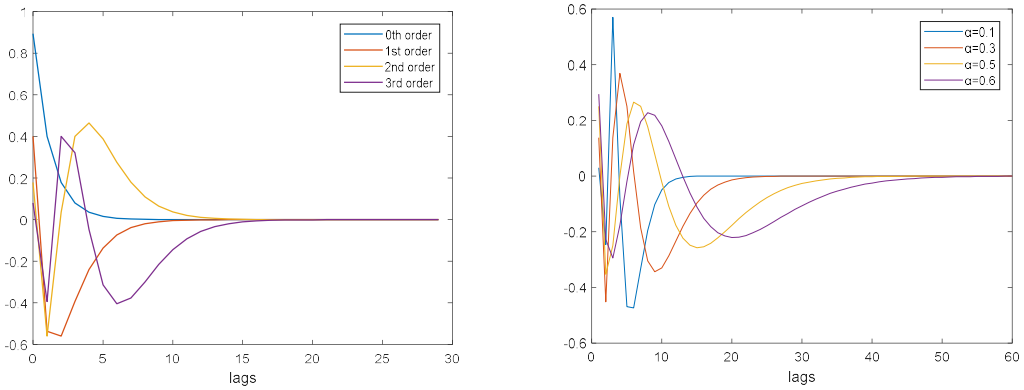
In practice, the expansion is truncated at order L and thus,

$$f_L(m) = \sum_{j=0}^{L-1} c_j\, b_j(m) \tag{2.12}$$

Based on the aforementioned, the Volterra kernels of Equation (2.5) can be expressed as follows,

$$k_Q(m_1, \dots, m_Q) = \sum_{j_1=0}^{L-1} \dots \sum_{j_Q=0}^{L-1} c_Q(j_1, \dots, j_Q) b_{j_1}(m_1) \dots b_{j_Q}(m_Q) \tag{2.13}$$

As explained earlier, the Laguerre parameter a defines the rate of exponential asymptotic decline of the DLFs (Figure 2.2). The smaller (larger) the values of the Laguerre parameter the faster (slower) the system dynamics.



(a) (b)

Figure 2.2 a) Graphical representation of the first 4 Laguerre basis functions for $\alpha = 0.2$. The number of zero crossings (roots) of each DLF equals its order. The higher the order, the longer the significant values of a DLF spread over time and the time separation between zero crossings. b) 3rd order DLF for $\alpha = 0.1$, $\alpha = 0.3$ and $\alpha = 0.5$. Increasing α results in longer spread of significant values and zero crossings. Thus, kernels with longer memory may require a larger a for efficient representation.

By combining Equations (2.5) and (2.8) it follows,

$$\boldsymbol{y} = \boldsymbol{V}\boldsymbol{c} + \boldsymbol{\varepsilon} \tag{2.14}$$

where $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ is the output vector, N is the length of the data set, $\boldsymbol{c} \in \boldsymbol{R}^{d\times 1}$ is the vector of the unknown expansion coefficients and $\boldsymbol{\varepsilon} \in \boldsymbol{R}^{N\times 1}$ is assumed to be a zero-mean white noise vector. The total number of the expansion coefficients, d, is equal to $((L+Q)!/L!\,Q!)$. $\boldsymbol{V}$ is a $N \times d$ matrix containing the convolution of all inputs with the Laguerre functions $v = x * b_j$, as well as nonlinear products between them. The variables v_j can be estimated using the following autorecursive relation [34],

$$v_j(n) = \sqrt{\alpha} v_j(n-1) + \sqrt{\alpha} v_{j-1}(n) - v_{j-1}(n-1) \tag{2.15}$$

$$v_0(n) = \sqrt{\alpha} v_0(n-1) + T\sqrt{1-\alpha} x(n) \tag{2.16}$$

where T is the sampling interval. The expansion coefficients $\boldsymbol{c}$ (Equation (2.14)) can be computed using ordinary least-squares (OLS) estimation based on the input and output data,

$$\hat{\boldsymbol{c}}_{OLS} = (\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V})^{-\mathbf{1}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{y} \tag{2.17}$$

The estimate of Equation (2.17) is an unbiased (i.e., its expectation is equal to the true coefficient), consistent (i.e., it converges in probability to the true coefficient) and efficient (i.e., it has minimum variance among all linear estimators) estimate if the residuals are zero-mean, white and Gaussian. If the residuals, however, are not white then the generalized least-squares (GLS) [35] can be used,

$$\hat{\boldsymbol{c}}_{GLS} = (\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{\Omega}^{-\mathbf{1}}\boldsymbol{V})^{-\mathbf{1}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{\Omega}^{-\mathbf{1}}\boldsymbol{y} \tag{2.18}$$

where $\boldsymbol{\Omega}$ is the covariance matrix of the residuals and $\boldsymbol{\Omega} = \boldsymbol{G}\boldsymbol{G}^T$. GLS is equivalent to applying OLS to a linearly transformed version of the data. By multiplying both sides of Equation (2.14) with $\boldsymbol{G}^{-1}$ it follows that,

$$\boldsymbol{y} = \boldsymbol{V}\boldsymbol{c} + \boldsymbol{\varepsilon} \Rightarrow \boldsymbol{G}^{-1}\boldsymbol{y} = \boldsymbol{G}^{-1}\boldsymbol{V}\boldsymbol{c} + \boldsymbol{G}^{-1}\boldsymbol{\varepsilon} \xrightarrow[\boldsymbol{V}^* = \boldsymbol{G}^{-1}\boldsymbol{V}]{\boldsymbol{y}^* = \boldsymbol{G}^{-1}\boldsymbol{y}} \boldsymbol{y}^* = \boldsymbol{V}^*\boldsymbol{c} + \boldsymbol{\varepsilon}^* \tag{2.20}$$

where $cov(\boldsymbol{\varepsilon}^*) = \boldsymbol{G}^{-1}\boldsymbol{\Omega}(\boldsymbol{G}^{-1})^T = \boldsymbol{I}_{n\times n}$ is the identity matrix**.** Thus, $\boldsymbol{c}$ now can be estimated efficiently by applying OLS to the transformed data (Equation (2.20)). Note here that, $\boldsymbol{\Omega}$ is not usually known in practice**.** A special case of GLS is the so called, Weighted Least Squares (WLS), where the off-diagonal elements of $\boldsymbol{\Omega}$ are assumed to be zero. WLS is applied when the residuals are white but heteroskedastic, i.e. the variance of the residuals is non-constant.

Practical implications also arise if $\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V}$ of Equation (2.17) is singular or ill-conditioned [1]. Direct inversion of $\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V}$ becomes numerically ill-posed leading to large values in the estimated coefficients. This type of ill-conditioning arises mainly if the input is not persistently exciting [27] (i.e., relatively broadband) in the frequency range that spans the systems bandwith. One solution is to apply regularization techniques like ridge regression [36]–[38]. An analytic solution can be obtained as follows,

$$\hat{\boldsymbol{c}}_{REG} = (\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V} + \gamma\boldsymbol{I})^{-\mathbf{1}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{y} \tag{2.21}$$

where γ is a positive scalar, known also as the regularization parameter. The addition of the term $\gamma \boldsymbol{I}$ to the matrix $\boldsymbol{V}^T\boldsymbol{V}$, in Equation (2.21), improves the condition number of $\boldsymbol{V}^T\boldsymbol{V}$ producing more numerically behaved estimates.

2.2.1.2 Model Order Selection

In order to estimate the expansion coefficients of Equation (2.14), it is necessary first to specify the size of the regressor matrix $\boldsymbol{V}$. The size of $\boldsymbol{V}$ depends on the structural parameters L_{x_i} and Q (i.e. the number of DLFs and the order of nonlinearity) which define the LV model complexity. The more complex the model the higher the chances of overfitting the noise and overestimating the true dynamics of the system. On the other hand, an oversimplified model lacks the flexibility to explain accurately the data. Model order determination is based either on cross-validation techniques (i.e. using training and testing datasets) or model order selection criteria like BIC (Bayesian Information Criterion) [39] and AIC (Akaike Information Criterion) [40],

$$BIC(d) = \frac{N}{2}\log\left(\frac{J}{N}\right) + \frac{d}{2}\log N \tag{2.22}$$

$$AIC(d) = \frac{N}{2}\log\left(\frac{J}{N}\right) + d \tag{2.23}$$

where d is the total number of unknown parameters, N is the sample size and J is the sum of squared errors between actual $\boldsymbol{y}$ and predicted output $\hat{\boldsymbol{y}}$,

$$e(n) = y(n) - \hat{y}(n) \tag{2.24}$$

$$J = \sum_{n=1}^{N} e^2(n) \tag{2.25}$$

Along with the model complexity, it is necessary to optimize the Laguerre parameter α_i which defines the memory of the system. Depending on the prior knowledge one has of the system, choosing α_i is a difficult problem to solve in general. This problem has been considered in many publications on the use of Laguerre functions for approximation and system identification [17], [41]–[43]. In practice, however, tuning of α_i (as well as L_{x_i}, Q) is realized using grid or exhaustive search. Typically, for a predefined model order, α_i is selected by minimizing Equation (2.25).

2.2.1.3 Model Interpretation

LV models have been applied successfully in modeling various physiological systems (from neuronal networks to cerebrovascular/cardiovascular systems) and they have been used to deduce model-based biomarkers under several pathological conditions [9], [12], [16], [23], [24], [44]–[56]. Physiological systems usually exhibit an asymptotically exponential structure in their Volterra kernels and therefore the exponential shape of the DLFs renders the LV models the ideal candidates for reproducing complex biological dynamics. Interpretation of the obtained models relies mainly on extracting kernel characteristics either in time or frequency domain (e.g. gain, phase) and relating them to specific physiological mechanisms that are known qualitatively. For example, CA is a mechanism that retains blood flow to the brain constant despite variations in pressure or other external factors. CA is mainly characterized by the relationship between Mean Arterial Blood Pressure (MABP) and Cerebral Blood Flow Velocity (CBFV) (Figure 2.3a) and it exhibits high-pass filter characteristics [57]. That is, coherence and gain are generally lower, and phase is higher at lower-frequency ranges, indicating that CA is more effective with slower fluctuations in MABP. Thus, a phase-shift between MABP and CBFV towards zero in the low frequency (LF) range (0.04 – 0.15 Hz) indicates impaired autoregulation (i.e. CBFV becomes pressure-dependent) [58], [59]. In the case of the LV models this phase-shift can be computed from the phase response of the 1st order kernel and can be used, along with other model features, to assess CA functionality under various conditions and pathologies [7], [19], [44], [51], [53]. In [44], for example, we examined the effects of concussion on the CA characteristics of young athletes. Prior to the start of the season, players completed baseline testing. 7 players who experienced concussions were brought in for repeated testing at 72 hours, 2 weeks, and 1 month post-injury. By extracting LV models and computing the phase response of the 1st order kernels, we observed an initial drop in the LF phase in all subjects 72 hours post-injury and a return around its baseline values by the first month. This implies that concussion leads to impaired autoregulatory function which requires a period of at least one month for complete restoration.

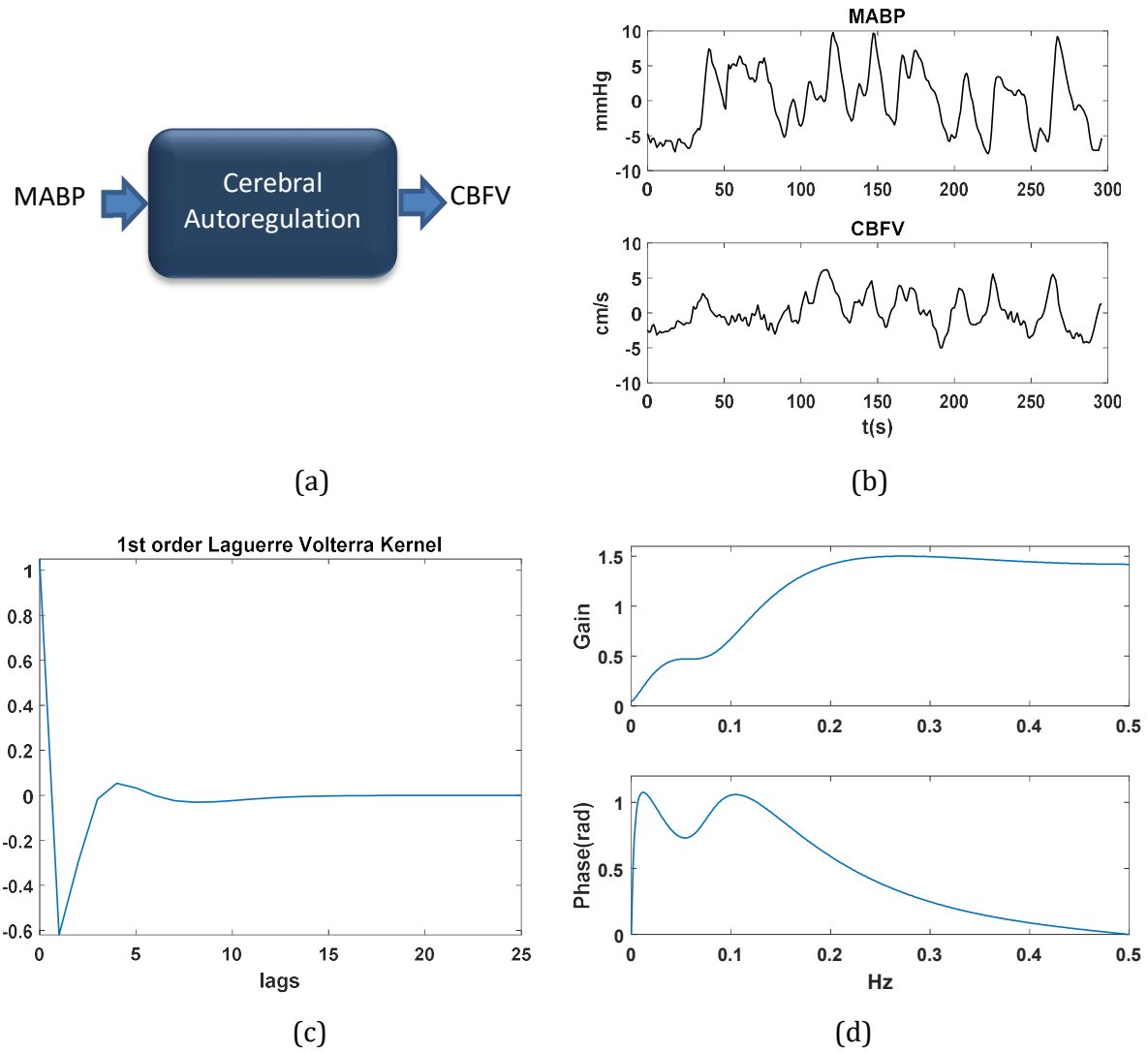


Figure 2.3 Representative example of a physiological system (a) CA is a complex homeostatic mechanism that maintains a constant CBF despite variations in arterial blood pressure (ABP) and is frequently assessed by the dynamic relation between MABP and CBFV. Hence, MABP is assumed to be the system input and CBFV the system output. (b) Representative segments of MABP and CBFV experimental data from one subject. The sampling rate is 1Hz. (c) 1st order LV kernel, in the time domain, extracted from the signals presented in (b). (d) Gain and phase response of the LV kernel depicted in (c) in the frequency domain.

2.2.2 Laguerre-Volterra Network (LVN)

2.2.2.1 Methodology

LET can be combined with feedforward artificial neural networks and hidden units with polynomial activation functions in the form of the LVN [14], [17], [30]. The input is initially convolved with a set of linear filters (Laguerre Filter Bank (LFB)). The obtained counterparts of the input are then fed into the units of the hidden layer (also known as neurons) giving rise to nonlinear interactions. Summation of the outputs of the units produces the final model output. Specifically, the *j*-th order DLF of the LFB is expressed as,

$$b_j(m) = \alpha^{(m-j)/2}(1-\alpha)^{1/2}\sum_{k=0}^{j}(-1)^k\binom{m}{k}\binom{j}{k}\alpha^{j-k}(1-\alpha)^k \tag{2.26}$$

where $j = 1 \dots L$ and and α is the Laguerre parameter described at Section 2.2.1.1. The corresponding filter output $v_j(n)$ is estimated by convolving $b_j(m)$ with the input $x(n)$. The variables v_j are given by the autorecursive relations of Equations (2.15) and (2.16). Functional equivalence with the Volterra model is achieved by employing polynomial activation functions on the units of the hidden layer. Each hidden unit k receives as input the weighted sum of the LFB outputs,

$$u_k(n) = \boldsymbol{w}_k{}^T\boldsymbol{v}(n) \tag{2.27}$$

where $k = 1 \dots K$, K is the total number of hidden units, $\boldsymbol{w}_k \in \boldsymbol{R}^{L\times 1}$ are the weights and $\boldsymbol{v}(n) \in \boldsymbol{R}^{L\times 1}$ is the output of each filter bank at time step n**.** $u_k(n)$ is then preprocessed and transformed into,

$$z_k(n) = \boldsymbol{c}_k{}^T\tilde{\boldsymbol{u}}_k(n) \tag{2.28}$$

where $\boldsymbol{c}_k \in \boldsymbol{R}^{Q\times 1}$**,** Q is the degree of the polynomial activation functions ($Q = 1$ refers to a linear model, whereas Q>1 to a Q-th order nonlinear model) and $\tilde{\boldsymbol{u}}_k(n) = [u_k(n) \quad \dots \quad u_k{}^Q(n)]^T$. The final LVN output $y(n)$ is simply the nonweighted summation of the hidden-unit outputs $z_k(n)$ and a trainable offset y_0,

$$y(n) = y_0 + \sum_{k=1}^{K} z_k(n) \tag{2.29}$$

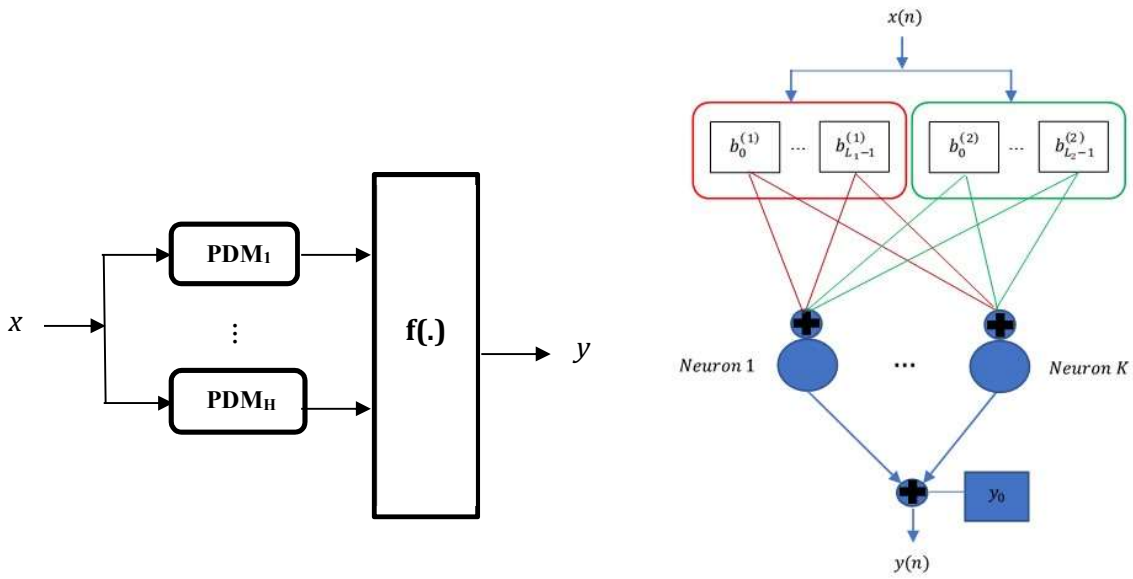
The total number of free parameters in the LVN is equal to $(L + Q) \cdot K$. The Volterra kernels of the system can be easily expressed in terms of the network parameters [30]. A convenient and more interpretable LVN representation is the Principal Dynamic Modes (PDM) model, which

consists of a set of parallel filters (known as PDMs), defined by the number of hidden units, followed by a static nonlinearity [1] (Figure 2.4a). The PDMs reflect the most significant input dynamics that give rise to the output and can be expressed as,

$$PDM_k(m) = \boldsymbol{w}_k{}^T \boldsymbol{b}(m) \tag{2.30}$$

where $\boldsymbol{b}(m) \in \boldsymbol{R}^{L\times 1}$ consists of all the DLFs $b_j(m)$.

In [14] Mitsis et al. extended the original LVN model by introducing two LFBs in the input layer instead of one. The role of the two LFBs is to differentiate the slow from the fast dynamics and alleviate modelling challenges due to the presence of multiple time scales in the system dynamics. The LFBs can be assigned with different Laguerre parameters and a different number of Laguerre functions (filters) (Figure 2.4b). This can significantly reduce the total number of parameters needed compared to the case of a single LFB. The proposed model can also capture effectively nonlinear interactions between slow and fast dynamics.



(a) (b)

Figure 2.4 (a) PDM model composed of H filters (PDMs) that span the dynamics of the system. The PDM outputs are fed into a H-input static nonlinearity to generate the final output (x is the input and y is the output). (b) A SI fully connected LVN network with K neurons. Each input is convolved with two sets of linear filters (LFBs) (i.e. red and green boxes). The LFBs are characterized by different Laguerre parameters and may contain different numbers of Laguerre functions (i.e. $[L_1, a_1]$ and $[L_2, \alpha_2]$).

2.2.2.2 Training and Model Order Selection

A critical practical issue for the successful application of the LVN is the proper selection of its structural parameters, i.e., the size of the Laguerre filter banks L_i (number of DLFs), the number K of hidden units and the degree Q of the polynomial activation functions. Selecting an excessively complex model may lead to overfitting and poor generalization. In other words, the model will fit perfectly the training data but it will perform poorly on new data. This is usually dealt with cross-validation techniques or model order selection criteria, as in the case of the LV models. Lately, regularization has also become a trending topic [9], [38], [52], [60], [61]. The idea behind regularization is to penalize complex models that have too many parameters. This is usually achieved by adding a complexity term to the loss function. The complexity term is controlled by the so-called regularization parameter λ. A large λ rules out overcomplex models, whereas a small λ provides higher model flexibility. Regularization parameter selection is a challenging problem and it relies either on cross validation methods or other more sophisticated techniques. In the LVN models, regularization along with pruning of unnecessary network elements has shown promising results [61]. Once the model structure has been defined, training of the network parameters (i.e. the weights $\boldsymbol{w}_k^{(i)}$, the Laguerre parameters a_i, the coefficients of the polynomial functions $\boldsymbol{c}_k$ and the output bias y_0) is performed using iterative estimation methods such as the backpropagation algorithm [14], [17]. Gradient based techniques however are known to suffer from convergence problems and local minima trapping. Thus, stochastic global optimization methods are a viable alternative solution for LVN training.

2.2.2.3 Model Interpretation

Compared to the LV models, LVNs require a lower total number of unknown parameters. In the LV models, model complexity depends on the order Q of the system exponentially. For the LVN models, this dependence is linear. Therefore, LVNs are more practical for higher-order systems and they exhibit superior performance even in small datasets. The LV kernels can be easily reconstructed from the network parameters. Model interpretation is based on the same principles as those outline for LV models (Section 2.2.1.3). In addition, the equivalent PDM form of the LVN is a compact and parsimonious representation of the system under examination, providing wealth of information regarding the main input dynamics that give rise to the output. The LVNs have been successfully used in various biomedical applications [6]–[8], [17], [61],

[62]. In Chapter 5, we have developed LVN models that produce binary output predictions. Binary response models are useful for examining neuronal dynamics (Figure 2.5).

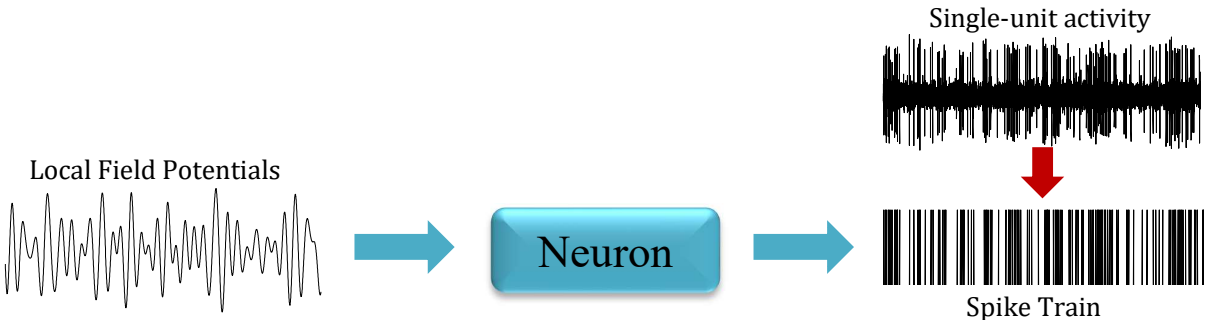


Figure 2.5 An illustrative example of a system with binary output. The neuron responds to the elicited Local Field Potentials (LFP) with Action Potentials (AP) in the form of the Single-Unit Activity (SUA). Since the time of arrival of the AP carries all the information, the shape of the AP is ignored and the SUA signal is transformed to a binary signal (through amplitude thresholding) with '1' indicating AP occurrence and '0' AP nonoccurrence. The binary signal is usually referred to as a spike train.

2.2.3 Multivariate Autoregressive Model with exogenous input (MVARX)

2.2.3.1 Methodology

In contrast to LV/LVN models, MVARX models are purely linear models. They consist of AR terms that take into account also the past history (past lags) of the outputs. An MVARX of order (n_a, n_b) [63] describes the data as follows,

$$\boldsymbol{y}(n) = \sum_{i=1}^{n_a} \boldsymbol{A_i}\boldsymbol{y}(n-i) + \sum_{j=0}^{n_b} \boldsymbol{B_j}\boldsymbol{u}(n-j) + \boldsymbol{\varepsilon}(t) \tag{2.31}$$

where $\boldsymbol{y}(n) \in \boldsymbol{R}^{M\times 1}$ is the vector of M response time series variables at time n, $\boldsymbol{u}(n) \in \boldsymbol{R}^{K\times 1}$ is the vector of K exogenous input time series variables at time n, $\boldsymbol{A_i} \in \boldsymbol{R}^{M\times M}$ is an autoregressive matrix for each order i, $\boldsymbol{B_j} \in \boldsymbol{R}^{M\times K}$ is a matrix representing exogenous terms for each order j and $\boldsymbol{\varepsilon}(n)$ is assumed to be a zero-mean white noise vector. In this dissertation, we focus mainly on the MVAR model which is the MVARX without the exogenous inputs i.e.

$$\boldsymbol{y}(t) = \sum_{i=1}^{n_a} \boldsymbol{A_i}\boldsymbol{y}(t-i) + \boldsymbol{\varepsilon}(t) \tag{2.32}$$

Equation (2.32) can be expresses in a matrix form as follows,

$$\boldsymbol{Y} = \boldsymbol{A\Phi} + \boldsymbol{\varepsilon} \tag{2.33}$$

where $\boldsymbol{Y} \in \boldsymbol{R}^{M\times N}$, N is the total number of samples, $\boldsymbol{A} \in \boldsymbol{R}^{M\times M\cdot n_a}$, $\boldsymbol{\Phi} \in \boldsymbol{R}^{M\cdot n_a\times N}$ and $\boldsymbol{\varepsilon} \in \boldsymbol{R}^{M\times N}$. The coefficient matrix $\boldsymbol{A}$ can be computed using OLS,

$$\widehat{\boldsymbol{A}} = \boldsymbol{Y\Phi^T}(\boldsymbol{\Phi\Phi}^T)^{-1} \tag{2.34}$$

Other estimation methods include the Yule-Walker correlation, the Vieira-Morf partial correlation and the Nutall-Strand partial correlation to name a few examples [64]. The predicted output is obtained from the following equation,

$$\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{A}}\boldsymbol{\Phi} \tag{2.35}$$

2.2.3.2 Model Order Selection

The most common approach for MVAR model order selection involves the minimization of one or more information criteria evaluated over a range of possible model orders (n_a). For the MVARs the multivariate versions of BIC and AIC (Equations (2.22) and (2.23)) are commonly used,

$$BIC(n_a) = N\log\left(|\widehat{\boldsymbol{\Sigma}}|\right) + \log(N)d \tag{2.36}$$

$$AIC(n_a) = N\log\left(|\widehat{\boldsymbol{\Sigma}}|\right) + 2d \tag{2.37}$$

where N is the length of the data set, $d = M^2n_a$ is the total number of parameters and $|\widehat{\boldsymbol{\Sigma}}|$ is the determinant of the estimated covariance of the error terms, i.e. $\widehat{\boldsymbol{\Sigma}} = cov(\boldsymbol{Y} - \widehat{\boldsymbol{Y}})$. A main issue with the MVAR formulation is that it considers all lagged variables up to a selected order. Conventional regularization techniques like Least Absolute Shrinkage and Selection Operator (LASSO) and ridge regression do not take into account the lagged dependent structure of the data. Therefore, other alternatives such as Group LASSO have been proposed [65], [66]. Various dimension reduction techniques have also been developed to mitigate with the inclusion of unnecessary model regressors [67], [68].

2.2.3.3 Model Interpretation

MVAR models can capture linear interdependencies among multiple time series and identify causality in time and frequency domain which is highly essential when studying physiological systems. The human organism is by itself a complex network of interacting physiologic organ systems. At the microscale level for example, information is propagated through interconnected neuronal structures allowing different brain areas to communicate with each other. At the

macroscale level, cardiovascular, cerebrovascular and respiratory systems cooperate to ensure that enough oxygen will be supplied to organ tissues. From a signal processing point of view, communication between these systems is mainly achieved through couplings in different frequency bands. Hence, MVAR models are the ideal candidates for studying the coupling strength and the directionality of information flow in physiological interconnected systems and this has been proven by a number of studies [65], [69]–[78].

The most popular measures extracted by the MVAR models are the Coherence (COH) [79], the Partial Coherence (PC) [80], [81], the Directed Coherence (DC) [82], the Partial Directed Coherence (PDC) [76]. By taking the Fourier transform of Equation (2.27) it follows,

$$\boldsymbol{y}(n) = \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{y}(n-k) + \boldsymbol{\varepsilon}(n) \xrightarrow{FT} [\boldsymbol{I} - \boldsymbol{A}(f)]\boldsymbol{Y}(f) = \boldsymbol{E}(f) \Rightarrow \boldsymbol{Y}(f) = \boldsymbol{H}(f)\boldsymbol{E}(f) \quad (2.38)$$

where $\boldsymbol{A}(f) = -\sum_{k=1}^{\boldsymbol{p}} \boldsymbol{A_k} e^{-i2\pi f k T} \in \boldsymbol{R}^{M\times M}$ is the *coefficient matrix* in the frequency domain, $\boldsymbol{Y}(f) = \sum_{\boldsymbol{n}=-\infty}^{+\infty} \boldsymbol{y}(n) e^{-i2\pi f n T}$ and $\boldsymbol{H}(f) = [\boldsymbol{I} - \boldsymbol{A}(f)]^{-\mathbf{1}} = \overline{\boldsymbol{A}}(f)^{-1} \in \boldsymbol{R}^{M\times M}$ is the *transfer matrix* in the frequency domain. Given the assumption of whiteness and uncorrelation of the input processes, the *spectral power density matrix* of $\boldsymbol{y}(n)$ is,

$$\boldsymbol{S}(f) = \boldsymbol{Y}(f)\boldsymbol{Y}^H(f) \overset{(2.38)}{\Longrightarrow} \boldsymbol{S}(f) = \boldsymbol{H}(f)\boldsymbol{\Sigma}\boldsymbol{H}^H(f) \quad (2.39)$$

The *inverse spectral power density matrix* of $\boldsymbol{y}$ can be expressed as,

$$\boldsymbol{P}(f) = \boldsymbol{S}^{-1}(f) = \overline{\boldsymbol{A}}^H(f)\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{A}}(f) \quad (2.40)$$

where the superscript H stands for Hermitian transpose and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_M^2 \end{bmatrix} \in \boldsymbol{R}^{M\times M}$ is the diagonal covariance matrix of $\boldsymbol{\varepsilon}$. The elements of the spectral density, transfer function, and coefficient matrices can be used to measure coupling and causality in the frequency domain. Specifically,

$$COH_{TD}(f) = \frac{S_{TD}(f)}{\sqrt{S_{TT}(f)}\sqrt{S_{DD}(f)}} \quad (2.41)$$

$$DC_{TD}(f) = \frac{\sigma_D H_{TD}(f)}{\sqrt{\sum_{m=1}^{M} \sigma_m^2 |H_{Tm}(f)|^{\mathbf{2}}}} \quad (2.42)$$

$$PCOH_{TD}(f) = \frac{P_{TD}(f)}{\sqrt{P_{TT}(f)}\sqrt{P_{DD}(f)}} \quad (2.43)$$

$$GPDC_{TD}(f) = \frac{\frac{1}{\sigma_D}\bar{A}_{TD}(f)}{\sqrt{\sum_{m=1}^{M}\frac{1}{\sigma_m^2}|\bar{A}_{mD}(f)|^2}} \tag{2.44}$$

where COH_{TD} describes the *Coherence* from D (driver) to T (target), DC_{TD} the *Directed Coherence*, $PCOH_{TD}$ the *Partial Coherence* and $GPDC_{TD}$ the *Generalized Partial Directed Coherence*. These four measures describe in the frequency domain the time domain concepts of coupling (COH), direct coupling (PCOH), causality (DC), and direct causality (PDC). PCOH and PDC measure direct connectivity between two processes, while COH and DC account for both direct and indirect connections. Regarding causality, DC and PDC, only, contain information related with the directionality of the interaction. It holds that, $0 \leq |DC_{TD}(f)|^2 \leq 1$, $\sum_{m=1}^{M}|DC_{Tm}(f)|^2 = 1$ and $0 \leq |GPDC_{TD}(f)|^2 \leq 1$, $\sum_{m=1}^{M}|GPDC_{mD}(f)|^2 = 1$. Note that DC uses the elements of the transfer function matrix H while GPDC uses those of $\bar{A}$. Since the computation of GPDC does not involve any matrix inversion, it is computationally more efficient and more robust than DC. Furthermore, GPDC is normalized with respect to the total inflow of information, whereas DC is normalized with respect to the total outflow of information.

In order to facilitate the readers understanding of the MVAR measures described in Equations (2.39-2.44), the following three dimensional ($M = 3$) MVAR process is considered [83],

$$y_1(n) = 2\rho\cos(2\pi f_1)\, y_1(n-1) - \rho^2 y_1(n-2) + \varepsilon_1(n) \tag{2.45}$$

$$y_2(n) = y_1(n-1) + 2\rho\cos(2\pi f_2)\, y_2(n-1) - 0.6y_1(n-2) - \rho^2 y_2(n-2) + \varepsilon_2(n) \tag{2.46}$$

$$y_3(n) = y_2(n-1) + 2\rho\cos(2\pi f_3)\, y_3(n-1) - \rho^2 y_3(n-2) + \varepsilon_3(n) \tag{2.47}$$

where $\rho = 0.9$, $f_1 = 0.3$, $f_2 = 0.1$, $f_3 = 0.05$ and $\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \boldsymbol{\varepsilon_3}$ are white noise signals (with variance $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$) driving the process. Representative simulated signals can be found in Figure 2.6. The sampling rate is assumed to be 1Hz. The process described in Equations (2.45-2.47) is one possible MVAR realization of the directed graph presented in Figure 2.7. The following observations can be made,

a) The *direct causality relations* are $y_1 \to y_2$, $y_2 \to y_3$, $y_1 \to y_1$, $y_2 \to y_2$, $y_3 \to y_3$ and this is captured by GPDC (Figure 2.8a) which is nonzero only for these cases.

b) The *indirect causality relations* are $y_1 \overset{y_2}{\to} y_3$ (i.e. from y_1 to y_3 through y_2). Since DC (Figure 2.8b) refers to both direct and indirect causal connections, DC is nonzero only for the relations described in a) and b).

c) The *direct couplings* are $y_1 \leftrightarrow y_2$ and $y_2 \leftrightarrow y_3$. PCOH (Figure 2.8c) is thus zero for $y_1 \leftrightarrow y_3$. The diagonal elements in Figure 2.8c represent the inverse spectra (P) of the time series.

d) The *indirect couplings* are $y_1 \overset{y_2}{\leftrightarrow} y_3$ (i.e. there is an indirect relationship between y_1 and y_3 mediated through y_2). COH (Figure 2.8d) represents both direct and indirect coupling thus it is nonzero for cases c) and d). The diagonal elements of COH are the spectra of the three time series.

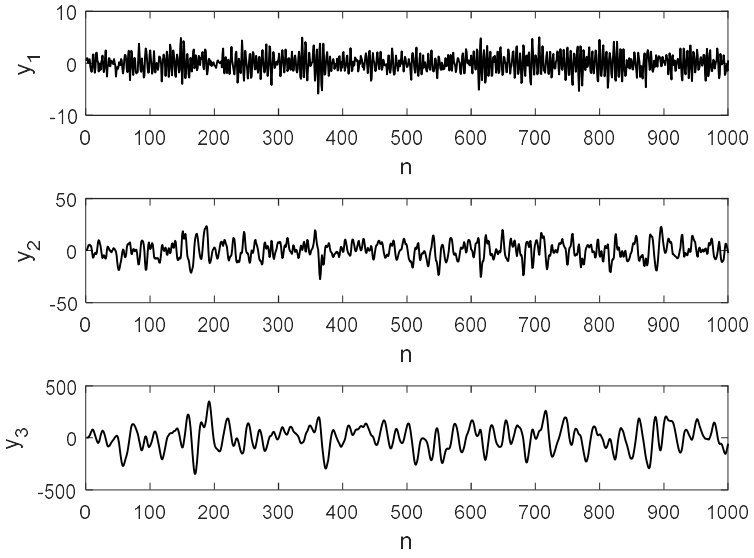


Figure 2.6 Simulated time series based on the MVAR process described in Equations (2.45-2.47).

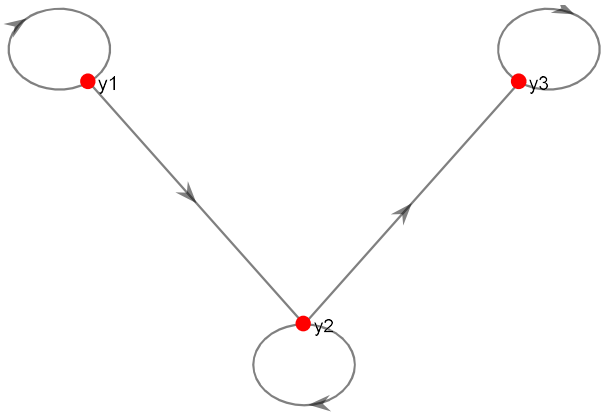


Figure 2.7 Directed graph representing the interactions described in Equations (2.45-2.47). The nodes correspond to processes and the connecting arrows depict direct causal connections.

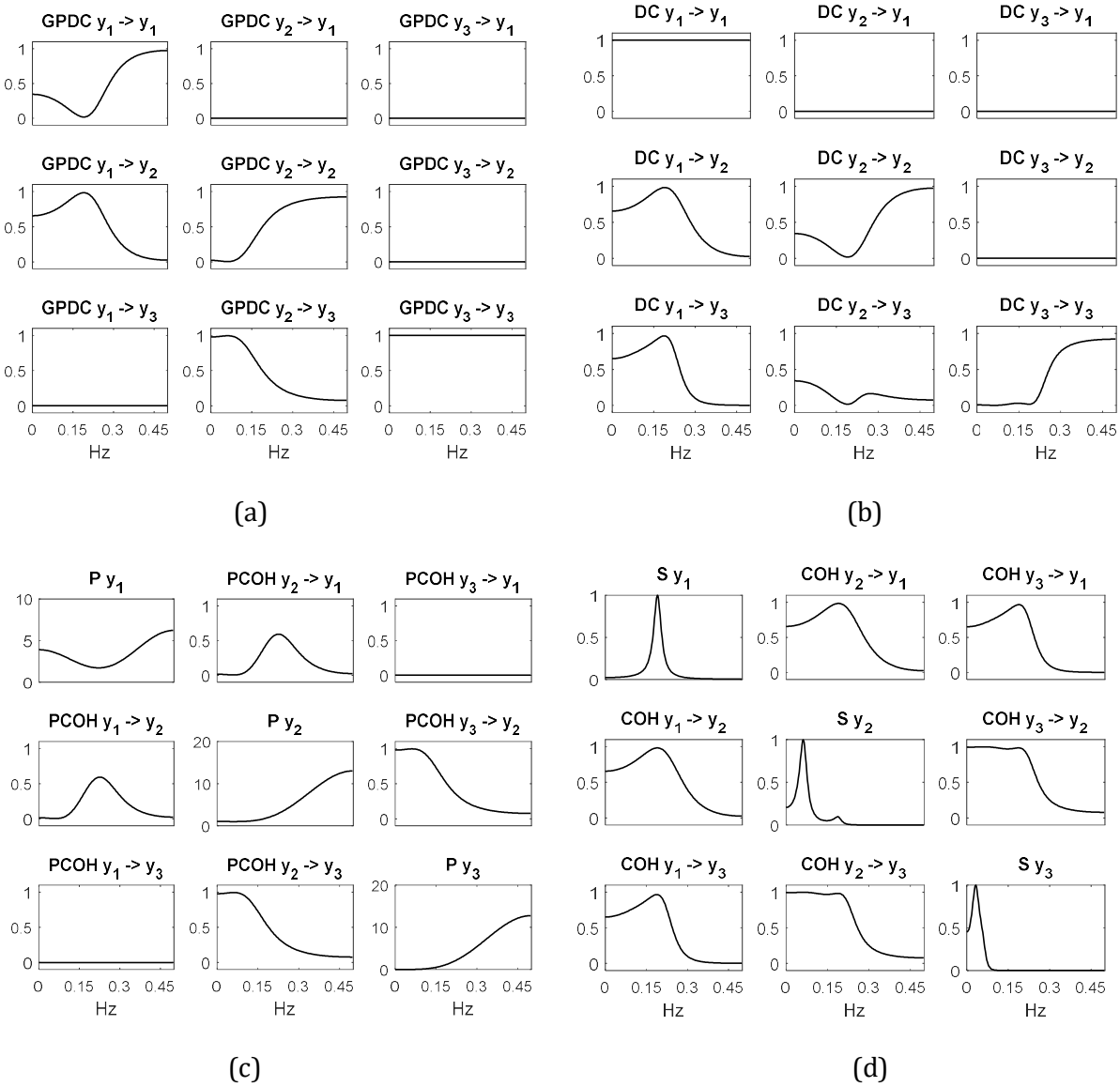


Figure 2.8 (a) GPDC, (b) DC, (c) Spectra and COH, (b) Inverse spectra and PCOH for the MVAR process described in Equations (2.45-2.47). y_D -> y_T denotes the driver and the target time series. COH and PCOH are symmetric measures, whereas DC and GPDC are asymmetric. Thus, COH y_D -> y_T is equal to COH y_T -> y_D and PCOH y_D -> y_T is equal to PCOH y_T -> y_D, whereas DC y_D -> y_T ≠ DC y_T -> y_D and GPDC y_D -> y_T ≠ GPDC y_T -> y_D. The diagonal components of COH and PCOH are the spectra (S) and the inverse specta (P), respectively, of each time series. Notice how GPDC and DC adds up to one, at each frequency, columnwise and rowise respectively.

2.3 Time-Varying (TV) models

Time-invariant models that are linear in the parameters express the input-output relationship as follows (assuming a SI system for simplification),

$$\boldsymbol{y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon} \tag{2.48}$$

where $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ is the output signal, N is the length of the data set, $\boldsymbol{Z} \in \boldsymbol{R}^{N\times d}$ is the regressor matrix (i.e. the $\boldsymbol{V}$ matrix for the LV models and the $\boldsymbol{\Phi}$ matrix for the MVAR models), d is the total number of model parameters, $\boldsymbol{\beta} \in \boldsymbol{R}^{d\times 1}$ contains the model parameters and $\boldsymbol{\varepsilon} \in \boldsymbol{R}^{N\times 1}$ is assumed to be zero-mean white noise. $\boldsymbol{\beta}$ can be easily estimated using OLS,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z^T Z})^{-1}\boldsymbol{Z^T y} \tag{2.49}$$

On the other hand, TV models assume that the model parameters are a function of time and thus Equation (2.48) becomes,

$$y(n) = \boldsymbol{z}^T(n)\boldsymbol{\beta}(n) + \varepsilon(n) \tag{2.50}$$

where $\boldsymbol{z}(n) \in \boldsymbol{R}^{d\times 1}$ is the regressor and $\boldsymbol{\beta}(n) \in \boldsymbol{R}^{d\times 1}$ are the model parameters at time point n. Several approaches have been adopted for the estimation of the TV parameters $\boldsymbol{\beta}(n)$. The quasistationary approach relies on identifying piecewise stationary models in sliding windows of data [17], [20], [84], [85]. Recursive approaches update the parameter estimates at each time point based on the data [19], [86]–[90]. Temporal expansion methods expand the TV parameters onto a set of basis sequences. The problem then becomes time-invariant with respect to the coefficients of the expansion [91]–[95]. Last but not least, ensemble methods use an ensemble of input-output realizations to produce a model of the system at each time point [21], [95], [96]. In the following subsections, we focus on the quasistationary and the recursive approaches.

2.3.1 Quasistationary approaches

One of the most popular techniques for TV parameter estimation is the quasistationary approach. The procedure fits piecewise stationary models to the input-output data producing this way piecewise-constant parameter trajectories. Initially, the data is segmented into widows of equal length (with or without overlap). Subsequently, time-invariant models are estimated in each window. The obtained parameters are either linearly interpolated (if no window overlap is used) or averaged (in case of overlapped windows). Applying window

overlapping increases significantly computation time but improves time resolution. The window length in which the system is assumed to remain stationary is of critical importance. Too small of a window may lead to noisy estimates, too big of a window may underestimate the magnitude and the rate of changes (Figure 2.9). Usually, the window length is selected empirically based on some prior knowledge regarding the TV characteristics of the system or visually based on the variability of the obtained estimates.

2.3.2 Recursive approaches

Contrary to quasistationary approaches, recursive algorithms update the model parameters continuously on the basis of new data. Some of the most common recursive estimation methods are the RLS and the KF technique [86]–[89].

2.3.2.1 Recursive Least Squares (RLS)

RLS is an adaptive technique that recursively estimates model parameters by minimizing a weighted LS cost function [86]–[88]. The general form of RLS can be summarized in the following recursive equations,

$$e(n) = y(n) - \boldsymbol{z}^{\boldsymbol{T}}(n)\widehat{\boldsymbol{\beta}}(n-1) \tag{2.51a}$$

$$r(n) = \boldsymbol{z}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1)\boldsymbol{z}(n) \tag{2.51b}$$

$$\boldsymbol{K}(n) = \frac{\boldsymbol{P}(n-1)\boldsymbol{z}(n)}{\lambda + r(n)} \tag{2.51c}$$

$$\boldsymbol{P}(n) = \frac{1}{\lambda}[\boldsymbol{P}(n-1) - \boldsymbol{z}^{\boldsymbol{T}}(n)\boldsymbol{K}(n)\boldsymbol{P}(n-1)] \tag{2.51d}$$

$$\widehat{\boldsymbol{\beta}}(n) = \widehat{\boldsymbol{\beta}}(n-1) + \boldsymbol{K}(n)e(n) \tag{2.51e}$$

where $e(n)$ is the output prediction error, $\boldsymbol{K}(n) \in \boldsymbol{R}^{d\times 1}$ is a gain matrix, $\boldsymbol{P}(n) \in \boldsymbol{R}^{d\times d}$ is the parameter covariance matrix corresponding to time n initialized as $\boldsymbol{P}(0) = P_0\boldsymbol{I}_{d\times d}$ and λ is the so-called forgetting factor (FF). Equations (2.51a-e) are derived by replacing the LS error cost function with an exponential weighted sum of the squared error signal,

$$J(n) = \sum_{k=0}^{n} \lambda^{n-k} e^2(k) \tag{2.52}$$

The role of the FF ($0 < \lambda < 1$) is to reduce the influence of old data by assigning exponentially lower weight to less recent error samples. In the classical RLS, λ is set to one and the covariance matrix $\boldsymbol{P}(n)$ (Equation (2.51d)) vanishes to zero with time, losing its capability to keep track of

changes in the parameters. In Equation (2.51d), however, $\boldsymbol{P}(n)$ is divided by $\lambda < 1$ slowing down this way the fading out of the covariance matrix. The lower the value of the FF (closer to 0) the faster the adaptation and the higher the sensitivity to noise. On the other hand, higher FF values (closer to 1) enable slower adaptation leading to less variable estimates. Therefore, the value of the FF is a trade-off between bias and variance of the parameter estimates (Figure 2.10). Usually the FF is set empirically between 0.98 and 0.995 but in Chapter 3 we outline, for the first time, all the factors that affect its value and we present a framework for optimal FF selection. In addition, under the hypothesis that the parameters follow different rate of variations, we propose the use of multiple TV FFs that adjust their values automatically based on the data.

2.3.2.1 Kalman Filter (KF)

TV parameters can also be adaptively estimated using the Kalman filter (KF) [87]–[89]. In contrast to RLS, KF assumes prior knowledge of the true parameter variations. The parameters change over time according to a random walk process driven by GWN with diagonal covariance matrix $\boldsymbol{R_1} = R_1\boldsymbol{I}_{d\times d}$ where $R_1 = \sigma_w^2$. The size of R_1 dictates the rate of the random walks and the sensitivity of the estimator to noise (Figure 2.11). The Kalman filter algorithm can be summarized as follows,

$$e(n) = y(n) - \boldsymbol{z^T}(n)\widehat{\boldsymbol{\beta}}(n-1) \tag{2.53a}$$

$$r(n) = \boldsymbol{z^T}(n)\boldsymbol{P}(n-1)\boldsymbol{z}(n) \tag{2.53b}$$

$$\boldsymbol{K}(n) = \frac{\boldsymbol{P}(n-1)\boldsymbol{z}(n)}{R_2 + r(n)} \tag{2.53c}$$

$$\boldsymbol{P}(n) = \boldsymbol{P}(n-1) + \boldsymbol{R_1} - \boldsymbol{K}(n)\boldsymbol{z^T}(n)\boldsymbol{P}(n-1) \tag{2.53d}$$

$$\widehat{\boldsymbol{\beta}}(n) = \widehat{\boldsymbol{\beta}}(n-1) + \boldsymbol{K}(n)e(n) \tag{2.53e}$$

where $\boldsymbol{K}$ is the Kalman gain matrix, which minimizes the *a posteriori* error covariance $\boldsymbol{P}$, and R_2 is the variance of the measurement noise which is also assumed to be GWN. $\boldsymbol{P}$ is initialized with a diagonal matrix of the form $\boldsymbol{P}(0) = P_0\boldsymbol{I}_{d\times d}$. For $R_1 = 0$ and $R_2 = 1$, KF and RLS with $\lambda = 1$ become identical. As in the case of RLS, the accuracy of the obtained estimates depends on the values of R_1 and R_2. In Chapter 3 and Chapter 4 we present modified versions of the KF algorithm that take into account the fact that each model parameter may follow a different TV pattern.

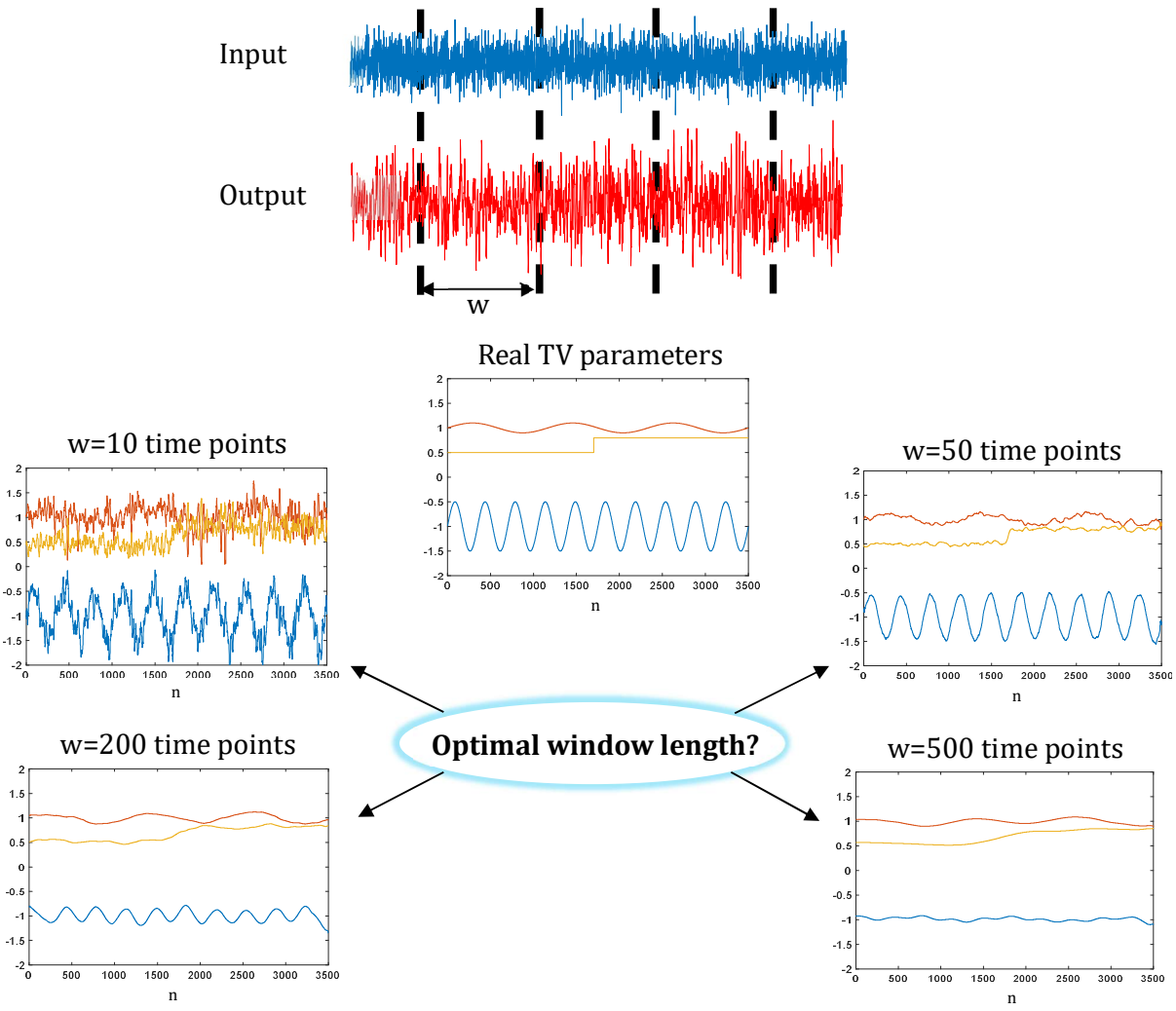


Figure 2.9 The effect of the window length (w) in quasistationary approaches for TV system identification. The input was GWN of unit variance and length of 3500 points. The output was simulated using Laguerre Volterra models ($L = 3,\ a = 0.5, Q = 1$) with TV parameters as depicted in the plot "Real TV parameters". Independent GWN was added to the output with 20dB signal to noise ratio (SNR). The input-output data was initially segmented into windows of equal length w (top panel) and overlap of w-1 time points. At each window, a stationary system was identified and the model parameters were estimated (bottom panel). Depending on w, the parameter estimates were either too noisy (w=10) or too smooth (w=500). The window length w=50 time points was selected as optimal. The amplitude of the blue parameter (sinusoid) was well approximated and the step change of the yellow parameter was followed more accurately. Nevertheless, fast tracking induced some variability to both red and yellow parameters.

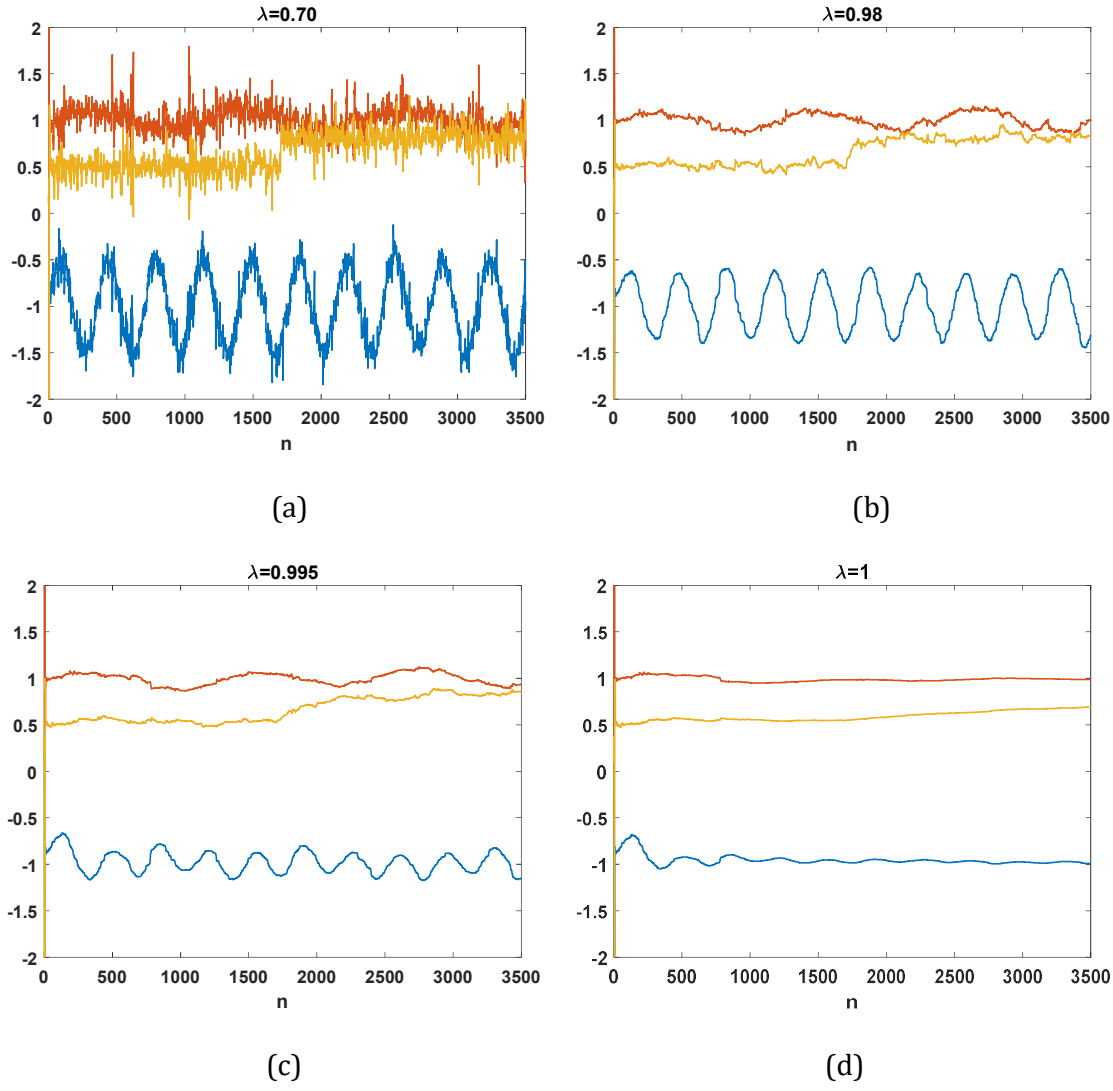


Figure 2.10 The effect of the FF in RLS (a) $\lambda = 0.7$ (β) $\lambda = 0.98$ (c) $\lambda = 0.995$ (d) $\lambda = 1$. The input-output data was simulated as described in Figure 2.9. The real TV parameters can also be found in Figure 2.9. The initial value for the covariance matrix was set to $\boldsymbol{P}(0) = 10000\boldsymbol{I}_{3\times3}$. For smaller values of λ, the estimator displays fast tracking but also higher sensitivity to noise. On the other hand, higher FF values produce more smooth estimates. For $\lambda = 1$ the parameters converge towards constant values and the system is assumed to be stationary. To increase accuracy all the parameters can be assigned with unique FFs. The red parameter requires a high valued FF since the observed variations are slow and small in magnitude. On the other hand, the blue parameter would require a smaller FF due to faster and larger changes. Finally, a TV FF would be more appropriate for the yellow parameter. The FF would initially be set close to 1, decrease during the step change and recover again to values close to 1.

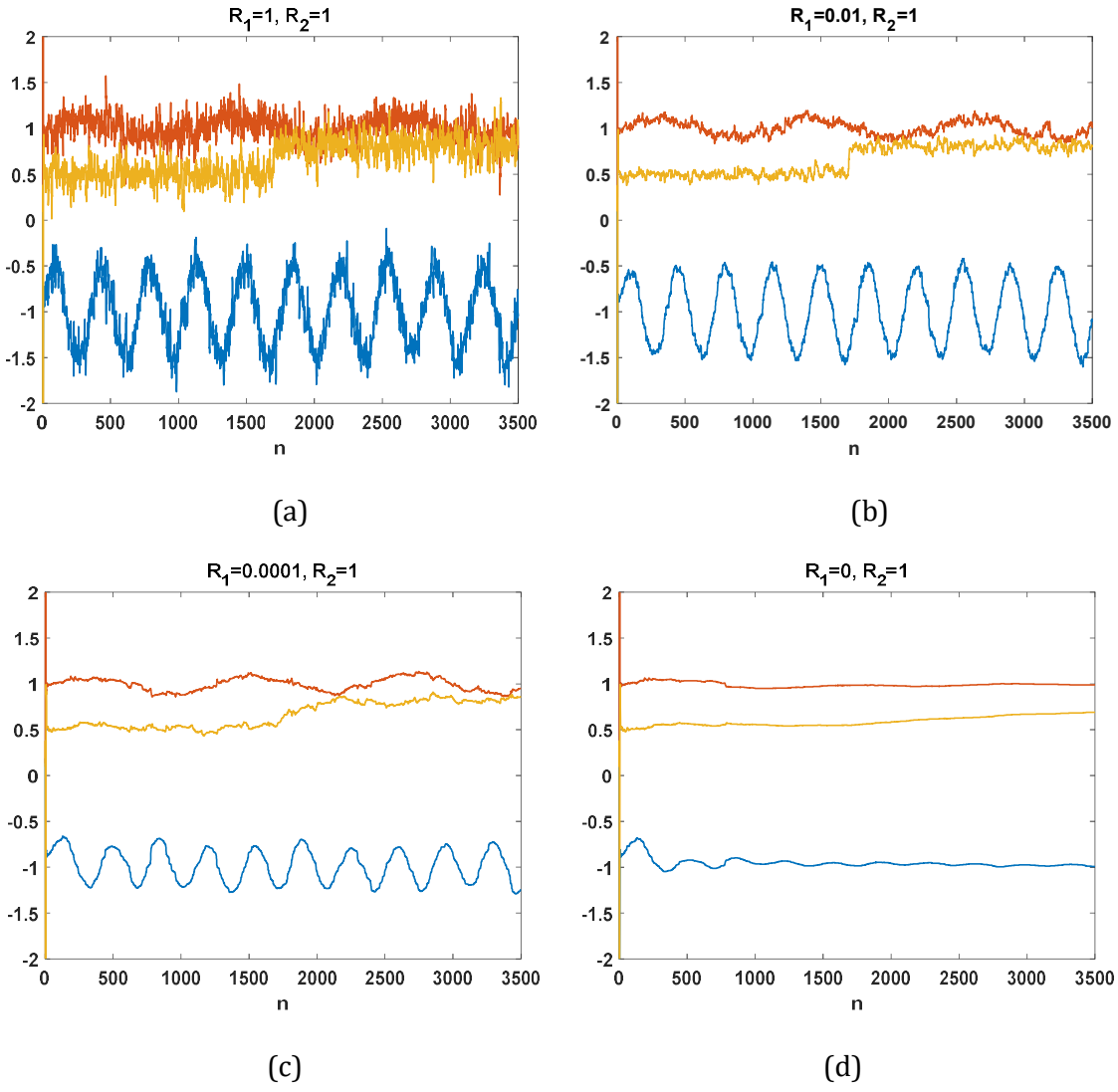


Figure 2.11 The effect of R_1 and R_2 in KF (a) $R_1 = 1, R_2 = 1$ (β) $R_1 = 0.01, R_2 = 1$ (c) $R_1 = 0.0001, R_2 = 1$ (d) $R_1 = 0, R_2 = 1$. The input-output data was simulated as described in Figure 2.9. The real TV parameters can also be found in Figure 2.9. The initial value for the covariance matrix was set to $\boldsymbol{P}(0) = 10000\boldsymbol{I}_{3\times3}$. Higher values of R_1 resulted into better tracking of the fast parameter variations at the cost of more sensitivity to noise. For $R_1 = 0$ and $R_2 = 1$ the KF estimator coincides with the RLS algorithm with unit FF. As in the RLS case, more accurate estimates would be obtained if each parameter was assigned with a unique R_1 value.

2.3.3 Model Order Selection

Model order selection in TV systems is not a straightforward task. In fact, the literature on model order selection methodologies under TV enviroments in general appears to be limited.

Recursive approaches, such as RLS and KF, require a predefined model structure. Assuming that the model structure remains constant over time, cross-validation techniques are not readily applicable. Cross-validation is based on the assumption that the system under consideration is stationary. In practice, model structure can be determined using quasistationary approaches combined with model order selection criteria like BIC and AIC. A median or maximum model order (from all windows) representative of the whole dataset is usually extracted and used for recursive parameter estimation [90], [97], [98]. Some studies rely on the model order that minimizes BIC or AIC based on the *a priori* model prediction error of the recursive approach on the whole dataset [99], [100]. Modified variants of AIC and BIC that take into account the memory of the recursive estimator have also been developed [101], [102]. Similarly, with the case of the time-invariant models, model order determination is still subject to time consuming grid-search procedures. In Chapter 3 and Chapter 4 we propose a model order selection framework based on GAs that significantly improves accuracy and computation time. Moreover, we explore in detail the behavior of BIC and AIC under nonstationary conditions.

2.4 Model Optimization Techniques

System identification consists of three main steps; the selection of an appropriate model structure, the optimization of the model hyperparameters (e.g. the Laguerre parameter in the LV models, the network weights in the LNVs) and the estimation of the model parameters. We have already described all three steps, and especially methods for estimating both time invariant and TV parameters (e.g., LS, RLS, KF etc.). Herein, we present analytically techniques that are employed in the following chapters for model order determination and hyperparameter optimization as alternatives to grid-search procedures that are usually computationally intractable (Note here that in the TV case, when we refer to model hyperparameters we also refer to the hyperparameters related with the recursive estimators e.g., forgetting factor in the RLS and innovation and measurement noise variance in the KF). Specifically, we propose heuristic optimization techniques such as mixed-integer [103] and hybrid local search-based Genetic Algorithms (GA). The basic concept is to transform the problem of model optimization into a problem of combinatorial optimization and then use the aforementioned techniques to realize simultaneously model structure and parameter identification. Recently there have been many studies on evolutionary algorithms such as

particle swarm optimization [104], ant colony optimization [105], artificial bee cologny algorithm [106], evolution strategies [107] and differential evolution [108]. However, we focus on the GA implementation of Matlab, Mathworks due to its availability, its ease of use and its flexibility with mixed-integer variables (integer variables are needed when optimizing the model order).

2.4.1 Basic Genetic Algorithm (GA)

GA is an adaptive heuristic algorithm that solves both constrained and unconstrained hard optimization problems [109]–[111]. The algorithm encodes a potential solution to a specific problem on a chromosome-like data structure and imitates the process of natural selection and evolution, making use of techniques such as selection, mutation and crossover. The evolution begins with a random initial population of chromosomes. The GA evaluates their fitness in respect to a specific objective function, known also as fitness function (J), and then generates new sample points in the search space by allocating higher reproductive opportunities to chromosomes (i.e., parents) that represent a better solution to the target problem. During the reproduction process, new chromosomes (i.e., children) are created using crossover and mutation operators. Crossover combines the information from two individual chromosomes into one chromosome, while mutation creates diversion by altering one chromosome to produce a new one. Over successive generations (i.e., iterations), the population evolves towards an optimal solution.

In more detail, let $\boldsymbol{\theta}_j = [\theta_{j_1} \quad \theta_{j_2} \quad \dots \quad \theta_{j_m}]$ be an individual chromosome (i.e., a combination of variables that consist a possible solution to the optimization problem) for $j = 1, \dots, N$, where N represents the total number of chromosomes in the population (i.e., population size). The elements θ_i for $i = 1, \dots, m$ are called genes. The search space $\Omega_{\boldsymbol{\theta}_j}$ for $\boldsymbol{\theta}_j$ is defined as,

$$\Omega_{\boldsymbol{\theta}_j} = \left\{ \boldsymbol{\theta}_j \in \boldsymbol{R}^m ; \middle|\ \theta_{j_1 min} \le \theta_{j_1} \le \theta_{j_1 max}, \dots, \theta_{j_m min} \le \theta_{j_m} \le \theta_{j_m max} \right\} \tag{2.54}$$

The outline for the basic GA is given as,

1. **[Initialize]** Generate an initial random population of N chromosomes.
2. **[Fitness]** Evaluate the fitness $J(\boldsymbol{\theta}_j)$ of each chromosome $\boldsymbol{\theta}_j$ in the current population.
3. **[Generate new population]** Repeat the following steps:
    3.1 **[Elite Selection]** Select some of the fittest chromosomes (i.e. elite chromosomes) from the current population and pass them immediately to the new population.

3.2 **[Selection]** Select parent chromosomes from the current population (other than the elite chromosomes) based on their fitness value (see section 2.4.1.1).

3.3 **[Crossover]** Form several new children (the total number of children depends on the crossover fraction for each population) from the selected parents of step 3.2 using a crossover operator (see section 2.4.1.2).

3.4 **[Mutation]** Form several new children (the fraction of mutated children for each population is defined as 1-crossover fraction) from some of the selected parents of step 3.2 by making small changes to the parent genes (see section 2.4.1.3).

4. **[Replace]** Replace the current population with the new population and go to step 2.
5. **[Test]** If one of the stopping criteria is met (e.g. number of generations, time limit, average change in the fitness function value over a specific number of generations), stop the algorithm and return the fittest chromosome in the current population.

In the following subsections, the selection, crossover and mutation operators are explained in more detail.

2.4.1.1 Selection

The selection operator is responsible for selecting parents for the next generation. Some well-known selection methods are the following,

- **Stochastic uniform selection** – Each parent chromosome corresponds to a section of a line of length proportional to its fitness value. The algorithm moves along the line in steps of equal size (*step*) and depending on the section it lands it selects a parent chromosome. The first step (*first_step*) is a random number (smaller than the step size) selected from a uniform distribution (Figure 2.12a).
- **Roulette selection** – A circular wheel is divided into N segments, where N is the population size. Each chromosome is allocated a portion of the circle which is proportional to its fitness value. A fixed point is then chosen on the wheel circumference and the wheel is rotated. The region of the wheel which lands in front of the fixed point determines the selected parent (Figure 2.12b).
- **Tournament selection** – In K-Way tournament selection, K chromosomes are selected randomly from the population. The best chromosome out of the set is chosen to be a parent. The same process is repeated for selecting the next parent. If K is large, weak chromosomes have a smaller chance to be selected (Figure 2.12c).

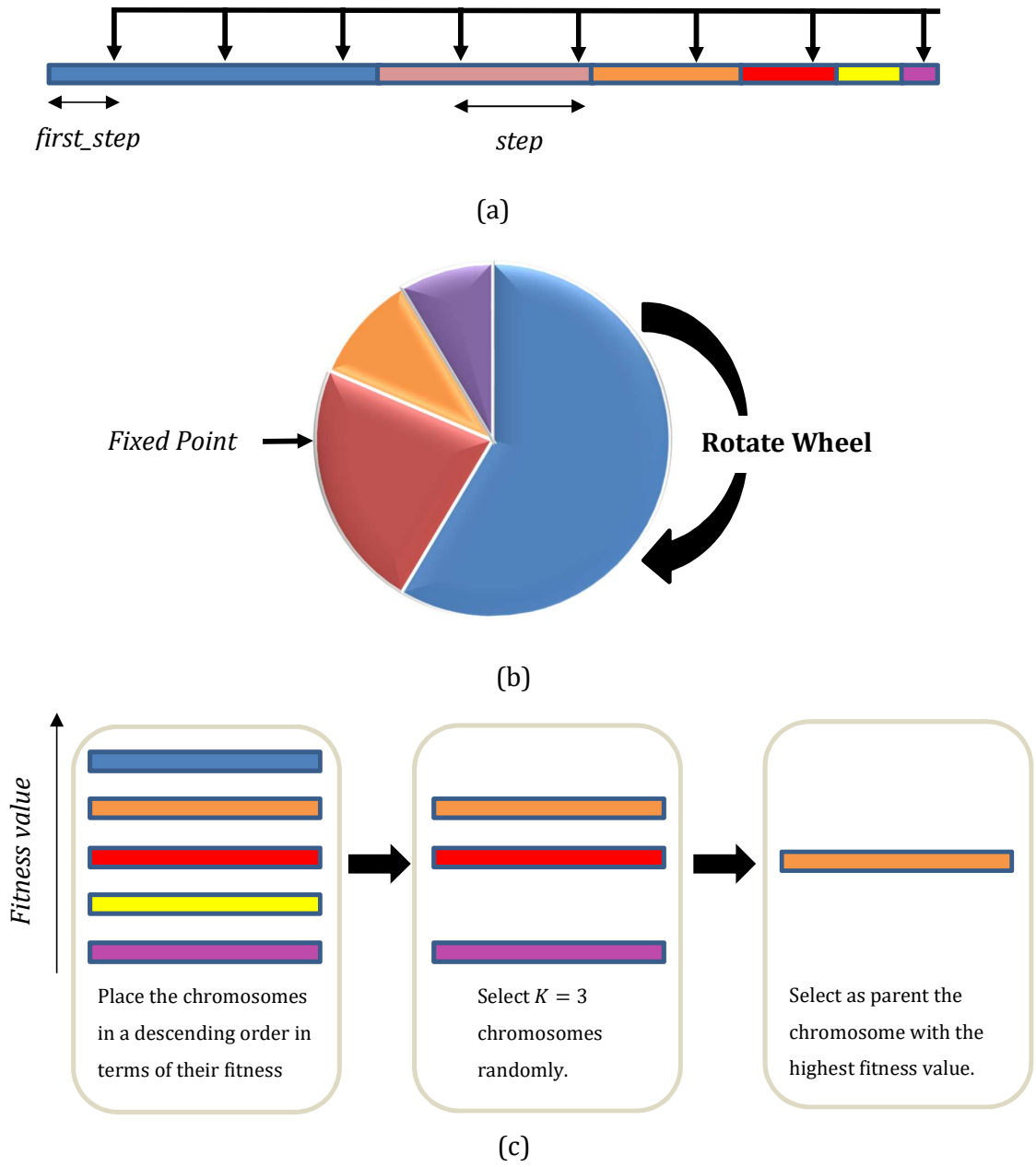


Figure 2.12 Selection operator (a) Stochastic uniform selection (b) Roulette selection (c) Tournament selection with $K = 3$.

2.4.1.2 Crossover

Crossover combines the information of two chromosomes (i.e., parents) to produce one new chromosome (i.e., child), attracting the population towards a local minimum (Figure 2.13). There are various implementations of the crossover operator; single-point, two-point, uniform, heuristic and the arithmetic crossover to name a few examples. Let $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$be the selected parent chromosomes,

- **Single-point crossover** – A single crossover point is selected. A new chromosome is produced by combining the genes from the beginning of $\boldsymbol{\Theta}_1$ to the crossover point and from the crossover point up to the final gene of $\boldsymbol{\Theta}_2$ (Figure 2.14a).
- **Two-point crossover** – Two crossover points are selected. A new chromosome is produced by combining the genes from the beginning of $\boldsymbol{\Theta}_1$ to the first crossover point, from the first to the second crossover point of $\boldsymbol{\Theta}_2$ and the second crossover point up to the final gene of $\boldsymbol{\Theta}_1$ (Figure 2.14b).
- **Uniform crossover** - A random binary vector is generated. A new chromosome is produced by selecting the genes where the vector is a 1 from $\boldsymbol{\Theta}_1$ and the vector is a 0 from $\boldsymbol{\Theta}_2$ (Figure 2.14c).
- **Heuristic crossover** – A new chromosome is produced by taking a weighted average of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ as follows,

$$if\ J(\boldsymbol{\Theta}_1) > J(\boldsymbol{\Theta}_2)\ then\ \boldsymbol{\Theta}_{new} = \boldsymbol{\Theta}_1 + r(\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2) \tag{2.55}$$

$$if\ J(\boldsymbol{\Theta}_2) > J(\boldsymbol{\Theta}_1)\ then\ \boldsymbol{\Theta}_{new} = \boldsymbol{\Theta}_2 + r(\boldsymbol{\Theta}_2 - \boldsymbol{\Theta}_1) \tag{2.56}$$

  where $r \in Uniform[0,1]$ and $J(\boldsymbol{\Theta}_j)$ is the fitness value of the chromosome $\boldsymbol{\Theta}_j$ (a higher fitness value means that the specific chromosome is a better solution to the target problem). Note that $\boldsymbol{\Theta}_{new}$ may not be a feasible solution if its genes reside outside the allowable upper and lower bounds. Thus, after a predefined number of tries in selecting r, if a feasible chromosome is not produced then the parent with the best fitness function is returned (Figure 2.14d).
- **Arithmetic crossover** – A new chromosome is produced by taking a weighted average of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ as follows,

$$\boldsymbol{\Theta}_{new} = r\boldsymbol{\Theta}_1 + (1 - r)\boldsymbol{\Theta}_2 \tag{2.57}$$

  where $r \in Uniform[0,1]$ (Figure 2.14e).

2.4.1.3 Mutation

Mutation creates diversity by making small changes in the genes of the chromosomes. The role of mutation is to occasionally pull the population out of a local optimum (minimum or maximum depending on the target problem) and direct it towards a better search space (Figure 2.13). As with crossover, there are various mutation types, e.g. gaussian, uniform and adaptive feasible. Let $\boldsymbol{\Theta}_1$ be the selected parent chromosome,

- **Gaussian mutation** – A new chromosome is produced by adding a random number taken from a Gaussian distribution with mean 0 to each gene of the parent chromosome

$\boldsymbol{\Theta}_1$. The Gaussian distribution is controlled by varying the *scale* and *shrink* parameters. *Scale* represents a measure of variance of the initial population, whereas *shrink* controls how the standard deviation shrinks along with generations,

$$\sigma_k = \sigma_{k-1}\left(1 - Shrink\frac{k}{Generations}\right) \tag{2.58}$$

where σ_k and σ_{k-1} is the standard deviation of the current and previous generation, k is the number of the current generations and $Generations$ the total number of generations.

- **Uniform mutation** – The algorithm selects a fraction of the genes of the parent chromosome $\boldsymbol{\Theta}_1$ for mutation, where each gene has a probability p_m of being mutated. A new chromosome is generated by replacing each selected gene by a random number uniformly sampled in the range of that gene.
- **Adaptive feasible mutation** – The direction and the step-size of the mutation is adapted with respect to the last generation, satisfying the lower and upper bounds of each gene, as well as the inequality constraints. The mutation process starts with the generation of a random mutation direction and an initial step size. If the mutated child resides into the infeasible region then the step size is adjusted to a smaller value, until the generated child is within the feasible region.

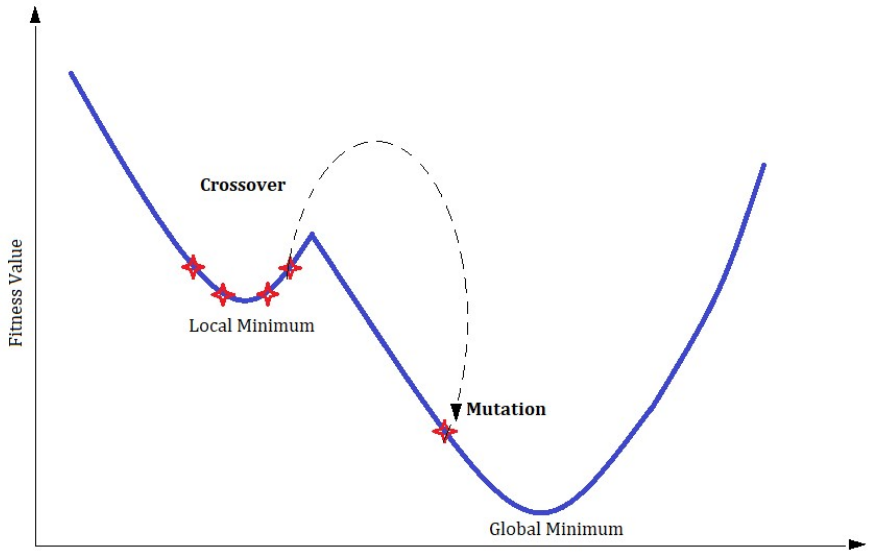


Figure 2.13 Crossover attracts the population towards a local optimum (minimum or maximum), whereas mutation occasionally pulls the population out of a local minimum and directs it towards a better search space.

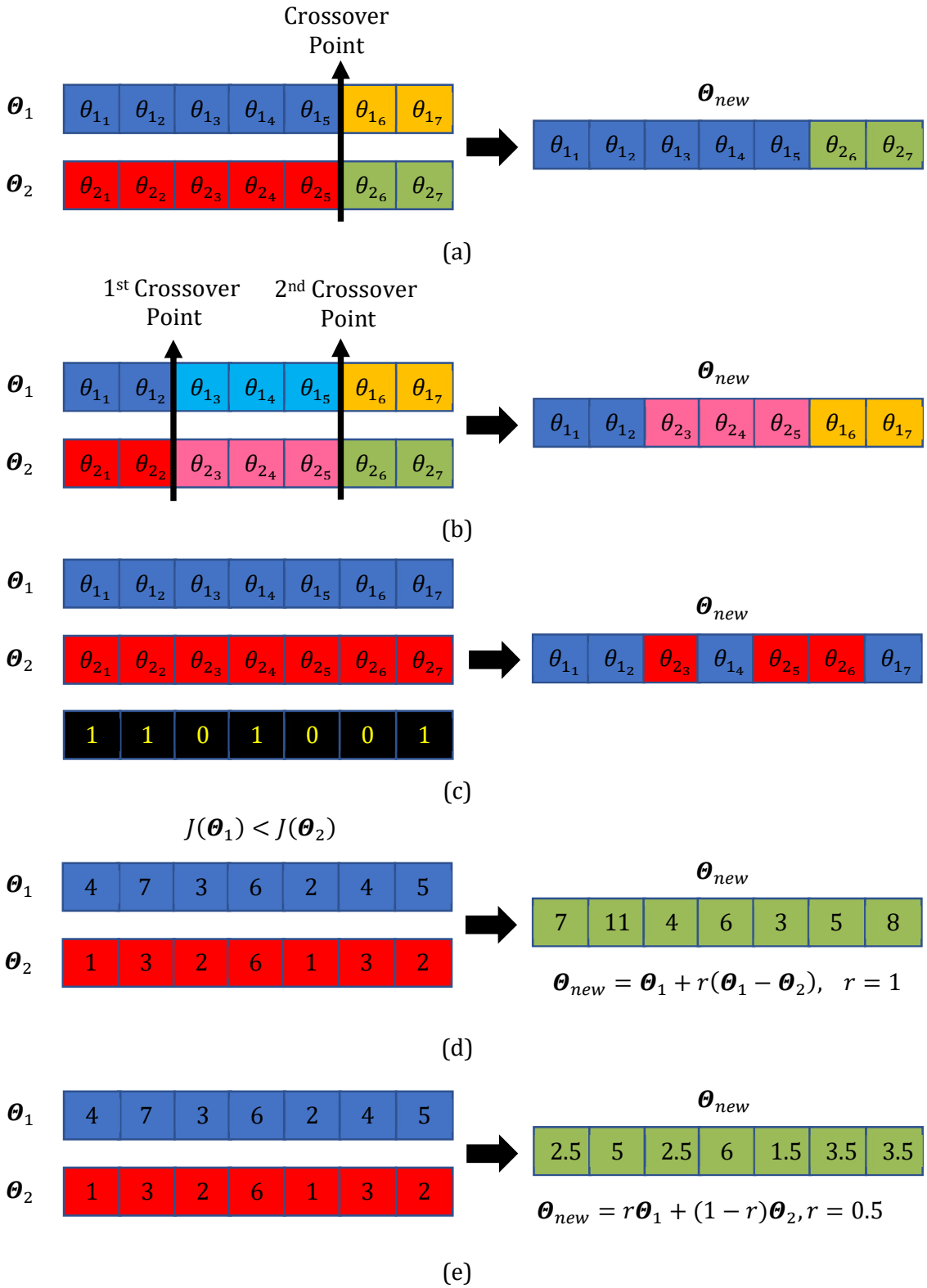


Figure 2.14 Crossover operator (a) Single-point crossover (b) Two-point crossover (c) Uniform crossover (d) Heuristic crossover with $J(\boldsymbol{\Theta}_1) < J(\boldsymbol{\Theta}_2)$ and $r = 1$ (e) Arithmetic crossover with $r = 0.5$.

2.4.1.4 Tuning the parameters of the GA

The performance of the GA may be affected from various factors. Herein, we briefly present some of these factors.

- **Generations** – Increasing the maximum number of generations (i.e. iterations) oftenly leads to improved final results. However, when the mutation operator is Gaussian, the amount of mutation is decreased in each generation by a factor that depends on the total number of generations (see Equation (2.58)), leading usually to worse results. In this work, a total number of 100-200 generations were found to be enough for our optimization problems.
- **Population size** – If the population is too small, the improvement per generation in the fitness function will be low. Increasing the population size allows the GA to explore more points in the search space [112]–[115]. However, the runtime of the algorithm may increase significantly. The GA Matlab implementation uses a default population size of 50 if the total number of variables (i.e. genes) is less than 5 and 200 otherwise. A 200 population size was found to be adequate in our work.
- **Crossover fraction** – The crossover fraction determines the fraction of the population that will be generated using crossover. A crossover fraction of 0 or 1 is not an effective strategy [112]–[117]. A crossover fraction of 1 implies that all children are generated only by recombining chromosomes from the initial population. Without mutation, the GA will not be able to cover the search space and will be restricted around a local optimum. On the other hand, a crossover fraction of 0 means that all children are mutation children and the GA becomes a random search technique where good chromosomes are perturbed constantly. Practically, a crossover fraction between 0.7 and 0.9 is desirable. The Matlab GA implementation has a default crossover fraction value of 0.8 which was found to be an appropriate value in our model optimization problems described on subsequent chapters.
- **Selection operator** – Although the selection operator does not affect significantly the results, it has been shown that tournament selection has better convergence and less computational complexity compared to other selection operators [118].

Note that in this work, the default settings of the Matlab GA implementation overall led to satisfactorily results.

2.4.2 Mixed integer Genetic Algorithm

The basic GA is developed for continuous variables. In case of mixed-integer variables (i.e., some variables are integer-valued), several modifications are made in order for the basic GA to handle integer restrictions [103].

- **Crossover operator** – Laplace crossover [119] is the main crossover operator. A number β, that satisfies the Laplace distribution, is generated as follows,

$$\beta = \begin{cases} a - bln(u), & if\ u \leq 0.5 \\ \\ a + bln(u), & if\ u > 0.5 \end{cases} \tag{2.59}$$

  where $u \in Uniform[0,1]$, a is a location parameter and b is a scaling parameter which is different than its default value when dealing with integer variables. If $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$are the selected parent chromosomes, then the child is generated as,

$$\boldsymbol{\Theta}_{new} = \boldsymbol{\Theta}_1 + \beta|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2| \ \ or \ \ \boldsymbol{\Theta}_{new} = \boldsymbol{\Theta}_2 + \beta|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2| \tag{2.60}$$

  Based on Equation (2.60), small values of β lead to children that are close to their parents and large values of β produce children further from their parents.

- **Mutation operator** – Power mutation [120] is defined as the main mutation operator. Let $\boldsymbol{\Theta}_1$ be the selected parent. Then a mutated child is created as follows,

$$\boldsymbol{\Theta}_{new} = \begin{cases} \boldsymbol{\Theta}_1 - s(\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_{1L}), & if\ t < r \\ \\ \boldsymbol{\Theta}_1 + s(\boldsymbol{\Theta}_{1U} - \boldsymbol{\Theta}_1), & if\ t \geq r \end{cases} \tag{2.60}$$

  where $\boldsymbol{\Theta}_{1L}$ and $\boldsymbol{\Theta}_{1U}$ are the lower and upper bounds of $\boldsymbol{\Theta}_1$ respectively, $t = \frac{\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_{1L}}{\boldsymbol{\Theta}_{1U} - \boldsymbol{\Theta}_1}$, $r \in Uniform[0,1]$, $s = (s_1)^p$ with $s_1 \in Uniform[0,1]$ and p being different for the case of integer or continouous variables. The strength of mutation is determined by p. For small values of p the genes of the child chromosome are less perturbed, whereas for large p values the child chromosome becomes more diverse compared to its parent.

- **Truncation for integer restrictions** – After crossover and mutation, if the integer-defined genes are not integers values anymore then these genes are truncated to $[\theta_{j_i}]$ (the integer part of the gene) or $[\theta_{j_i}] + 1$ each with 0.5 probability. This truncation

procedure avoids the generation of the same integers whenever a gene lyes between the same two consecutive integers.

- **Fitness function** – The GA optimizes the fitness function plus a penalty term for infeasibility. Specifically, the fitness of a feasible solution is its fitness values, whereas the fitness of an infeasible solution is equal to the fitness value of the worst feasible solution in the current population and the amount of constraint violation. This implies that in case of two infeasible solutions the one having the smallest constraint violation is preferred. This procedure ensures that infeasible solutions are directed towards the feasible region.

The mixed-integer GA described in this subsection will be used in Chapters 3 and 4 for simultaneous model order selection (note here that the model order is an integer variable) and hyperparameter optimization in LV and MVAR models.

2.4.3 Hybrid local search-based Genetic Algorithm

An hybrid GA (known also as memetic GA [121]) is a basic GA that incorporates local searh techniques to improve its performance. The basic GA is known to perform well as a global search technique for solving large-scale problems with multiple local optima. It can be used efficiently in the case of both non-convex and discontinuous objective functions. However, depending on the target problem, it may take a relatively long time to converge to a global optimum. On the other hand, local search algorithms converge faster to a local optimum, but they lack global perspective and they are sensitive to the initial guess of the solution. Hybridizing the GA with local search mechanisms ensures faster convergence. In practice, the GA is run for a small number of generations, performing this way an initial search of decision space and identifying solutions close to the global optimum. Then the GA solution is fed as an initial point to a faster and more efficient local search algorithm. Note here that for the mixed-integer GA, local search techniques are not supported.

Local search algorithms can be grossly discriminated in gradient-based and nongradient-based approaches. Some examples of gradient-based techniques are: conjugate gradient, steepest descent, Levenberg-Marquadt, Quasi-Newton, Gauss-Newton, interior point and sequential quadratic programming (SQP) [122]–[124]. Nongradient-based techniques include: Nelder-Meade simplex, Powell's method, Hook-Jeeves method and Pattern search [125], [126]. Gradient-based approaches usually converge faster but they require a reasonably continuous

and differentiable objective function. On the contrary, nongradient-based techniques are more flexible when dealing with discontinuities in the decision space even at a local scale.

In Chapter 5, we will apply a hybrid GA combined with interior point constrained nonlinear method to train LVN models. The interior point method (known also as barrier method) is used to solve both linear and nonlinear convex optimization problems. The main idea of the interior point method is to iteratively approach the optimal solution from the interior of the feasible set. In the following subsection, we describe in detail the interior-point approach to constrained minimization.

2.4.3.1 Interior-point Algorithm

Consider the following nonlinear optimization problem,

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad subject\ to\ h_j(\boldsymbol{x}) = 0\ and\ g_i(\boldsymbol{x}) \leq 0 \tag{2.61}$$

where $i = 1, \dots, m\ and\ j = 1, \dots, k$, $\boldsymbol{x} \in \boldsymbol{R}^{n\times 1}$, $f: \boldsymbol{R}^n \to \boldsymbol{R}$, $h_j: \boldsymbol{R}^n \to \boldsymbol{R}$ and $g_i: \boldsymbol{R}^n \to \boldsymbol{R}$ are smooth functions. The approximate problem of Equation (2.61) is,

$$\min_{\boldsymbol{x},\boldsymbol{s}} f_\mu(\boldsymbol{x},\boldsymbol{s}) = \min_{\boldsymbol{x},\boldsymbol{s}} \left\{ f(\boldsymbol{x}) - \mu \sum_{i=1}^{m} \ln(s_i) \right\} subject\ to\ h_j(\boldsymbol{x}) = 0\ and\ g_i(\boldsymbol{x}) + s_i = 0 \tag{2.62}$$

where $\boldsymbol{s} = [s_1 \quad \cdots \quad s_m]^T$ is the vector of slack variables s_i (where $s_i > 0$) for each inequality constraint g_i and $\mu > 0$. The logarithmic term in Equation (2.62) is known as the barrier function [127], [128] and it forces the algorithm to remain in the feasible set. The Lagrangian of Equation (2.62) is,

$$L(\boldsymbol{x},\boldsymbol{s},\boldsymbol{\lambda}_g,\boldsymbol{\lambda}_h) = f(\boldsymbol{x}) - \mu \sum_{i=1}^{m} \ln(s_i) + \sum_{j=1}^{k} \lambda_{h_j} h_j(\boldsymbol{x}) \sum_{i=1}^{m} \lambda_{g_i} [g_i(\boldsymbol{x}) + s_i] \tag{2.63}$$

where $\boldsymbol{\lambda}_g = \left[\lambda_{g_1} \quad \cdots \quad \lambda_{g_m}\right]^T$ and $\boldsymbol{\lambda}_h = [\lambda_{h_1} \quad \cdots \quad \lambda_{h_k}]^T$ are vectors of Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) optimality conditions of the problem of Equation (2.62) are,

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{s},\boldsymbol{\lambda}_g,\boldsymbol{\lambda}_h) = \nabla f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_{h_j} \nabla h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\boldsymbol{x}) = 0 \tag{2.64}$$

$$\nabla_{\boldsymbol{s}} L(\boldsymbol{x},\boldsymbol{s},\boldsymbol{\lambda}) = -\mu \sum_{i=1}^{m} \frac{\lambda_i}{s_i} = 0 \tag{2.65}$$

$$h_j(\boldsymbol{x}) = 0 \quad for\ j = 1, \dots, k \tag{2.66}$$

$$g_i(\boldsymbol{x}) + s_i = 0 \quad for\ i = 1, \dots, m \tag{2.67}$$

At each iteration, Equation (2.62) is solved using one of the following steps,

- **Newton step** – By applying Newton's method to the Karush-Kuhn-Tucker (KKT) conditions of Equations (2.64-2.67) the following step is obtained [127]–[129],

$$\begin{bmatrix} \boldsymbol{H} & \boldsymbol{0} & \boldsymbol{J_h}^{\boldsymbol{T}} & \boldsymbol{J_g}^{\boldsymbol{T}} \\ \boldsymbol{0} & \boldsymbol{S\Lambda} & \boldsymbol{0} & -\boldsymbol{S} \\ \boldsymbol{J_h} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{J_g} & -\boldsymbol{S} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta x} \\ \boldsymbol{\Delta s} \\ -\boldsymbol{\Delta \lambda_h} \\ -\boldsymbol{\Delta \lambda_g} \end{bmatrix} = - \begin{bmatrix} \nabla f(\boldsymbol{x}) - \boldsymbol{J_h}^{\boldsymbol{T}} \boldsymbol{\lambda_h} - \boldsymbol{J_g}^{\boldsymbol{T}} \boldsymbol{\lambda_g} \\ \boldsymbol{S\lambda_g} - \mu \boldsymbol{e} \\ h(\boldsymbol{x}) \\ g(\boldsymbol{x}) + \boldsymbol{s} \end{bmatrix} \tag{2.68}$$

  where $\boldsymbol{S} = diag(\boldsymbol{s})$, $\boldsymbol{\Lambda} = diag(\boldsymbol{\lambda_g})$, $\boldsymbol{e} \in \boldsymbol{R}^{m \times 1}$ is a vector of ones, $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{J_h}$ and $\boldsymbol{J_g}$ are the Jacobians of the constraint functions $h(\boldsymbol{x})$ and $g(\boldsymbol{x})$ respectively,

$$\boldsymbol{J_h} = \begin{bmatrix} \frac{\partial h_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial h_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_k(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial h_k(\boldsymbol{x})}{\partial x_n} \end{bmatrix} and\ \boldsymbol{J_g} = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial g_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \tag{2.69}$$

  $\boldsymbol{H}$ is the hessian of Equation (2.63),

$$\boldsymbol{H} = \nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{\lambda_g}, \boldsymbol{\lambda_h}) = \nabla^2 f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_{h_j} \nabla^2 h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_{g_j} \nabla^2 g_i(\boldsymbol{x}) \tag{2.70}$$

  Note, that $\boldsymbol{H}$ is positive definite if f is convex, each g_i is concave, each h_j is affine and each $\lambda_{h_j}, \lambda_{g_i} \geq 0$ (i.e., the nonlinear programming problem is convex). In order to solve Equation (2.68) for $(\boldsymbol{\Delta x}, \boldsymbol{\Delta s})$ LDL factorization is applied. If the projected Hessian matrix is not positive definite then the conjugate gradient step (described below) is applied. The solution $(\boldsymbol{\Delta x}, \boldsymbol{\Delta s})$ determines the direction for movement of the primal and dual variables, respectively.

- **Conjugate Gradient step** – The problem of Equation (2.62) is solved using the conjugate gradient approach [127]–[129]. The main goal is to minimize a quadratic approximation to the approximate problem of Equation (2.62) in a trust region, subject to linearized constraints. Specifically, the search direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta s})$ is given by solving the quadratic problem,

$$\min_{\boldsymbol{\Delta x}, \boldsymbol{\Delta s}} \left\{ \nabla^T f(\boldsymbol{x}) \boldsymbol{\Delta x} + \frac{1}{2} \boldsymbol{\Delta x}^T \boldsymbol{H} \boldsymbol{\Delta x} + \mu \boldsymbol{e}^T \boldsymbol{S^{-1}} \boldsymbol{\Delta s} + \frac{1}{2} \boldsymbol{\Delta s}^T \boldsymbol{S^{-1}} \boldsymbol{\Lambda} \boldsymbol{\Delta s} \right\}$$

$$\text{subject to } g(\boldsymbol{x}) + \boldsymbol{J}_g \boldsymbol{\Delta x} + \boldsymbol{\Delta s} = 0, \qquad h(\boldsymbol{x}) + \boldsymbol{J}_h \boldsymbol{\Delta x} = 0 \tag{2.71}$$

The KKT conditions associated with the problem of Equation (2.71) are exactly specified by the system of Equation (2.68) and, thus, a solution ($\boldsymbol{\Delta x}, \boldsymbol{\Delta s}$) of the system of Equation (2.68) is an optimal solution to the problem of Equation (2.71). The desirable characteristic of the conjugate gradient procedure is that there is no need to invert or factorize $\boldsymbol{H}$.

Chapter 3 Identification of multiple input, linear and nonlinear, time-varying systems for biomedical applications

Physiological systems are prone to time variations either due to unobservable inputs or inherent variability in the system dyamics. In this chapter we describe novel recursive schemes for estimating single- (SI) and multiple-input (MI) TV systems that can be applied to different type of models that are linear in the parameters. Contrary to the conventional recursive methods, the proposed schemes take into account the fact that each model parameter may exhibit different rates and amplitude variations, achieving this way superior performance under slow, fast or even mixed-mode system changes. In addition, we propose a model order selection and hyperparameter optimization framework based on mixed integer GAs that can replace computationally intractable exhaustive or grid search procedures. Most importantly, however, we explore for the first time the link between the hyperparameter values of the proposed recursive estimators, the model complexity and the systems noise and TV characteristics. The performance of our proposed methodology was evaluated using both simulated and experimental data.

Identification of multiple input, linear and nonlinear time-varying systems for biomedical applications

Kyriaki Kostoglou[1], Ronald Schondorf[2], and Georgios D. Mitsis[3]

[1] Department of Electrical and Computer Engineering, McGill University, Montreal, Canada.

[2] Department of Neurology, McGill University, Montreal, Canada.

[3] Department of Bioengineering, McGill University, Montreal, Canada.

(*Submitted to Elsevier Signal Processing*)

Abstract **—** We present novel computational schemes for estimating single- and multiple-input time-varying (TV) systems, combining a Laguerre-Volterra model formulation with improved recursive schemes based on conventional Recursive Least Squares (RLS) and Kalman Filtering (KF). The proposed recursive estimators achieve superior performance, particularly in the case of TV systems with multiple-inputs or systems that exhibit mixed-mode nonstationarities. RLS-based schemes were found to perform better in the case of TV linear systems, while the KF-based schemes were found to perform considerably better in the case of TV nonlinear systems. Model order selection and tuning of the estimator hyperparameters were implemented using Genetic Algorithms (GA), significantly improving performance and reducing computation time. Furthermore, exploiting the search efficiency in hyperparameter space yielded by the proposed GA, we rigorously examined the correlations between the hyperparameter values, the model complexity and the nonstationary characteristics of the true underlying system. The performance of the proposed algorithms was assessed using simulations and experimental data from patients undergoing head-up tilt testing for the diagnosis of vasovagal syncope.

Index Terms **—** Nonstationary systems, Recursive Least Squares, Kalman Filter, Multiple Forgetting Factors, Laguerre Expansion Technique.

3.1 Introduction

A challenging problem in the field of system identification is modeling physical phenomena that are time-varying (TV). TV system dynamics, also termed nonstationarities, may reflect inherent variability in the system dynamics, the effect of additional modulating factors (e.g. unobserved or unmodeled inputs or disturbances) or both. Identification of nonstationary systems is usually addressed by applying either quasistationary approaches, i.e. piecewise stationary models, or recursive estimation methods, i.e. updating a selected model at each time step. The Recursive Least Squares (RLS) algorithm with constant forgetting factor (FF) [86]–[88] and the Kalman Filtering (KF) technique [87]–[89] are the most widely used adaptive methods to identify TV systems.

Several variations of the aforementioned recursive identification techniques have been proposed. Under low system excitation, the RLS algorithm with constant FF may become extremely sensitive to noise. In order to tackle this problem, Fortescue et al. [130] proposed a TV FF based on the residual variance chosen to maintain the information content of the estimator constant at each time step. Sripada and Fisher [131] suggested a TV FF along with an on/off estimation criterion depending on the levels of excitation of the system under consideration. In another work, Saldago et al. [132] proposed an exponential resetting of the RLS algorithm whenever the excitation was insufficient. On the other hand, [133]–[135] introduced the concept of "directional forgetting", whereby old information is discounted non-uniformly for different parameters. In [136] and [137], the application of parallel multiple RLS algorithms with different constant FFs was proposed in order to identify systems with irregular parameter changes. Jiang and Zhang [84] proposed a sliding window blockwise least squares approach with automatically adjustable window length, whereas Warwick et al. [138] combined the classical RLS algorithm with genetic algorithms (GAs) to achieve improved performance. Studies [139]–[143] derived various TV FFs by using either gradient based techniques or recovering the true output noise from the *a priori* model prediction error. Additionally, in order to track efficiently parameters with different rates of variations, [144]–[147] suggested matrix FFs which assign unique FFs to each parameter. Adaptive methods for estimating online the noise statistics of the KF algorithm were introduced in [148]–[150] whereas, Almagbile et al. [151] reviewed and evaluated the performance of different adaptation scenarios for the process and noise covariance matrices. Most of the aforementioned recursive approaches rely on tunable parameters that are usually selected in an *ad hoc* or empirical

manner, possibly based on some prior knowledge. This may affect the accuracy of the obtained estimates and makes the comparison between different algorithms difficult. Furthermore, an important aspect in system identification is model order selection. This issue has been investigated extensively under the assumption of stationarity; however there are rather few studies focusing on TV systems [100]–[102], [152]–[154]. The performance of various model order selection criteria is relatively unknown when modeling nonstationary relationships.

In the present paper, we propose novel data-driven computational schemes that can be used to identify a wide class of TV systems (linear/nonlinear, single-input (SI) and multiple-input (MI) systems with finite memory) efficiently. These schemes were applied in combination with Laguerre-Volterra models, which can capture a wide range of nonlinear dynamic input-output causal interrelationships; however, they can be easily extended to other type of dynamic models which are linear in the parameters. In Section 3.2, we introduce the proposed TV Laguerre-Volterra formulation [19], [90] along with improved recursive computational schemes for its estimation. Eight different algorithms are considered. These include RLS with a single constant FF, an improved version of RLS with a single adaptive FF first presented in [35] and [36], as well as new versions of the RLS suitable for MI nonlinear systems employing multiple constant and adaptive FFs. They also include the conventional KF technique and two novel adaptive KF algorithms for both SI-TV and MI-TV systems. Importantly, we propose a scheme to simultaneously perform model order selection and optimization of the recursive algorithm hyperparameters, based on a mixed-integer GA, significantly improving performance and reducing running time. In Section 3.3, we describe the simulated SI and MI TV systems that were used to evaluate the proposed methods. In Section 3.4, we investigate rigorously, for the first time to our knowledge, the relationship between the hyperparameter values of the aforementioned recursive estimators and the nonstationary characteristics (magnitude, frequency) of the true underlying system. In Section 3.5, we evaluate the performance of the proposed algorithms under different types of nonstationarities and noise levels. Finally, in Section 3.6, we apply our methodology to experimental data in order to detect changes in dynamic Cerebral Autoregulation (dCA) in patients suffering from Vasovagal Syncope (VVS) during a Head-Up Tilt (HUT) protocol.

3.2 Methods

3.2.1 Multiple Input - Single Output discrete-time Volterra Model

The input-output relationship of a Q-th order, nonlinear, MI, single-output, causal dynamic system can be expressed by the following Volterra model in discrete time [1],

$$y(n) = k0 + \sum_{i=1}^{I}\sum_{m} k_{1_{(x_i)}}(m)x_i(n-m) + \sum_{i_1,i_2=1}^{I}\sum_{m_1,m_2} k_{2_{(x_{i_1},x_{i_2})}}(m_1,m_2)x_{i_1}(n-m_1)x_{i_2}(n-m_2) + \cdots + \sum_{i_1..i_Q=1}^{I}\sum_{m_1..m_Q} k_{Q_{\left(x_{i_1},\ldots,x_{i_Q}\right)}}(m_1,\ldots,m_Q)x_{i_1}(n-m_1)\ldots x_{i_Q}(n-m_Q) + \varepsilon(n) \quad (3.1)$$

where I is the total number of inputs, $x_i(n)$ is the i-th input, $y(n)$ is the output, $\varepsilon(n)$ is zero-mean white noise and k_Q are the Q-th order Volterra kernels of the system. Volterra kernels can be interpreted as weighting functions that describe the effect of past values of input x_i ($k_{1_{(x_i)}}$; linear kernels), as well as the effect of the Q-th order products between past values of each input ($k_{Q_{\left(x_{i_1},\ldots,x_{i_Q}\right)}}$ for $x_{i_1} = \cdots = x_{i_Q}$; nonlinear self-kernels) and past values of different inputs ($k_{Q_{\left(x_{i_1},\ldots,x_{i_Q}\right)}}$ for some $x_{i_1},\ldots,x_{i_Q}$ different; nonlinear cross-kernels) in order to generate the output signal. The zeroth-order Volterra kernel k_0 is the output of the system when all inputs are absent. An efficient way to estimate these kernels is the Laguerre expansion technique (LET) [29]. Specifically, the discretized Volterra kernels of the system can be expanded in terms of the orthonormal basis of discrete-time Laguerre functions (DLFs) as follows:

$$k_{Q_{\left(x_{i_1},\ldots,x_{i_Q}\right)}}(m_1,\ldots,m_Q) = \sum_{j_1=0}^{L_{x_{i_1}}-1} \cdots \sum_{j_Q=0}^{L_{x_{i_Q}}-1} c_{Q_{(x_{i_1},\ldots,x_{i_Q})}}(j_1,\ldots,j_Q)\, b_{j_1(x_{i_1})}(m_1)\ldots b_{j_Q\left(x_{i_Q}\right)}(m_Q) \quad (3.2)$$

where $b_{j_{(x_i)}}(m)$ is the j-th order Discrete-time Laguerre function (DLF) corresponding to input i,

$$b_{j_{(x_i)}}(m) = \alpha_i^{(m-j)/2}(1-\alpha_i)^{1/2}\sum_{k=0}^{j}(-1)^k\binom{m}{k}\binom{j}{k}\alpha_i^{j-k}(1-\alpha_i)^k \quad (3.3)$$

The Laguerre parameter α_i lies between zero and one and determines the rate of exponential decay of the DLFs. The effects of inputs with different dynamic characteristics can be captured more accurately and parsimoniously (lower number of DLFs) by using different Laguerre parameters for each input. Therefore, choosing an appropriate value for these parameters is important as it affects model accuracy and complexity. By combining Equations (3.1) and (3.2) we can write,

$$\boldsymbol{y} = \boldsymbol{V}\boldsymbol{c} + \boldsymbol{\varepsilon} \tag{3.4}$$

where $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ is the output observation vector, $\boldsymbol{c} \in \boldsymbol{R}^{d\times 1}$ is the vector of the unknown expansion coefficients and $\boldsymbol{\varepsilon} \in \boldsymbol{R}^{N\times 1}$ is assumed to be a zero-mean white noise vector. The total number of the expansion coefficients, d, is equal to $((L+Q)!/L!\,Q!)$ where $L = L_{x_1} + \cdots + L_{x_I}$. $\boldsymbol{V} \in \boldsymbol{R}^{N\times d}$ is a matrix containing the convolution of all inputs with the Laguerre functions $v_j^{(i)} = x_i * b_j{}^{(i)}$, as well as nonlinear products between them.

When modeling stationary dynamic systems, the expansion coefficients $c_{Q_{(x_{i_1},\ldots,x_{i_Q})}}(j_1,\ldots,j_Q)$ are constant over time and can be estimated using ordinary least-squares (OLS) based on the input and output data [29],

$$\hat{\boldsymbol{c}} = (\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V})^{-\boldsymbol{1}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{y} \tag{3.5}$$

In the case of nonstationary systems, the relationship between the inputs and output is time-varying. Thus, the unknown model parameters (expansion coefficients) should also vary over time.

3.2.2 Recursive Estimation Schemes

3.2.2.1 Recursive Least Squares with Constant Forgetting Factor (RLSC)

One common technique used to estimate and track parameters in TV systems is the RLS algorithm [86]–[88] summarized as,

$$e(n) = y(n) - \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\hat{\boldsymbol{c}}(n-1) \tag{3.6a}$$

$$r(n) = \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n) \tag{3.6b}$$

$$\boldsymbol{K}(n) = \frac{\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n)}{\lambda + r(n)} \tag{3.6c}$$

$$\boldsymbol{P}(n) = \frac{1}{\lambda}[\boldsymbol{P}(n-1) - \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{K}(n)\boldsymbol{P}(n-1)] \tag{3.6d}$$

$$\hat{\boldsymbol{c}}(n) = \hat{\boldsymbol{c}}(n-1) + \boldsymbol{K}(n)e(n) \tag{3.6e}$$

where $e(n)$ is the output prediction error, $\boldsymbol{\varphi}(n) \in \boldsymbol{R}^{d\times 1}$ is the n-th transposed row of the matrix $\boldsymbol{V}$ (4-6), $\boldsymbol{K}(n) \in \boldsymbol{R}^{d\times 1}$ is a gain matrix and $\boldsymbol{P}(n) \in \boldsymbol{R}^{d\times d}$ is the parameter covariance matrix corresponding to time n. The initial value for this matrix is usually a diagonal matrix $\boldsymbol{P}(0) = P_0\boldsymbol{I}_{d\times d}$. Equation (3.6e) updates the current estimates by adding a correction term to the previous estimates. The correction term is the error between the predicted and observed output multiplied by a gain factor. The forgetting factor (FF) λ lies between 0 and 1. It assigns exponentially lower weight to less recent error samples and adapts the size of the matrix $\boldsymbol{P}(n)$ in order to track the TV parameters of the system. The LS error cost function is replaced with an exponential weighted sum of the squared error signal,

$$J(n) = \sum_{k=0}^{n} \lambda^{n-k} e^2(k) \tag{3.7}$$

where the FF λ determines the exponentially decaying memory of the algorithm. The corresponding effective number of samples is given by,

$$n_{ef} = \sum_{k=0}^{n-1} \lambda^n \tag{3.8}$$

Asymptotically, Equation (3.8) becomes,

$$n_{ef}{}^{\infty} = \lim_{n\to\infty} n_{ef} = \frac{1}{1-\lambda} \tag{3.9}$$

Based on Equation (3.9), when λ gets closer to 1 more weight is given to less recent error samples. On the other hand, when λ gets closer to 0, the memory length of the algorithm decreases and less recent information is discarded, enabling faster adaptation and tracking. However, for smaller values of λ, the estimator exhibits higher sensitivity to noise and leads to more variable parameter estimates. Therefore, the choice of the FF is crucial for good performance. Here we select λ using the GA scheme described in Section 3.2.3.

3.2.2.2 Recursive Least Squares with Adaptive Forgetting Factor (RLSA)

When using a constant FF, the parameter estimates are computed based on time windows of length $n_{ef}{}^{\infty}$ approximately. Hence, it is implicitly hypothesized that changes occur with a constant rate. However, this is not always the case. The system parameters may exhibit intervals of slow, fast or even abrupt changes. Subsequently the value of the FF should be able

to vary accordingly. To achieve this, we modified and applied the computational scheme proposed in [155], [156]. The proposed procedure detects parameter changes by estimating the influence of new observations using Cook's distance [157]. The algorithm follows the structure given in Equation (3.6). The key difference is that λ is now $\lambda(n)$; i.e. the FF is a function of time. The update equations for the TV FF are as follows,

$$C(n) = \frac{\boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n)\, e^2\,(n)}{\big(1 + \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n)\big)\hat{\sigma}^2(n-1)} \tag{3.10a}$$

$$S(n) = P\Big(\chi_d^2 > C(t)\Big) \tag{3.10b}$$

$$\lambda(n) = \min\big[max\big(\lambda_{min}\,, S(n)\big)\,, \lambda_{max}\big] \tag{3.10c}$$

$$\hat{\sigma}^2(n) = \lambda_e \hat{\sigma}^2(n-1) + (1-\lambda_e)e^2(n) \tag{3.10d}$$

where we have modified the scheme proposed in [155], [156] by recursively estimating the variance of the prediction error $\hat{\sigma}^2(n)$ (Equation (3.10d)) to achieve further adaptability. Equation (3.10d) can be described as an exponential moving average, where λ_e is a constant smoothing factor between 0 and 1 and, importantly, its precise value is automatically selected by the proposed GA scheme (Section 3.2.3). The influence of new data is translated into a FF by comparing Cook's distance $C(n)$ with a χ^2 distribution with d degrees of freedom (Equation (3.10b)), where d is the total number of unknown coefficients and λ_{min}, λ_{max} are lower and upper bounds for the FF respectively. For instance, if a parameter change is detected, $C(n)$ will increase whereas $S(n)$ will become closer to zero. Consequently, based on Equation (3.10c), $\lambda(n)$ will acquire smaller values. The bounds λ_{min} and λ_{max} are also optimized using the GA (Section 3.2.3).

3.2.2.3 Recursive Least Squares with Multiple Adaptive Forgetting Factors (RLSMA)

In the case of MI systems, different inputs typically exhibit different dynamics, suggesting that the corresponding model parameters may vary with different rates. A single constant FF may therefore compromise overall performance, since all model parameters are forced to vary in the same manner. For this reason, we propose a multiple forgetting factor scheme, whereby we allow all FFs to be adaptive. The resulting update equations for the unknown coefficient vector at time point *n* are written as,

$$e(n) = y(n) - \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\hat{\boldsymbol{c}}(n-1) \tag{3.11a}$$

$$r(n) = \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n) \tag{3.11b}$$

$$\boldsymbol{K_a}(n) = \boldsymbol{P}(n-1)\boldsymbol{\varphi}(n) = \begin{bmatrix} K_{a_1}(n) \\ \vdots \\ K_{a_d}(n) \end{bmatrix} \tag{3.11c}$$

$$\boldsymbol{K}(n) = \begin{bmatrix} \dfrac{K_{a_1}(n)}{\lambda_1(n) + r(n)} \\ \vdots \\ \dfrac{K_{a_d}(n)}{\lambda_d(n) + r(n)} \end{bmatrix} \tag{3.11d}$$

$$\hat{\boldsymbol{c}}(n) = \hat{\boldsymbol{c}}(n-1) + \boldsymbol{K}(n)e(n) \tag{3.11e}$$

$$\boldsymbol{W}(n) = \boldsymbol{P}(n-1) - \boldsymbol{K}(n)\boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1) \tag{3.11f}$$

$$\boldsymbol{P}(n) = \boldsymbol{\Lambda}(n)\boldsymbol{W}(n)\boldsymbol{\Lambda}(n) \tag{3.11g}$$

$$\boldsymbol{\Lambda}(t) = \begin{bmatrix} \sqrt{\dfrac{1}{\lambda_1(n)}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sqrt{\dfrac{1}{\lambda_d(n)}} \end{bmatrix} \tag{3.11h}$$

where $\boldsymbol{\Lambda}(t) \in \boldsymbol{R}^{d \times d}$ is a diagonal matrix and each element of the diagonal corresponds to distinct model parameters. For instance, for a two-input linear system, two FFs are used ($\lambda_{1,2\ldots,L_{x_1}}(n) = \lambda_{x_1}(n)$ and $\lambda_{L_{x_1}+1,\ldots,d}(n) = \lambda_{x_2}(n)$). In the case of a two-input, second-order system we assign two extra FFs to the coefficients of the second-order self-kernels of each input ($\lambda_{x_{1s}}(n)$ and $\lambda_{x_{2s}}(n)$ respectively) and one extra FF to the coefficients of the cross-kernels ($\lambda_{x_{cr}}(n)$), as we observed that these may vary with considerably different rates compared to their first-order counterparts (typically, $\lambda_{x_{cr}}(n)$ acquires values within the range between $\lambda_{x_{1s}}(n)$ and $\lambda_{x_{2s}}(n)$). In turn, this resulted in greatly improved tracking accuracy. The gain vector $\boldsymbol{K}(n)$ proposed in [144] and [158] was modified as follows: Instead of dividing the gain of each coefficient with $1 + r(n)$, we used the scalar quantity $\lambda_j(n) + r(n)$ instead, where j corresponds to the j-th model coefficient, in order to take into account the effect of the FFs (Equation (3.11c-d)). The proposed gain vector yielded better performance compared to [144] and [158]. The influence of new data is assessed using Cook's distance and the update equations for the FFs become,

$$C_{x_i}(n) = \frac{\boldsymbol{\varphi}_{\boldsymbol{x_i}}{}^{\boldsymbol{T}}(n)\boldsymbol{P}_{\boldsymbol{x_i}}(n-1)\boldsymbol{\varphi}_{\boldsymbol{x_i}}(n)\, e^2\,(n)}{\left(1+\boldsymbol{\varphi}_{\boldsymbol{x_i}}{}^{\boldsymbol{T}}(n)\boldsymbol{P}_{\boldsymbol{x_i}}(n-1)\boldsymbol{\varphi}_{\boldsymbol{x_i}}(n)\right)\hat{\sigma}^2(n-1)} \tag{3.12a}$$

$$S_{x_i}(n) = P\left(\chi^2_{d_i} > C_{x_i}(n)\right) \tag{3.12b}$$

$$\lambda_{x_i}(n) = \min\left[max\left(\lambda_{min_i}, S_{x_i}(n)\right), \lambda_{max_i}\right] \tag{3.12c}$$

$$\hat{\sigma}^2(n) = \lambda_e\hat{\sigma}^2(n-1) + (1-\lambda_e)e^2(n) \tag{3.12d}$$

where d_i is the total number of parameters for input x_i and $\boldsymbol{\varphi}_{\boldsymbol{x_i}}(n)$, $\boldsymbol{P}_{\boldsymbol{x_i}}(n)$ contain elements that correspond only to input x_i and $\lambda_{min_i}, \lambda_{max_i}$ are the lower and upper bounds respectively for λ_{x_i}. In the case of nonlinear systems $\boldsymbol{\varphi}_{\boldsymbol{x_i}}(n)$, $\boldsymbol{P}_{\boldsymbol{x_i}}(n)$ contain elements that correspond only to the first-order components of input x_i, and $\lambda_{x_i}(n)$ corresponds to the first-order kernel coefficients. Additionally, $\lambda_{x_{is}}(n)$ corresponds to the second-order self-kernel coefficients of input x_i and $\lambda_{x_{cr}}(n)$ to the cross-kernel coefficients. All the upper and lower bounds of the FFs and λ_e are selected using the GA optimization scheme presented in Section 3.2.3. Note that if the FFs are time-invariant, the algorithm will be referred to as *Recursive Least Squares with Multiple constant Forgetting Factors (RLSM).*

3.2.2.4 Kalman Filter (KF)

Another approach that has been extensively used to identify nonstationary systems is the Kalman filter (KF) [87]–[89]. KF can be described as a statistical adaptive approach that, in contrast to RLS, assumes prior knowledge of the true parameter variations. Specifically, parameter changes are modeled as a random walk under Gaussian white driving noise with covariance matrix equal to $\boldsymbol{R_1} = R_1\boldsymbol{I}_{d\times d}$ where $R_1 = \sigma_w^2$. Small values of R_1 indicate that small changes are expected. If R_1 values are large, better tracking of fast variations can be achieved but at the cost of more sensitivity to noise. The KF formulation is based on the hypothesis that the measurement noise (i.e. innovations) is also Gaussian and white with variance equal to R_2. The Kalman filter algorithm can be summarized by the following set of equations,

$$e(n) = y(n) - \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\hat{\boldsymbol{c}}(n-1) \tag{3.13a}$$

$$r(n) = \boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n) \tag{3.13b}$$

$$\boldsymbol{K}(n) = \frac{\boldsymbol{P}(n-1)\boldsymbol{\varphi}(n)}{R_2 + r(n)} \tag{3.13c}$$

$$\boldsymbol{P}(n) = \boldsymbol{P}(n-1) + \boldsymbol{R_1} - \boldsymbol{K}(n)\boldsymbol{\varphi}^{\boldsymbol{T}}(n)\boldsymbol{P}(n-1) \tag{3.13d}$$

$$\hat{\boldsymbol{c}}(n) = \hat{\boldsymbol{c}}(n-1) + \boldsymbol{K}(n)e(n) \tag{3.13e}$$

where $\boldsymbol{K}$ is the Kalman gain matrix, which minimizes the *a posteriori* error covariance $\boldsymbol{P}$. The initial value for $\boldsymbol{P}$ is a diagonal matrix $\boldsymbol{P}(0) = P_0\boldsymbol{I}_{d\times d}$. We optimally select R_1, R_2 and P_0 using the GA (Section 3.2.3). For $R_1 = 0$ and $R_2 = 1$, the KF and RLS with unit FF become identical, i.e. the system is assumed to be stationary.

3.2.2.5 Adaptive Kalman Filter (KFA)

The KFA algorithm follows the same concept as RLSA. To accurately track parameter variations, the innovation variance R_2 is held fixed to a constant value and we have modified the standard KF by allowing the process noise covariance matrix $\boldsymbol{R_1}(n) = R_1(n)\boldsymbol{I}_{d\times d}$ to be updated adaptively at each time step based on the following equation:

$$R_1(n) = \lambda_w R_1(n-1) + (1-\lambda_w)e^2(n) \tag{3.14}$$

where λ_w is a smoothing factor between 0 and 1, the value of which is selected by the proposed GA optimization scheme (Section 3.2.3). Equation (3.14) is based on the assumption that system changes are reflected on the prediction error. Hence, when the rate of change of the corresponding model parameters increases, the prediction error variance will also tend to increase. We examined all possible combinations i.e. adapting both R_1 and R_2 or holding R_1 fixed and adapting R_2 and vice versa and we found that the above scheme yielded the best results.

3.2.2.6 Adaptive Kalman Filter for Multiple Inputs (KFMA)

In the case of MI systems, we extended the proposed updating scheme for the diagonal process noise covariance matrix $\boldsymbol{R_1}$ (Eq. (3.14)) as follows,

$$\boldsymbol{R_1}(n) = diag\left\{R_{1_j}(n)\right\}, \qquad j = [1\ 2\ \dots\ d] \tag{3.15}$$

where $R_{1_j}(n) = \sigma_{w_j}^2(n)$ is an adaptive variance estimate for the j-th coefficient. Hence Equation (3.15) becomes,

$$\boldsymbol{R_1}(n) = \boldsymbol{\Lambda_w}\boldsymbol{R_1}(n-1) + (\boldsymbol{I}-\boldsymbol{\Lambda_w})e^2(n) \tag{3.16a}$$

$$\boldsymbol{\Lambda_w} = diag\left\{\lambda_{w_j}(n)\right\}, \qquad j = [1\ 2\ \dots\ d] \tag{3.16b}$$

Similarly to the RLSMA case, for MI linear systems the parameters associated with different inputs are likely to vary with different rates over time. For instance, two TV process noise

variances are estimated recursively ($R_{1_{x_1}}(n) = R_{1_{1,2,\ldots,L_{x_1}}}(n)$ and $R_{1_{x_2}}(n) = R_{1_{L_{x_1+1},\ldots,d}}(n)$) in the case of a two-input system,

$$R_{1_{x_1}}(n) = \lambda_{w_{x_1}} R_{1_{x_1}}(n-1) + \left(1 - \lambda_{w_{x_1}}\right) e^2(n) \quad (3.17a)$$

$$R_{1_{x_2}}(n) = \lambda_{w_{x_2}} R_{1_{x_2}}(n-1) + \left(1 - \lambda_{w_{x_2}}\right) e^2(n) \quad (3.17b)$$

The quantities $\lambda_{w_{x_1}}$ and $\lambda_{w_{x_2}}$ in Equations (3.17a,b) are smoothing factors assigned to the two inputs. For a two-input nonlinear system, additional variance terms are used for the estimation of the cross- and self-kernel coefficients ($R_{1_{cr}}(n)$ and $R_{1_{x_{is}}}(n)$ respectively). The hyperparameters λ_{w_j} and R_2 are chosen optimally using the proposed GA methodology in Section 3.2.3. Note that when $\boldsymbol{R_1}$ is time-invariant, the algorithm will be referred to as *Kalman Filter for Multiple Inputs (KFM).*

3.2.3 Model order selection and optimization of the recursive estimator hyperparameters

Model order selection and tuning of the recursive estimators hyperparameters can be realized using e.g. exhaustive search. However, this can be a very time consuming or impractical procedure. To mitigate this, we applied heuristic techniques, specifically Genetic Algorithms (GA) [109], [159]. GA can be described as adaptive search algorithms inspired from natural evolution. In each generation, the fitness function of a population of candidate solutions, called individuals, is evaluated. The fittest solutions are selected and used in the next iteration. Usually, the algorithm is terminated if a maximum number of iterations or a predetermined fitness level has been reached. In the present study a mixed integer GA [103] was used to optimize the model order complexity L_{x_i} for each input, the values of the Laguerre parameters α_i as well as all the hyperparameters of each recursive estimator. The candidate solutions evaluated by the GA, depending on the estimator, are of the form presented in Table 3.1. The GA evaluates the performance of a specific recursive estimator based on the entire dataset using different combinations of hyperparameters (candidate solutions) until it reaches a global minimum for a predefined fitness function.

In order to examine the behavior of different model order selection criteria in the case of nonstationary systems, the fitness functions that we considered were the negative Variance Accounted For (nVAF), the Bayesian Information Criterion (BIC) [39] and the Akaike

Information Criterion (AIC) [40], evaluated over the entire training dataset. For each candidate solution, the procedure can be summarized as follows,

- Compute the matrix $\boldsymbol{V}$ based on the selected values of L_{x_i}, a_i.
- Obtain the TV estimates $\hat{\boldsymbol{c}}$ using one of the recursive estimators described in Section 3.2.2.
- Compute the predicted output $\hat{y}(n) = \boldsymbol{V}\hat{\boldsymbol{c}}(n)$.
- Compute the error between the actual (noisy) and the predicted output, $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$.
- Evaluate the selected fitness function (nVAF, BIC or AIC) for the entire training dataset,

$$nVAF(d) = -\left(1 - \frac{\sigma_{\boldsymbol{e}}^2}{\sigma_{\boldsymbol{y}}^2}\right) \tag{3.18}$$

$$BIC(d) = \frac{N}{2}\log\left(\frac{J}{N}\right) + \frac{d}{2}\log N \tag{3.19}$$

$$AIC(d) = \frac{N}{2}\log\left(\frac{J}{N}\right) + d \tag{3.20}$$

where N is the length of the training data set, d the total number of coefficients (i.e. total number of Laguerre functions used), $\sigma_{\boldsymbol{e}}^2$ the variance of the prediction error, $\sigma_{\boldsymbol{y}}^2$ the variance of the actual output and $J = \sum_{n=1}^{N} e^2(n)$.

Table 3.1 Hyperparameters optimized by the GA

Method	Two input linear system	Two input nonlinear system
RLSC	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda, P_0]$
RLSA	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{min}, \lambda_{max}, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{min}, \lambda_{max}, P_0]$
KF	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, R_1, R_2, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, R_1, R_2, P_0]$
KFA	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_w, R_2, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_w, R_2, P_0]$
RLSM	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{x_1}, \lambda_{x_2}, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{x_1}, \lambda_{x_2}, \lambda_{x_{1s}}, \lambda_{x_{2s}}, \lambda_{x_{cr}}, P_0]$
RLSMA	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{min_1}, \lambda_{min_2}, \lambda_{max_1}, \lambda_{max_2}, \lambda_e, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{min_1}, \lambda_{min_2}, \lambda_{min_{1S}}, \lambda_{min_{2S}}, \lambda_{min_{cr}}, \dots$ $, \lambda_{max_1}, \lambda_{max_2}, \lambda_{max_{1S}}, \lambda_{max_{2S}}, \lambda_{max_{cr}}, \lambda_e, P_0]$
KFM	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, R_{1_{x_1}}, R_{1_{x_2}}, R_2, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, R_{1_{x_1}}, R_{1_{x_2}}, R_{1_{x_{1s}}}, R_{1_{x_{2s}}}, R_{1_{x_{cr}}}, R_2, P_0]$
KFMA	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{w_{x_1}}, \lambda_{w_{x_2}}, R_2, P_0]$	$[L_{x_1}, L_{x_2}, \alpha_1, \alpha_2, \lambda_{w_{x_1}}, \lambda_{w_{x_2}}, \lambda_{w_{x_{1s}}}, \lambda_{w_{x_{2s}}}, \lambda_{w_{x_{cr}}}, R_2, P_0]$

In order to acquire feasible solutions and reduce the computational complexity bounds were set for each unknown variable; i.e. $L \in [1\ 15]$, $\alpha \in [0.1\ 0.9]$, $\lambda \in [0.1\ 1]$, and $R_1, R_2, P_0 \in [0\ Inf]$.

3.2.4 Practical Issues

3.2.4.1 Persistence of excitation

When applying recursive schemes to experimental data, the covariance matrix $\boldsymbol{P}(n)$ may become close to singular. Moreover, when the excitation of the system is poor, the traditional FF scheme may lead to a phenomenon referred to as the covariance "wind-up" problem [130], [144]. During low excitation periods, the covariance matrix is continuously divided by a fixed FF and thus grows exponentially in time. If excitation reoccurs then the gain will be already very large and the parameter estimates will change abruptly affecting not only the overall performance of the estimator but also the interpretation of the underlying variations of the system dynamics. Therefore, during low excitation, it would be appropriate for the FF to take values close to 1. It is also possible that even under uniform excitation, different rates of parameter changes may lead to "wind up" problems [144]. All these problems corroborate the need for applying multiple variable FFs. Comparing Equations (3.6d) and (3.13d), under excitation failures ($\boldsymbol{\varphi}(n) \approx 0$), the KF is considerably more robust compared to RLS. Covariance blowup for the KF follows a linear growth rate, whereas this rate is exponential for RLS. By applying KFM and KFMA, the probability of estimator "wind-up" problems decreases considerably. We suggest imposing additional mechanisms to ensure that $\boldsymbol{P}(t)$ will not grow abruptly for both the RLS and KF schemes [135], [160]–[162]. For example, the regularized constant trace (RCT) technique [160], [161] in combination with the proposed algorithms results into more accurate and stable estimates. RCT scales the $\boldsymbol{P}$-matrix by keeping its trace constant at each time step and adding a unit matrix scaled by a small positive constant,

$$\boldsymbol{P}(n) = \frac{q\boldsymbol{P}(n)}{trace[P(n)]} + r\boldsymbol{I} \tag{3.21}$$

The additional unknown variables (q, r) are also tuned during the model order selection and optimization step (Section 3.2.3) using the GA.

3.2.4.2 Smoothing

Fast adaptation or low SNR levels may lead to parameter estimates with high variance, therefore a smoothing procedure can be used, particularly when the system descriptors are not needed in real time. A smoother utilizes future and past data when computing the estimates at a given time point; thus, it is applied independently once the TV model parameters have been computed.

RLS Smoothing: For all the RLS-type estimators, the basic smoothing recursions for the parameter estimates are given in [163]. We have extended these to accommodate adaptive or multiple FFs,

$$\hat{c}_j^s(n) = \hat{c}_j(n) + \lambda_j(n)\left[\hat{c}_j^s(n+1) - \hat{c}_j(n)\right] \tag{3.22}$$

where $j = [1\ 2\ \dots\ d]$, the superscript s refers to the smoothed estimates and $\lambda_j(n)$ is the FF that was either assigned by the GA to the j-th parameter in the RLSM case (and it is constant through time) or adaptively estimated in the RLSMA case. The smoothed estimates are obtained by updating the extracted estimates $\hat{\boldsymbol{c}}$ backwards in time, starting at $n = N-1$ up to $n = 1$, with initial values $\hat{c}_j^s(N) = \hat{c}_j(N)$.

KF Smoothing: For the KF-type estimators, we adopted the Rauch-Tung-Striebel [164], [165] fixed-interval equations,

$$\boldsymbol{A}(n) = \boldsymbol{P}(n)[\boldsymbol{P}(n+1) + \boldsymbol{R_1}(n+1)]^{-1} \tag{3.23a}$$

$$\hat{\boldsymbol{c}}^{\boldsymbol{s}}(n) = \hat{\boldsymbol{c}}(n) + \boldsymbol{A}(n)[\hat{\boldsymbol{c}}^{\boldsymbol{s}}(n+1) - \hat{\boldsymbol{c}}(n)] \tag{3.23b}$$

As before, the smoothed estimates are obtained by updating the extracted estimates $\hat{\boldsymbol{c}}$ backwards in time, starting at $n = N-1$ up to $n = 1$ with initial values $\hat{\boldsymbol{c}}^{\boldsymbol{s}}(N) = \hat{\boldsymbol{c}}(N)$. $\boldsymbol{P}(n)$ and $\boldsymbol{R_1}(n)$ refer to the matrices acquired during the forward sweep. Note here that RLS smoothing combined with RCT cannot be applied directly; hence, for the RLS schemes we can use the KF smoothing approach (where the $\boldsymbol{P}$-matrices in Equation (3.23a) are scaled as $\frac{q\boldsymbol{P}(n)}{trace[P(n)]}$ and $\boldsymbol{R_1}$ is the small unit matrix $r\boldsymbol{I}$ added in Equation (3.21)).

3.3 Simulations

SI and MI TV system kernels were initially simulated using linear combinations of DLFs (TV DLF kernels) with smooth periodic (i.e. sinusoidal) and mixed-mode TV expansion coefficients ($\boldsymbol{c}(n)$). The mixed-mode case was characterized by model coefficients that exhibited occasional jumps, which may correspond to abrupt changes of the system operating point, and sinusoidal changes with varying rates and amplitudes (see Appendix). We also generated kernels given by sinusoidally modulated oscillating exponential functions (SME kernels), whereby the true model order is not known *a priori*. Nonstationarities were introduced by modulating in time the kernel amplitude according to smooth periodic or aperiodic signals. All the systems were simulated for Gaussian white noise inputs of unit variance and length of 2000 points. The effect

of output-additive noise was evaluated by adding 100 realizations of independent gaussian white noise signals to the output, for SNRs of 0, 10 and 20dB.

3.4 Effect of nonstationarity and noise characteristics on the hyperparameters

To rigorously examine the effect of nonstationarity type and noise characteristics on the algorithm hyperparameters, we used sinusoidally varying SI TV systems with DLF kernels (Appendix A1; Equation (A.1)), whereby the ground truth is known. A simulation set is characterized by three factors:

- The true system complexity (Q_{true} and L_{true}) and Laguerre parameter (a).
- The amplitude (A) and frequency (F) of the sinusoidally oscillating expansion coefficients.
- The true system output SNR (SNR $= 0, 10, 20\ dB$ – 100 independent realizations per case).

For each simulation set, the model order complexity was assumed to be known to avoid the confounding effects of incorrect model order selection. Since the model structure was known *a priori*, all hyperparameters except Q, L and α were optimized using nVAF as the fitness function. Four recursive schemes were examined; RLS with single constant FF (RLSC), RLS with single adaptive FF (RLSA), Kalman Filter (KF) and Adaptive Kalman Filter (KFA). The effect of different factors and the relationship between hyperparameters were quantified using Spearman's rank correlation matrices (Figure 3.1). Note that the following observations apply for MI systems; however, in the interest of space, we focus on SI systems.

For the RLSC algorithm, negative correlations (Figure 3.1a) were observed between the values of the FF and the magnitude and frequency of the expansion coefficient oscillations (A, F). When the system exhibits a high degree of nonstationarity (large A) that is also fast (large F), the value of the FF decreases to achieve faster tracking. However, the parameter estimates become more sensitive to measurement noise, deteriorating the accuracy of the estimation result ($VAF_t(\%)$, $VAF_{k1}(\%)$, $VAF_{k2}(\%)$). A significant correlation was also obtained between the FF and the noise level; the lower the SNR, the higher the value of the FF. This was expected since in the presence of noise, RLS requires more samples to provide more stable parameter estimates. On the other hand, the value of the FF was positively correlated with the total number of expansion coefficients. The FF is related with the effective memory of the RLS algorithm (Equations (3.8-3.9)). Specifically, in the case of LS estimation, the minimum window size should not be smaller

than the total number of parameters. Consequently, it is expected that the FF will be bounded below by,

$$n_{ef}{}^{\infty}{}_{min} = \frac{1}{1-\lambda_{min}} \geq d \Rightarrow \lambda_{min} \geq 1 - \frac{1}{d} \tag{3.24}$$

where d is the total number of parameters. In the case of high SNR and pronounced nonstationarities, the FF reaches its minimum value. For the same number of DLFs, nonlinear models require a larger number of free parameters compared to linear models. This results in a higher FF value, which may sometimes prevent the accurate tracking of fast changes and deteriorate estimation accuracy compared to the linear case. Another interesting observation is that P_0 was negatively correlated with A (and in turn positively correlated with FF values) and the SNR. In other words, if the system is weakly nonstationary (small A) or the noise levels are high, a large for P_0 is selected by the GA and vice versa.

The RLSA algorithm exhibited similar behavior to RLSC in general (Figure 3.1b). The FF bounds λ_{max} and λ_{min} were negatively correlated with the degree of nonstationarity and the SNR, while they were positively correlated with the total number of parameters in the nonlinear case. The smoothing factor λ_e was also positively correlated with the FF bounds. When the system becomes more nonstationary, the prediction error variance changes faster, thus the smoothing factor λ_e decreases.

For the KF algorithm, we optimized the innovation (R_2) and process noise variance (R_1) using the GA. R_1 was positively correlated with A and SNR (Figure 3.1c). This is due to that larger changes can be modeled as a random walk with higher driving noise variance, whereas more noise in the data restricts the variance to smaller values. As was the case for the relation between FF values and total number of parameters, R_1 decreased when the model order complexity increased. However, the positive correlation between R_1 and R_2 suggests that KF copes with this complexity issue by decreasing R_2. Thus, it is expected that KF will perform better than RLSC and RLSA in case of models with a large number of parameters (see also below – Figures 3.2, 3.3 and Section 3.5).

In the case of KFA, the smoothing factor λ_w, which essentially defines the window length over which R_1 changes, exhibited negative correlation with the SNR, A and R_2 (Figure 3.1d). The more nonstationary the system becomes (increased A) the higher the prediction error variance; thus, the smoothing factor λ_w decreases in order to be able to accurately capture fast or large

parameter variations affecting the prediction error variance. On the other hand, in case of low SNR levels, λ_w attains high values close to 1 to avoid overfitting the noise.

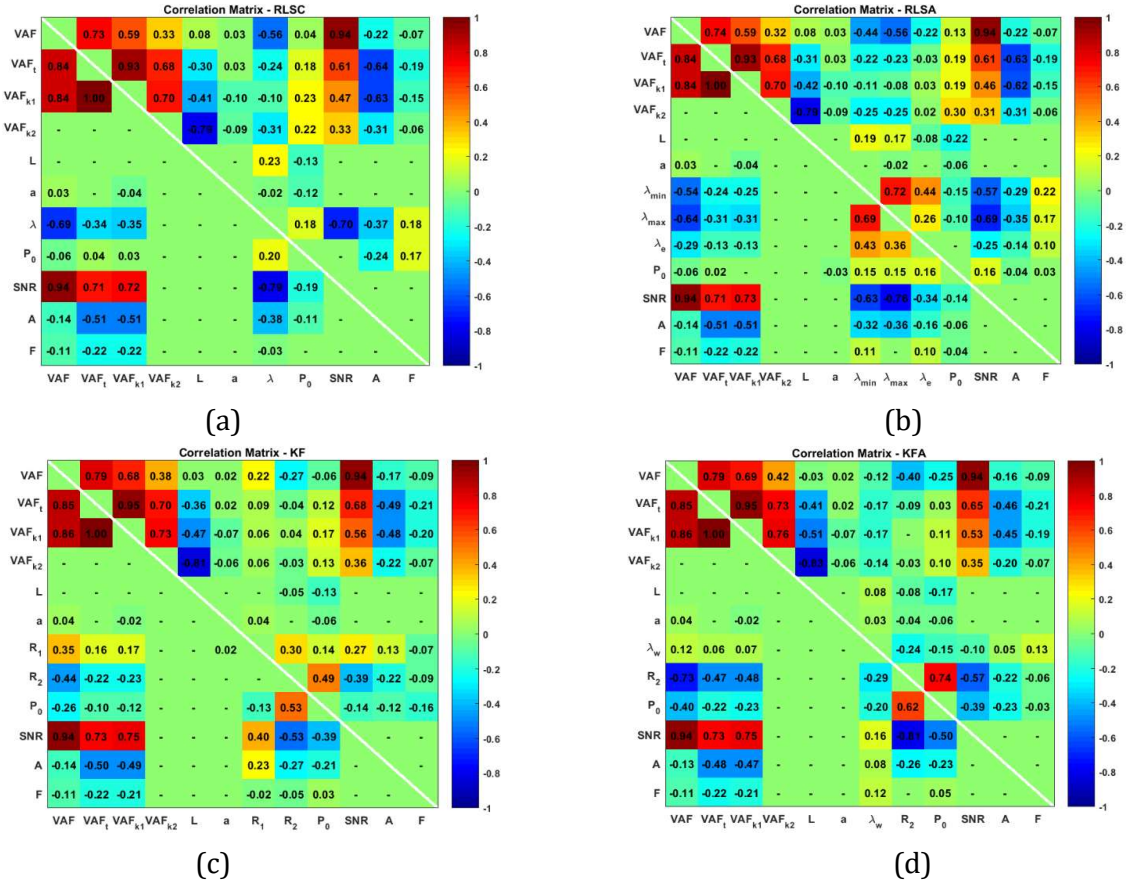


Figure 3.1 Matrices depicting correlations between the hyperparameters of different recursive estimators, the degree of nonstationarity and the levels of noise for simulated SI-TV linear (lower triangular) and 2nd order nonlinear systems (upper triangular) (a) RLSC (b) RLSA (c) KF (d) KFA (VAF**:** VAF(%) between noisy and predicted outputs, VAF_t**:** VAF(%) between noise-free and predicted outputs, VAF_{k1}**:** VAF(%) between the true and predicted 1st order kernels, VAF_{k2}**:** VAF(%) between the true and predicted 2nd order kernels, L: number of Laguerre functions used to construct the true TV DLF kernels, α: Laguerre parameter of the true TV DLF kernels, λ: RLSC FF, λ_{min}: RLSA lower bound of the FF, λ_{max}: RLSA upper bound of the FF, λ_e: RLSA smoothing factor for the prediction error variance, R_1**:** KF variance of the driving noise, R_2**:** KF/KFA variance of the measurement noise, SNR: Signal to Noise Ratio, λ_w: KFA smoothing factor for the process noise variance, P_0: Initial value of the diagonal elements of the covariance matrix **P**, A: magnitude of the simulated sinusoidally TV DLF kernel coefficients, F: frequency of the simulated sinusoidally TV DLF kernel coefficients). Note that only significant correlations ($p < 0.01$) are shown.

3.5 Performance of Recursive Estimators

3.5.1 Performance of Recursive Estimators under known model order complexity

The proposed recursive estimators were initially tested in systems with TV DLF kernels (see Appendix). As a measure of performance, we used the VAF (%) between noise free output and predicted output (VAF_t %))[166]. Model order complexity was assumed to be known (SI case: $[L, \alpha] = [4, 0.2]$; MI case: $[L_1, \alpha_1, L_2, \alpha_2] = [5, 0.2,3,0.5]$)

3.5.1.1 SI-TV systems with sinusoidal variations

All recursive estimators performed similarly for linear systems (not shown here). However, in the case of highly nonstationary (i.e. large A and F) nonlinear systems, KF and KFA outperformed RLSC and RLSA (Figure 3.2a,b). As mentioned above, the values of the RLSC and RLSA FFs are positively correlated with the number of model parameters. Given the same number of DLFs, nonlinear models require a larger number of free parameters compared to their linear counterparts. This results in larger FF values, which may sometimes prohibit the accurate tracking of fast or large changes and deteriorate estimation performance. KF and KFA, on the other hand, are not restricted by the minimum window LS rule and can therefore cope with the increase of the innovation (R_2) variance by decreasing the process noise (R_1) variance.

3.5.1.2 SI-TV systems with mixed-mode variations

The proposed RLSA and KFA recursive estimators performed better compared to the conventional RLS and KF techniques (Figure 3.2c,d) indicating that under mixed-mode nonstationarities, constant FFs (RLS) or process noise variances (KF) do not sufficiently capture non-periodic and/or abrupt changes in the true system. Figure 3.2c,d depicts the VAFt(%) achieved by each estimator under different SNR levels. RLSA yielded the best performance in the case of linear systems. However, in the case of nonlinear systems KFA outperformed RLSA. As mentioned earlier, the larger number of parameters required for nonlinear models prevents the FFs from acquiring low values and tracking the underlying fast changes accurately. Since KFA is based on a more stochastic approach, it performs better in such cases.

3.5.1.3 MI-TV systems with sinusoidal variations

In the case of MI-TV systems the dynamics associated with different inputs may exhibit different variations over time, suggesting that the corresponding model parameters may vary with different rates. The performance of RLSC, RLSA, KF and KFA was found to be inferior compared to the performance of the algorithms designed specifically for MI systems (RLSM, RLSMA, KFM and KFMA; Figure 3.3a,b). KFM and KFMA yielded more accurate estimates, particularly in the nonlinear case. As before, MI nonlinear models require a larger number of free parameters and, consequently, the tracking abilities of RLSM or RLSMA are limited to time windows determined by the total number of model parameters.

3.5.1.4 MI-TV systems with mixed-mode variations

As before, MI recursive approaches yielded superior performance compared to RLSC, RLSA, KF and KFA. (Figure 3.3c,d). Furthermore, RLSMA and KFMA yielded more accurate estimates as their hyperparameters (i.e. FFs and innovation/process noise variances) were updated at each time step. For the reasons mentioned in the case of sinusoidal variations, KFMA performs better than RLSMA when modeling MI nonlinear systems.

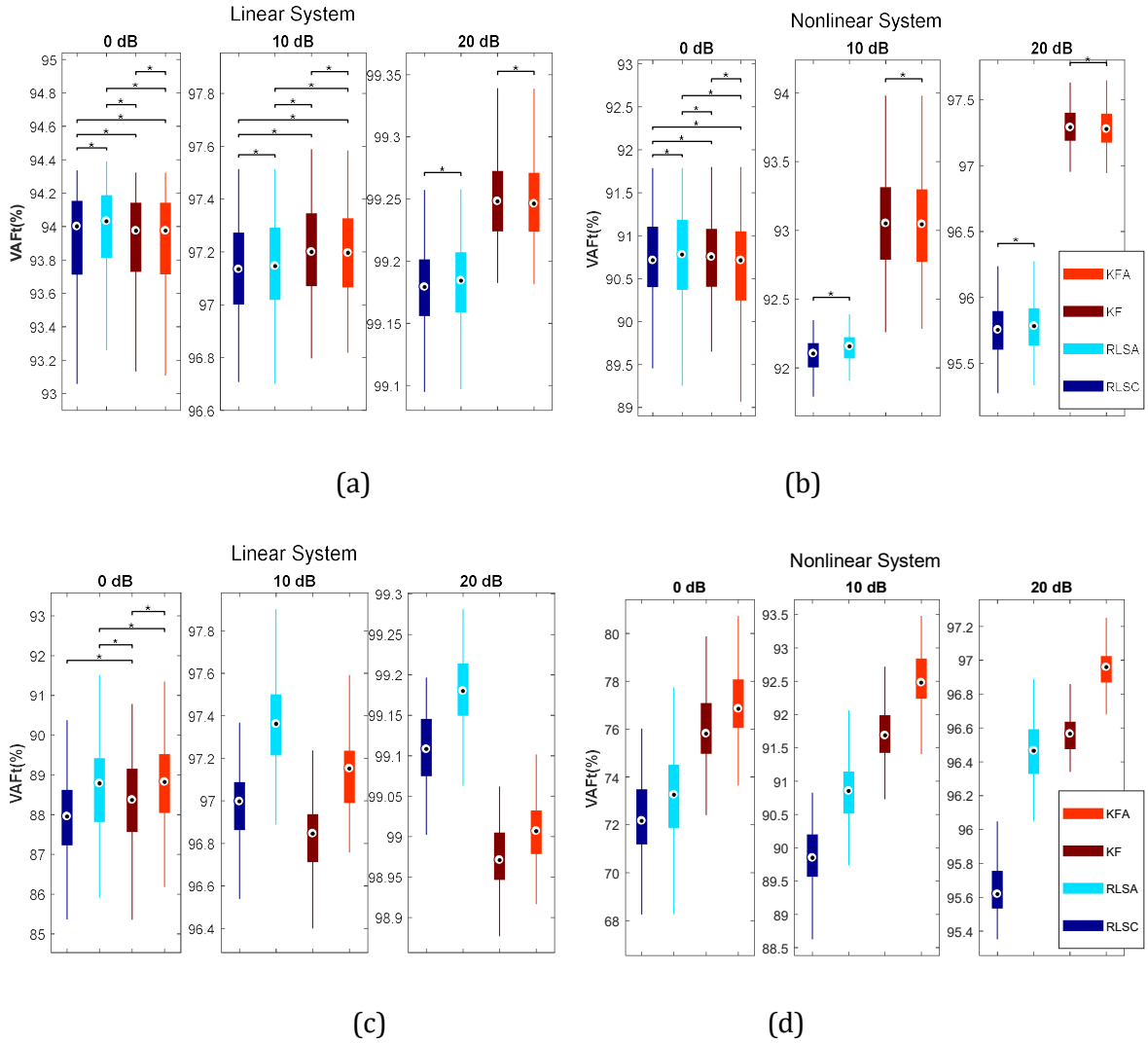


Figure 3.2 Performance (VAFt (%)) of RLSC, RLSA, KF, KFA for (a, c) linear (Q=1) and (b, d) nonlinear (Q=2) SI-TV systems under different SNR levels (0, 10, 20 dB). Top panel: Systems with TV DLF kernels ($[L_{true}, a_{true}] = [4,0.2]$) exhibiting sinusoidal variations (Appendix A.1; Equation (A.1)) of amplitude $A = 0.2$ and frequency $F = 10$. Bottom panel: SI-TV systems with TV DLF kernels ($[L_{true}, a_{true}] = [4,0.2]$) exhibiting mixed mode variations (Appendix A.2; Equation (A.6)). *: denotes non-significant differences between estimators (p>0.001). KFA, RLSA were found to perform similarly with the conventional KF and RLS respectively in the case of smooth, sinusoidal variations (a,b). KF estimators outperformed the RLS estimators in the case of nonlinear systems. In the case of mixed-mode variations (c,d), RLSA and KFA achieved better performance compared to the conventional RLS and KF respectively. RLSA was found to be superior in the linear case, whereas KFA yielded the best performance in the nonlinear case. Note here that the model order of the system was assumed to be known.

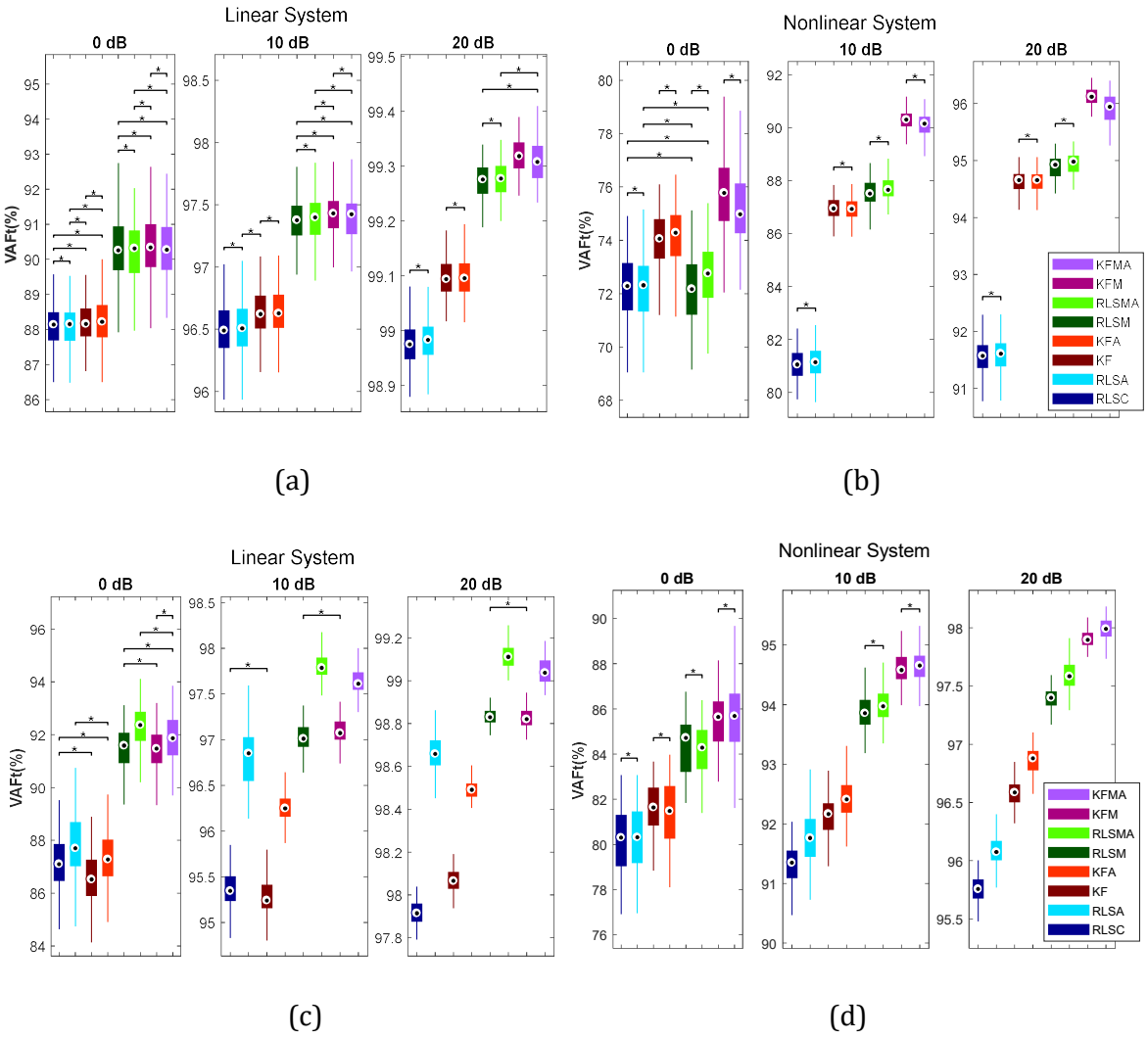


Figure 3.3 Performance (VAFt (%)) of RLSC, RLSA, KF, KFA, RLSM, RLSMA, KFM, KFMA for (a, c) linear (Q=1) and (b,d) nonlinear (Q=2) MI-TV systems under different SNR levels (0, 10, 20 dB). Top panel: System with TV DLF kernels ($[L_{1true}, \alpha_{1true}, L_{2true}, \alpha_{2true}] = [5, 0.2, 3, 0.5]$) exhibiting sinusoidal variations (Appendix A.3; Equations (A.8-A.12)). Bottom panel: SI-TV systems with TV DLF kernels ($[L_{1true}, \alpha_{1true}, L_{2true}, \alpha_{2true}] = [5, 0.2, 3, 0.5]$) exhibiting mixed-mode variations (Appendix A.4; Equations (A.20-A.22)).* denotes non-significant differences between estimators (p>0.001). The MI recursive estimators (RLSM, RLSMA, KFM, KFMA) yielded better performance in all cases. Under the presence of mixed-mode variations (c,d) the adaptive MI estimators yielded better performance compared to the estimators that used constant hyperparameters. Note that the model order of the system was assumed to be known.

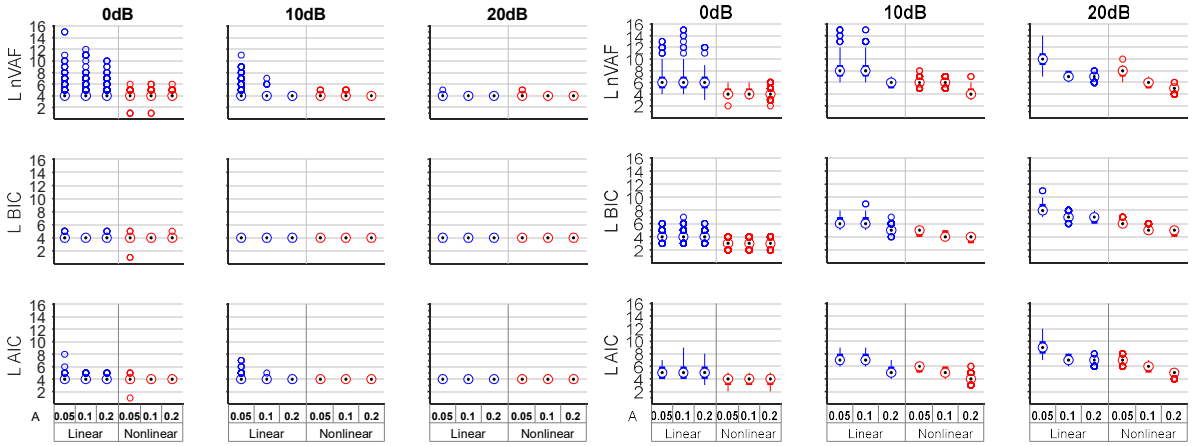
3.5.2 Performance of Recursive Estimators under uknown model order complexity

We have already examined the performance of the recursive estimators assuming that the underlying system is exactly known. In contrast to Section 3.5.1, where we optimized only the hyperparmeters of each estimator, here we also performed model order selection (L and α) using BIC, AIC and nVAF as the GA fitness functions (Section 3.2.3). The purpose of this subsection was twofold. First, we aimed to examine the behavior of different model order selection criteria under nonstationary conditions. We initially considered the SI systems with TV DLF and SME kernels described in Section 3.5.1 along with the mixed integer GA optimization scheme (Section 3.2.3). The simulated kernels were modulated in time by sinusoidal waves of different amplitudes. In the DLF case, the ground truth ($L_{true} = 4$ and $\alpha_{true} = 0.2$) is already known, contrary to the SME case. We considered all cases: TV SI systems with sinusoidal and mixed-mode variations (Appendix A1, A2) as well as MI systems (Appendix A3, A4). Second, we aimed to investigate the performance of the recursive estimators under unknown model complexity and assess the extent to which the model order selection procedure may be affected by the choice of a particular estimator.

In order to examine the behavior of different model order selection criteria under nonstationary conditions, we used as point of reference SI systems with sinusoidal variations (Appendix A1), whereby we varied the amplitude and frequency of these variations. In Figure 3.4, results for the simulated linear (blue) and nonlinear (red) systems were ordered based on A ($A_1 < A_2 < A_3$, where $A_1 = 0.05,\ A_2 = 0.1, A_3 = 0.2$), while the frequency of the sinusoidal variations was kept constant ($F = 5$). In the DLF case (Figure 3.4a), all three metrics yielded the correct model order ($L = 4$ and $\alpha = 0.2$), particularly when SNR and A increased. As the system gradually became more nonstationary (larger A), overestimating the true order became less likely. For example, for the RLS-type estimators, the effective memory of the algorithm (and consequently the FF) is bounded below by the total number of free parameters. Overestimating the model order leads to an increase in the amount of the free parameters, a longer RLS effective memory, higher values for the FFs and subsequently inefficient tracking, which is reflected on the total estimation error. Nonlinear systems were less prone to overfitting with increased A and SNR, since an increase in the number of DLFs resulted into a larger number of free parameters compared to the linear case (Figure 3.4a). Some overfitting was evident mainly in the linear case for low SNR values combined with weak nonstationarities, especially when using

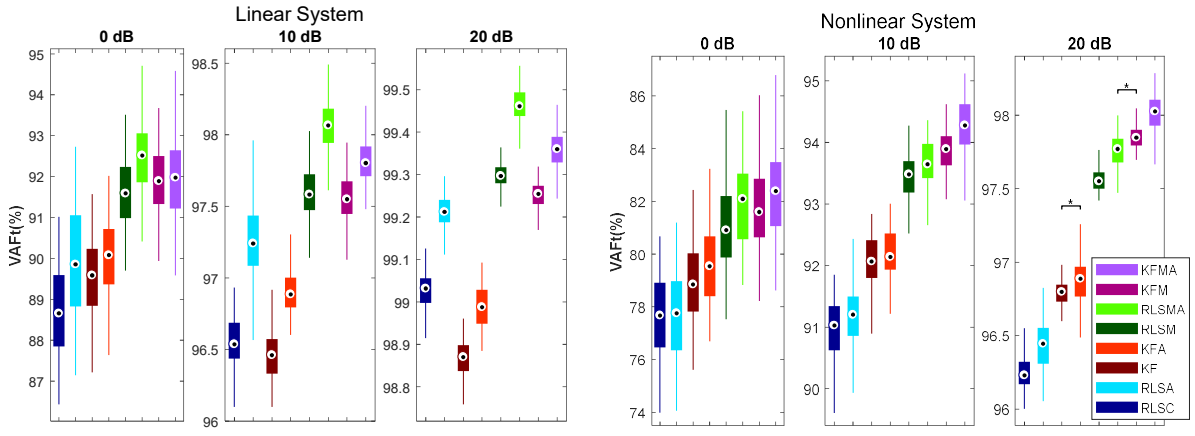
VAF as a fitness function (Figure 3.4a). However, for increased SNR and A, VAF yielded the correct model order. BIC detected successfully, under all conditions, the true model order L (and Laguerre parameter α – not shown); this was expected as BIC assumes an underlying "true" model amongst a set of candidates. AIC performed similarly with BIC. In the SME case (Figure 3.4b) we observed the same pattern as for DLF systems, i.e. a gradual decrease in the selected number of DLFs as A increased (under the same SNR). Note the agreement between AIC, BIC and nVAF ($L = 7$) for both linear and nonlinear systems, SNR=20dB and A=0.20 (Figure 3.4b). On the other hand, increased SNR resulted in the selection of a larger model order (with BIC being more conservative compared to AIC and VAF). This can be explained as follows: selecting a more complex model when the SNR is high offers the flexibility to reconstruct in a more accurate manner small details of the kernel shape (e.g. low amplitude oscillations at the tail of the kernel). The same observations can be extended to MI systems.

The results obtained using DLF kernels (Figure 3.4a) suggest that the model order selection procedure was mostly affected by the criterion type (AIC, BIC, nVAF) used as fitness function rather than the particular choice of the recursive estimator. In general, the same observations apply in the case of SME kernels (Figure 3.4b). Overall, the comparative performance of different estimators was found to be similar in the DLF and SME cases (Figures 3.2, 3.3 and 3.5 respectively). Specifically, Figure 3.5 illustrates the performance of all estimators using the AIC in the case of a MI system with mixed-mode TV SME kernels. Note the resemblance with Figure 3.3c,d, where the model complexity of the system was assumed to be known. In Figure 3.6 we present the true and obtained 1st-order kernels of a linear MI TV SME system (Appendix A.4) using RLSA for an SNR of 20dB.



(a) (b)

Figure 3.4 Model order (L) as selected by the GA for linear (Q=1, blue) and 2nd-order nonlinear (Q=2, red) SI-TV systems with (a) DLF kernels ($[L_{true}, a_{true}] = [4,0.2]$) and (b) SME kernels with sinusoidal frequency variations (F=5) and different amplitude values (A=0.05, 0.1 and 0.2 – Appendix A1; Equations (A.1-A.5)) under different SNR levels (0, 10, 20 dB). nVAF, BIC, AIC were used as model order selection criteria (top, middle and lower panel respectively). We present results yielded by all the estimation algorithms (RLSC, RLSA, KF, KFA). Note that the agreement between all three criteria in terms of model order as A becomes larger and SNR increases for both linear and nonlinear systems.



(a) (b)

Figure 3.5 Performance (VAFt (%)) of RLSC, RLSA, KF, KFA, RLSM, RLSMA, KFM, KFMA for (a, c) linear (Q=1) and (b, d) nonlinear (Q=2) MI-TV systems with SME kernels with mixed-mode variations (Appendix A4; Equations (A.23-A.24)) under different SNR levels (0, 10, 20 dB). * denotes non-significant differences between estimators (p>0.001). Note that the model order of the system was assumed to be unknown. Model order selection was done using the GA and AIC as fitness function. Note the similarities with Figure 3.3c and 3.3d.

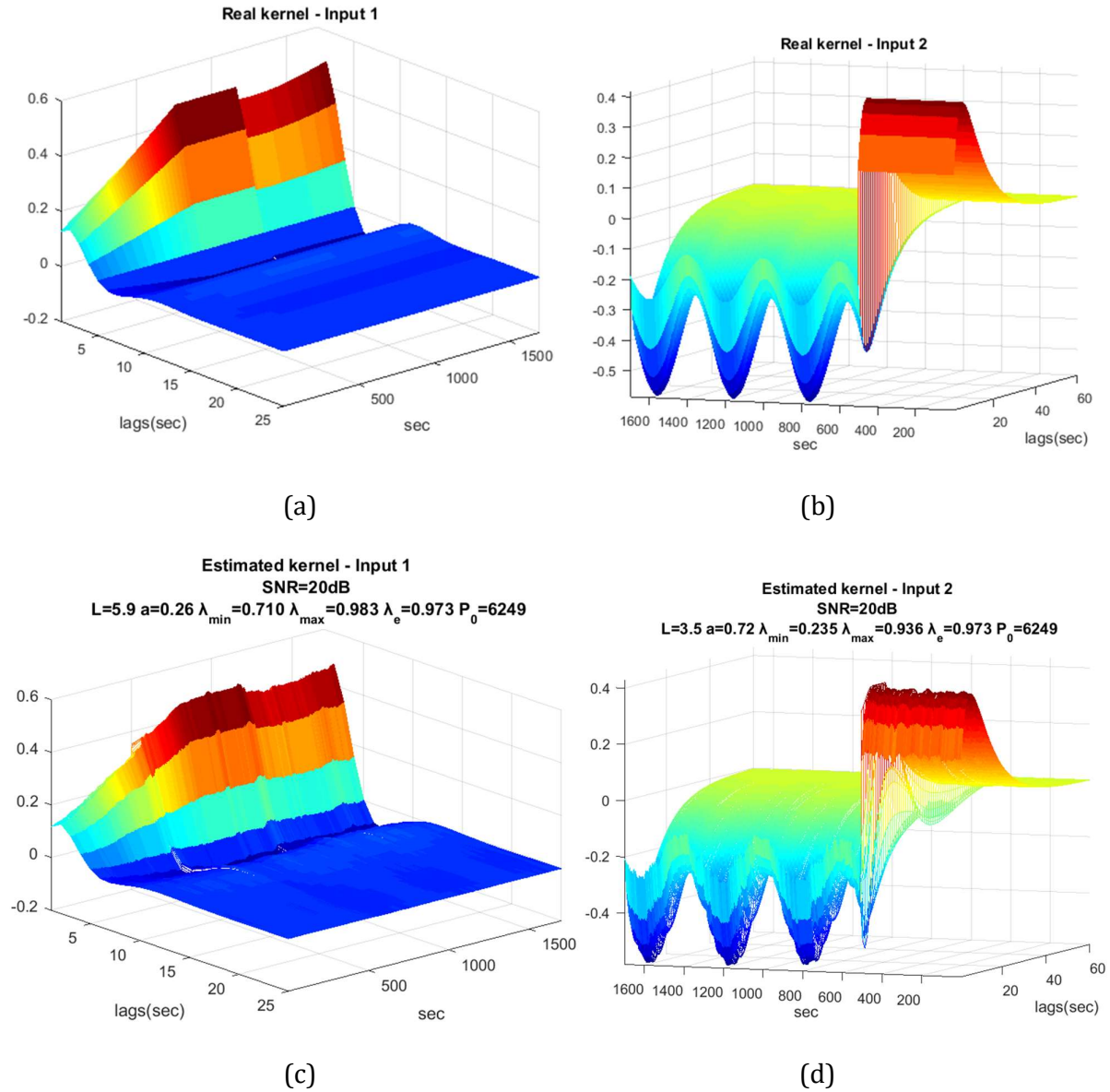


Figure 3.6 (a,b) Simulated MI TV 1st–order kernels with mixed mode variations for the case described in Appendix A4; Equations (A.23-A.24). (c,d) TV first-order kernels estimated by RLSA (after applying smoothing to the extracted TV model coefficients), averaged over all 100 realizations for an SNR of 20dB. In the bottom panel, on top of each plot we also show the average values of the hyperparameters, as optimized by the GA (Table I; RLSA – Two input linear system). The estimated kernel associated with Input 2 (d) exhibited higher variability between realizations compared to the estimated kernel of Input 1 (a). This is due to the abrupt and faster changes of (b) compared to (a).

3.6 Experimental Data

Syncope is a common clinical problem that affects up to 3.5% of the general population. VVS is one of the most frequent causes of transient loss of consciousness and its pathophysiology remains unclear. HUT is a common diagnostic tool for VVS. The procedure involves initially the patient lying strapped in supine position on a tilt table and then being suspended at an angle of 60°-80°. VVS patients initially adjust normally to the upright position but, within 20-30 minutes, arterial blood pressure (ABP) and/or heart rate (HR) drops perilously low, leading to reduced blood flow to the brain and ultimately to a transient loss of consciousness. The time-varying underlying physiological mechanisms remain poorly understood.

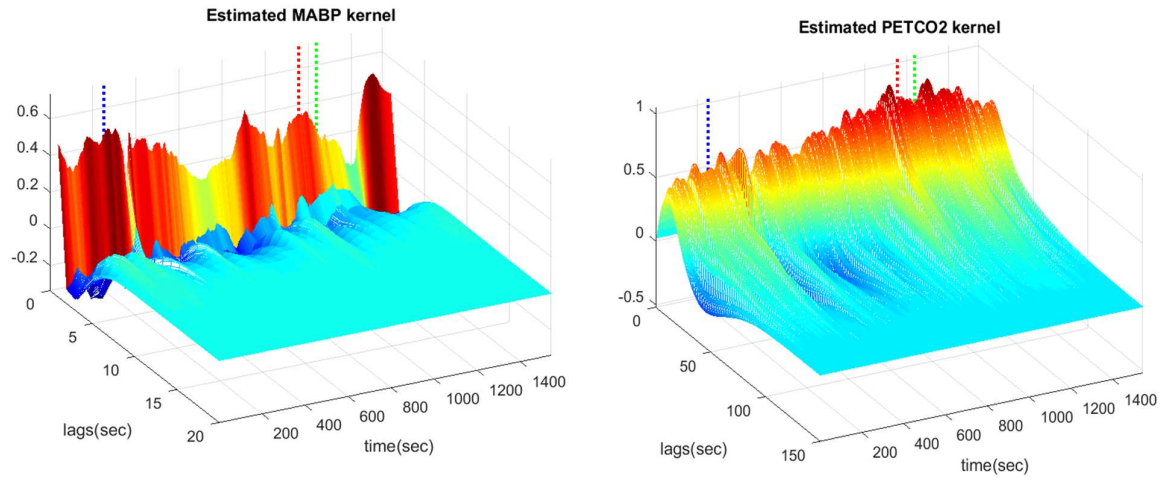
In this context, we examined these mechanisms in 14 VVS patients, undergoing HUT in the Jewish General Hospital Autonomic Reflex Laboratory, by using the proposed methods to quantify the time-varying interactions between mean ABP (MABP), end-tidal partial pressure of carbon dioxide ($PETCO_2$) and cerebral blood flow velocity (CBFV). The HUT protocol involves different stages; thus, we expect its underlying hemodynamic and cardiovascular effects to change over time. Our main objective was to track changes in physiological mechanisms and particularly dCA before and after syncope occurrence. dCA is a complex homeostatic mechanism that maintains a constant CBF despite variations in ABP and is frequently assessed by the dynamic relation between MABP and CBFV [57]. When dCA is impaired, ABP changes have a more pronounced effect on CBF. dCA is known to exhibit high-pass filter characteristics, thus a phase-shift (drop) towards zero in the low frequency (LF) range (0.04 – 0.15 Hz) may indicate impaired autoregulation [167]. Carbon dioxide (CO_2) and its well-known vasodilating effects can also lead to CBF changes, even during resting conditions [168], [169]. Both one-input (MABP) and two-input (MABP, $PETCO_2$) single output (CBFV) models have been used successfully to study dCA in a stationary and nonstationary context [7], [19], [20], [90], [168].

Prior to modeling, CBFV and MABP beat to beat values from HUT recordings were interpolated (spline interpolation) and resampled at 1 Hz to obtain equally spaced time series. Breath-by-breath values of $PETCO_2$ were linearly interpolated and time-matched to the resampled data for analysis. The resulting time series were demeaned. The HUT protocol consisted of a 10-min resting period in the supine position and a maximum of 40 min of 80° HUT which ended if a subjective sensation of impending syncope was associated with a clear precipitous drop in ABP or HR.

We extracted two-input (MABP, $PETCO_2$) single output (CBFV) linear and nonlinear TV Laguerre-Volterra models using the schemes presented in Section 3.2. Model parameters were estimated using the multiple input versions of RLS and KF (RLSM, RLSMA, KFM, KFMA) combined with the RCT technique (Section 3.2.4.1). In addition to the variables given in Table I, the following variables were optimized using the GA: a constant bias k_0, a pure time delay del for the $PETCO_2$ effects on CBFV (which is known to be around 4 sec [170]), and two variables for the RCT (26) (q, r). We used AIC as cost function and smoothing was applied to all extracted estimates.

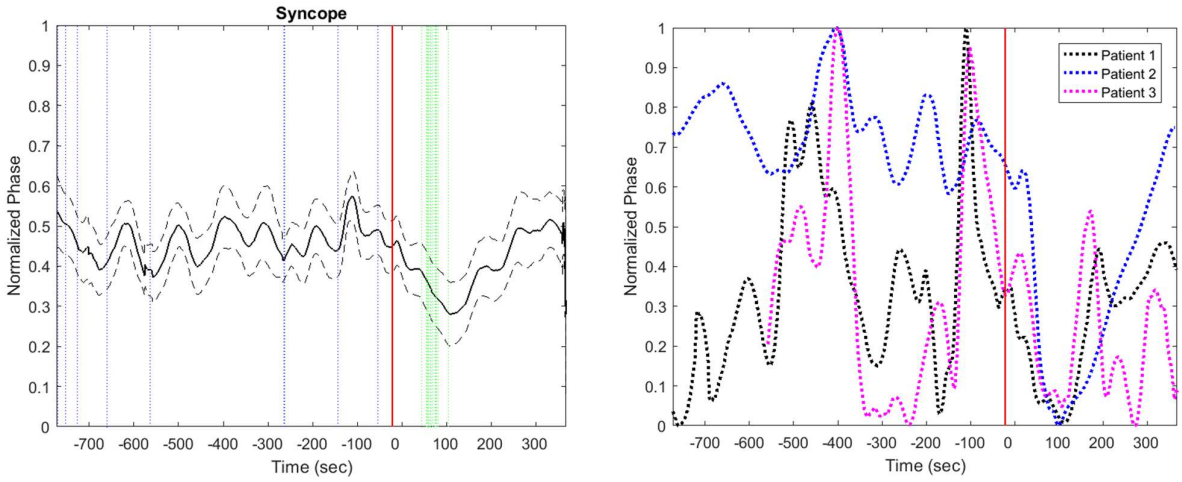
To illustrate the reduction in computation time achieved by the proposed GA scheme, we also performed exhaustive search using RLSM and linear models (DLF number: 1-10, Laguerre parameters: 0.1-0.9 with a step of 0.1, FFs: 0.9 to 1 with a step of 0.002 and all the remaining tuning variables set constant). The required computation time was 5 hours to find an optimal model for one subject even when parallel processing was used, while the grid of the exhaustive search is of relatively low resolution (note for other algorithm versions, e.g. RLSMA, exhaustive search would take even longer). On the other hand, the mean runtime per subject based on the proposed GA scheme was 24.15s/43.53s (linear/nonlinear models) for RLSM, 28.3s/57.99s for RLSMA, 23.04s/29.06s for KFM and 23.48s/38.08s for KFMA.

Based on the results, MI estimators (i.e. RLSM, RLSMA, KFM, KFMA) outperformed the estimators that used a single hyperparameter to track all TV coefficients (i.e. RLS, RLSA, KF, KFA). On the other hand, constant and adaptive realizations were found to perform similarly, implying periodic or slow variations. Linear models were found to be sufficient according to the AIC and BIC. Specifically, we selected as optimal the linear models estimated using RLSMA and AIC. The resulting TV MABP and $PETCO_2$ kernels are shown in Figure 3.7. After tilting, time to syncope varied considerably between patients. Hence the results from all subjects were aligned around the time of the syncopal event and the phase values were averaged (Figure 3.8). A significant phase decrease was observed after the onset of syncope but not before that time point. These results are in accordance with [171], where it was shown that dCA in VVS patients was preserved during the 3 min preceding syncope. However, we detected consistent LF phase oscillations in all patients, which could be related to fluctuations in other physiological signals. Whether these oscillations are physiologic or pathologic is an interesting question but lies outside the scope of the present work.



(a) (b)

Figure 3.7 (a) TV MABP and (b) $PETCO_2$ first-order kernels from a representative subject during HUT (Section 3.6). The blue dashed vertical lines define the onset of the tilting phase. The red vertical dashed lines denote the time of syncope occurrence and the green dashed lines denote the time point when MABP reached its minimum value.



(a) (b)

Figure 3.8 (a) Mean±standard error (standard deviation divided by the square root of the total number of subjects at each time point) of the normalized phase between MABP and CBFV across subjects in the LF range, centered around the time of the syncopal event (red vertical line). The blue vertical dashed lines in (a) define the onset of the tilting phase for different subjects. The green vertical dashed lines denote the time point when MABP reached its minimum value in different subjects. (b) Representative normalized LF phase between MABP and CBFV from 3 representative patients.

3.7 Conclusions

TV system identification is a challenging problem; nonstationarities may arise due to inherent TV dynamics or missing information. The ability to accurately track TV characteristics may help interpret the underlying mechanisms more precisely and in a timely manner especially in cases were timing plays a crucial role (e.g. event-related system changes). Extracting correctly the TV properties of a system may also elucidate the source of the observed variations. For example, in the case of Linear Parameter Varying systems (LPV) [172], where the dynamics vary as a function of (possibly unknown) TV signals termed scheduling variables (SVas), a TV model can be initially estimated and used to identify plausible SVas. In the present work, we have developed novel recursive computational schemes that are applicable to a wide class of TV systems (nonlinear, MI systems with finite memory) that yield excellent performance under different nonstationarity types (slow, fast/abrupt variations) by extending the conventional RLS and KF algorithms and using GA to efficiently select model order and estimator hyperparameters. The proposed estimators were able to track more accurately the true underlying system variations, leading to considerably improved performance, particularly in the case of fast/abrupt changes. Taking into account also that in MI-TV systems the dynamics associated with different inputs are likely to exhibit different rates of variations, we developed MI RLS and KF estimators for both periodically and aperiodically varying systems, which yielded more accurate estimates compared to the conventional RLS and KF methods. Our proposed TV system identification framework was based on finite memory Laguerre-Volterra models but it can be easily applied to other type of models (e.g. univariate/multivariate autoregressive models). Model order selection and estimation was merged in one single procedure; by using a mixed-integer GA we were able to simultaneously select model order and tune all the hyperparameters linked to the proposed recursive estimators. This led to greatly improved performance in terms of both accuracy and required computation time. Additionally, it allowed us to explore in detail the link between hyperparameters, model complexity and TV system characteristics. Model validation using conventional techniques (e.g. cross-validation) is not readily applicable under TV conditions, as the system may undergo significant changes. However, in many practical cases the model structure can be assumed to remain similar at different times, therefore the identified model structure can be used either from an initial data segment or previous experiment realizations in order to achieve real-time tracking. Finally, application of the proposed schemes to experimental data from VVS patients during HUT

demonstrated their ability to track time-varying changes in the case of a physiological system. In general, the proposed approaches yield promise for accurate real-time tracking of physiological signal couplings, perhaps using wearable devices, in a number of pathophysiologies.

Chapter 4 Modeling time-varying couplings between time series for biomedical applications

Couplings between various biological systems (e.g., couplings between neural networks in the brain, brain-heart interactions, cortico-cardio-respiratory couplings) are considered as one of the most important communication mechanisms in physiological function. One method for characterizing such interactions is the MVAR analysis. MVAR models can capture linear interdependencies between multiple time series and identify the directionality of flow of information in both time and frequency domain. In this chapter we develop TV-MVAR models, based on the recursive approaches proposed in Chapter 3, that can track efficiently TV changes in coupling strength and directionality. We also elucidate the source of TV variance in the MVAR residuals when dealing with biosignals that are obtained during experimental protocols consisting of phase transitions or events and we propose methods to deal with the obsesrved heteroskedasticity. Our proposed modeling methodology was applied to both simulated and experimental data.

Modeling time-varying couplings between time series for biomedical applications

K. Kostoglou[1], A. Robertson[2], B. J. Macintosh[2] and G. D. Mitsis[3]

[1] Department of Electrical and Computer Engineering, McGill University, Montreal, Canada.

[2] Canadian Partnership for Stroke Recovery, Sunnybrook Research Institute, University of Toronto, Toronto, Canada.

[3] Department of Bioengineering, McGill University, Montreal, Canada.

(Submitted to IEEE Transactions on Biomedical Engineering)

Abstract — We present a modeling framework for identifying time-varying (TV) couplings between time series of biomedical relevance. Our proposed methodology is based on multivariate autoregressive models (MVAR) that have been extensively used in the literature to study biosignal interactions and information flow. Contrary to standard estimation methods that assume time-invariant relationships we tracked changes in the model parameters through a modified recursive Kalman filtering (KF) technique. Model determination and hyperparameter optimization was achieved with the use of Genetic Algorithms (GA) improving accuracy and computation time. Based on simulations, our recursive estimator performed significantly better than the conventional KF algorithm especially when model parameters exhibited different rate of variations and amplitude changes throughout time. In addition, we addressed the issue of biased TV-MVAR connectivity measures (i.e. coherence, partial coherence, directed coherence and partial directed coherence) due to heteroskedasticity in the model residuals associated with event-related changes or phase transitions during experimental protocols. Our proposed modeling methodology was applied in experimental data to detect hemodynamic changes during exercise in young and old healthy adults, as well as individuals with chronic stroke.

Index Terms — MVAR, time-varying, Kalman Filter, heteroskedasticity, coherence, exercise

4.1 Introduction

Biological functions are usually regulated by coupling of individual oscillators. Oscillations have been identified for example as a fundamental property of neural responses [173] and couplings between different oscillatory neural populations are considered as one of the most important mechanisms of information processing and communication between brain regions [174], [175]. Frequency-derived identification and characterization of the propagation of information under different brain states [176] and pathological conditions [177] has drawn growing attention of late. Coherent oscillations between electroencephalogram (EEG), electromyogram (EMG) and magnetoencephalogram (MEG) have been increasingly used to elucidate the role of the motor cortex in movement control [178]. On a macroscale level, cardiovascular and respiratory system interactions and their directionality have been the subject of much debate among researchers [179]. Based on the abovementioned examples, couplings are a reality in physiological systems and the development of tools to accurately identify and quantify them is essential. From a system's identification point of view, one method of characterizing such interactions is multivariate autoregressive (MVAR) analysis [63]. MVAR models capture linear interdependencies among multiple time series and identify causality in time and frequency domain. However, the coupling strength, the type of coupling, and the coupling function may not necessarily remain constant throughout time, and the rate of change may vary as well. To address this temporal variability various studies have focused on identifying time-varying (TV) MVARs [180]–[188]. The main assumption behind the estimation of a TV-MVAR is that the parameters of the model are no longer constant but rather a function of time. Two main approaches exist in the literature for estimating TV parameters. The first approach segments the data into windows and tracks parameters variations in a piecewise stationary manner [180]. The second approach is based on recursive techniques, like Recursive Least Squares (RLS) and Kalman Filter (KF) [181]–[187]. In the quasistationary methodology, choosing an appropriate window length is of critical importance. If the window is too small, then the obtained estimates are too noisy. On the other hand, if the window is too big then the estimates are too smooth. Recursive approaches usually rely on the optimal tuning of hyperparameters i.e. the forgetting factor in the RLS case that defines the memory of the estimator. Conventionally, both KF and RLS assume constant rate of variations, which is not always valid. Moreover, all parameters do not necessarily change in the same manner. Parameters may exhibit intervals of slow, fast or even abrupt changes. Thus, an accurate

representation of TV interactions requires the development of algorithmic schemes that can track different type of TV patterns. Another major issue that may affect the estimation of a TV-MVAR model is heteroskedasticity of the residuals. Classic system identification methods assume normality in the error terms, with zero mean and constant variance (homoscedasticity). However, unequal variances can arise due to changes in measurement noise, unobserved external factors or event-related changes and experimental phase transitions that are associated with different signal to noise ratio (SNR) levels. Heteroskedasticity, depending on its degree, may impact significantly the TV-MVAR parameter estimation procedure especially when applying recursive techniques. The most striking effect though is the overestimation/underestimation of the strength and the directionality of interactions due to the dependence of MVAR connectivity measures on the statistical properties of the model residuals.

Herein, we have developed a novel recursive scheme for estimating TV-MVAR models. We show using simulations that the proposed implementation uncovers accurately the TV interacting properties of a system. In addition, we examined the effects of heteroskedasticity on the representation of various TV-MVAR connectivity measures and we propose corrections that reduce coupling biases due to changes in variance of the error terms. Finally, we applied the developed TV framework to investigate cerebrovascular regulation during exercise across healthy subjects and stroke survivors.

4.2 Methods

4.2.1 Time Varying Multivariate Autoregressive Models (TV-MVAR)

In MVAR models, each variable is regressed on its past values and the past values of the other variables in the system. Under the assumption of time-invariance, the influence of one time series to the other and to itself is constant throughout time. However, in order to capture changes in interactions we allow the parameters of the model to evolve through time. Specifically, a TV-MVAR model of order p describes the data as follows,

$$\boldsymbol{y}(n) = \sum_{k=1}^{p} \boldsymbol{A}_k(n)\boldsymbol{y}(n-k) + \boldsymbol{\varepsilon}(n) = \boldsymbol{A}(n)\boldsymbol{\Phi}(n) + \boldsymbol{\varepsilon}(n) \tag{4.1}$$

where $\boldsymbol{y}(n) \in \boldsymbol{R}^{M\times 1}$ is the vector of M response time series variables at time n, $\boldsymbol{A}_k(n) \in \boldsymbol{R}^{M\times M}$ is an autoregressive matrix for each order k at time n and $\boldsymbol{\varepsilon}(n)$ is assumed to be a zero-mean

white noise vector. $\boldsymbol{A}(n)$ and $\boldsymbol{\Phi}(n)$ are defined as, $\boldsymbol{A}(n) = [\boldsymbol{A_1}(n) \quad \dots \quad \boldsymbol{A}_p(n)] \in \boldsymbol{R}^{M\times Mp}$ and $\boldsymbol{\Phi}(n) = [\boldsymbol{y}(n-1) \quad \dots \quad \boldsymbol{y}(n-p)]^{\boldsymbol{T}} \in \boldsymbol{R}^{Mp\times 1}$. To facilitate parameter estimation, we reformulate Equation (4.1) as,

$$\boldsymbol{y}(n) = [\boldsymbol{I}_{M\times M} \otimes \boldsymbol{\Phi}^{\boldsymbol{T}}(n)]\, \boldsymbol{\Gamma}(n) + \boldsymbol{\varepsilon}(n) = \widetilde{\boldsymbol{\Phi}}(n)\, \boldsymbol{\Gamma}(n) + \boldsymbol{\varepsilon}(n) \tag{4.2}$$

where $\boldsymbol{I}_{M\times M}$ is the identity matrix, $\otimes$ denotes the Kronecker delta product, $\widetilde{\boldsymbol{\Phi}}(n) \in \boldsymbol{R}^{M\times M^2p}$ is an extended regressor matrix and $\boldsymbol{\Gamma}(n) \in \boldsymbol{R}^{M^2p\times 1}$ is the vectorized $\boldsymbol{A}(n)$ containing all MVAR parameters at time n.

4.2.2 TV-MVAR parameter estimation

The TV model parameters $\boldsymbol{\Gamma}(n)$ are estimated using a modified Kalman filtering (KF) technique. The conventional KF assumes that parameter changes are modeled as a random walk driven by Gaussian white noise with covariance matrix equal to $\boldsymbol{R_1} = R_1\boldsymbol{I}_{d\times d}$**.** In effect, R_1 basically defines the size of the expected parameter variations. The KF formulation is based on the hypothesis that the measurement noise is also Gaussian and white with variance equal to R_2. The KF algorithm combined with the MVAR formulation presented in Equation (4.2) consists of the following recursive steps,

$$\boldsymbol{e}(n) = \boldsymbol{y}(n) - \hat{\boldsymbol{y}}(n) = \widetilde{\boldsymbol{\Phi}}(n)\widehat{\boldsymbol{\Gamma}}(n-1) \tag{4.3}$$

$$\boldsymbol{r}(n) = \widetilde{\boldsymbol{\Phi}}(n)\boldsymbol{P}(n-1)\widetilde{\boldsymbol{\Phi}}(n) \tag{4.4}$$

$$\boldsymbol{K}(n) = \frac{\boldsymbol{P}(n-1)\widetilde{\boldsymbol{\Phi}}^{\boldsymbol{T}}(n)}{R_2 + \boldsymbol{r}(n)} \tag{4.5}$$

$$\boldsymbol{P}(n) = \boldsymbol{P}(n-1) + \boldsymbol{R_1} - \boldsymbol{K}(n)\widetilde{\boldsymbol{\Phi}}(n)\boldsymbol{P}(n-1) \tag{4.6}$$

$$\widehat{\boldsymbol{\Gamma}}(n) = \widehat{\boldsymbol{\Gamma}}(n-1) + \boldsymbol{K}(n)\boldsymbol{e}(n) \tag{4.7}$$

where $\widehat{\boldsymbol{\Gamma}}(n)$ are the estimated parameters at time point n, $\boldsymbol{P}(n) \in \boldsymbol{R}^{M^2p\times M^2p}$ is the parameter estimation error covariance matrix, $\boldsymbol{K}(n) \in \boldsymbol{R}^{M^2p\times M}$ is the Kalman gain, R_2 is the measurement noise variance and $\boldsymbol{R_1} \in \boldsymbol{R}^{M^2p\times M^2p}$ is the diagonal process noise covariance matrix.

The issue with the conventional KF technique is that it assumes that all parameters follow a random walk with the same statistical properties. However, parameters that exhibit large and fast variations will need to be assigned with higher R_1 values compared to those that follow slow and small changes. Moreover, the TV patterns may not be periodic and they may followan alternative mode of slow, fast or even abrupt alterations. To this end, the diagonal process noise

covariance matrix $\boldsymbol{R_1}$ is updated adaptively at each time step based on the following proposed equations,

$$\boldsymbol{R_1}(n) = \boldsymbol{\Lambda_w}\boldsymbol{R_1}(n-1) + (\boldsymbol{I} - \boldsymbol{\Lambda_w})e^2(n) \tag{4.8}$$

$$\boldsymbol{\Lambda_w} = diag\{\boldsymbol{\lambda_w}\}, \qquad \boldsymbol{\lambda_w} = [\lambda_{w_1} \quad \dots \quad \lambda_{w_d}] \tag{4.9}$$

where $d = M^2p$ is the total number of parameters and $\boldsymbol{\Lambda_w} \in \boldsymbol{R}^{d\times d}$ is a diagonal matrix of smoothing factors λ_{w_j} for $\lambda = 1 \dots d$, that lie between 0 and 1. The rationale behind this procedure is that changes in the parameters can be detected by tracking the variations in the error terms. Equation (4.8) can be described as a multivariate moving average (MA) that follows the mean of the variance of the residuals with different rates (due to the different smoothing factors assigned to the diagonal elements of $\boldsymbol{\Lambda_w}$). Through Equations (4.8) and (4.9), we are indirectly assigning unique adaptive update coefficients to each model parameter. In addition, we assume that the noise measurement variance is different for each time series and therefore R_2 from Equation (4.5) becomes a M-dimensional diagonal matrix $\boldsymbol{R_2} = diag\{\boldsymbol{r_2}\}$, $\boldsymbol{r_2} = [r_{2_1} \quad \dots \quad r_{2_M}]$.

4.2.3 Model order selection and hyperparameters optimization

Model order selection and optimization of the hyperparameters $\boldsymbol{\lambda_w}, \boldsymbol{r_2}$ was realized using a mixed integer Genetic Algorithm (GA). The candidate solutions $\boldsymbol{X}_i$ evaluated by the GA, were of the following form,

$$\boldsymbol{X}_i = [p \quad \boldsymbol{\lambda_w} \quad \boldsymbol{r_2}] \tag{4.10}$$

where $p \in [1 \dots p_{max}]$ is the selected MVAR model order and p_{max} is the maximum order that we are interested in. The value of each element of $\boldsymbol{\lambda_w}$ resides between 0 and 1, whereas $\boldsymbol{r_2}$ values are simply constrained to be positive. The total number of hyperparameters optimized by the GA are $1 + M^2p + M$. In the case of the conventional KF technique the only hyperparameters that are needed to be tuned are R_1 and R_2. Thus, the candidate solutions take the form,

$$\boldsymbol{X}_i = [p \quad R_1 \quad R_2] \tag{4.11}$$

GAs can be described as adaptive search algorithms inspired from natural evolution. In each generation, the fitness function (i.e. objective function) of a population of candidate solutions is evaluated. The fittest solutions are then selected and used in the next iteration of the algorithm. Crossover, mutation and selection are the basic search mechanisms of the GA. Crossover

combiness two individuals to produce a new one, while mutation creates diversion by altering one solution to produce a new one. Selection, on the other hand, reduces the search area by discarding poor solutions. Usually, the algorithm is terminated if a maximum number of iterations or a fitness level has been reached. We selected the Akaike Information Criterion (AIC) [40] on the entire dataset as the fitness function. Thus, for each candidate solution $\boldsymbol{X}_i$ the GA estimates recursively the TV-MVAR model based on all the available data and then it evaluates the following quantity,

$$AIC(\boldsymbol{X}_i) = N\log\left(|\widehat{\boldsymbol{\Sigma}}|\right) + 2d \tag{4.12}$$

where N is the length of the data set, $d = M^2p$ is the total number of parameters and $|\widehat{\boldsymbol{\Sigma}}|$ is the determinant of the estimated covariance of the error terms, i.e. $\widehat{\boldsymbol{\Sigma}} = cov\left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\right) = cov(\boldsymbol{e})$. The optimal candidate solution is the one that achieves the lowest AIC value.

4.2.4 Smoothing

Smoothing is applied to the extracted TV model parameters independently in order to deal with noisy estimates either due to fast tracking or excessive noise. For the KF-type estimators, we adopt the Rauch-Tung-Striebel [164], [165] fixed-interval equations,

$$\boldsymbol{W}(n) = \boldsymbol{P}(n)[\boldsymbol{P}(n+1) + \boldsymbol{R_1}(n+1)]^{-\mathbf{1}} \tag{4.13}$$

$$\widehat{\boldsymbol{\Gamma}}^{\boldsymbol{s}}(n) = \widehat{\boldsymbol{\Gamma}}(n) + \boldsymbol{W}(n)\left[\widehat{\boldsymbol{\Gamma}}^{\boldsymbol{s}}(n+1) - \widehat{\boldsymbol{\Gamma}}(n)\right] \tag{4.14}$$

By applying Equation (4.14) and updating $\widehat{\boldsymbol{\Gamma}}(n)$ in a backward manner starting from time point $n = N-1$ up to $n = 1$ with initial values $\widehat{\boldsymbol{\Gamma}}^{\boldsymbol{s}}(N) = \widehat{\boldsymbol{\Gamma}}(N)$, the smoothed estimates $\widehat{\boldsymbol{\Gamma}}^{\boldsymbol{s}}$ are obtained. In Equation (4.13) the matrices $\boldsymbol{P}(n)$ and $\boldsymbol{R_1}(n)$ have been computed already during the forward sweep.

4.2.5 TV-MVAR measures of coupling and causality

By taking the Fourier transform of Equation (4.1) the TV spectral power density matrix of $\boldsymbol{y}$ can be computed as,

$$\boldsymbol{S}(f,n) = \boldsymbol{H}(f,n)\boldsymbol{\Sigma}\boldsymbol{H}^H(f,n) \tag{4.15}$$

where $\boldsymbol{H}(f,n) = [\boldsymbol{I} - \boldsymbol{A}(f,n)]^{-\mathbf{1}} = \overline{\boldsymbol{A}}(f,n)^{-1} \in \boldsymbol{R}^{M\times M}$ is the *transfer matrix* in the frequency domain at time point n, $\boldsymbol{A}(f,n) = -\sum_{k=\mathbf{1}}^{\boldsymbol{p}} \boldsymbol{A_k}(n)e^{-i2\pi f} \quad \in \boldsymbol{R}^{M\times M}$ is the *coefficient matrix* in the

frequency domain at time point n and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_M^2 \end{bmatrix} \in \boldsymbol{R}^{M\times M}$ is the diagonal covariance matrix of $\boldsymbol{\varepsilon}$. The superscript H stands for Hermitian transpose. Similarly, the *inverse spectral power density matrix* of $\boldsymbol{y}$ can be written in the following form,

$$\boldsymbol{P}(f,n) = \boldsymbol{S}^{-1}(f,n) = \overline{\boldsymbol{A}}^{H}(f,n)\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{A}}(f,n) \tag{4.16}$$

The elements of the TV spectral density, transfer function, and coefficient matrices can be used to measure TV couplings and causality in the frequency domain. Specifically,

$$COH_{TD}(f,n) = \frac{S_{TD}(f,n)}{\sqrt{S_{TT}(f,n)}\sqrt{S_{DD}(f,n)}} \tag{4.17}$$

$$DC_{TD}(f,n) = \frac{\sigma_D H_{TD}(f,n)}{\sqrt{\sum_{m=1}^{M}\sigma_m^2|H_{Tm}(f,n)|^2}} \tag{4.18}$$

$$PCOH_{TD}(f,n) = \frac{P_{TD}(f,n)}{\sqrt{P_{TT}(f,n)}\sqrt{P_{DD}(f,n)}} \tag{4.19}$$

$$GPDC_{TD}(f,n) = \frac{\frac{1}{\sigma_D}\bar{A}_{TD}(f,n)}{\sqrt{\sum_{m=1}^{M}\frac{1}{\sigma_m^2}|\bar{A}_{mD}(f,n)|^2}} \tag{4.20}$$

where COH_{TD} describes the *Coherence* from D (driver) to T (target), DC_{TD} the *Directed Coherence*, $PCOH_{TD}$ the *Partial Coherence* and $GPDC_{TD}$ the *Generalized Partial Directed Coherence*. COH describes the linear relationship between two processes, whereas PCOH differentiates the direct from the indirect connections in the frequency domain. COH and PCOH are symmetric measures and they can be decomposed into subcomponents that express directionality (i.e. DC and PDC respectively). DC refers to the direct and indirect causal links between two time series. PDC on the other hand detects only directional influences without considering indirect paths of information flow. It holds that, $0 \le |DC_{TD}(f,n)|^2 \le 1$, $\sum_{m=1}^{M}|DC_{Tm}(f,n)|^2 = 1$ and $0 \le |GPDC_{TD}(f,n)|^2 \le 1$, $\sum_{m=1}^{M}|GPDC_{mD}(f,n)|^2 = 1$.

All the aforementioned are based on the assumption that the residuals are white and uncorrelated. However, we observed residuals with TV variances when datasets containing event-related changes or phase transitions were fitted to TV-MVAR models. We speculate that these transitions are translated into changes in the operating point of the system where the SNR levels may be different. The main issue with heteroskedasticity is that it leads to

underestimated error covariances ($\boldsymbol{\Sigma}$) which in turn bias the connectivity measures extracted from the MVAR models (Equations (4.17-4.20)). In practice, the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ is estimated either based on the residuals acquired from all the dataset or by applying windows and estimating an average value. In the latter case and in piecewise stationary approaches, these windows usually coincide with the windows used to estimate the TV model parameters. In recursive approaches, these windows are comparable in size with the recursive estimator's memory. Heteroskedasticity however implies that the covariance matrix $\boldsymbol{\Sigma}$ is TV in nature and does not necessarily follow the same TV patterns as the model parameters. Hence, we propose tracking the TV covariance matrix $\widehat{\boldsymbol{\Sigma}}(n)$ using Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models [189]. Subsequently, elements that belong to $\boldsymbol{\Sigma}$ in equations (4.15-4.20) are no longer constant and acquire different values at each time step.

4.2.6 Heteroskedasticity in the error terms

GARCH models are used to describe a changing, possibly volatile variance. Under heteroskedasticity, the variance of each error term is a function of time thus it follows that,

$$\varepsilon_m(n) \sim N(0, \sigma_m^2(n)) \qquad for\ m = 1 \dots M \tag{4.21}$$

A GARCH model of order (r, q) can model these changes by expressing the current conditional variance as a linear combination of past conditional variances and past squared errors,

$$\sigma_m^2(n) = \omega_m + \sum_{i=1}^{r} \gamma_{im} \varepsilon_m^2(n-i) + \sum_{j=0}^{q} \delta_{jm} \sigma_m^2(t-j) \tag{4.22}$$

Usually a GARCH(1,1) is enough to model the variance σ_m^2, however the optimal model order is selected based on the Bayesian (BIC) [39] or Akaike Information Criterion (AIC) [40]. In a more refined approach, this methodology can be extended to multivariate GARCH (MV-GARCH) models [190] taking into account interactions between different error time series and thus obtaining TV covariance matrices with off-diagonal elements. To summarize, based on the estimated TV parameters of the MVAR models and the covariance of the error terms, TV coupling measures can be computed from Equations (4.17-4.20). In case of heteroskedasticity, the time-invariant estimate $\widehat{\boldsymbol{\Sigma}}$ can be replaced at each time point with the TV covariance $\widehat{\boldsymbol{\Sigma}}(n)$ extracted by the GARCH model.

4.3 Simulations

4.3.1 Case 1: Homoskedastic error terms

We simulated fifty realizations of the following 3-dimensional (i.e. $M = 3$) TV-MVAR process of order 2 (i.e. $p = 2$),

$$\begin{bmatrix} y_1(n) \\ y_2(n) \\ y_3(n) \end{bmatrix} = \sum_{k=1}^{2} \begin{bmatrix} a_{11}^{(k)}(t) & a_{12}^{(k)}(t) & a_{13}^{(k)}(t) \\ a_{21}^{(k)}(t) & a_{22}^{(k)}(t) & a_{23}^{(k)}(t) \\ a_{31}^{(k)}(t) & a_{32}^{(k)}(t) & a_{33}^{(k)}(t) \end{bmatrix} \begin{bmatrix} y_1(n-k) \\ y_2(n-k) \\ y_3(n-k) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(n) \\ \varepsilon_2(n) \\ \varepsilon_3(n) \end{bmatrix} \tag{4.23}$$

TV relationships were achieved by allowing the model parameters to vary through time as depicted in Figure 4.1. The TV-MVAR process was driven by white noise signals $\boldsymbol{\varepsilon}$ of equal variances,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.24}$$

We estimated the TV model parameters using both the conventional KF technique (Equations (4.3-4.7)) and our proposed implementation (Equations (4.3-4.9)). Model order selection and hyperparameter optimization was achieved in both cases using the GA scheme described in Section 4.2.3. The GA algorithm was able to identify the right model order in all cases. The performance of the two estimators was assessed by computing the mean squared error (MSE(%)) between the actual and the extracted parameters before and after smoothing,

$$No\ smoothing: \qquad MSE(\%) = \frac{1}{Nd} \sum_{j=1}^{d} \sum_{n=1}^{N} \left[\Gamma_j(n) - \hat{\Gamma}_j(n) \right]^2 \tag{4.25}$$

$$With\ smoothing: \qquad MSE(\%) = \frac{1}{Nd} \sum_{j=1}^{d} \sum_{n=1}^{N} \left[\Gamma_j(n) - \hat{\Gamma}_j^{\,s}(n) \right]^2 \tag{4.26}$$

where $\Gamma_j(n)$ is the j-th parameter of the real parameter vector $\boldsymbol{\Gamma}$ at time n, $\hat{\Gamma}_j(n)$ and $\hat{\Gamma}_j^{\,s}(n)$ is the j-th parameter of the estimated unsmoothed ($\hat{\boldsymbol{\Gamma}}$) and smoothed ($\hat{\boldsymbol{\Gamma}}^{\boldsymbol{s}}$) parameter vector, respectively, at time n. In Figure 4.2 we present the real (simulated) and the estimated TV COH (averaged over all 50 realizations) using our proposed implementation. Note that TV COH was computed based on Equation (4.17). The covariance matrix $\hat{\boldsymbol{\Sigma}}$ was computed based on the residuals acquired from the whole dataset. Based on Figure 4.3-Homoskedastic errors, our

approach led to significantly lower MSE(%) values compared to the conventional KF scheme indicating superior tracking performance.

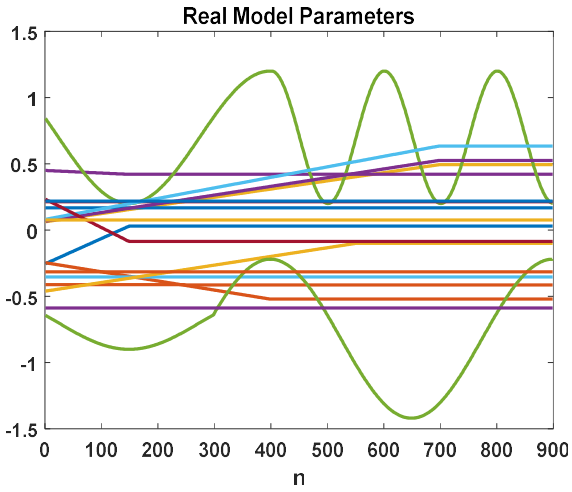


Figure 4.1 Real (i.e. simulated) TV model parameters. The parameters either remain constant throughout time or evolve as ramps or sinusoids with TV amplitude and frequency content.

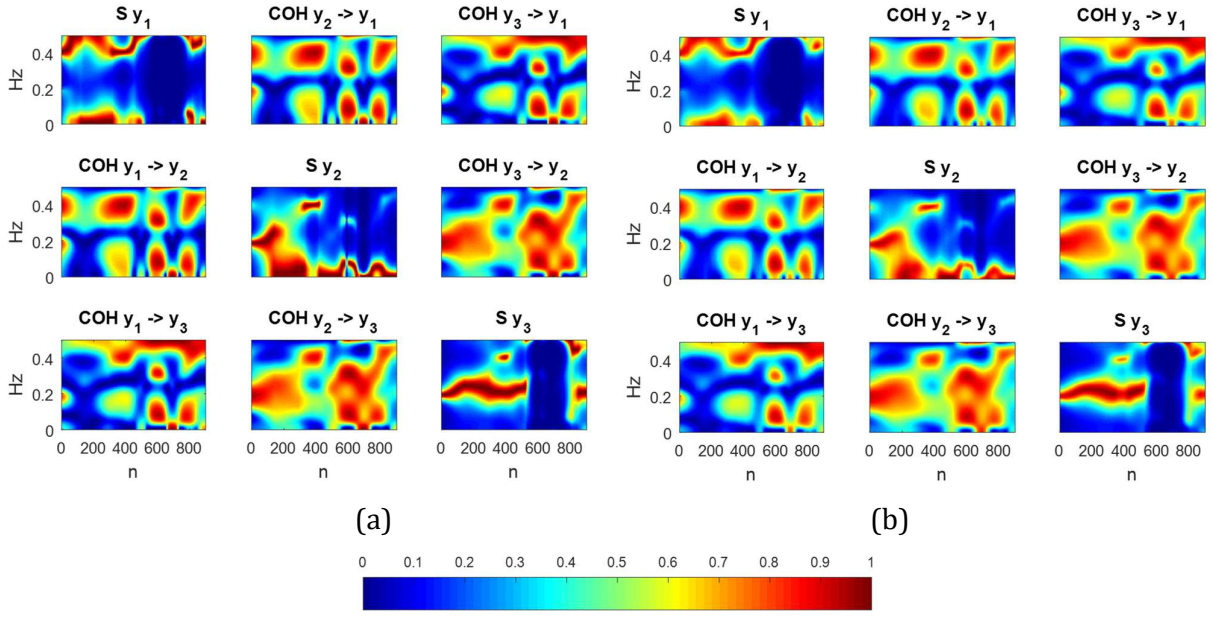


Figure 4.2 Heatmaps depicting the (a) real and (b) estimated TV COH under homoskedastic error terms (Case 1) using our proposed implementation (average from all 50 realizations). The x-axis represents time and the y-axis, frequency. y_D -> y_T denotes the driver and the target time series. COH is a symmetric measure that describes direct and indirect couplings (no directionality is assumed) between two time series. Thus, COH y_D -> y_T is equal to COH y_T -> y_D. The diagonal components of COH are the spectra (S) of each time series.

4.3.2 Case 2: Heteroskedastic error terms

Fifty realizations of the TV-MVAR process of Equation (4.23) were generated with parameters varying as in Figure 4.1. The process was driven by white noise signals of which were initially of equal variance, but that changed over time,

$$\boldsymbol{\Sigma_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad for\ n \leq 450 \tag{4.27}$$

$$\boldsymbol{\Sigma_2} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} for\ n > 450 \tag{4.28}$$

We estimated the TV parameters as in Section 4.3.1 and we additionally parametrized the TV covariance $\widehat{\boldsymbol{\Sigma}}$ using the GARCH approach described in Section 4.2.6. TV-MVAR measures were computed based on Equations (4.15-4.20) combined with Equation (4.22). Again, our proposed implementation exhibited significantly better performance than the conventional KF algorithm (Figure 4.3-Heteroskedastic errors). Further, based on Figure 4.3, the significantly higher MSE values under the heteroskedastic case compared to the homoskedastic case implies that heteroskedasticity affects in general the parameter estimation procedure. However, our proposed KF algorithm combined with smoothing performed similarly in both cases proving the superiority of our approach.

We also estimated the normalized mean squared error (NMSE) between the simulated and the predicted TV-MVAR measures of Equations (4.15-4.20),

$$NMSE(\%) = \frac{100}{M^2} \sum_{T=1}^{M} \sum_{D=1}^{M} \left\{ \frac{\sum_{n=1}^{N} \sum_{f} \left[Z_{TD}(f,n) - \hat{Z}_{TD}(f,n) \right]^2}{\sum_{n=1}^{N} \sum_{f} [Z_{TD}(f,n)]^2} \right\} \tag{4.29}$$

where $Z_{TD}(f,n)$ and $\hat{Z}_{TD}(f,n)$ refer to one of the simulated and estimated TV-MVAR measures, respectively, of Equations (4.15-4.20). In Figure 4.4a and 4.4b, we present the obtained NMSE(%) from Case 2, using the conventional and the proposed KF technique (after applying smoothing to the obtained parameters), when we either ignore or take into account heteroskedasticity. As expected, our proposed technique resulted into lower NMSE values for all TV-MVAR measures. Figure 4.5 depicts the simulated and estimated TV GPDC between all three time series. If we ignore the changes in the variance of the error terms then GPDC is either overestimated or underestimated in some time intervals (Figure 4.5b). This may lead to wrong

conclusions when interpreting the obtained results. For example, Figure 4.5b shows a larger flow of information from y_3 to y_2 at the frequency range of [0.2 0.3Hz] after time point 450, when, in reality, the causality strength between these two time series is higher before that time point (Figure 4.5a). This actuality is captured better when applying our proposed GARCH approach (Figure 4.5c).

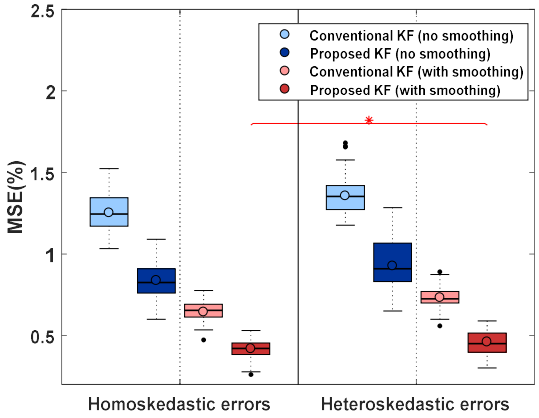


Figure 4.3 Boxplots depicting the MSE(%) between real and estimated parameters, before and after applying smoothing, using the conventional KF and our proposed implementation under homoskedastic (Case 1) and heteroskedastic (Case 2) error terms. *: denotes statistically nonsignificant differences (assessed using Analysis of Variance along with multiple comparison tests). Within and between each case all comparisons were found to be statistically significant except of the one denoted with a star.

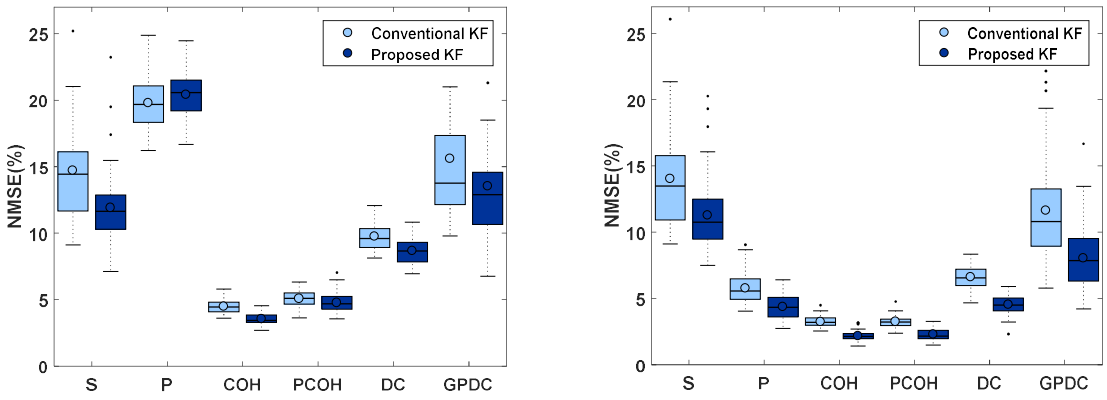


Figure 4.4 Boxplots depicting the NMSE(%) between real and estimated TV-MVAR measures using the conventional KF and our proposed implementation under heteroskedastic (Case 2) error terms (a) ignoring heteroskedasticity (b) taking into account heteroskedasticity.

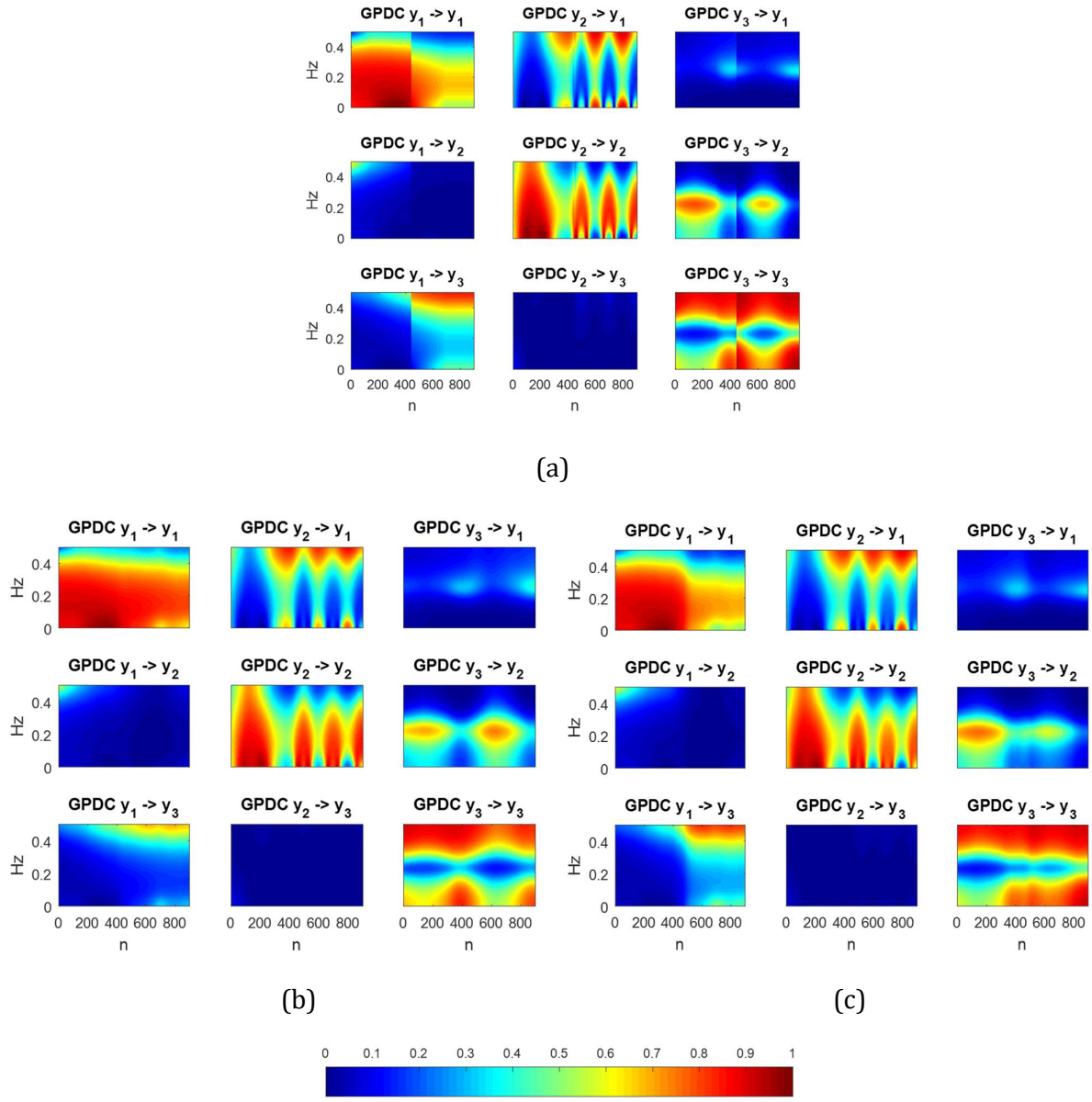


Figure 4.5 Heatmaps depicting the (a) real and (b) estimated TV GPDC for Case 2 using our proposed implementation ignoring heteroskedasticity (average from all 50 realizations) and (c) estimated TV GPDC for Case 2 using our proposed implementation taking into account heteroskedasticity (average from all 50 realizations). The x-axis represents time and the y-axis frequency. The variance of the driving noise changes at time point 450. y_D -> y_T denotes the driver and the target. GPDC is an asymmetric measure that describes direct causality between two time series. Thus, GPDC y_D -> y_T is not the same as GPDC y_T -> y_D.

4.4 Experimental Data

Physical exercise is known to boost physical and mental health, through a series of cardiac, vascular and metabolic adaptations. Growing evidence suggests that even a single exercise session enhances cerebrovascular function in healthy adults and patients with stroke [191]–[193]. In order to better understand the cerebrovascular/cardiovascular changes that occur during physical activity we applied our proposed TV-MVAR methodology using stroke volume (SV), pulse pressure (PP) and middle cerebral artery pulsatility index (PI) as signals of interest from 3 groups; stroke survivors (17 participants), young (12 participants) and old healthy adults (10 participants), performing 20 min of moderate intensity stationary cycling. Prior to presenting the obtained results, we briefly explain the nature of the investigated signals.

SV is the volume of blood pumped through the left heart ventricle with each contraction. We obtained a non-invasive estimation of SV based on a Modelflow prediction algorithm using the finger arterial pulse waveform (Beatscope Software, Finapres Medial Systems, Amsterdam NL) [194]. PP describes the differential in arterial blood pressure (ABP) between the systolic and diastolic phases of the cardiac cycle, where systolic and diastolic reflect the maximum and minimum values, respectively. We reconstructed systolic and diastolic pressures within the brachial artery using a generalized waveform filter on ABP obtained using finger-cuff photoplethysmography (Finapres Medial Systems) [195]. PP is a function of the volume of blood ejected from the heart, as well as the compliance of the central arteries. As more blood is ejected from the left ventricle to the arteries, PP increases. With age, arteries become stiffer leading to increases in systolic pressure and decreases in diastolic pressure [196], [197]. It has been shown that, along with the observed PP increase after age 50, there is a prominent age-related decrease in SV. The ratio of SV to PP (SV/PP) remains constant between 17 and 50 years before declining abruptly. Hence, SV/PP is regarded as a simplified index of arterial compliance [198] and has been proposed as a biomarker of cardiovascular risk [199]. PP itself and arterial stiffness are positively associated with increased risk of stroke and coronary heart disease [200]. A number of studies have shown that physical activity can lead to short- and long-term increases in arterial compliance, reducing the negative effects of arterial stiffness [201]–[206]. PI reflects the intra-beat pulsatile characteristics of cerebral blood flow velocity (CBFV) and is calculated as the difference between the peak systolic and minimum diastolic velocities divided by the mean velocity during one cardiac cycle. We measured CBFV in the bilateral middle cerebral arteries using 2-MHz transcranial Doppler ultrasound with a secure headframe to

maintain positioning throughout exercise (ST3 Transcranial Doppler, Spencer Technologies, Redmond WA USA). PI also increases with age due to changes in PP and arterial stiffness [207]. Excessive pulsatility may expose the brain to harmful levels of pressure and flow provoking detrimental effects to cerebral structure [208].

In this work, we are interested in examining the short-term relationship between PP, SV and PI before, during and after exercise. Studying these TV interactions may help illuminate the exercise-induced mechanisms of cerebrovascular/cardiovascular regulation and detect pathological or age-related deficits. The experimental protocol consisted of the following six phases,

- Phase 1 – 5 minutes of seated rest
- Phase 2 – 3 minutes warm-up with an increase in workload each minute
- Phase 3 – 20 minutes of steady state exercise
- Phase 4 – 2 minutes cool-down with a decrease in workload each minute
- Phase 5 – 1 minute cool-down in cycling position without pedaling
- Phase 6 – 5 minutes of post-exercise recovery and seated rest

The experimental protocol was approved by the Sunnybrook Health Sciences Research Ethics Board. All participants provided written informed consent. Beat-to-beat data was extracted from continuous ABP and CBFV waveforms, and interpolated to 1Hz for analysis. By applying our proposed methodology, we extracted TV measures of COH, PCOH, DC and PDC in all subjects. No preprocessing (e.g. mean removal, standardization) was applied to the data but we did add a constant offset vector to Equation (4.1) that was optimized by the GA along with the rest hyperparameters. Note that we detected heteroskedasticity on the residuals associated with phase transitions. Hence, we applied the GARCH approach presented in Section 4.2.6 to parametrize the changing variance of the error terms. For each group, i.e. healthy young, healthy old and stroke survivors, we computed median TV patterns of COH, PCOH, DC and PDC (Figure 4.6a,b,c). We also estimated the TV changes of each measure from baseline (Figure 4.6d,e,f) as follows,

$$dZ_{TD}(f,n) = Z_{TD}(f,n) - \text{median}[Z_{TD}(f,T_{baseline})] \tag{4.30}$$

where $Z_{TD}(f,n)$ is one of the TV-MVAR measures (Equations (4.17-4.20)) at time point n and frequency f, $T_{baseline}$ is the baseline period, and median$[Z_{TD}(f,T_{baseline})]$ is the median baseline value of Z_{TD} at frequency f. We focused on four frequency bands of interest, very low

(VLF; 0.01-0.04 Hz), low (LF; 0.04-0.15 Hz), high (HF; 0.15-0.3 Hz) and very high (VHF;0.3-0.4Hz). We extracted bandlimited median values of each TV measure for all subjects during baseline (Phase 1), exercise (Phase 3) and post-exercise recovery (Phase 6).

All groups exhibited the same hemodynamic profile; a progressive dissociation between SV and PP after the onset of the exercise and a return to baseline or higher than baseline COH values during post-exercise recovery (Figure 4.6 and Figure 4.7). The loss of coherence between SV and PP is in accordance with previous studies that observed exercise-induced decreases in the SV/PP ratio [209] and the coherence between SV and SBP [210]. This implies that arterial stiffness increases during exercise. This may be explained by the fact that exercise stress activates the sympathetic nervous system [211] leading to ABP and heart rate (HR) increases [212] that temporarily induce vasoconstriction and arterial stiffness. By examining Figure 4.6 one can see a progressive decrease in COH and dCOH values from young to elderly and to stroke survivors. This is expected since SV/PP ratio is an indirect measure of arterial compliance and is known to decrease with age and stroke occurrence. Overall the young group exhibited higher VLF and LF COH levels, whereas stroke survivors and elderly displayed higher HF COH values (Figure 4.6a,b,c and Figure 4.7a,b,c). During cycling, there were significant HF COH decreases from baseline in all groups (Figure 4.6d,e,f and Figure 4.7e). The most striking changes though occurred in the young group during the recovery phase. LF COH increased significantly compared to the other groups (Figure 4.6a,d and Figure 4.7a,d). In elderly and stroke survivors COH returned to its baseline values.

DC (not shown here) revealed that the main directionality effects were from SV to PP, indicating that changes in SV lead to changes in PP and this is physiologically expected. Based on the GPDC (SV->PP) estimates, direct TV effects from SV to PP in the young group were more pronounced in the HF and VHF range (Figure 4.8a,d). This component was frequency specific and it was observed around 0.25-0.35Hz. On the other hand, in the elderly and stroke survivors, SV directly affected PP more strongly in the LF range. Based on the aforementioned, interactions in the LF and HF range seem to represent two distinct mechanisms with the HF component being more active in the younger subjects. The LF and HF range in cardiovascular systems are known to represent sympathetic and parasympathetic activity respectively [213], [214]. Physiological ageing and stroke occurrence is linked with autonomic dysregulation and more specifically with a reduction in parasympathetic control and upsurge in sympathetic tone [215]–[217]. This could explain the absence of a HF GPDC component during baseline in the

elderly and stroke group (Figure 4.8b,c). In [218], the authors detected a lesser decrease in parasympathetic tone during exercise in healthy old compared to young subjects due to lower parasympathetic tone levels in the elderly. This is in accordance with the TV patterns seen in dGPDC (Figure 4.8d,e,f). The young group exhibited a larger drop in HF GPDC from baseline. Both elderly and stroke survivors displayed significant increases in LF and HF/VHF dGPDC (statistics not shown here) during recovery indicating probable increases in both sympathetic and parasympathetic activity. This implies that exercise induces partial restoration of parasympathetic tone that is lost post-stroke and with age. It should be noted here that COH describes both the direct and indirect influences on the SV-PP relationship. GPDC however expresses the directional influence of one signal to the other and thus reflects different regulatory actions. No group differences or exercise-related changes were found in associations between PI and SV or PP. The relationship between PI and the other signals was stationary and more prominent in the VLF/LF range. PP and PI are linked to Cerebral Autoregulation (CA) [57], a mechanism that retains blood flow to the brain constant despite variations in ABP. Based on transfer function analysis, it has been previously shown that CA is maintained during low and moderate intensity exercise in the LF range [219]. This has also been demonstrated in healthy aging [220].

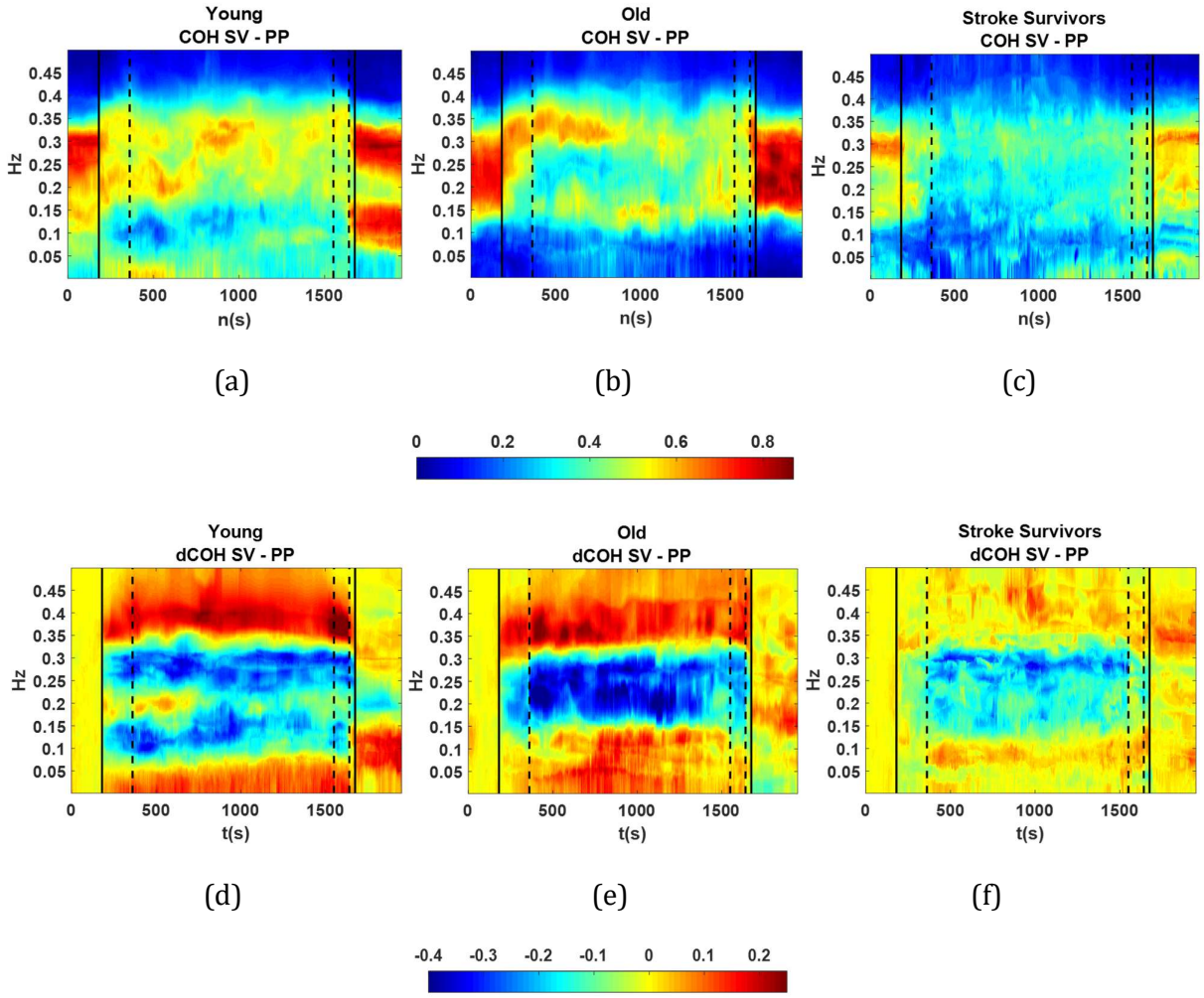


Figure 4.6 Heatmaps depicting the median TV COH and dCOH (changes from baseline) between SV and PP obtained from (a,d) the young, (b,e) the old and (c,f) the stroke survivors group respectively. The solid black lines denote the onset and the offset of the cycling phase (onset of Phase 2 and onset of Phase 6). Dashed lines denote intermediate phases.

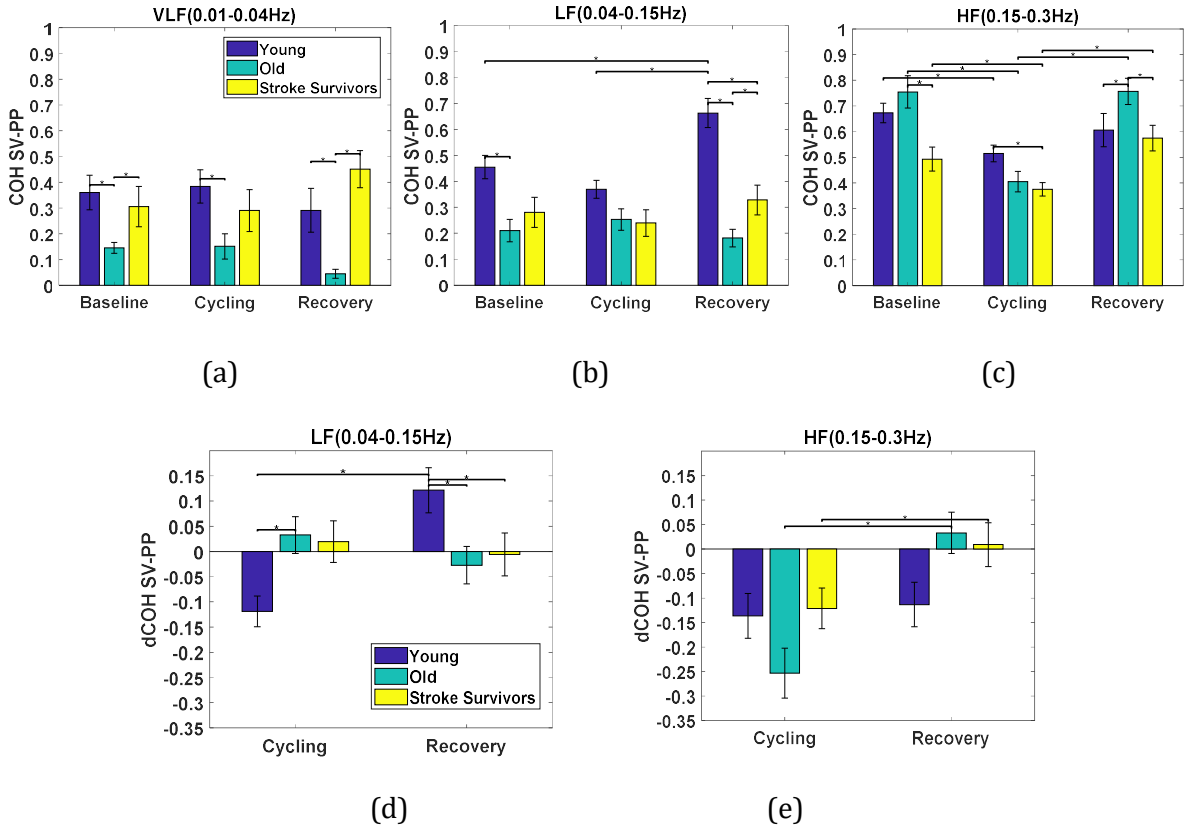


Figure 4.7 Error bars depicting median COH and dCOH values between SV and PP obtained from all groups during baseline, cycling and post-exercise recovery in the (a) VLF, (b,D) LF and (c,e) HF band. *: denotes statistically significant differences (assessed using Analysis of Variance along with multiple comparison tests). The VHF band of dCOH and COH was omitted due to nonsignificant differences. The same applies for the VLF band of dCOH. In dCOH baseline phase was not taken into account since dCOH measures changes from baseline.

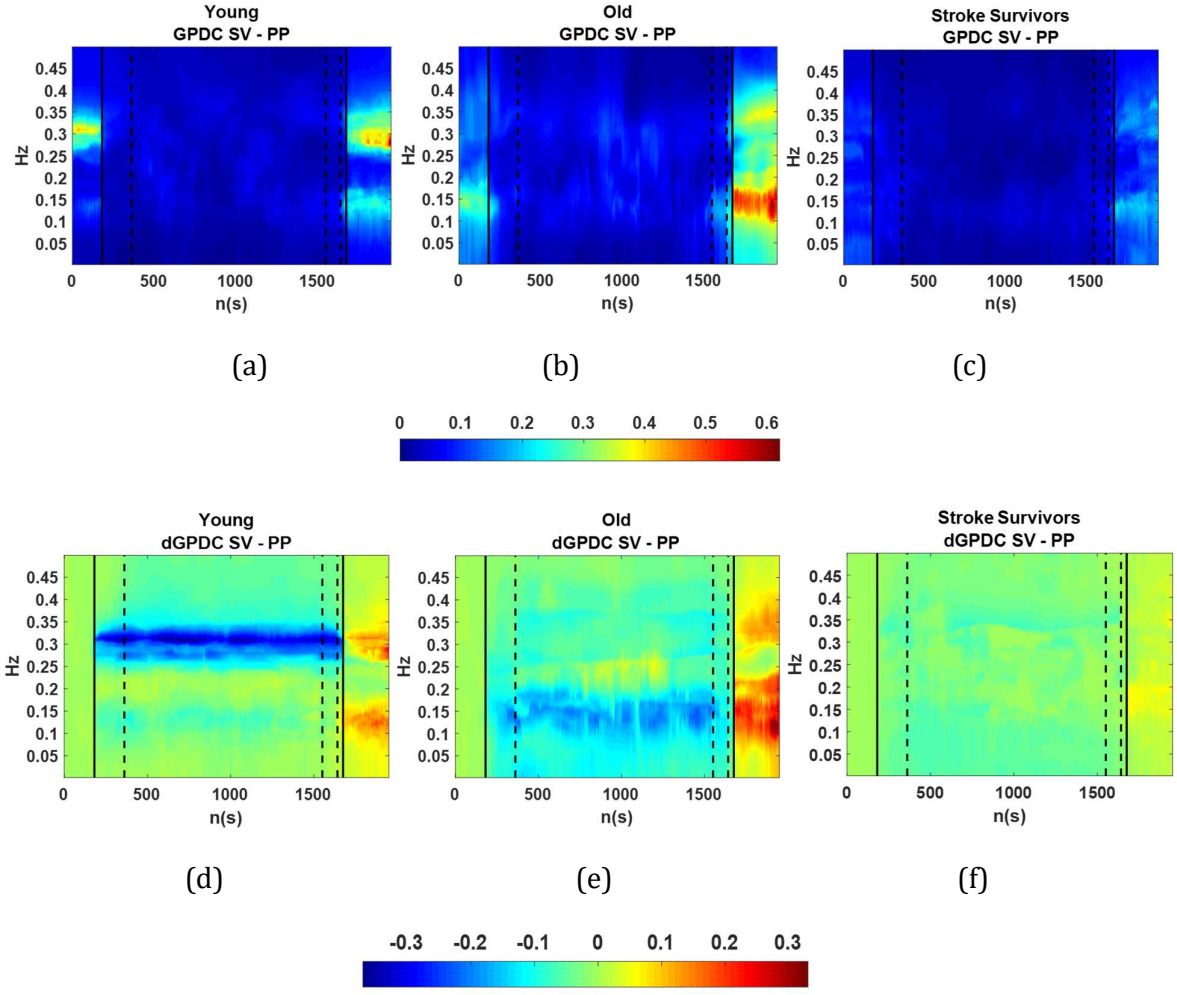


Figure 4.8 Heatmaps depicting TV GPDC SV->PP and dGPDC SV->PP (changes from baseline) between SV and PP obtained from (a,d) the young, (b,e) the old and (c,f) the stroke survivors group respectively. The solid black lines denote the onset and the offset of the cycling phase (onset of Phase 2 and onset of Phase 6). Dashed lines denote intermediate phases.

4.5 Conclusions

In this paper, we have developed a TV-MVAR methodology that tracks accurately TV interactions between time series. Our estimation method is based on the KF technique modified to accommodate multiple adaptive update coefficients allowing independent tracking of each model parameter. A mixed integer GA was assigned to simultaneously select the optimal model order and tune the hyperparameters of the proposed recursive scheme avoiding this way exhaustive search procedures. Heteroskedasticity in the error terms was parametrized using GARCH models. Based on simulations, our approach achieved superior performance compared to conventional techniques and led to more accurate representations of the true underlying TV interactions in both time and frequency domain. We applied our methodology in real experimental data in order to track exercise-induced hemodynamic changes in young, old and stroke survivors. Our goal was not to conclude any clinical findings, but demonstrate the capabilities of the proposed method. Future work involves the incorporation of a recursive GARCH estimation methodology to the proposed recursive scheme in order to simultaneously track the heteroskedasticity in the error terms and update the model parameters. One of our main priorities however is the development of a statistical framework for significance testing of TV COH, PCOH, DC and PDC. Although there are existing statistical techniques for the stationary case, only few studies have attempted to address this issue under TV conditions [221]–[224].

Chapter 5 Modeling systems with binary output for biomedical applications

In the biomedical field, reaserchers usually analyse continuous variables. Chapters 3 and 4 focused on physiological systems with continuous output. However, there are applications where event-related dynamics are of main concern (e.g. prediction of the neuronal spiking activity in the brain or the assessment of heartbeat dynamics). In such type of systems, the output is expressed as a binary signal where '1' indicates the occurrence of an event and '0' the absence of an event in time. Herein, we have developed a LVN network, with and without autoregressive terms, that can produce binary output predictions. The modified LVN model was used to predict neuronal spikes from Local Field Potentials (LFP) in the parkinsonian subthalamic nucleus (STN) in an attempt to understand the underlying mechanisms that affect the responsiveness of a neuron to the LFP.

Prediction of the spiking activity in the Parkinsonian subthalamic nucleus using Local Field Potentials and Laguerre Volterra Networks

Kyriaki Kostoglou[1], Konstantinos P. Michmizos[2], Pantelis Stathis[3], Damianos Sakas[3], Konstantina S. Nikita[4] and Georgios D. Mitsis[5]

[1]Department of Electrical and Computer Engineering, McGill University, Montreal, Canada.

[2]Department of Computer Science, Rutgers University, New Brunswick, NJ, USA.

[3]Department of Neurosurgery, National and Kapodistrian University of Athens, Athens, Greece.

[4]School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece.

[5]Department of Bioengineering, McGill University, Montreal, Canada.

(To be submitted to Journal of Neuroscience)

Abstract **—** In this study, we utilized a data-driven approach, without relying on feature selection, to model the causal and dynamic relationship between Local Field Potentials (LFP) and spike trains in the Subthalamic Nucleus (STN) of 20 Parkinson's Disease (PD) patients undergoing Deep Brain Stimulation (DBS). Specifically, we employed Laguerre-Volterra Networks (LVN) with and without autoregressive terms modified to produce probabilities of firing in their output. Network parameters were trained by applying a hybrid optimization scheme (genetic algorithm and interior nonlinear method) for faster and more efficient training. The results indicate that LFPs can predict the PD STN spiking activity in a window of ±1ms. Based on the predictive performance of each LFP-spike pair we detected three neuronal clusters with distinct spatiotemporal characteristics linked indirectly with "off" state UPDRS improvement following DBS surgery.

Index terms **—** Spike Prediction, Local Field Potentials, Laguerre Volterra Network, Parkinson's Disease, Subthalamic Nucleus.

5.1 Introduction

Extracellular microelectrode recordings (MERs) of electrical activity in the brain are typically separated into two major counterparts that represent different aspects of neural signaling. The low frequency component, known as Local Field Potentials (LFP), is mostly generated by the collective activity of aggregates of neurons around the recording site (0.5 – 3mm) [225] and it is believed to comprise the synchronised input into the observed region. On the other hand, action potentials (AP) of nearby neurons (in a distance of 100-300μm from the electrode tip) [226] as well as smaller sub-noise level spikes, which are referred to as Multiunit Activity (MUA), are reflected in the high frequency content of the MER signals and represent the output from the observed region. LFPs and their spatiotemporal characteristics have been used extensively in various applications (e.g. studying brain rhythms [173], detecting pathological network activity [227], [228] and brain-machine interface applications [229]). On the other hand, the temporal structure of spiking activity has been of particular interest from the neural coding point of view, since it has been shown that information can also be carried in terms of the precise timing of spikes [230]–[233]. A number of studies have shown that temporal coding is achieved on a millisecond time scale [233]. Although spikes have been treated as the main "carriers" of information it has been shown that LFP can be as efficient as a individual neuron in terms of decoding behaviors and brain activity [234], [235]. Given the entangled but yet different nature of the LFP and the MUA one might ask whether it is possible to predict spikes solely on the basis of information conveyed in the LFP. Previous studies, have suggested that there is indeed a relationship between LFPs and spike timing in both animals [236]–[239] and humans [54], [240]–[242].

In this work, we followed a data-driven approach to capture the causal relationship between LFPs and spikes in the Parkinsonian Subthalamic Nucleus (STN). MERs were obtained from 20 Parkinson's disease (PD) patients undergoing Deep Brain Stimulation (DBS) [243]. The specific dataset has been used previously to classify PD patients based on their response to DBS and predict their improvement [228], [244]. Herein, we focus on understanding the LFP triggering mechanisms of the STN spiking activity and quantify the amount of LFP information that affects multiple or even individual neurons. In order to map the relationship between LFP and spikes we used Laguerre-Volterra Networks (LVN) that have been previously successfully applied in identifying physiological systems [14], [17]. The network was modified to produce probabilities of "firing" in its output and model parameters were trained using a hybrid genetic

algorithm (GA) – nonlinear interior point optimization technique overcoming this way gradient-based local minima and convergence problems. We extended the model by adding autoregressive terms (LNV-ARX) and we compared its performance to the initial network. Various time and frequency domain LFP/MUA features were examined in an effort to understand the observed variability in spike timing predictability between recordings. We extracted three neuronal clusters with different spatiotemporal and spectral characteristics linked with high, moderate and low predictive accuracy. These clusters are very similar to neuronal populations that have been observed previously in the Parkinsonian STN [245] and in the temporal cortex of epileptic patients [54].

5.2 Methods

5.2.1 Data acquisition and Signal Processing

MERs were obtained from 20 "off" state PD patients undergoing DBS surgery at the Neurosurgery Clinic, Evangelismos General Hospital, Athens, Greece. MERs were performed in 0.5mm steps, starting 5mm above the MRI-defined target, through the STN and towards the substantia nigra, using 5 parallel microelectrodes in a cross "Ben Gun" configuration [246]. During the recording phase (10sec) no electrical stimulation was applied.

MERs exhibiting spontaneous STN activity were only used in this study. LFPs were acquired by low-pass filtering (FIR equiripple) the raw signals (sampling rate: 12k HZ) with a cutoff frequency of 200 Hz. MUA was extracted by applying a high-pass filter to the raw signals with cutoff frequency around 500 Hz. The exact procedure is described elsewhere [228], [240]. 50 Hz power-line noise and its harmonics were removed from the LFP signal by fitting sine/cosine waves at the specified frequencies and subsequently subtracting the estimated components. All LFPs were normalized to zero mean and unit variance. Spike detection was achieved through amplitude thresholding of the MUA. The threshold was set to,

$$threshold = 4\sigma_{\boldsymbol{n}}, \qquad \sigma_{\boldsymbol{n}} = \frac{median(|\boldsymbol{h}|)}{0.6745} \tag{5.1}$$

where $\boldsymbol{h}$ is the MUA signal and $\sigma_{\boldsymbol{n}}$ is an estimate of the standard deviation of the background noise. Time points where the signal exceeded the $threshold$ and exhibited a maximum value in a window of ±1ms where selected as the timestamps of the multi-unit spike train. Based on this vector of timestamps, a binary signal was produced, indicating the absence ('0') or presence

('1') of a spike. Both LFP and spike trains were downsampled at a frequency of 1kHz in order to reduce the computational complexity. The new sampling rate was chosen based on the minimum interspike interval (ISI) observed amongst all recordings (2ms). Spike trains were produced by shifting each spike to the nearest sampling point after applying downsampling. The original waveforms acquired from each detected spike where analysed using *wave_clus* [247], a spike sorting Matlab toolbox, which applies superparamagnetic clustering for the identification of single-cell neuronal sources. Both single- (SUA) and multi-unit (MUA) activity was included in our analysis.

A large variety of MER features were extracted in the effort to correlate them later with spike prediction accuracy. Low-frequency (<200Hz) background unit activity (BUA) envelopes were extracted from the MUA signals following the procedure described in [245], [248]. As mentioned earlier, MUA is a mixture of APs of neurons very close to the electrode tip as well as smaller sub-noise level spikes from nearby populations that comprise the BUA. BUA characteristics can give us insight about the activity of neurons in closer proximity than those recorded from each one of the 5 microelectrodes. We computed power band ratios, peak-to-average power ratios, and mean/max coherence between LFPs, SUA/MUA and BUA in the delta (D; 1–4Hz), theta (T; 4–12 Hz), low beta (LB; 12–30 Hz), high beta (HB; 30-45Hz), gamma (G; 45–100 Hz) and high gamma (HG; 100–200 Hz) frequency bands. Power spectral densities and magnitude squared coherences were estimated by applying Welch's method to consecutive 1.5s segments with 50% overlap. We also included information theory measures [249] like entropy and mutual information (MIn) in order to quantify the stochastic variability of the signals and their pairwise association respectively. Temporal coordination of neural activity was described using cross-frequency coupling (CFC) and phase-locking (PL) indices. LFP phase-amplitude CFC between different subbands was assessed using the mean vector length modulation index methodology [250]. Phase synchronisation between spiking units and LFP was computed using a similar approach as CFC. PL index in a specific frequency range was defined as the modulus of the average value of the analytical bandpass filtered LFP signal, extracted using Hilbert Transform [251], during spiking events. The angle of this complex number represents the phase preference of the neuronal unit. Features related with the neuronal firing characteristics were also examined (i.e. mean firing rate, interspike interval distribution, bursting index based on the Poisson Surprise method [252], [253]).

5.2.2 Modeling the relationship between LFP and Spikes

5.2.2.1 Laguerre-Volterra Network (LVN)

The main aim of this study was to predict spikes from LFPs. In order to achieve this however, the precise knowledge of the relationship between the two signals is needed. System identification is the process of building dynamical models from measured data to determine and express the predictive relationships between them. One commonly used nonparametric model structure is the Volterra model [1]. The main aim of the identification procedure is to extract the Volterra kernels of the system. The Volterra kernels can be viewed as weighting functions that describe the effect of past input values (linear kernel), as well as the effect of the Q-th order products between past values of the input (nonlinear kernel) in order to generate the output signal. An efficient way to estimate these kernels is the Laguerre expansion technique (LET) [1]. Specifically, the discretized Volterra kernels of the system can be expanded in terms of the orthonormal basis of discrete-time Laguerre functions (DLFs) achieving this way model compactness and estimation accuracy.

LET can be combined with feedforward artificial neural networks and hidden units with polynomial activation functions in the form of the Laguerre – Volterra network (LVN) [30]. LVNs require a low total number of unknown parameters and thus produce good results even in small datasets. The input signal is initially convolved with a set of linear filters to allow the identification of both slow and fast dynamics. These counterparts of the input are then fed into the units of the hidden layer inducing this way nonlinearities. The summation of the units outputs produces the model prediction. In more detail, the j-th order DLF corresponding to filter bank i is defined as,

$$b^{(i)}{}_j(m) = \alpha_i^{(m-j)/2}(1-\alpha_i)^{1/2}\sum_{k=0}^{j}(-1)^k\binom{m}{k}\binom{j}{k}\alpha_i^{j-k}(1-\alpha_i)^k \tag{5.2}$$

where $j = 1 \dots L_i$ and α_i is the Laguerre parameter ($0 < \alpha_i < 1$) which determines the rate of exponential decay of the DLFs. The two filter banks are characterized by different Laguerre parameters, describing this way different interdependent input dynamics (e.g. slow and fast dynamics), and may contain different numbers of Laguerre functions (filters) L_i. The corresponding filter output $v_j^{(i)}(n)$ is the convolution of $b^{(i)}{}_j(m)$ with the input $x(n)$. The variables $v_j^{(i)}$ can be estimated using the following autorecursive relation [34],

$$v_j^{(i)}(n) = \sqrt{\alpha_i} v_j^{(i)}(n-1) + \sqrt{\alpha_i} v_{j-1}^{(i)}(n) - v_{j-1}^{(i)}(n-1) \tag{5.3}$$

initialized by,

$$v_0^{(i)}(n) = \sqrt{\alpha_i} v_0^{(i)}(n-1) + T\sqrt{1-\alpha_i} x(n) \tag{5.4}$$

where T is the sampling interval. The output of each filter bank at each time step n is $\boldsymbol{v}^{(i)}(n) \in \boldsymbol{R}^{L_i \times 1}$. The hidden units in the second layer employ polynomial activation functions in order to make the network functionally equivalent to a Volterra model. The input of each hidden unit k is the weighted sum of the DLF filter bank outputs,

$$u_k(n) = \sum_{i=1}^{2} \boldsymbol{w}_k^{(i)^T} \boldsymbol{v}^{(i)}(n) \tag{5.5}$$

where $k = 1 \ldots K$ and $\boldsymbol{w}_k^{(i)} \in \boldsymbol{R}^{L_i \times 1}$ are the weights**.** $u_k(n)$ then undergoes a Q-th order nonlinear transformation of the following form,

$$z_k(n) = \boldsymbol{c}_k{}^T \tilde{\boldsymbol{u}}_k(n) \tag{5.6}$$

where $\boldsymbol{c}_k \in \boldsymbol{R}^{Q \times 1}$**,** Q is the degree of the polynomial activation functions ($Q=1$ refers to a linear model, whereas $Q>1$ to a Q-th order nonlinear model) and $\tilde{\boldsymbol{u}}_k(n) = [u_k(n) \quad \ldots \quad u_k{}^Q(n)]^T$. The LVN output is given by the nonweighted summation of the hidden-unit outputs including a trainable offset y_0,

$$y(n) = y_0 + \sum_{k=1}^{K} z_k(n) \tag{5.7}$$

The Volterra kernels of the system can be easily expressed in terms of the network parameters [1]. However, a more interpretable representation of the system is the extracted Principal Dynamic Modes (PDMs) model. The PDM model, which is equivalent to the LVN, consists of a minimum set of parallel filters (PDMs), adequate to represent the kernels of the system, followed by a static nonlinearity [1]. In the LVN case the number of PDMs is defined by the number of hidden units (Fig.1). The PDMs are computed as,

$$PDM_k(m) = \sum_{i=1}^{2} \boldsymbol{w}_k^{(i)^T} \boldsymbol{b}^{(i)}(m) \tag{5.8}$$

where $\boldsymbol{b}^{(i)}(m) \in \boldsymbol{R}^{L_i \times 1}$ consists of all the DLFs $b^{(i)}{}_j(m)$.

5.2.2.2 Autoregressive Laguerre-Volterra Network with exogenous input (LVN-ARX)

The LVN, in its conventional form, describes a dynamic causal effect from input to output. However, there are cases where feedbacks from the output enter the system. To this end, we extended the LVN model in an autoregressive with exogenous input (ARX) formulation, where the autoregressive (AR) and the exogenous terms represent the history of the output and input respectively (Figure 5.1). Two extra filter banks are added in the network to account for the AR terms. For the AR terms, Equations (5.3) and (5.4) become,

$$v_{y_j}^{(i)}(n) = \sqrt{\alpha_{y_i}} v_{y_j}^{(i)}(n-1) + \sqrt{\alpha_{y_i}} v_{y_{j-1}}^{(i)}(n) - v_{y_{j-1}}^{(i)}(n-1) \tag{5.9}$$

initialized by,

$$v_{y_0}^{(i)}(n) = \sqrt{\alpha_{y_i}} v_{y_0}^{(i)}(n-1) + T\sqrt{1-\alpha_{y_i}}\, y(n-1) \tag{5.10}$$

The two extra filter banks are assigned with different Laguerre parameters α_{y_i} and may contain different numbers of Laguerre functions L_{y_i}. Equation (5.5) can be written as,

$$u_k(n) = \sum_{i=1}^{2} \boldsymbol{w}_k^{(i)^T} \boldsymbol{v}^{(i)}(n) + \sum_{i=1}^{2} \boldsymbol{w}_{y_k}^{(i)^T} \boldsymbol{v}_y^{(i)}(n) \tag{5.11}$$

where $k = 1 \dots K$, $\boldsymbol{w}_k^{(i)} \in \boldsymbol{R}^{L_i \times 1}$ and $\boldsymbol{w}_{y_k}^{(i)} \in \boldsymbol{R}^{L_{y_i} \times 1}$.

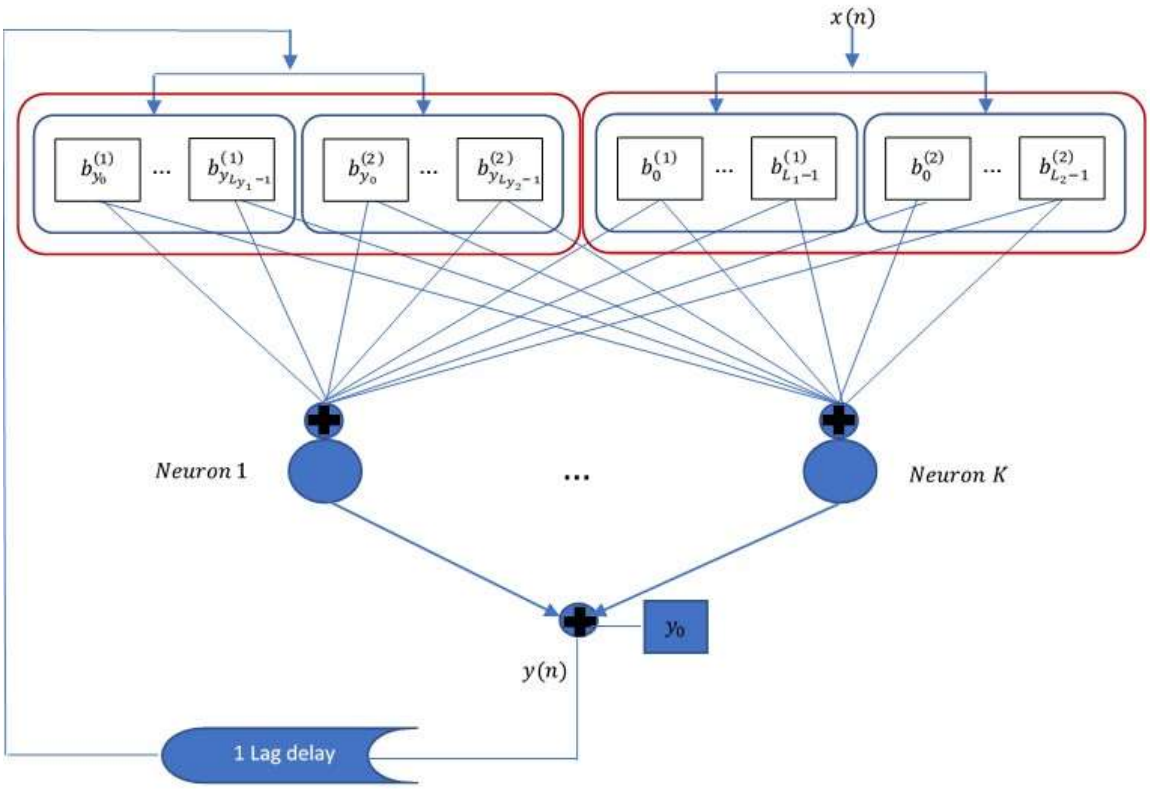


Figure 5.1 LVN-ARX schematic representation. A fully connected network with two Laguerre filter banks $\{b_j^{(1)}\}$ and $\{b_j^{(2)}\}$ that preprocess the input $x(n)$ and two Laguerre filter banks $\{b_{y_j}^{(1)}\}$ and $\{b_{y_j}^{(2)}\}$ that preprocess the output $y(n)$. The neurons (or else hidden units) in the hidden layer, described by polynomial activation functions, receive input from the filter banks. The output $y(n)$ is given by the nonweighted summation of the hidden-unit outputs including a trainable offset y_0. Note that for the simple LVN there is no feedback from the output back to the network.

5.2.2.3 LVN training and Spike prediction

Based on Equation (5.7), the LVN (or LVN-ARX) output is a continuous signal whereas the output signal in our dataset is binary ('1': spike event – '0': nonspike event). The MSE (Mean Squared Error) loss is typically not a good one for classification, as it forces the model to exactly predict the values imposed by the targets. Moreover, gaussianity assumption of the target data in classification is not valid, due to its discrete nature. Instead, a more commonly used, probabilistic objective is the negative log-likelihood. To minimize a negative log-likelihood, we first need to turn the predictions of our models into properly normalized log-probabilities. To this end, we used a sigmoidal activation function at the output layer of the LVN, transforming the predicted continuous output into probabilities of firing,

$$p(n) = \frac{1}{1 + e^{-(s\hat{y}(n)+y_0)}} \tag{5.12}$$

where $\hat{y}(n)$ is the output of the LVN at time n, s and y_0 the slope and the bias of the sigmoidal function, respectively. We then trained our model using the cross-entropy loss function defined as

$$J = -\frac{1}{N}\left\{\sum_{t=1}^{N} y(n)\ln[p(n)] + \sum_{t=1}^{N}[1 - y(n)]\ln[1 - p(n)]\right\} \tag{5.13}$$

where N is the total number of time samples, $y(n)$ is the observed binary spike signal at time n and $p(n)$ the probability of firing. Gradient based techniques (e.g. back-propagation) are usually used to train networks with fixed topologies. However, they may be easily trapped into local minima if they are not initialised properly. To this end, we applied a hybrid scheme, a genetic algorithm (GA) combined with interior point constrained nonlinear method, to minimize Equation (5.13) and estimate the LVN parameters (i.e. $\boldsymbol{w}_k^{(i)}$, $\boldsymbol{c}_k$, α_i, s and y_0). GAs are adaptive search algorithms based on the idea of natural evolution. A population of candidate solutions, called individuals, evolves toward better solutions by operators such as crossover, mutation and selection. GAs can efficiently search large and complex spaces to find nearly global minima and can be used as a complement to gradient-based techniques. In our case, the "optimal" solution obtained from the GA was fed into the nonlinear optimization method as an initial starting point for a more efficient local search. Note here that neural networks, in general, suffer from permutation problems (i.e. symmetry in representation). For example, two networks with different ordering of their hidden units might be structurally different but computationally they

can perform the same. In order to facilitate the optimization procedure, we imposed structural constraints to the problem (for networks with more than 2 hidden units) by forcing the root mean square (RMS) of the hidden units output in a descending order; i.e. the first unit should produce the highest RMS output and the last one the lowest RMS. This was achieved by adding a random penalty to the cost function every time that the order was violated.

Training the LVN involves the estimation of the network parameters using a specific model order complexity (number of inputs, number of filter banks, number of Laguerre functions for each bank, number of hidden units, and degree of nonlinearity). Model selection on the other hand is the task of selecting a specific model structure from a set of candidate models in order to avoid overfitting and poor predictive performance in new datasets. For this reason, we selected the model structure that minimized the Akaike Information Criterion (AIC) [40],

$$AIC(d) = 2J + 2d \tag{5.14}$$

where J is the cross-entropy loss function (Equation (5.13)) and d the total number of parameters of the model, which in the case of the simple LVN is,

$$d = \left(\sum_{i=1}^{2} L_i + Q \right) \cdot K \tag{5.15}$$

while for the LVN-ARX,

$$d = \left[\sum_{i=1}^{2} (L_i + L_{y_i}) + Q \right] \cdot K \tag{5.16}$$

5.2.2.4 LVN performance

In order to quantify the performance of the LVN (or LVN-ARX) we transformed the obtained probabilities of firing into spikes by selecting an “optimal” threshold between 0 and 1. Under the assumption of equal consequences of misclassification, a classifier tends to bias toward the larger groups that have more observations in the training sample. The imbalanced nature of the spike signal (more 0’s than 1’s) can affect greatly the results leading to an imperative need of class skew insensitive measures. Matthews correlation coefficient (MCC) [254] is usually used to overcome such problems. MCC values were calculated according to the relationship,

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5.17}$$

where TN (TP) and FN (FP) are the numbers of correctly and incorrectly predicted non-spike events (spike events). A MCC value of 1 corresponds to an excellent prediction, while a value of -1 indicates a total disagreement between prediction and observation. Random results (50 % of both negative and positive correctly predicted) give a value of 0. In our case, the threshold value that yielded the maximum MCC was selected as the optimal. The time scale of toleration for deciding if there was a hit or miss was 1ms. In addition to MCC we computed other statistical measures of performance like True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR) and False Negative Rate (FNR) which can be derived easily from the obtained confusion matrices used to estimate MCC.

For visualization purposes we used the cumulative sum of spike timings (CDF) and the van Rossum distance (VRD) [255]. CDFs are useful in detecting ISIs that are missed or closely followed by the model prediction. VRD is defined as the integrated squared difference between two spike trains (one flipped) convolved with an exponential function of a time constant t_c. Here, we used the curve produced by the difference of the spike trains before integrating in order to track the errors in time.

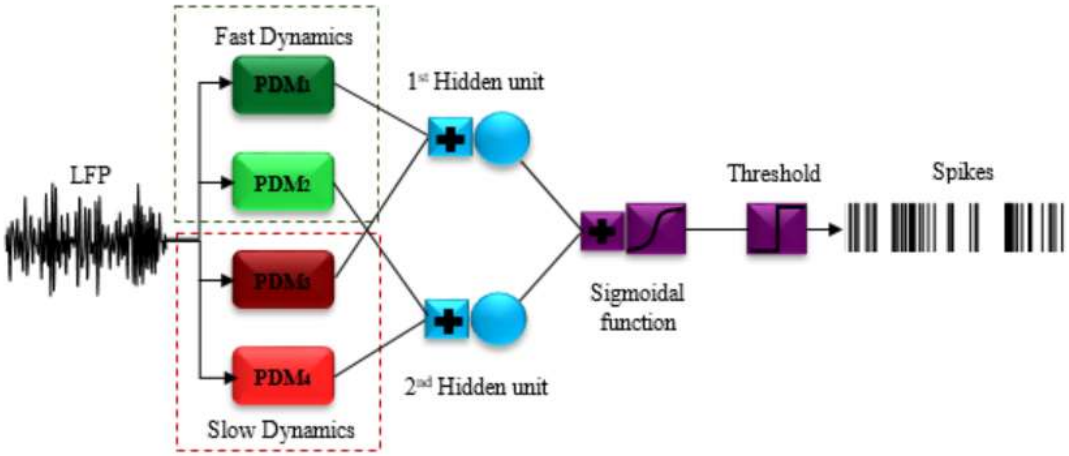
5.2.2.5 Single- and multi-unit LVN analysis

Spike waveforms would either be clearly discriminated or would be difficult to cluster. In this study, *Single-unit activity (SUA) recordings* consisted of low noise and high amplitude spikes belonging clearly to one or more neuronal sources. On the other hand, *Multi-unit activity (MUA) recordings* comprised of action potentials from a population of neurons that were difficult to discriminate based on their shape or their ISI distribution characteristics. *Single-unit (SU) analysis* refers to the identification of LVN models in order to predict SU spike trains from LFPs of both MUA and SUA recordings, whereas *Multi-unit (MU) analysis* is the procedure of extracting LVN models that predict MU spike trains from both MUA and SUA recordings with more than one firing unit. Specifically, in the SU analysis we kept only the spike trains that belonged to the dominant cluster (i.e. highest RMS power and cluster size) whereas in the MU analysis we considered all detected clusters. We hypothesized that the relationship between LFP and MUA spikes would be different from that between LFP and SUA spikes. Subpopulations of neurons within a local area may exhibit different functionality and therefore respond to LFP oscillations in a dissimilar manner.

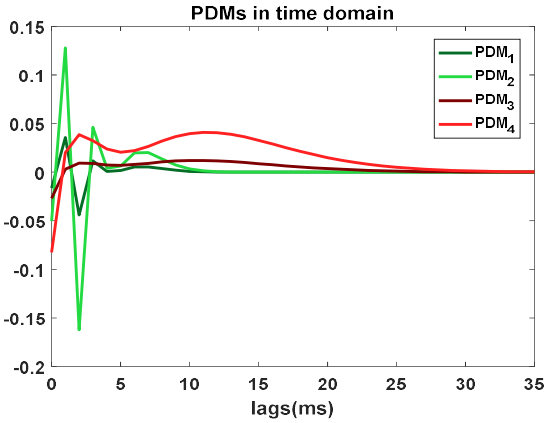
5.3 Results

5.3.1 Estimated LVN model mapping the relationship between LFP and Spikes

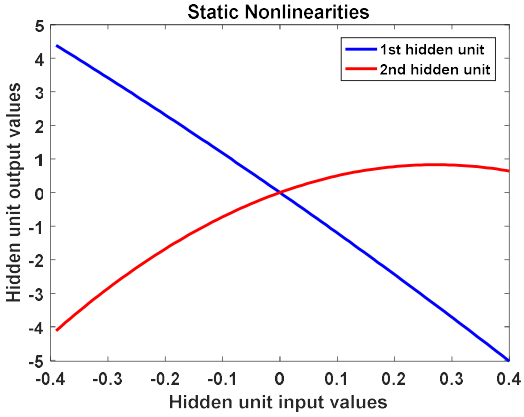
We extracted a total of 270 pairs of LFP-spike trains (85 SUA and 185 MUA recordings) from all PD subjects. In 70/85 SUA cases, 2nd-order nonlinear LVN models ($Q = 2$) exhibited the highest predictive performance (MCC) when applying SU analysis. For the MUA recordings, we trained 2 LVN models; one using the MU spike train as output (MU analysis) and one using the spiking activity of the dominant neuron (SU analysis). In terms of the MU analysis, in 127/185 cases 2nd-order nonlinear models explained more accurately the relationship between LFPs and spikes. The rest 58/182 led to the selection of linear models ($Q = 1$). On the other hand, in the SU analysis 2nd-order nonlinear models were superior in 147/185 cases (more compared to MU analysis). We estimated also 3rd-order nonlinear models ($Q = 3$) however they were found to be redundant. The results imply that the LFP – spike relationship is overall nonlinear in nature and this is in line with previous studies [54], [236], [240]. Variability in the magnitude of linear and nonlinear contributions has been previously observed [54], [256] and it has been attributed to the degree of phaselocking in the lower frequencies combined with the amplitude of high frequency oscillations [54]. Based on the SU and MU analysis results, in our case we believe, first of all, that the noisy nature of MUA is the main reason that leads to the selection of linear models especially if the SNR levels are very low. And indeed, we found a positive correlation between the spike signal SNR and model order complexity. Second, we observed that increased cross-frequency couplings and phaselocking led to decreases in the entropy of the LFP signal (due to synchronized oscillatory activity [257], [258]) and the emergence of 2nd -order nonlinearities validating the findings of [54]. A representative example of an identified LVN nonlinear model with two hidden units is shown in Figure 5.2a in its PDM form. The PDMs that express the fast and the slow dynamics are found in Figure 5.2b whereas the shape of the polynomial activation functions is shown in Figure 5.2c.



(a)



(b)



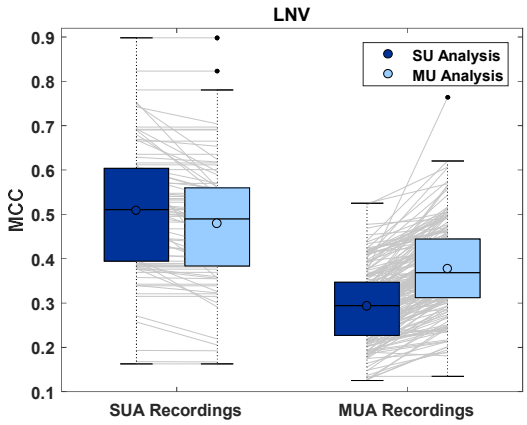
(c)

Figure 5.2 (a) A representative estimated LVN model from a SU recording in its PDM form $[L_1, L_2, Q, K, \alpha_1, \alpha_2] = [5,6,2,2,0.51,0.04]$. The PDMs are given by Equation (5.8). The green PDMs (PDM_1 and PDM_2) correspond to the fast dynamics of the system, whereas the red one (PDM_3 and PDM_4) correspond to the slow dynamics. The LFP is convolved with the PDMs and fed into a static nonlinearity defined by the polynomial activation functions of the hidden units. A sigmoidal function transforms the continuous output from the hidden units into probabilities of firing and a threshold is then applied for spike generation. The optimal threshold is selected by maximizing the MCC between observed and predicted spikes. (b) PDMs in time domain (c) Hidden unit's activation functions.

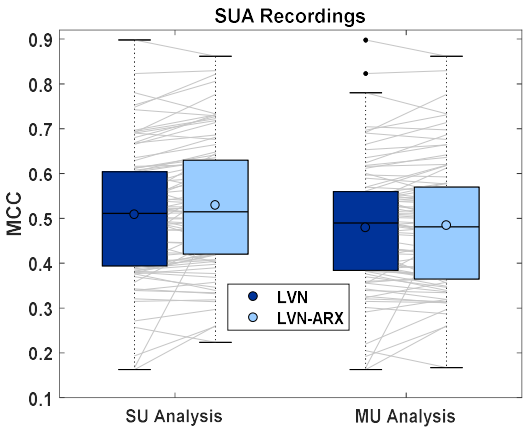
5.3.2 Predicting Spikes from LFPs using LVN models

In Figure 5.3a we present the LVN MCC values acquired between observed and predicted spikes for both SUA and MUA recordings using both SU and MU analysis. SUA recordings combined with SU analysis resulted in a median MCC of 0.511 ([0.162 0.898]), whereas MU analysis led to a median MCC value of 0.489 ([0.162 0.898]). In the case of MUA recordings, a median MCC of 0.294 ([0.125 0.525]) and 0.368 ([0.135 0.764]) was obtained applying SU and MU analysis respectively. In Figure 5.3a, the grey lines provide a visual representation of how MCC values change depending on the analysis used for each LFP-spike pair. It is obvious that SU analysis favours SUA recordings, whereas MUA is more accurately represented using MU analysis, validating our hypothesis that single neurons respond to LFP changes in unique ways compared to the collective activity of a population of neurons. The low MCC values observed for MUA recordings and SU analysis could be due to inaccurate clustering of the detected spike waveforms. MUA recordings consisted of clusters that could not be easily discriminated, hence the extracted dominant clusters may have been erroneous. Comparing the results obtained using LVN and LVN-ARX models (Figure 5.3b,c), there were cases where LVN-ARX models were superior and recordings where LVN models had higher predictive performance. However, there seems to be an overall increase in predictability using LVN-ARX models for SUA recordings combined with SU analysis. Thus, LVN-ARX models are better suited for single-units.

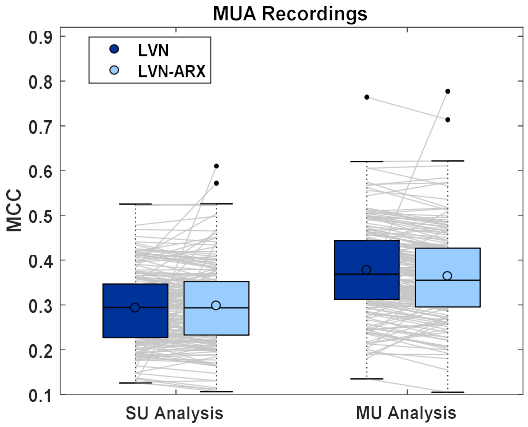
In order to understand visually the relationship between the magnitude of the MCC and the prediction we present in Figure 5.4 the observed and the predicted spikes from representative recordings with MCC values of 0.294, 0.504 and 0.898 (maximum acquired MCC in the dataset). It is obvious that both 0.898 and 0.504 are good predictions. In Figure 5.4c (MCC: 0.294) the VRD plot reveals a stream of errors of the same amplitude. This indicates that the model is oftenly either missing spikes or predicting false ones. Based on the CDF plot, and since the blue line (cumulative spike counts of the observed spike train) exceeds the red one (cumulative spike counts of the predicted spike train) we conclude that the model is actually missing spikes. Note here that MCC values were computed with a strict temporal resolution of ±1ms. No correlations were found between number of spikes in a recording and MCC.



(a)



(b)



(c)

Figure 5.3 (a) Boxplots of LVN MCC values acquired between observed and predicted activity in SUA and MUA recordings using both SU and MU analysis. The grey lines provide a visual feedback of how the MCC changes between MU and SU analysis for each one of the recordings (b) Boxplots that compare LVN and LVN-ARX MCC values for SUA recordings using both SU and MU analysis. (c) Boxplots that compare LVN and LVN-ARX MCC values for MUA recordings using both SU and MU analysis.

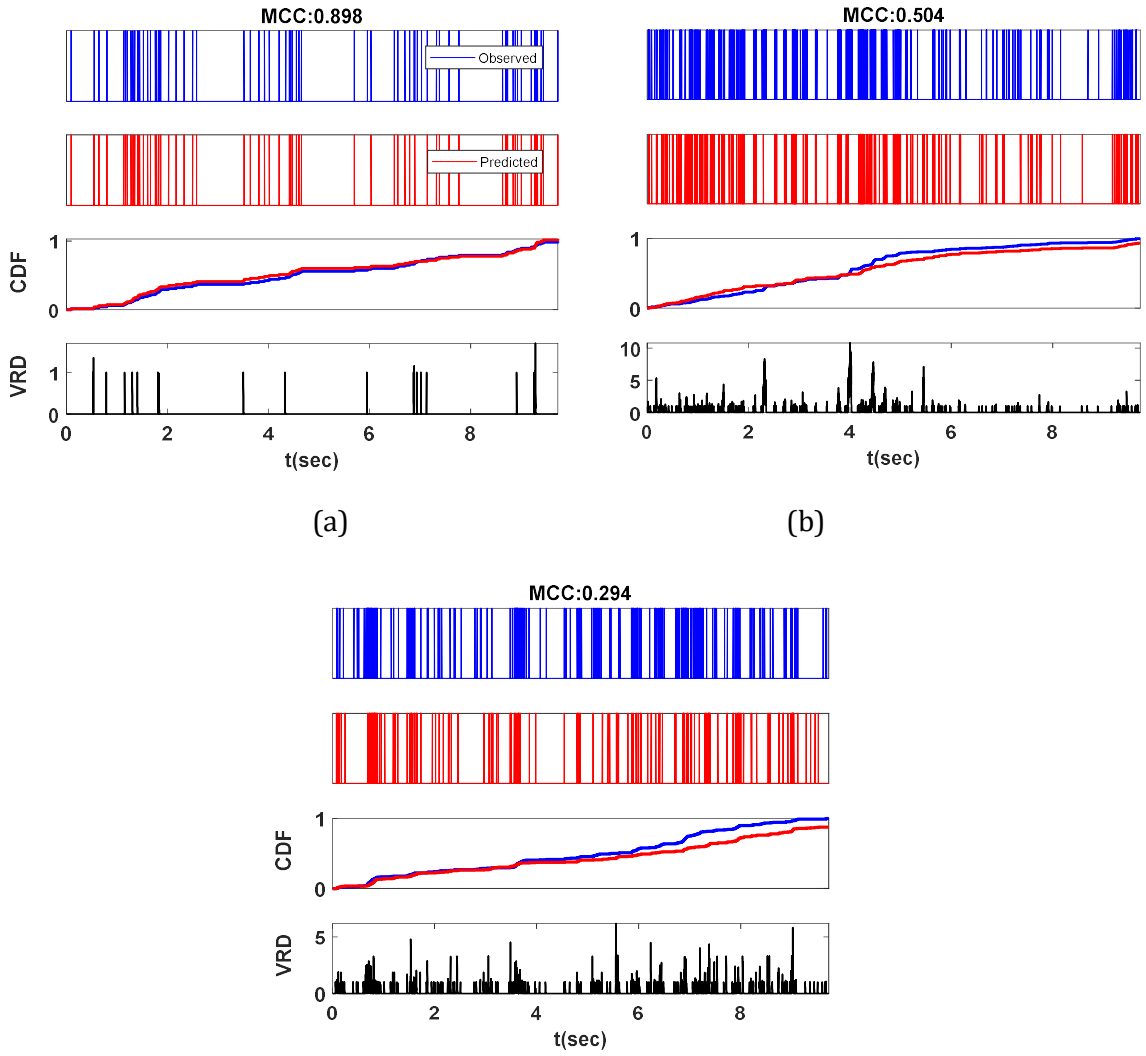
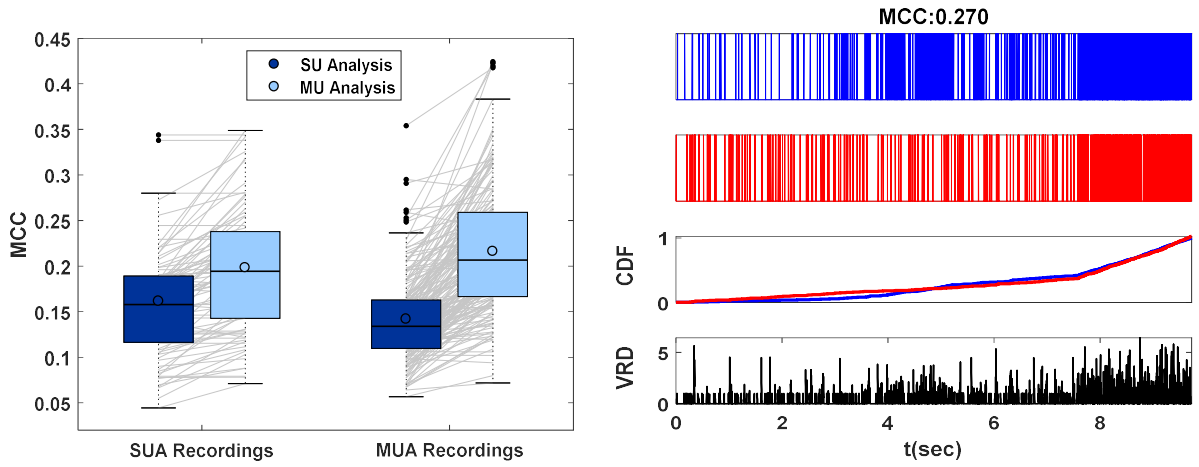


Figure 5.4 Observed (blue) and predicted (red) spike trains (first and second horizontal panels) with MCC values of (a) 0.898 (b) 0.504 and (c) 0.294. The third horizontal panel depicts the cumulative spike counts (CDF) for both observed and predicted spikes. The 4th horizontal panel is the van Rossum distance (VRD) before integrating. Note here that no correlation was found between MCC and number of spikes in a recording.

5.3.3 Predicting Spikes from LFPs of neighboring electrodes using LVN models

The Ben Gun configuration allows the simultaneous recording of activity in 5 different sites. The distance between any peripheral electrode and the central one was 2 mm. We used LFPs from neighboring electrodes to predict the spiking activity of a target electrode inside the STN. LVN models were only considered. Based on Figure 5.5a, predictions acquired from neighboring LFPs were inferior and as expected MCC values decreased with increasing distance from the target electrode (not shown here). MU analysis led to superior performance for both SUA and MUA recordings. This was expected since MUA activity captured from the target electrode may contain spikes from neurons firing in closer proximity to other electrodes. In Figure 5.5b we present a representative prediction.

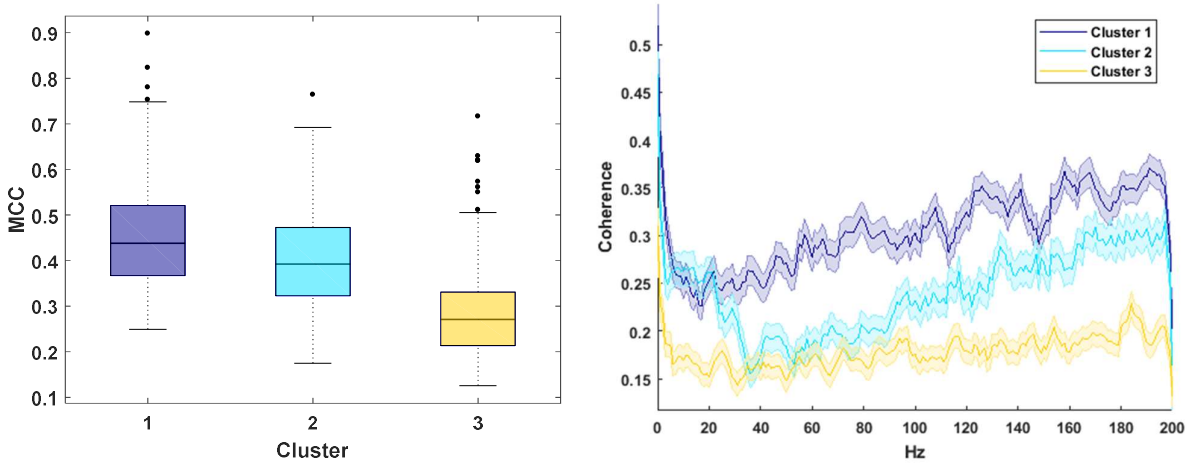


(a) (b)

Figure 5.5 (a) LVN MCC values using LFPs from neighbouring electrodes to predict the spiking activity of a target electrode. (b) A representative prediction using spikes from a target electrode and LFP from a neighbouring electrode.

5.3.4 LFP and Spike features related with spike timing

We assigned to each MER its own unique feature vector related with the LFP, SUA/MUA and BUA characteristics and we clustered the data into 3 meaningful groups (Figure 5.6a) using K-medoids and correlation as distance measure. Significant discriminative features were selected based on the Kruskal-Wallis test (p<0.05). We ran the clustering procedure twice. Initially, MCC values obtained from the LVN models were not considered. The second time we included them as features. Surprisingly, the clustering results were very similar and MCC was a significant discriminative feature in the latter case. This means that there was a clear *a priori* structure in the data related with spike timing predictability.



(a) (b)

Figure 5.6 (a) MCC boxplots of the extracted data clusters (number of recordings in each cluster: $n_1 = 197, n_2 = 147, n_3 = 195$) ; median firing rate (SUA/MUA) of each cluster: $MFR_1 = 27/48\ Hz, MFR_2 = 27/53\ Hz, MFR_3 = 15/30\ Hz$) (b) Coherence between actual and predicted spike trains for each cluster separately.

Another important feature that could distinguish the three clusters was the coherence between LFP and SUA/MUA (Figure 5.7). The cluster with the highest spiking predictability (cluster 1) exhibited increased coherence between LFP and spikes in the higher frequencies (HB – HG), whereas LFPs in clusters 2 and 3 shared more information with the spiking units in the lower frequencies (Figure 5.7). In more detail, cluster 1 comprised of recordings with high frequency characteristics. LFP power was mainly accumulated in the HB-HG range. Cross-frequency coupling between HB and HG was dominant and SUA/MUA and BUA were firing synchronously in that range. Based on these characteristics, these neurons probably belong to

a local oscillatory network with a high frequency operating point and with partial response to low frequency global information. Cluster 2, on the other hand, had enhanced low frequency LFP components, increased bursting activity (and thus mean firing rate) coherent with the BUA in the D-LB range. Spike prediction accuracy using neighboring electrodes was significantly higher for this cluster (followed by cluster 1 and 3) indicating a synchronized oscillatory population phaselocked mainly in the LB. Cluster 3 exhibited a broadband shift in LFP power. Due to low SNR levels (i.e. increased background noise) and coherence between LFP and spikes, in a large number of recordings the overall predictive performance was poor. Spikes in this group responded mainly to the T component of the LFP in a phaselocked manner. The coherence between BUA and SUA/MUA was low implying that neurons from this group were probably firing asynchronously with their surrounding population [248]. In addition, depending on the SNR levels and the mean firing rate, spike prediction would be either poor or moderate. The spike count was low and this may have also led to poor training of the models. By examining the coherence between actual and predicted spike trains (Figure 5.6b), we observed that cluster 2 predicted more accurately the lower frequency components of the SUA/MUA in the range between T and LB, whereas cluster 1 showed superior performance in the HB-HG band. This was expected due to the differences observed in coherence between LFP and spikes in each cluster. Representative MERS from each cluster along with their respective MCC values can be found in Figure 5.8.

There is a great similarity between the clusters we extracted and the so called “Low Frequency” (LF) and “High Frequency” (HF) cells described in [54]. The authors detected two different neuronal populations based on their predictive characteristics. In the LF cases, there was increased phaselocking and spike field coherence (SFC) in the lower frequencies. Spikes tended to occur at low HF LFP amplitudes and linear models were more predictive in this group. On the other hand, in the HF case, SFC was weak for lower frequencies and spikes were less phaselocked, occurring usually when HF LFP had high amplitude. The nonlinear model contribution was more significant in this case. We examined all the abovementioned characteristics and indeed we found similarities between cluster 1 and HF cells and cluster 2 and LF cells. The authors corroborate that the differences accounted by these two types of cells could be attributed to their projections towards more local or more distant populations i.e. LF cells may be communicating with more distant populations, whereas HF cells may be projecting to more proximal populations. Cluster 2 and cluster 3 resemble the so called “High frequency

band" (HFB) and "Tremor frequency band" (TFB) neurons presented in [245]. HFB and TFB neurons displayed oscillatory activity at the A/LB and T band (known also as tremor band), respectively. The main difference between these two groups was that TFB neurons exhibited high SUA/MUA power, low BUA power and in general low coherence between these two signals. The opposite applies for the HFB neurons. Indeed, in Figure 5.7 we can see the same pattern for cluster 2 and cluster 3. Moreover, it was found that TFB neurons tended to oscillate in a more sporadic manner compared to the continuous oscillatory activity of HFB neurons. This phenomenon was projected in the mean firing rate of each cluster. Cluster 3 had a mean firing rate of 15 and 30 Hz for SUA and MUA recordings respectively, whereas the mean firing rate for cluster 2 was 27 and 53 Hz.

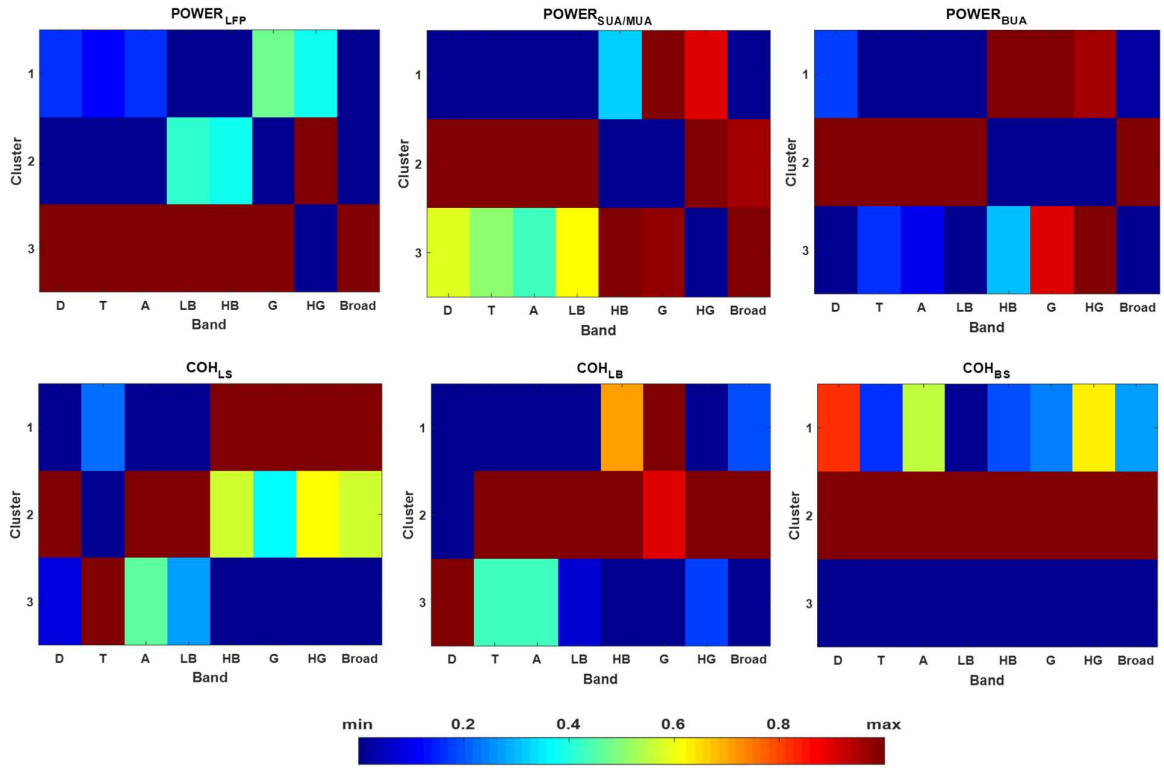


Figure 5.7 Heatmaps representing median features in different frequency bands (D, T, LB, HB, G, HG and Broadband) for each cluster ($POWER_X$: power band ratio of X, where X is either the LFP, SUA/MUA depending the type of recording or the BUA envelope; COH_{XY}: mean coherence between X and Y – L refers to LFP, S to SUA/MUA and B to the BUA envelope). Note here that for each matrix columns were normalized between 0 and 1. Values are comparable between clusters only and not between frequency bands i.e. cluster 2 exhibited the highest $POWER_{HG}$ compared to cluster 1 and cluster 3. However, this does not necessarily imply that this feature is stronger in that specific frequency band in the same cluster.

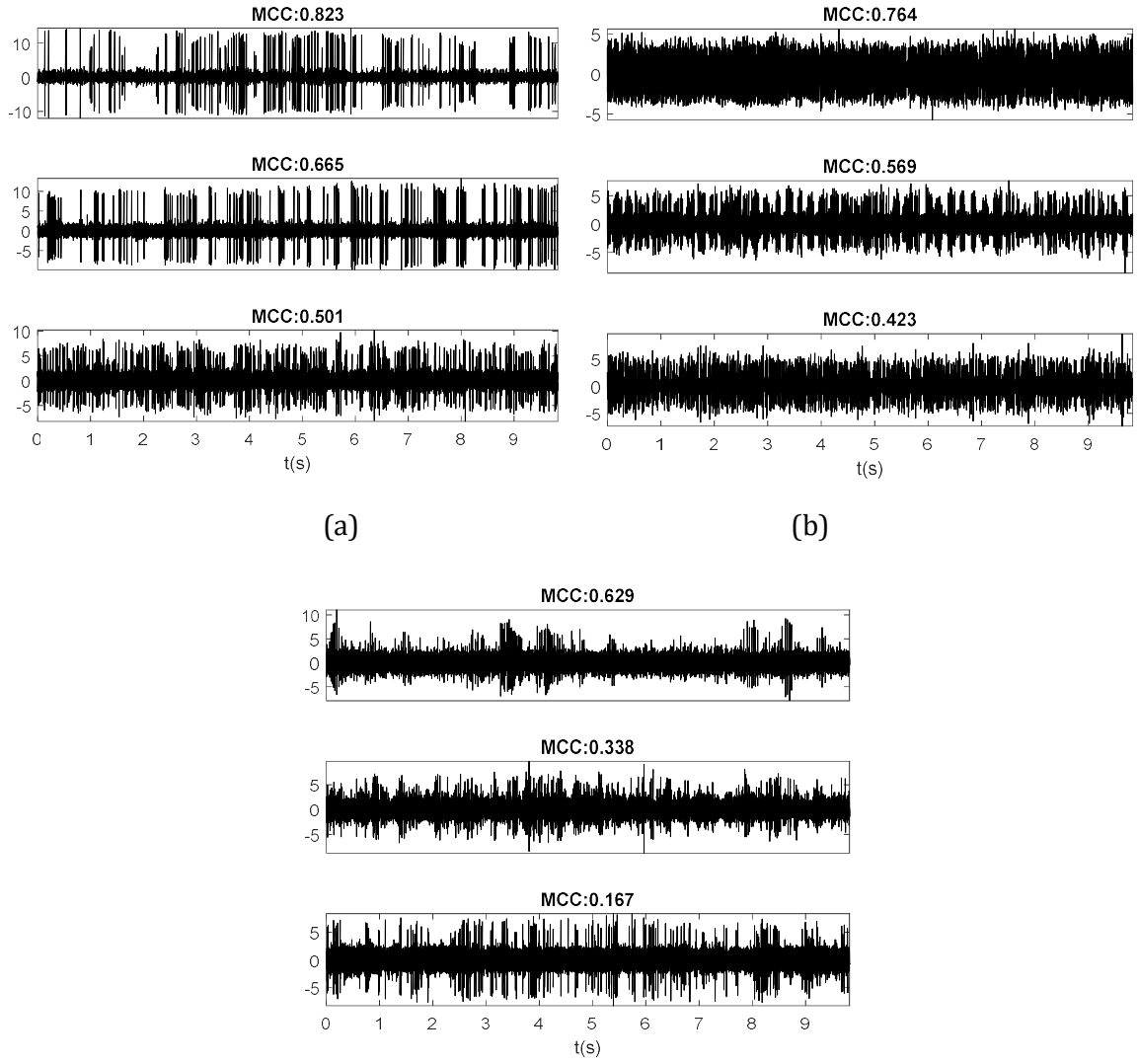


Figure 5.8 Representative MERS that belong to (a) cluster 1, (b) cluster 2 and (c) cluster 3. On top of each subplot we denote the MCC value acquired for each recording.

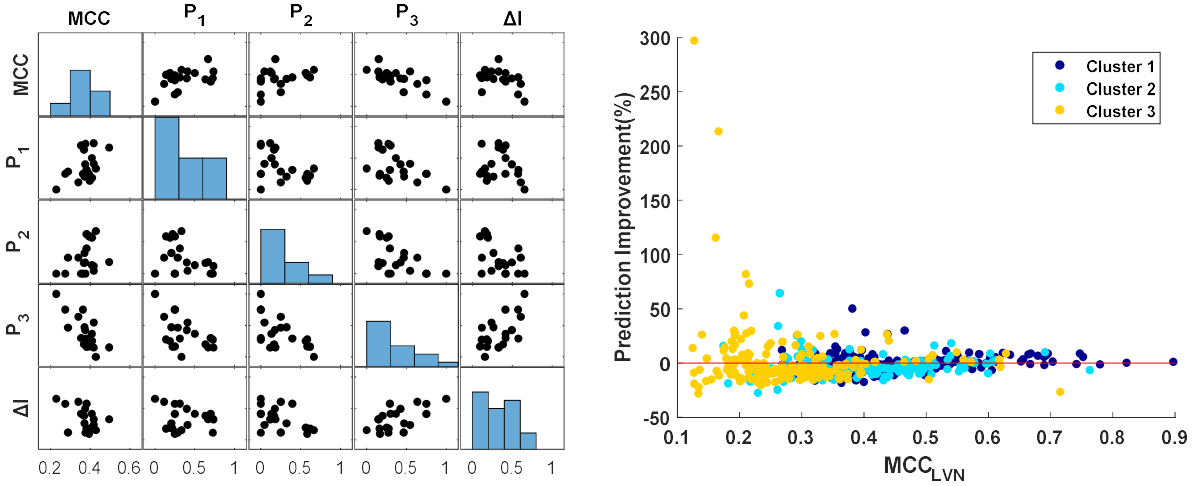
In contrast to [245], we observed a dependence between clusters and "off" state UPDRS improvement (ΔI) following DBS. The Kruskall Wallis test returned a significant p-value (p=3.6915e-06) indicating statistically significant differences between the three clusters. Follow up multiple comparisons detected significant differences between cluster 1/cluster 3 and cluster 2/cluster 3 with cluster 3 and 2 exhibiting the highest and smallest median ΔI value, respectively. For each patient, we then computed from all MERS a mean MCC value and a cluster preference probability based on the number of recordings found in each one of the three clusters (P1, P2 and P3 respectively). Indeed, we found significant correlations (p<0.05)

between MCC and ΔI (ρ=-0.45), MCC and P1 (ρ=0.47), MCC and P3 (ρ=-0.75), P1 and P3 (ρ=-0.51), P2 and ΔI (ρ=-0.51), P2 and P3 (ρ=-0.58), P3 and ΔI (ρ=0.62) (Figure 5.9a). Overall, there was a decrease in MCC with increased improvement and a negative and positive relationship between improvement and the number of MERS detected in cluster 2 and 3 respectively. In [259], neuronal STN oscillations were found to be selectively associated with specific PD motor symptoms. Thus, in the same manner, spike predictability may also be a biomarker of PD phenotype or symptom severity or perhaps a predictive feature of DBS response.

We proceeded by comparing the performance between LVN and LVN-ARX models for all 3 clusters. The prediction improvement (%) obtained using LVN-ARX models was defined as,

$$Prediction\ Improvement = \frac{MCC_{LVN-ARX} - MCC_{LVN}}{MCC_{LVN}}\% \tag{5.18}$$

where MCC_{LVN} and $MCC_{LVN-ARX}$ is the MCC acquired using LVN and LVN-ARX models respectively. 38%, 23% and 32% of the recordings in cluster 1, 2 and 3 respectively showed a positive prediction improvement (Figure 5.9b). LVN-ARX models resulted into smaller FPR for cluster 1 and 3 but higher FPR for cluster 2. There was also an overall increase in the FNR. This imples that the extra AR terms were inhibiting spike generation. Large improvements were mainly observed in cluster 3. We speculate that the AR terms facilitate the prediction of bursting episodes. The start of a burst may be easily predicted if it is phaselocked to the LFP. The exact timing of the subsequent spikes though may be difficult to foresee. Consequently, the role of the AR terms is to extract patterns based on the spiking history of the cell. Probably recordings with sparse bursting patterns, like those in cluster 3, can be more easily tracked based on their prior history. Another hypothesis is that LVN-ARX models may be more efficient in the case of locally generated oscillations that cannot be inferred from the LFP [248].



(a) (b)

Figure 5.9 (a) Matrix scatter plot depicting relationships between UPDRS improvement (ΔI), cluster preference probability (P_i where i is either cluster 1, 2 or 3) and MCC values. (b) Scatter plot depicting the prediction improvement (%) achieved using LVN-ARX models as a function of the MCC of the conventional LVN model. Different dot colors correspond to different clusters. The red line separates the plot into positive and negative prediction improvement areas. Positive prediction means that LVN-ARX models performed better than the simple LVN models.

5.4 Discussion and Conclusions

In this study we validated previous findings that suggest that it is possible to predict the exact timing of spikes based on past information of the LFP in the STN [240]–[242] by using models specifically optimized for binary response systems. We extracted SUA and MUA recordings from the STN of 20 PD patients undergoing DBS and we trained different LVNs for each case. In SU analysis 2nd order nonlinear models exhibited higher predictive performance indicating that the relationship between LFP and individual neuron is nonlinear. On the other hand, in the MU analysis we observed that in 32% of the recordings linear models were selected over the nonlinear ones. This could be due to the noisy nature of the MU spike train or it could indicate less complex dynamics associated with more asynchronous activity. Overall, predictions from SUA recordings were more accurate compared to MUA recordings and this again may be related to the fact that SUA recordings exhibited clearer spiking patterns and higher SNR levels. Our results are in agreement with [240] where it was shown that in the PD STN neural activity could be predicted by the LFP with high to moderate accuracy depending on the number of neurons present in a recording. We also examined the effect of the LFP of neighboring electrodes on the spiking activity of a target electrode. As expected the performance overall was inferior and we observed a decrease in MCC with distance. However, there were time windows where the observed and the predicted response coincided, indicating that part of the spiking information could be explained by LFPs of electrodes in close proximity. In order to account for past spiking history, we modified the LVN models by adding AR terms. Up to this point, we only observed an improvement for the SU cases but later we detected the conditions under which such model structures are more efficient.

In our effort to understand the reasons of variability in spike timing predictability, we isolated 3 type of neuronal populations that exhibited high (cluster 1), moderate (cluster 2) and low (cluster 3) prediction accuracy. The main discriminative feature was the coherence between LFP and spikes. The first group displayed increased coherence in the HF range. In the other two groups, coherence was mainly detected in lower frequencies. Cluster 1 could be described as a local oscillatory population operating in the HB-HG range. Cluster 2 was characterized by excessive synchronized T-LB activity, whereas cluster 3 exhibited asynchronous A/T activity. LVN-ARX models improved the prediction performance of SU bursting cells. Large improvements were detected especially in cluster 3 implying possibly locally generated oscillations that could not be inferred from the LFP. There was a striking resemblance between

our detected clusters and the so called LF and HF cells in [54] and the TFB and HFB cells in [245]. The most interesting finding in this study was the link between spike timing predictability and “off” state UPDRS improvement following DBS surgery. The higher the incidence of finding MERS belonging to the cluster with low prediction accuracy (cluster 3) the higher the UPDRS improvement. On the other hand, MERS belonging to cluster 2 were more likely to belong to patients with low improvement. An explanation could be that the asynchronous activity of cluster 3 acts as neural noise that decorrelates and weakens the pathologic hypersynchronization detected in cluster 2, facilitating this way DBS or indicating a predisposition to good DBS response [260], [261].

Chapter 6 Summary and Conclusions

6.1 Thesis summary

It is well established that physiological systems are difficult to analyze due to their nonlinear and/or nonstationary characteristics. In this research work we focused on the algorithmic development of modeling techniques that facilitate physiological system identification and provide accurate representation of the underlying system dynamics.

In Chapter 3, we concentrated on MI linear and nonlinear TV systems and we provided effective tools for their identification. Initially, we highlighted the pitfalls of existing TV modeling estimation methods regarding the assumption of constant rate of parameter variations and we emphasized the need for optimal hyperparameter selection strategies. To this end, we developed LV based recursive schemes, inspired by the traditional RLS and KF techniques, that can track slow, fast or even mixed-mode variations combined with abrupt changes. In addition, we proposed a model order selection and hyperparameter optimization framework based on mixed integer GAs that can replace computationally intractable exhaustive search procedures. We also investigated thoroughly, for the first time, the link between the hyperparameter values of the proposed recursive estimators and the TV characteristics of the true system, as well as the behavior of different model order selection criteria when the stationarity assumption is violated. The performance of our proposed methodology was assessed using simulations. By applying our technique on real experimental data, we clearly elucidated the role of CA on VVS occurrence and we concluded that impaired CA is not associated with abrupt loss of consciousness in patients suffering from VVS.

In Chapter 4, the TV analysis presented in Chapter 3 was extended to MVAR models. Although MVAR models assume linear interrelationships, they provide useful information regarding time series interactions and couplings. One of the proposed TV estimation techniques of Chapter 3 was adapted to this type of models and we extracted TV-MVAR measures of connectivity between time series in the frequency domain. Initially, the TV-MVAR framework was applied on experimental data to track exercise-induced hemodynamic changes in healthy subjects and stroke survivors during cycling. However, we observed error terms with TV variances associated with switching from resting to exercise initiation or changes in cycling intensities. This led us to examine rigorously the effect of heteroskedasticity of the residuals on the parameter estimation procedure, as well as on the extracted TV-MVAR measures of connectivity. Based on simulations, heteroskedasticity resulted into significantly more variable parameter estimates when using conventional recursive schemes. On the other hand, our approach exhibited a more robust behavior. Nonetheless, due to the dependence of the MVAR measures on the statistical properties of the residuals, heteroskedasticity still gave rise to overestimated or underestimated time periods of coupling and causality between the signals of interest. Therefore, we parametrized the TV covariance of the error terms using GARCH models and corrected the observed biases on the TV-MVAR connectivity measures. Regarding the experimental data, we extracted exercise-induced TV coupling patterns, linked with different regulatory mechanisms, that could clearly differentiate the three groups. We concluded that immediately after exercise there is a partial restoration of parasympathetic nervous system activity (which is known to be reduced post-stroke or due to ageing) on both elderly and stroke survivors.

Finally, in Chapter 5, the attention was shifted towards linear and nonlinear systems with binary responses. We employed time-invariant LVN and LVN-ARX models modified to produce probabilities of event occurrence in their output. Network parameters were trained using a hybrid optimization scheme based on GA's and interior point constrained nonlinear methods. Specifically, the optimal solution obtained from the GA was fed into the nonlinear optimization method as an initial starting point for a more efficient local search. Instead of the conventional MSE function, we minimized the cross-entropy loss function and the AIC was used for model order selection. The continuous output predicted probabilities were transformed into '0's and '1's by selecting the threshold that maximized the MCC. The MCC is a class skew insensitive measure and is usually used to mitigate issues arising due to unbalanced datasets. The

proposed LVN and LVN-ARX models were applied to predict spikes from LFPs in parkinsonian patients undergoing STN DBS in order to understand the functionality of the PD STN neuron. Spike timing predictability was found to be associated with the characteristics of distinct neuronal populations, as well as motor improvement following DBS surgery, indicating the possible use of spike timing predictability as a predictive biomarker for DBS response.

6.2 Future work

Although this work addresses some of the main issues of physiological system identification, there is still room for improvement in our proposed identification schemes. Herein, we briefly summarize potential avenues for future research based on each chapter separately.

Chapter 3: The TV identification methodology described in this chapter assumes zero-mean white noise. A potential future scope could be the investigation of the effect of colored or impulsive noise on the TV parameter estimation procedure and the development of robust recursive schemes for systems contaminated by non-white noise. A different technique for identifying TV systems are the so called Linear Parameter Varying (LPV) models [172]. In the LPV formulation the relationship between input and output is modulated externally by one or more TV variables, known as scheduling variables (SVa). In terms of model estimation, the LPV approach is mathematically simple since it treats the system as time-invariant. The model parameters are assumed to be functions of the SVas and usually the dependence is static. In LPV models the parameters appear to evolve in time. When dealing with models with TV parameters, the following question usually arises: Is the model time-varying or parameter-varying? In practice, if the SVa is unknown or unmeasurable then TV models should be used. On the other hand, if the scheduling signal is available then we can apply a LPV model. An interesting research question could be the following: Is it possible to model the LPV systems as TV and if yes, can we reconstruct the scheduling signal using the extracted TV parameters of the TV model? One straightforward method to evaluate if a signal could be considered as SVa is by estimating its correlation with the extracted TV parameters. The higher the correlation, the higher the probability that this signal affects the model dynamics. Nonetheless, more advanced methods can be used to reconstruct the SVa without any prior knowledge.

Chapter 4: A main problem with the conventional MVAR model is that it includes all lags of a variable up to a selected order. However, some of these terms may be unnecessary and uninformative, adding extra complexity to the extracted models. Our proposed TV-MVAR

formulation allows the parameters to evolve with time and thus it permits the zeroing out of some of the model parameters when necessary. Nevertheless, complete zeroing out is not possible due to error fluctuations. Hence, there is an increasing need of developing algorithms that can either selectively remove in time some of the parameters from the estimation procedure depending on rules and thresholds, or apply recursive regularization techniques which is quite a challenging task. Another potential refinement on our TV-MVAR methodology could be the incorporation of recursive formulas for GARCH process estimation, achieving simultaneous recursive parameter estimation and parametrization of heteroskedasticity in the error terms. Since heteroskedasticity in general affects the estimation procedure, we anticipate that the suggested approach will lead to more accurate parameter estimates especially in cases where the variances of the error terms fluctuate significantly. We believe, however, that the most pressing matter is the establishment of a validated statistical framework for testing significance of the MVAR-based connectivity measures and therefore it should be certainly considered for future research.

Chapter 5: The LVN-ARX models presented in this chapter, for identification of either binary or continuous output systems, closely resemble the polynomial NARX models [32] (Section 2.1.2.2). In polynomial NARX models, the output is expressed as a nonlinear (polynomial) function of past input and output terms. The challenging task of model order determination (due to the curse of dimensionality of polynomial expansions) and NARX estimation has led to the development of a plethora of identification methods. The LVN-ARX models on the other hand, can represent efficiently and with a smaller number of parameters high order nonlinearities. Therefore, a promising future research topic is the comparison between LVN-ARX and polynomial or other type of NARX models in terms of both performance and computational complexity.

References

[1] V. Z. Marmarelis, *Nonlinear dynamic modeling of physiological systems*, vol. 10. John Wiley & Sons, 2004.

[2] D. T. Westwick and R. E. Kearney, *Identification of nonlinear physiological systems*, vol. 7. John Wiley & Sons, 2003.

[3] P. Z. Marmarelis and K.-I. Naka, "Identification of multi-input biological systems," *IEEE Trans. Biomed. Eng.*, no. 2, pp. 88–101, 1974.

[4] G. D. Mitsis and V. Z. Marmarelis, "Nonlinear modeling of physiological systems with multiple inputs," in *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, 2002, vol. 1, pp. 21–22.

[5] D. T. Westwick, E. A. Pohlmeyer, S. A. Solla, L. E. Miller, and E. J. Perreault, "Identification of multiple-input systems with highly coupled inputs: application to EMG prediction from multiple intracortical electrodes," *Neural Comput.*, vol. 18, no. 2, pp. 329–355, 2006.

[6] G. D. Mitsis, M. J. Poulin, P. A. Robbins, and V. Z. Marmarelis, "Nonlinear modeling of the dynamic effects of arterial pressure and CO2 variations on cerebral blood flow in healthy humans," *IEEE Trans.Biomed.Eng*, vol. 51, no. 0018–9294 (Print), pp. 1932–1943, 2004.

[7] G. D. Mitsis, R. Zhang, B. D. Levine, and V. Z. Marmarelis, "Cerebral hemodynamics during orthostatic stress assessed by nonlinear modeling," *J. Appl. Physiol.*, pp. 354–366, 2006.

[8] G. D. Mitsis, R. Zhang, B. D. Levine, E. Tzanalaridou, D. G. Katritsis, and V. Z. Marmarelis, "Autonomic neural control of cerebral hemodynamics.," *IEEE Eng. Med. Biol. Mag.*, vol. 28, no. 6, pp. 54–62, 2009.

[9] D. Song, B. Robinson, R. Hampson, V. Marmarelis, S. Deadwyler, and T. Berger, "Sparse Large-Scale Nonlinear Dynamical Modeling of Human Hippocampus for Memory Prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2016.

[10] E. J. Perreault, R. F. Kirsch, and A. M. Acosta, "Multiple-input, multiple-output system identification for characterization of limb stiffness dynamics," *Biol. Cybern.*, vol. 80, no. 5, pp. 327–337, 1999.

[11] G. D. Mitsis and V. Z. Marmarelis, "Nonlinear modeling of glucose metabolism: comparison of parametric vs. nonparametric methods," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 5967–5970.

[12] G. D. Mitsis, M. G. Markakis, and V. Z. Marmarelis, "Nonlinear modeling of the dynamic effects of infused insulin on glucose: comparison of compartmental with Volterra models," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 10, pp. 2347–2358, 2009.

[13] G. D. Mitsis, A. S. French, U. Höger, S. Courellis, and V. Z. Marmarelis, "Principal dynamic mode analysis of action potential firing in a spider mechanoreceptor," *Biol. Cybern.*, vol. 96, no. 1, pp. 113–127, 2007.

[14] G. D. Mitsis and V. Z. Marmarelis, "Modeling of nonlinear physiological systems with fast and slow dynamics. I. Methodology," *Ann. Biomed. Eng.*, vol. 30, no. 2, pp. 272–281, 2002.

[15] D. T. Westwick and R. E. Kearney, "Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics," *Ann. Biomed. Eng.*, vol. 29, no. 8, pp. 707–718, 2001.

[16] V. Z. Marmarelis, G. D. Mitsis, D. C. Shin, and R. Zhang, "Multiple-input nonlinear modelling of cerebral haemodynamics using spontaneous arterial blood pressure, end-tidal CO2 and heart rate measurements," *Phil. Trans. R. Soc. A*, vol. 374, no. 2067, p. 20150180, 2016.

[17] G. D. Mitsis, R. Zhang, B. D. Levine, and V. Z. Marmarelis, "Modeling of nonlinear physiological systems with fast and slow dynamics. II. Application to cerebral autoregulation," *Ann. Biomed. Eng.*, vol. 30, no. 4, pp. 555–565, 2002.

[18] S. Cerutti, L. T. Mainardi, and A. M. Bianchi, "Time-frequency and time-varying analysis for assessing the dynamic responses of cardiovascular control," *Crit. Rev. Biomed. Eng.*, vol. 30, no. 1–3, 2002.

[19] K. Kostoglou, C. T. Debert, M. J. Poulin, and G. D. Mitsis, "Nonstationary multivariate modeling of cerebral autoregulation during hypercapnia," *Med. Eng. Phys.*, vol. 36, no. 5, pp. 592–600, 2014.

[20] V. Z. Marmarelis, D. C. Shin, M. Orme, and R. Zhang, "Time-varying modeling of cerebral hemodynamics," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 694–704, 2014.

[21] J. B. MacNeil, R. E. Kearney, and I. W. Hunter, "Identification of time-varying biological systems from ensemble data (joint dynamics application)," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 12, pp. 1213–1225, 1992.

[22] H. Kim and J. Rosen, "Epileptic seizure detection-an AR model based algorithm for implantable device," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 5541–5544.

[23] K. Kostoglou, K. P. Michmizos, P. Stathis, D. Sakas, S. Nikita, and G. D. Mitsis, "Prediction of the Parkinsonian subthalamic nucleus spike activity from local field potentials using nonlinear dynamic models," *Proc. 2012 IEEE 12th Int. Conf. Bioinforma. Bioeng.*, vol. 16, no. November, pp. 11–13, 2012.

[24] G. D. Mitsis, S. Courellis, A. S. French, and V. Z. Marmarelis, "Principal dynamic mode analysis of a spider mechanoreceptor action potentials," in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, 2003, vol. 3, pp. 2051–2054.

[25] G. Valenza, L. Citi, E. P. Scilingo, and R. Barbieri, "Point-process nonlinear models with laguerre and volterra expansions: Instantaneous assessment of heartbeat dynamics," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2914–2926, 2013.

[26] L. Ljung, *System identification*. Springer, 1998.

[27] T. Söderström and P. Stoica, *System identification*. Prentice-Hall, Inc., 1988.

[28] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*. John Wiley & Sons, 2012.

[29] V. Z. Marmarelis, "Identification of nonlinear biological systems using Laguerre expansions of kernels," *Ann. Biomed. Eng.*, vol. 21, no. 6, pp. 573–589, 1993.

[30] K. Alataris, T. Berger, and V. Marmarelis, "A novel network for nonlinear modeling of neural systems with arbitrary point-process inputs," *Neural Networks*, vol. 13, no. 2, pp. 255–266, 2000.

[31] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems part I: deterministic non-linear systems," *Int. J. Control*, vol. 41, no. 2, pp. 303–328, 1985.

[32] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[33] P. S. C. Heuberger, P. M. J. van den Hof, and B. Wahlberg, *Modelling and identification with rational orthogonal basis functions*. Springer Science & Business Media, 2005.

[34] H. Ogura, "Estimation of Wiener kernels of a nonlinear system and a fast algorithm using digital Laguerre filters," in *15th NIBB Conference, Okazaki, Japan*, 1985, pp. 14–62.

[35] A. C. Aitken, "IV.—on least squares and linear combination of observations," *Proc. R. Soc. Edinburgh*, vol. 55, pp. 42–48, 1936.

[36] A. Tikhonov, *Solutions of ill-posed problems*. .

[37] J. Sjöberg, T. McKelvey, and L. Ljung, "On the use of regularization in system identification," in *Proceedings of the 12th IFAC World Congress, Sydney, Australia*, 1993, vol. 7, pp. 381–386.

[38] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[39] G. Schwarz and others, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[40] H. Akaike, "A new look at the statistical model identification," *{IEEE} Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.

[41] L. S. H. Ngia, "Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters," *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 48, no. 6, pp. 562–579, 2001.

[42] T. O. e Silva, "On the determination of the optimal pole position of Laguerre filters," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2079–2087, 1995.

[43] G. Clowes, "Choice of the time-scaling factor for linear system approximations using orthonormal Laguerre functions," *IEEE Trans. Automat. Contr.*, vol. 10, no. 4, pp. 487–489, 1965.

[44] K. Kostoglou, A. D. Wright, J. D. Smirl, K. Bryk, P. van Donkelaar, and G. D. Mitsis, "Dynamic cerebral autoregulation in young athletes following concussion," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 696–699.

[45] P. C. Prokopiou, K. Murphy, R. G. Wise, and G. D. Mitsis, "Estimation of voxel-wise dynamic cerebrovascular reactivity curves from resting-state fMRI data," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 1143–1146.

[46] G. Valenza, L. Citi, and R. Barbieri, "Instantaneous nonlinear assessment of complex cardiovascular dynamics by laguerre-volterra point process models," in *Engineering in Medicine and Biology Society*

(EMBC), 2013 35th Annual International Conference of the IEEE, 2013, pp. 6131–6134.

[47] D. Song, M. Harway, V. Z. Marmarelis, R. E. Hampson, S. A. Deadwyler, and T. W. Berger, "Extraction and restoration of hippocampal spatial memories with non-linear dynamical modeling," *Front. Syst. Neurosci.*, vol. 8, 2014.

[48] S. Saleem, Y.-C. Tzeng, W. B. Kleijn, and P. D. Teal, "Detection of impaired sympathetic cerebrovascular control using functional biomarkers based on principal dynamic mode analysis," *Front. Physiol.*, vol. 7, 2016.

[49] D. Song *et al.*, "Identification of functional synaptic plasticity from spiking activities using nonlinear dynamical modeling," *J. Neurosci. Methods*, vol. 244, pp. 123–135, 2015.

[50] M. Cole, S. Eikenberry, T. Kato, R. A. Sandler, S. M. Yamashiro, and V. Z. Marmarelis, "Nonparametric model of smooth muscle force production during electrical stimulation," *J. Comput. Biol.*, vol. 24, no. 3, pp. 229–237, 2017.

[51] V. Z. Marmarelis, D. C. Shin, M. E. Orme, and R. Zhang, "Model-based physiomarkers of cerebral hemodynamics in patients with mild cognitive impairment," *Med. Eng. Phys.*, vol. 36, no. 5, pp. 628–637, 2014.

[52] D. Song *et al.*, "Identification of sparse neural functional connectivity using penalized likelihood estimation and basis functions," *J. Comput. Neurosci.*, vol. 35, no. 3, pp. 335–357, 2013.

[53] V. Z. Marmarelis, D. C. Shin, M. E. Orme, and R. Zhang, "Model-based quantification of cerebral hemodynamics as a physiomarker for Alzheimer's disease?," *Ann. Biomed. Eng.*, vol. 41, no. 11, pp. 2296–2317, 2013.

[54] S. Zanos, T. P. Zanos, V. Z. Marmarelis, G. A. Ojemann, and E. E. Fetz, "Relationships between spike-free local field potentials and spike timing in human temporal cortex," *J. Neurophysiol.*, vol. 107, no. 7, pp. 1808–1821, 2012.

[55] W. X. Y. Li, R. H. M. Chan, W. Zhang, R. C. C. Cheung, D. Song, and T. W. Berger, "High-performance and scalable system architecture for the real-time estimation of generalized Laguerre-Volterra MIMO model from neural population spiking activity," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 1, no. 4, pp. 489–501, 2011.

[56] V. Z. Marmarelis, D. C. Shin, and R. Zhang, "Linear and nonlinear modeling of cerebral flow autoregulation using principal dynamic modes," *Open Biomed. Eng. J.*, vol. 6, p. 42, 2012.

[57] R. Zhang, J. H. Zuckerman, and B. D. Levine, "Spontaneous fluctuations in cerebral blood flow: insights from extended-duration recordings in humans," *Am. J. Physiol. Circ. Physiol.*, vol. 278, no. 6, pp. H1848--H1855, 2000.

[58] R. R. Diehl, D. Linden, and D. Liicke, "Spontaneous blood pressure oscillations and cerebral autoregulation," pp. 7–12, 1998.

[59] A. S. S. Meel-van den Abeelen *et al.*, "Between-centre variability in transfer function analysis, a widely used method for linear quantification of the dynamic pressure-flow relation: The CARNet study," *Med. Eng. Phys.*, vol. 36, no. 5, pp. 620–627, 2014.

[60] D. Song, H. Wang, and T. W. Berger, "Estimating sparse Volterra models using group L1-regularization," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*,

2010, pp. 4128–4131.

[61] K. Geng and V. Z. Marmarelis, "Methodology of Recurrent Laguerre-Volterra Network for Modeling Nonlinear Dynamic Systems," *IEEE Trans. neural networks Learn. Syst.*, 2016.

[62] K. Kostoglou *et al.*, "Accurate spike time prediction from LFP in monkey visual cortex: A non-linear system identification approach," 2014.

[63] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[64] A. I. N. Press, "A comparison of multivariate autoregressive estimators," vol. 86, pp. 2426–2429, 2006.

[65] S. Haufe, R. Tomioka, G. Nolte, K.-R. Müller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain sources for EEG/MEG," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1954–1963, 2010.

[66] W. Tang, S. L. Bressler, C. M. Sylvester, G. L. Shulman, and M. Corbetta, "Measuring Granger causality between cortical regions from voxelwise fMRI BOLD signals with LASSO," *PLoS Comput. Biol.*, vol. 8, no. 5, p. e1002513, 2012.

[67] I. Vlachos and D. Kugiumtzis, "Backward-in-Time Selection of the Order of Dynamic Regression Prediction Model," *J. Forecast.*, vol. 32, no. 8, pp. 685–701, 2013.

[68] E. Siggiridou and D. Kugiumtzis, "Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1759–1773, 2016.

[69] A. Korzeniewska, M. Mańczak, M. Kamiński, K. J. Blinowska, and S. Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," *J. Neurosci. Methods*, vol. 125, no. 1, pp. 195–207, 2003.

[70] L. Faes and G. Nollo, "Multivariate frequency domain analysis of causal interactions in physiological time series," in *Biomedical Engineering, Trends in Electronics, Communications and Software*, InTech, 2011.

[71] L. Faes, G. Nollo, and others, "Assessing frequency domain causality in cardiovascular time series with instantaneous interactions," *Methods Inf. Med.*, vol. 49, no. 5, pp. 453–457, 2010.

[72] L. Astolfi *et al.*, "Estimation of the effective and functional human cortical connectivity with structural equation modeling and directed transfer function applied to high-resolution EEG," *Magn. Reson. Imaging*, vol. 22, no. 10, pp. 1457–1470, 2004.

[73] M. Hu, M. Li, W. Li, and H. Liang, "Joint analysis of spikes and local field potentials using copula," *Neuroimage*, vol. 133, pp. 457–467, 2016.

[74] E. Lalo *et al.*, "Patterns of bidirectional communication between cortex and basal ganglia during movement in patients with Parkinson disease," *J. Neurosci.*, vol. 28, no. 12, pp. 3008–3016, 2008.

[75] D. A. Gutnisky and K. Josić, "Generation of spatiotemporally correlated spike trains and local field potentials using a multivariate autoregressive process," *J. Neurophysiol.*, vol. 103, no. 5, pp. 2912–2930, 2010.

[76] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biol. Cybern.*, vol. 84, no. 6, pp. 463–474, 2001.

[77] J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccalá, "Frequency domain connectivity identification: an application of partial directed coherence in fMRI," *Hum. Brain Mapp.*, vol. 30, no. 2, pp. 452–461, 2009.

[78] G. Wang, Z. Sun, R. Tao, K. Li, G. Bao, and X. Yan, "Epileptic seizure detection based on partial directed coherence analysis," *IEEE J. Biomed. Heal. informatics*, vol. 20, no. 3, pp. 873–879, 2016.

[79] S. M. Kay, *Modern spectral estimation*. Pearson Education India, 1988.

[80] K. Yacoub, "Relationship between multiple and partial coherence functions," *IEEE Trans. Inf. Theory*, vol. 16, no. 6, pp. 668–672, 1970.

[81] M. Eichler, R. Dahlhaus, and J. Sandkühler, "Partial correlation analysis for the identification of synaptic connections," *Biol. Cybern.*, vol. 89, no. 4, pp. 289–302, 2003.

[82] L. A. Baccala, K. Sameshima, G. Ballester, A. C. Do Valle, and C. Timo-Iaria, "Studying the interaction between brain structures via directed coherence and Granger causality," *Appl. Signal Process.*, vol. 5, no. 1, p. 40, 1998.

[83] L. Faes, S. Erla, A. Porta, and G. Nollo, "A framework for assessing frequency domain causality in physiological time series with instantaneous effects," *Phil. Trans. R. Soc. A*, vol. 371, no. 1997, p. 20110618, 2013.

[84] J. Jiang and Y. Zhang, "A revisit to block and recursive least squares for parameter estimation," *Comput. Electr. Eng.*, vol. 30, no. 5, pp. 403–416, 2004.

[85] J. Jiang and Y. Zhang, "A novel variable-length sliding window blockwise least-squares algorithm for on-line estimation of time-varying parameters," *Int. J. Adapt. Control Signal Process.*, vol. 18, no. 6, pp. 505–521, 2004.

[86] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, 1977.

[87] L. Ljung, "System identification: theory for the user," *Englewood Cliffs*, 1987.

[88] M. Nied{ź}wiecki, *Identification of Time-varying Processes*. John Wiley & Sons, Ltd, 2000.

[89] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems 1," *J. Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.

[90] M. M. Markou, M. J. Poulin, and G. D. Mitsis, "Nonstationary analysis of cerebral hemodynamics using recursively estimated multiple-input nonlinear models," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 2011, pp. 5768–5773.

[91] M. K. Tsatsanis and G. B. Giannakis, "Time-varying system identification and model validation using wavelets," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3512–3523, 1993.

[92] R. Ghanem and F. Romeo, "A wavelet-based approach for the identification of linear time-varying dynamical systems," *J. Sound Vib.*, vol. 234, no. 4, pp. 555–576, 2000.

[93] H.-L. Wei and S. A. Billings, "Identification of time-varying systems using multiresolution wavelet models," *Int. J. Syst. Sci.*, vol. 33, no. 15, pp. 1217–1228, 2002.

[94] B. I. Ikharia and D. T. Westwick, "Identification of time-varying Hammerstein systems using a basis expansion approach," in *Electrical and Computer Engineering, 2006. CCECE'06. Canadian Conference on*, 2006, pp. 1858–1861.

[95] B. I. Ikharia and D. T. Westwick, "On the identification of Hammerstein systems with time-varying parameters," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 6475–6478.

[96] M. Lortie and R. E. Kearney, "Identification of time-varying Hammerstein systems from ensemble data," *Ann. Biomed. Eng.*, vol. 29, no. 7, pp. 619–635, 2001.

[97] B. Schack, "Dynamic topographic spectral analysis of cognitive processes," *Anal. Neurophysiol. brain Funct.*, pp. 230–251, 1999.

[98] L. Astolfi *et al.*, "Estimation of the time-varying cortical connectivity patterns by the adaptive multivariate estimators in high resolution EEG studies," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 2446–2449.

[99] M. Havlicek, J. Jan, M. Brazdil, and V. D. Calhoun, "Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data," *Neuroimage*, vol. 53, no. 1, pp. 65–77, 2010.

[100] B. M. Potscher, "Model selection under nonstationarity: Autoregressive models and stochastic linear regression models," *Ann. Stat.*, pp. 1257–1274, 1989.

[101] S. Goto, M. Nakamura, and K. Uosaki, "On-line spectral estimation of nonstationary time series based on AR model parameter estimation and order selection with a forgetting factor," *Signal Process. IEEE Trans.*, vol. 43, no. 6, pp. 1519–1522, 1995.

[102] C. D. Giurcăneanu and S. A. Razavi, "AR order selection in the case when the model parameters are estimated by forgetting factor least-squares algorithms," *Signal Processing*, vol. 90, no. 2, pp. 451–466, 2010.

[103] K. Deep, K. P. Singh, M. L. Kansal, and C. Mohan, "A real coded genetic algorithm for solving integer and mixed integer optimization problems," *Appl. Math. Comput.*, vol. 212, no. 2, pp. 505–518, 2009.

[104] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*, Springer, 2011, pp. 760–766.

[105] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.

[106] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *J. Glob. Optim.*, vol. 39, no. 3, pp. 459–471, 2007.

[107] H.-G. Beyer and H.-P. Schwefel, "Evolution strategies--A comprehensive introduction," *Nat. Comput.*, vol. 1, no. 1, pp. 3–52, 2002.

[108] R. Storn and K. Price, "Differential evolution--a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[109] D. Lawrence and others, "Handbook of genetic algorithms," *Van No Strand Reinhold, New York*, 1991.

[110] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.

[111] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning, 1989," *Read. Addison-Wesley*, 1989.

[112] Y. Zhang, M. Sakamoto, and H. Furutani, "Effects of population size and mutation rate on results of genetic algorithm," in *Natural Computation, 2008. ICNC'08. Fourth International Conference on*, 2008, vol. 1, pp. 70–75.

[113] S. Sarmady, "An investigation on genetic algorithm parameters," *Sch. Comput. Sci. Univ. Sains Malaysia*, 2007.

[114] K. Deb and S. Agrawal, "Understanding Interactions among Genetic Algorithm Parameters.," in *FOGA*, 1998, pp. 265–286.

[115] J. Schaffer, "A study of control parameters affecting online performance of genetic algorithms for function

optimization," *San Meteo, Calif.*, 1989.

[116] D. Whitley, "An overview of evolutionary algorithms: practical issues and common pitfalls," *Inf. Softw. Technol.*, vol. 43, no. 14, pp. 817–831, 2001.

[117] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," *Computer (Long. Beach. Calif).*, vol. 27, no. 6, pp. 17–26, 1994.

[118] D. E. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *Found. Genet. algorithms*, vol. 1, pp. 69–93, 1991.

[119] K. Deep and M. Thakur, "A new crossover operator for real coded genetic algorithms," *Appl. Math. Comput.*, vol. 188, no. 1, pp. 895–911, 2007.

[120] K. Deep and M. Thakur, "A new mutation operator for real coded genetic algorithms," *Appl. Math. Comput.*, vol. 193, no. 1, pp. 211–230, 2007.

[121] P. Moscato and others, "On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms," *Caltech Concurr. Comput. program, C3P Rep.*, vol. 826, p. 1989, 1989.

[122] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[123] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[124] S. J. Wright and J. Nocedal, "Numerical optimization," *Springer Sci.*, vol. 35, no. 67–68, p. 7, 1999.

[125] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*. SIAM, 2009.

[126] C. Audet and W. Hare, "Derivative-free and blackbox optimization," *En préparation, prévu pour*, 2017.

[127] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optim.*, vol. 9, no. 4, pp. 877–900, 1999.

[128] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Math. Program.*, vol. 89, no. 1, pp. 149–185, 2000.

[129] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Math. Program.*, vol. 107, no. 3, pp. 391–408, 2006.

[130] T. R. Fortescue, "Brief Paper Implementation of Self-tuning Regulators with Variable Forgetting Factors*," vol. 17, no. 6, pp. 831–835, 1977.

[131] N. Rao Sripada and D. Grant Fisher, "Improved least squares identification," *Int. J. Control*, vol. 46, no. 6, pp. 1889–1913, 1987.

[132] M. E. Salgado, G. C. Goodwin, and R. H. Middleton, "Modified least squares algorithm incorporating exponential resetting and forgetting," *Int. J. Control*, vol. 47, no. 2, pp. 477–491, 1988.

[133] T. Hägglund, "Recursive estimation of slowly time varying parameters," in *7th IFAC Symposium on Identification and System Parameter Estimation*, 1985.

[134] R. Kulhav??, "Restricted exponential forgetting in real-time identification," *Automatica*, vol. 23, no. 5, pp. 589–600, 1987.

[135] J. E. Parkum, N. K. Poulsen, and J. Holst, "Recursive forgetting algorithms," *Int. J. Control*, vol. 55, no. 1, pp. 109–128, 1992.

[136] P. Andersson, "Adaptive forgetting in recursive identification through multiple models†," *Int. J. Control*, vol. 42, no. January 2015, pp. 1175–1193, 1985.

[137] K. Uosaki, M. Yotsuya, and T. Hatanaka, "Adaptive identification of non-stationary systems with multiple

forgetting factors," *Decis. Control. 1996., Proc. 35th IEEE Conf.*, vol. 1, pp. 851–856 vol.1, 1996.

[138] K. Warwick, Y.-H. Kang, and R. J. Mitchell, "Genetic least squares for system identification," *Soft Comput. - A Fusion Found. Methodol. Appl.*, vol. 3, no. 4, pp. 200–205, 1999.

[139] S. Song, J.-S. Lim, S. Baek, and K.-M. Sung, "Gauss Newton variable forgetting factor recursive least squares for time varying parameter tracking," *Electron. Lett.*, vol. 36, no. 11, pp. 988–990, 2000.

[140] S. H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, 2005.

[141] C. Paleologu, J. Benesty, and S. Ciochina, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.

[142] M. Z. A. Bhotto and A. Antoniou, "New improved recursive least-squares adaptive-filtering algorithms," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 60, no. 6, pp. 1548–1558, 2013.

[143] A. K. Kohli and A. Rai, "Numeric variable forgetting factor RLS algorithm for second-order volterra filtering," *Circuits, Syst. Signal Process.*, vol. 32, no. 1, pp. 223–232, 2013.

[144] S. Saelid and B. Foss, "Adaptive controllers with a vector variable forgetting factor," in *Decision and Control, 1983. The 22nd IEEE Conference on*, 1983, pp. 1488–1494.

[145] A. S. Poznyak and J. J. Medel Juarez, "Matrix forgetting factor," *Int. J. Syst. Sci.*, vol. 30, no. 2, pp. 165–174, 1999.

[146] A. Vahidi, A. Stefanopoulou, and H. Peng, "Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments," *Veh. Syst. Dyn.*, vol. 43, no. 1, pp. 31–55, 2005.

[147] J. Li, Y. Zheng, and Z. Lin, "Recursive identification of time-varying systems: Self-tuning and matrix RLS algorithms," *Syst. Control Lett.*, vol. 66, no. 1, pp. 104–110, 2014.

[148] A. P. Sage and G. W. Husa, "Adaptive filtering with unknown prior statistics," *Jt. Autom. Control Conf.*, vol. 7, pp. 760–769, 1969.

[149] R. Mehra, "On the identification of variances and adaptive Kalman filtering [J]," *IEEE Trans. On, Autom. Control.*, vol. 15, no. 2, pp. 175–184, 1970.

[150] K. A. Myers and B. D. Tapley, "Adaptive Sequential Estimation with Unknown Noise Statistics," *IEEE Trans. Automat. Contr.*, vol. 21, no. 4, pp. 520–523, 1976.

[151] A. Almagbile, J. Wang, and W. Ding, "Evaluating the Performances of Adaptive Kalman Filter Methods in GPS/INS Integration," *J. Glob. Position. Syst.*, vol. 9, no. 1, pp. 33–40, 2010.

[152] R. S. Tsay, "Order selection in nonstationary autoregressive models," *Ann. Stat.*, pp. 1425–1433, 1984.

[153] D. Tjøstheim and J. Paulsen, "Least squares estimates and order determination procedures for autoregressive processes with a time dependent variance," *J. Time Ser. Anal.*, vol. 6, no. 2, pp. 117–133, 1985.

[154] A. Karagrigoriou, "Asymptotically efficient order selection in nonstationary AR processes," *Test*, vol. 9, no. 2, pp. 371–391, 2000.

[155] I. Sánchez, "Recursive estimation of dynamic models using Cook's distance, with application to wind energy forecast," *Technometrics*, vol. 48, no. 1, pp. 61–73, 2006.

[156] I. Sanchez, "Short-term prediction of wind energy production," *Int. J. Forecast.*, vol. 22, no. 1, pp. 43–56, 2006.

[157] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.

[158] N. Yoshitani and A. Hasegawa, "Model-based control of strip temperature for the heating furnace in continuous annealing," *IEEE Trans. Control Syst. Technol.*, vol. 6, no. 2, pp. 146–156, 1998.

[159] K. Deb, "An efficient constraint handling method for genetic algorithms," *Comput. Methods Appl. Mech. Eng.*, vol. 186, no. 2–4, pp. 311–338, 2000.

[160] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2013.

[161] S. Gunnarsson, "Combining tracking and regularization in recursive least squares identification," *Proc. 35th IEEE Conf. Decis. Control*, vol. 3, no. December, pp. 2551–2552, 1996.

[162] A. Andersson and H. Broman, "A second-order recursive algorithm with applications to adaptive filtering and subspace tracking," *IEEE Trans. signal Process.*, vol. 46, no. 6, pp. 1720–1725, 1998.

[163] T. Kashima, K. Fukawa, and H. Suzuki, "Recursive Least Squares Channel Estimator with Smoothing and Removing for Iterative-MAP Receiver of MIMO-OFDM Mobile Communications."

[164] H. E. Rauch, C. T. Striebel, and F. Tung, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, 1965.

[165] R. S. Ers and E. S. Ers, "Estimation of Nonstationary EEG With Kalman Smoother Approach : An Application to Event- Estimation of Nonstationary EEG With Kalman Smoother Approach : An Application to," no. April, 2004.

[166] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–21, 1990.

[167] R. R. Diehl, D. Linden, D. Lücke, and P. Berlit, "Phase relationship between cerebral blood flow velocity and blood pressure a clinical test of autoregulation," *Stroke*, vol. 26, no. 10, pp. 1801–1804, 1995.

[168] G. D. Mitsis, M. J. Poulin, P. A. Robbins, and V. Z. Marmarelis, "Nonlinear Modeling of the Dynamic Effects of Arterial Pressure and CO 2 Variations on Cerebral Blood Flow in Healthy Humans," *IEEE Trans Biomed Engin*, vol. 51, no. 11, pp. 1932–1943, 2004.

[169] R. G. Wise, K. Ide, M. J. Poulin, and I. Tracey, "Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal," *Neuroimage*, vol. 21, no. 4, pp. 1652–1664, 2004.

[170] M. J. Poulin, P. J. Liang, and P. A. Robbins, "Dynamics of the cerebral blood flow response to step changes in end-tidal PCO2 and PO2 in humans," *J. Appl. Physiol.*, vol. 81, no. 3, pp. 1084–1095, 1996.

[171] R. Schondorf, R. Stein, R. Roberts, J. Benoit, and W. Cupples, "Dynamic cerebral autoregulation is preserved in neurally mediated syncope.," *J. Appl. Physiol.*, vol. 91, no. 6, pp. 2493–2502, 2001.

[172] R. Tóth, *Modeling and identification of linear parameter-varying systems*, vol. 403. Springer, 2010.

[173] G. Buzsáki and A. Draguhn, "Neuronal oscillations in cortical networks," *Science (80-.).*, vol. 304, no. 5679, pp. 1926–1929, 2004.

[174] R. Eckhorn *et al.*, "Coherent oscillations: A mechanism of feature linking in the visual cortex?," *Biol. Cybern.*, vol. 60, no. 2, pp. 121–130, 1988.

[175] P. Fries, "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence," *Trends Cogn. Sci.*, vol. 9, no. 10, pp. 474–480, 2005.

[176] T. H. Bullock, M. C. McClune, J. Z. Achimowicz, V. J. Iragui-Madoz, R. B. Duckrow, and S. S. Spencer, "Temporal

fluctuations in coherence of brain waves," *Proc. Natl. Acad. Sci.*, vol. 92, no. 25, pp. 11568–11572, 1995.

[177] A. Schnitzler and J. Gross, "Normal and pathological oscillatory communication in the brain," *Nat. Rev. Neurosci.*, vol. 6, no. 4, pp. 285–296, 2005.

[178] B. Yao, S. Salenius, G. H. Yue, R. W. Brown, and J. Z. Liu, "Effects of surface EMG rectification on power and coherence analyses: an EEG and MEG study," *J. Neurosci. Methods*, vol. 159, no. 2, pp. 215–223, 2007.

[179] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, "Identification of coupling direction: application to cardiorespiratory interaction," *Phys. Rev. E*, vol. 65, no. 4, p. 41909, 2002.

[180] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," *Biol. Cybern.*, vol. 83, no. 1, pp. 35–45, 2000.

[181] E. Möller, B. Schack, M. Arnold, and H. Witte, "Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models," *J. Neurosci. Methods*, vol. 105, pp. 143–158, 2001.

[182] L. Astolfi *et al.*, "Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 902–913, 2008.

[183] T. Milde *et al.*, "A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials," *Neuroimage*, vol. 50, no. 3, pp. 960–969, 2010.

[184] A. H. Omidvarnia, M. Mesbah, M. S. Khlif, J. M. O'Toole, P. B. Colditz, and B. Boashash, "Kalman filter-based time-varying cortical connectivity analysis of newborn EEG," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 1423–1426.

[185] A. Omidvarnia, M. Mesbah, J. M. O'Toole, P. Colditz, and B. Boashash, "Analysis of the time-varying cortical neural connectivity in the newborn EEG: A time-frequency approach," in *Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on*, 2011, pp. 179–182.

[186] L. Hu, Z. G. Zhang, and Y. Hu, "A time-varying source connectivity approach to reveal human somatosensory information processing," *Neuroimage*, vol. 62, no. 1, pp. 217–228, 2012.

[187] J. O. Garcia, J. Brooks, S. Kerick, T. Johnson, T. R. Mullen, and J. M. Vettel, "Estimating direction in brain-behavior interactions: Proactive and reactive brain states in driving," *Neuroimage*, vol. 150, pp. 239–249, 2017.

[188] T. Milde *et al.*, "Time-variant partial directed coherence in analysis of the cardiovascular system. A methodological study," *Physiol. Meas.*, vol. 32, no. 11, p. 1787, 2011.

[189] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econom.*, vol. 31, no. 3, pp. 307–327, 1986.

[190] T. Bollerslev, "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model," *Rev. Econ. Stat.*, pp. 498–505, 1990.

[191] A. D. Robertson *et al.*, "Exercise intensity modulates the change in cerebral blood flow following aerobic exercise in chronic stroke," *Exp. brain Res.*, vol. 233, no. 8, pp. 2467–2475, 2015.

[192] B. J. MacIntosh *et al.*, "Impact of a single bout of aerobic exercise on regional brain perfusion and activation responses in healthy young adults.," *PLoS One*, vol. 9, no. 1, p. e85163, 2014.

[193] A. S. Rajab, D. E. Crane, L. E. Middleton, A. D. Robertson, M. Hampson, and B. J. MacIntosh, "A single session

of exercise increases connectivity in sensorimotor-related brain networks: a resting-state fMRI study in young healthy adults," *Front. Hum. Neurosci.*, vol. 8, 2014.

[194] L. W. J. Bogert and J. J. van Lieshout, "Non-invasive pulsatile arterial pressure and stroke volume changes from the human finger," *Exp. Physiol.*, vol. 90, no. 4, pp. 437–446, 2005.

[195] I. Guelen *et al.*, "Finometer, finger pressure measurements with the possibility to reconstruct brachial pressure," *Blood Press. Monit.*, vol. 8, no. 1, pp. 27–30, 2003.

[196] S. S. Franklin, "Ageing and hypertension: the assessment of blood pressure indices in predicting coronary heart disease.," *J. Hypertens. Suppl. Off. J. Int. Soc. Hypertens.*, vol. 17, no. 5, pp. S29--36, 1999.

[197] M. E. Safar, B. I. Levy, and H. Struijker-Boudier, "Current perspectives on arterial stiffness and pulse pressure in hypertension and cardiovascular diseases," *Circulation*, vol. 107, no. 22, pp. 2864–2869, 2003.

[198] D. Chemla *et al.*, "Total arterial compliance estimated by stroke volume-to-aortic pulse pressure ratio in humans," *Am. J. Physiol. Circ. Physiol.*, vol. 274, no. 2, pp. H500--H505, 1998.

[199] G. de Simone, M. J. Roman, M. J. Koren, G. A. Mensah, A. Ganau, and R. B. Devereux, "Stroke volume/pulse pressure ratio and cardiovascular risk in arterial hypertension," *Hypertension*, vol. 33, no. 3, pp. 800–805, 1999.

[200] F. U. S. Mattace-Raso *et al.*, "Arterial stiffness and risk of coronary heart disease and stroke," *Circulation*, vol. 113, no. 5, pp. 657–663, 2006.

[201] A. W. Ashor, J. Lara, M. Siervo, C. Celis-Morales, and J. C. Mathers, "Effects of exercise modalities on arterial stiffness and wave reflection: a systematic review and meta-analysis of randomized controlled trials," *PLoS One*, vol. 9, no. 10, p. e110034, 2014.

[202] T. Okamoto, S. Min, and M. Sakamaki-Sunaga, "Arterial compliance and stiffness following low-intensity resistance exercise," *Eur. J. Appl. Physiol.*, vol. 114, no. 2, pp. 235–241, 2014.

[203] H. Kawano, H. Tanaka, and M. Miyachi, "Resistance training and arterial compliance: keeping the benefits while minimizing the stiffening," *J. Hypertens.*, vol. 24, no. 9, pp. 1753–1759, 2006.

[204] B. A. Kingwell, K. L. Berry, J. D. Cameron, G. L. Jennings, and A. M. Dart, "Arterial compliance increases after moderate-intensity cycling," *Am. J. Physiol. Circ. Physiol.*, vol. 273, no. 5, pp. H2186--H2191, 1997.

[205] A. E. DeVan, M. M. Anton, J. N. Cook, D. B. Neidre, M. Y. Cortez-Cooper, and H. Tanaka, "Acute effects of resistance exercise on arterial compliance," *J. Appl. Physiol.*, vol. 98, no. 6, pp. 2287–2291, 2005.

[206] H. Tanaka, F. A. Dinenno, K. D. Monahan, C. M. Clevenger, C. A. DeSouza, and D. R. Seals, "Aging, habitual exercise, and dynamic arterial compliance," *Circulation*, vol. 102, no. 11, pp. 1270–1275, 2000.

[207] T. Tarumi *et al.*, "Cerebral hemodynamics in normal aging: central artery stiffness, wave reflection, and pressure pulsatility," *J. Cereb. Blood Flow Metab.*, vol. 34, no. 6, pp. 971–978, 2014.

[208] A. J. S. Webb, M. Simoni, S. Mazzucco, W. Kuker, U. Schulz, and P. M. Rothwell, "Increased cerebral arterial pulsatility in patients with leukoaraiosis," *Stroke*, vol. 43, no. 10, pp. 2631–2636, 2012.

[209] T. Otsuki *et al.*, "Contribution of systemic arterial compliance and systemic vascular resistance to effective arterial elastance changes during exercise in humans," *Acta Physiol.*, vol. 188, no. 1, pp. 15–20, 2006.

[210] Z. Zhao, W. Lin, P. Yang, and Y. Zhang, "Spectral and Coherence Analysis of the Variabilities of Heart Rate, Stroke Volume, and Systolic Blood Pressure in Exercise Stress Tests," vol. 25, no. Bhi, pp. 771–774, 2012.

[211] R. Perini and A. Veicsteinas, "Heart rate variability and autonomic activity at rest and during exercise in

various physiological conditions," *Eur. J. Appl. Physiol.*, vol. 90, no. 3–4, pp. 317–325, 2003.

[212] J. Sung, S. H. Choi, Y.-H. Choi, D.-K. Kim, and W. H. Park, "The relationship between arterial stiffness and increase in blood pressure during exercise in normotensive persons," *J. Hypertens.*, vol. 30, no. 3, pp. 587–591, 2012.

[213] A. J. Camm *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[214] G. E. Billman, "Heart rate variability--a historical perspective," *Front. Physiol.*, vol. 2, 2011.

[215] A. McLaren *et al.*, "Autonomic function is impaired in elderly stroke survivors," *Stroke*, vol. 36, no. 5, pp. 1026–1030, 2005.

[216] M. Dütsch, M. Burger, C. Dörfler, S. Schwab, and M. J. Hilz, "Cardiovascular autonomic function in poststroke patients," *Neurology*, vol. 69, no. 24, pp. 2249–2255, 2007.

[217] D. R. Seals and M. D. Esler, "Human ageing and the sympathoadrenal system," *J. Physiol.*, vol. 528, no. 3, pp. 407–417, 2000.

[218] W. C. Levy *et al.*, "Effect of endurance exercise training on heart rate variability at rest in healthy young and older men," *Am. J. Cardiol.*, vol. 82, no. 10, pp. 1236–1241, 1998.

[219] S. Ogoh *et al.*, "Middle cerebral artery flow velocity and pulse pressure during dynamic exercise in humans," *Am J Physiol Hear. Circ Physiol*, vol. 288, no. 4, pp. H1526–H1531, 2005.

[220] J. P. Fisher, S. Ogoh, C. N. Young, P. B. Raven, and P. J. Fadel, "Regulation of middle cerebral artery blood velocity during dynamic exercise in humans: influence of aging," *J. Appl. Physiol.*, vol. 105, no. 1, pp. 266–273, 2008.

[221] C. Wilke, L. Ding, B. He, and others, "Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 11, pp. 2557–2564, 2008.

[222] P. L. C. Rodrigues and L. A. Baccalá, "Statistically significant time-varying neural connectivity estimation using generalized partial directed coherence," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 5493–5496.

[223] C. Wilke, L. Ding, and B. He, "An adaptive directed transfer function approach for detecting dynamic causal interactions," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 4949–4952.

[224] L. Faes, H. Zhao, K. H. Chon, and G. Nollo, "Time-varying surrogate data to assess nonlinearity in nonstationary time series: application to heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 685–695, 2009.

[225] U. Mitzdorf, "Properties of the evoked potential generators: current source-density analysis of visually evoked potentials in the cat cortex.," *Int. J. Neurosci.*, vol. 33, no. December, pp. 33–59, 1987.

[226] A. D. Legatt, J. Arezzo, and H. G. Vaughan, "Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials," *J. Neurosci. Methods*, vol. 2, no. 2, pp. 203–217, Apr. 1980.

[227] A. A. Kühn, A. Kupsch, G.-H. Schneider, and P. Brown, "Reduction in subthalamic 8--35 Hz oscillatory activity correlates with clinical improvement in Parkinson's disease," *Eur. J. Neurosci.*, vol. 23, no. 7, pp. 1956–1960,

2006.

[228] K. P. Michmizos, P. Frangou, P. Stathis, D. Sakas, and K. S. Nikita, "Beta-band frequency peaks inside the subthalamic nucleus as a biomarker for motor improvement after deep brain stimulation in Parkinson's disease," *IEEE J. Biomed. Heal. informatics*, vol. 19, no. 1, pp. 174–180, 2015.

[229] M. A. Lebedev and M. A. L. Nicolelis, "Brain--machine interfaces: past, present and future," *TRENDS Neurosci.*, vol. 29, no. 9, pp. 536–546, 2006.

[230] J. Huxter, N. Burgess, and J. O'keefe, "Independent rate and temporal coding in hippocampal pyramidal cells," *Nature*, vol. 425, no. 6960, pp. 828–832, 2003.

[231] R. VanRullen, R. Guyonneau, and S. J. Thorpe, "Spike times make sense," *Trends Neurosci.*, vol. 28, no. 1, pp. 1–4, 2005.

[232] A. K. Engel, P. König, A. K. Kreiter, T. B. Schillen, and W. Singer, "Temporal coding in the visual cortex: new vistas on integration in the nervous system," *Trends Neurosci.*, vol. 15, no. 6, pp. 218–226, 1992.

[233] D. A. Butts *et al.*, "Temporal precision in the neural code and the timescales of natural vision," *Nature*, vol. 449, no. 7158, pp. 92–95, 2007.

[234] B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, and R. A. Andersen, "Temporal structure in neuronal activity during working memory in macaque parietal cortex," *Nat. Neurosci.*, vol. 5, no. 8, pp. 805–811, 2002.

[235] C. Mehring, J. Rickert, E. Vaadia, S. C. de Oliveira, A. Aertsen, and S. Rotter, "Inference of hand movements from local field potentials in monkey motor cortex," *Nat. Neurosci.*, vol. 6, no. 12, pp. 1253–1254, 2003.

[236] M. J. Rasch, A. Gretton, Y. Murayama, W. Maass, and N. K. Logothetis, "Inferring spike trains from local field potentials," *J. Neurophysiol.*, vol. 99, no. 3, pp. 1461–1476, 2008.

[237] Y. Cui, L. D. Liu, J. M. McFarland, C. C. Pack, and D. A. Butts, "Inferring cortical variability from local field potentials," *J. Neurosci.*, vol. 36, no. 14, pp. 4121–4135, 2016.

[238] R. Storchi, A. G. Zippo, G. C. Caramenti, M. Valente, and G. E. M. Biella, "Predicting Spike Occurrence and Neuronal Responsiveness from LFPs in Primary Somatosensory Cortex," *PLoS One*, vol. 7, no. 5, p. e35850, 2012.

[239] E. E. Galindo-Leon and R. C. Liu, "Predicting stimulus-locked single unit spiking from cortical local field potentials," *J. Comput. Neurosci.*, vol. 29, no. 3, pp. 581–597, 2010.

[240] K. P. Michmizos, D. Sakas, and K. S. Nikita, "Prediction of the timing and the rhythm of the parkinsonian subthalamic nucleus neural spikes using the local field potentials," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 2, pp. 190–197, 2012.

[241] K. P. Michmizos and K. S. Nikita, "Local field potential driven Izhikevich model predicts a subthalamic nucleus neuron activity," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5900–5903.

[242] K. P. Michmizos, D. Sakas, and K. S. Nikita, "Parameter identification for a local field potential driven model of the Parkinsonian subthalamic nucleus spike activity," *Neural Networks*, vol. 36, pp. 146–156, 2012.

[243] A. L. Benabid, "Deep brain stimulation for Parkinson's disease," *Curr. Opin. Neurobiol.*, vol. 13, no. 6, pp. 696–706, 2003.

[244] K. Kostoglou *et al.*, "Classification and Prediction of Clinical Improvement in Deep Brain Stimulation from Intraoperative Microelectrode Recordings," *Trans. Biomed. Eng.*, 2016.

[245] A. Moran, H. Bergman, Z. Israel, and I. Bar-Gad, "Subthalamic nucleus functional organization revealed by parkinsonian neuronal oscillations and synchrony," *Brain*, vol. 131, no. 12, pp. 3395–3409, 2008.

[246] D. E. Sakas, A. T. Kouyialis, E. J. Boviatsis, I. G. Panourias, P. Stathis, and G. Tagaris, "Technical aspects and considerations of deep brain stimulation surgery for movement disorders BT - Operative Neuromodulation: Volume 2: Neural Networks Surgery," D. E. Sakas and B. A. Simpson, Eds. Vienna: Springer Vienna, 2007, pp. 163–170.

[247] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural Comput.*, vol. 16, no. 8, pp. 1661–1687, 2004.

[248] A. Moran and I. Bar-Gad, "Revealing neuronal functional organization through the relation between multi-scale oscillatory extracellular signals," *J. Neurosci. Methods*, vol. 186, no. 1, pp. 116–129, 2010.

[249] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nat. Neurosci.*, vol. 2, no. 11, pp. 947–957, 1999.

[250] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," *Trends Cogn. Sci.*, vol. 14, no. 11, pp. 506–515, 2010.

[251] M. L. V Quyen and et al., "Comparison of {Hilbert} transform and wavelet methods for the analysis of neuronal synchrony," *J. Neurosci. Methods*, vol. 111, pp. 83–98, 2001.

[252] C. R. Legendy and M. Salcman, "Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons," *J. Neurophysiol.*, vol. 53, no. 4, pp. 926–939, 1985.

[253] B. Gourévitch and J. J. Eggermont, "A nonparametric approach for detection of bursts in spike trains," *J. Neurosci. Methods*, vol. 160, no. 2, pp. 349–358, 2007.

[254] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *BBA - Protein Struct.*, vol. 405, no. 2, pp. 442–451, 1975.

[255] M. C. W. van Rossum, "A novel spike distance," *Neural Comput.*, vol. 13, no. 4, pp. 751–763, 2001.

[256] K. P. Michmizos, D. Sakas, and K. S. Nikita, "Toward relating the subthalamic nucleus spiking activity to the local field potentials acquired intranuclearly," *Meas. Sci. Technol.*, vol. 22, no. 11, pp. 114021-1–9, 2011.

[257] A. K. Sen and J. O. Dostrovsky, "Evidence of intermittency in the local field potentials recorded from patients with Parkinson's disease: a wavelet-based approach," *Comput. Math. Methods Med.*, vol. 8, no. 3, pp. 165–171, 2007.

[258] C. C. Chen *et al.*, "Complexity of subthalamic 13-35Hz oscillatory activity directly correlates with clinical impairment in patients with Parkinson's disease," *Exp. Neurol.*, vol. 224, no. 1, pp. 234–240, 2010.

[259] A. Sharott *et al.*, "Activity Parameters of Subthalamic Nucleus Neurons Selectively Predict Motor Symptom Severity in Parkinson's Disease," *J. Neurosci.*, vol. 34, no. 18, pp. 6273–6285, 2014.

[260] T. Radman, Y. Su, J. H. An, L. C. Parra, and M. Bikson, "Spike timing amplifies the effect of electric fields on neurons: implications for endogenous field effects," *J. Neurosci.*, vol. 27, no. 11, pp. 3030–3036, 2007.

[261] B. Voytek and R. T. Knight, "Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease," *Biol. Psychiatry*, vol. 77, no. 12, pp. 1089–1097, 201

Appendix

A.1. SI-TV systems with sinusoidal variations

Time-varying linear ($Q = 1$) and second-order nonlinear ($Q = 2$) kernels given by linear combinations of DLFs were constructed using 2, 4 and 6 functions ($L = \{2,\ 4,\ 6\}$). The Laguerre parameters were set to 0.2, 0.4 and 0.6 ($\alpha = \{0.2,\ 0.4,\ 0.6\}$). TV expansion coefficients were constructed in the following manner,

$$c_j(n) = A\sin(2\pi f n) + M_j, \qquad f = F/N \tag{A. 1}$$

where $j = 1, \dots, ((L+Q)!/L!\,Q!) - 1$, M_j is a random number with its absolute value defined in the range [0.25 1] and represents the mean coefficient value, A is the amplitude ($A = \{0.05\ , 0.1, 0.2\}$) and F is the frequency ($F = \{2\ ,5\ ,10\}$) of the sinusoidally oscillating coefficients respectively, and N is the total number of time samples. For the SME case, the first-order kernel was of the form,

$$l(m) = e^{-\frac{m}{3}} \sin\left(\frac{\pi m}{5}\right) \tag{A. 2}$$

Its amplitude was modulated in time as follows,

$$s(n) = A\sin(2\pi f n) + 1, \qquad f = F/N \tag{A. 3}$$

$$k_1(n, m) = l(n, m) = l(m)s(n) \tag{A. 4}$$

$$k_2(n, m_1, m_2) = l(n, m_1)l(n, m_2) \tag{A. 5}$$

where A and F (same values as in the DLF case) is the amplitude and frequency of the sinusoidal signal $s(n)$ and $k_1(n,m)$, $k_2(n,m_1,m_2)$ are the 1st- and 2nd- TV system kernels respectively.

A.2. SI-TV systems with mixed-mode variations

Mixed-mode nonstationarities were characterized by an initial period of slow sinusoidal kernel variations, an abrupt jump and a final period of fast sinusoidal changes. Specifically, in the DLF case the coefficients were varied in the following way,

$$c_j(n) = \begin{cases} 0.01\sin(2\pi f_1 n) + M_j, & f_1 = 2/N\,, \quad n = 1,\dots,1000 \\ \\ 0.4\sin(2\pi f_2 n) - M_j, & f_2 = 5/N\,, \; n = 1001,\dots,2000 \end{cases} \tag{A.6}$$

whereas in the SME case (Equation (A.2)) the modulating signal was of the form,

$$s(n) = \begin{cases} 0.01\sin(2\pi f_1 n) + 1, & f_1 = 2/N\,, \; n = 1,\dots,1000 \\ \\ 0.4\sin(2\pi f_2 n) - 1, & f_2 = 5/N\,, \quad n = 1001,\dots,2000 \end{cases} \tag{A.7}$$

TV kernels were created based on Equations (A.4-A.5).

A.3. MI-TV systems with sinusoidal variations

Two-input TV linear and nonlinear kernels were constructed using DTLF functions ($[L_1, L_2]$) with Laguerre parameters ($[\alpha_1, \alpha_2]$) set to specific values. TV expansion coefficients were created as follows,

$$c_j^{(1)}(n) = 0.4\sin\left(2\pi f^{(1)} n\right) + M_j^{(1)}, f^{(i)} = 5/N \tag{A.8}$$

$$c_j^{(2)}(n) = 0.05\sin\left(2\pi f^{(1)} n\right) + M_j^{(2)}, f^{(i)} = 2/N \tag{A.9}$$

$$c_j^{(1,1)}(n) = 0.4\sin\left(2\pi f^{(1,1)} n\right) + M_j^{(1,1)}, f^{(1,1)} = 5/N \tag{A.10}$$

$$c_j^{(2,2)}(n) = 0.05\sin\left(2\pi f^{(2,2)} n\right) + M_j^{(2,2)}, f^{(2,2)} = 2/N \tag{A.11}$$

$$c_j^{(1,2)}(n) = 0.2\sin\left(2\pi f^{(1,2)} n\right) + M_j^{(1,2)}, f^{(1,2)} = 3/N \tag{A.12}$$

where $c_j^{(i)}$(for $j = 1 \dots L_i$) and $c_j^{(i,i)}$ $\left(\text{for } j = 1 \dots \frac{L_i(L_i+1)}{2}\right)$ are the coefficients of the 1st-order and 2nd-order self-kernel of input i ($i = 1,2$), whereas $c_j^{(1,2)}$(for $j = 1 \dots L_1 L_2$) are the coefficients of the 2nd-order cross-kernel. For the SME case, systems were simulated with 1st-order kernels of the form,

$$l_1(m) = e^{-\frac{m}{3}} \sin\left(\frac{\pi m}{5}\right) \tag{A. 13}$$

$$l_2(m) = e^{-\frac{m}{10}} - e^{-\frac{m}{3}} \tag{A. 14}$$

Their amplitude was modulated in time by the signals,

$$s_1(n) = 0.4\sin(2\pi f_1 n) + 1, \qquad f_1 = 5/N \tag{A. 15}$$

$$s_2(n) = 0.05\sin(2\pi f_2 n) + 1, \quad f_2 = 2/N \tag{A. 16}$$

respectively obtaining the TV 1st- and 2nd-order self-kernels,

$$k_1^{(i)}(n, m) = l_i(n, m) = l_i(m) s_i(n) \tag{A. 17}$$

$$k_2^{(i,i)}(n, m_1, m_2) = l_i(n, m_1) l_i(n, m_2) \tag{A. 18}$$

for $i = 1,2$ and the cross-kernel,

$$k_2^{(1,2)}(n, m_1, m_2) = l_1(n, m_1) l_2(n, m_2) \tag{A. 19}$$

A.4. MI-TV systems with mixed-mode variations

In the DLF case coefficients exhibited the following variations in time,

$$c_j^{(1)}(n) = \begin{cases} 0.01\sin\left(2\pi f_1^{(1)} n\right) + M_j^{(1)}, & f_1^{(1)} = 5/N, \quad n = 1 \dots 1200 \\ 0.05\sin\left(2\pi f_2^{(1)} n\right) + M_j^{(1)}, & f_2^{(1)} = 5/N, \quad n = 1201 \dots 1600 \\ 0.1\sin\left(2\pi f_3^{(1)} n\right) + M_j^{(1)}, & f_3^{(1)} = \dfrac{1}{N}, \quad n = 1601 \dots 2000 \end{cases} \tag{A. 20}$$

$$c_j^{(2)}(n) = \begin{cases} 0.4\sin\left(2\pi f_1^{(2)} n\right) + M_j^{(2)}, & f_1^{(2)} = 1/N, \quad n = 1 \dots 700 \\ 0.2\sin\left(2\pi f_2^{(2)} n\right) - M_j^{(2)}, & f_2^{(2)} = 5/N, \quad n = 701 \dots 1400 \end{cases} \tag{A. 21}$$

$$0.4\sin\left(2\pi f_3^{(2)} n\right) + M_j^{(2)}, \qquad f_3^{(2)} = 1/N, \qquad n = 1400 \ldots 2000$$

$$c_j^{(1,2)}(n) = \begin{cases} 0.2\sin\left(2\pi f_1^{(1,2)} n\right) + M_j^{(1,2)}, & f_1^{(1,2)} = 2/N\,, \quad n = 1 \ldots 900 \\ 0.4\sin\left(2\pi f_2^{(1,2)} n\right) + M_j^{(1,2)}, & f_2^{(1,2)} = 3/N\,, \quad n = 901 \ldots 1700 \\ 0.01\sin\left(2\pi f_3^{(1,2)} n\right) + M_j^{(1,2)}, & f_3^{(1,2)} = 4/N, \quad n = 1701 \ldots 2000 \end{cases} \quad \text{(A. 22)}$$

Again, all the M_j's are random numbers with their absolute value defined in the range [0.25 1]. The coefficients of the 2nd-order self-kernels ($c_j^{(i,i)}$) were the same as the coefficients of their first order counterparts ($c_j^{(i)}$) but with different mean values (i.e. $M_j^{(i,i)}$). In the SME case, the amplitude of the kernels of Equations (A.13-A.14) were modulated in time by the signals,

$$s_1(n) = \begin{cases} n/1000, & n = 1 \ldots 1000 \\ 1, & n = 1001 \ldots 1400 \\ 0.1\sin\left(2\pi f_1^{(1)} n\right) + 1, & f_1^{(1)} = 1/\mathrm{N}, \quad n = 1401 \ldots 2000 \end{cases} \quad \text{(A. 23)}$$

$$s_2(n) = \begin{cases} 0.2\sin\left(2\pi f_1^{(2)} n\right) + 1, & f_1^{(2)} = 1/\mathrm{N}, \quad n = 1 \ldots 800 \\ \\ 0.4\sin\left(2\pi f_2^{(2)} n\right) - 1, & f_2^{(2)} = 5/\mathrm{N}, \quad n = 801 \ldots 2000 \end{cases} \quad \text{(A. 24)}$$

The TV 1st- and 2nd-order kernels were computed as in Equations (A.17-A.18).