A Machine Learning Methodology for Developing Microscopic Vehicular Fuel Consumption and Emission Models for Local Conditions using Real-World Measures

by

Ehsan Moradi

Department of Civil Engineering and Applied Mechanics

McGill University, Montreal

December 2020

A thesis submitted to McGill University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Ph.D.) in Civil Engineering

Author’s Declaration

I hereby declare that I am the sole author of this thesis. To the best of my knowledge, this thesis contains no material previously published by any other person except where due acknowledgment has been made. I further authorize McGill University to reproduce this thesis by photocopying, online sharing, or by other means, in total or in part, at the request of other institutions or individuals for scholarly research.

Abstract

Road transport is a major contributor to world energy consumption and emissions. The validity of models developed for environmental assessment of transport projects when used out of their origins is questionable as they are only validated for the prevailing conditions at their origin. This study starts by the validation of one of the most popular transportation environmental assessment models, MOVES, for use in non-U.S. regions such as Canada through performing on-road measurements. Distinct differences between the ground-truth and MOVES predictions are revealed. MOVES underestimates fuel and CO2 rates by 17% and 35%, respectively. *Nitrogen Oxides* (NOx) and *Particulate Matters* (PM) predictions set overestimation records of up to +420%. Furthermore, MOVES output is biased for vehicle groups with specific attributes.

The results of MOVES validation emphasized the need for using alternative local fuel and emission models. However, many of the existing vehicular fuel and emission modeling methodologies are criticized in aspects such as ignoring real-world training data, low diversity of test fleet, impracticality in real-world applications (such as instrument-independent eco-driving or use alongside with traffic microsimulation), and low prediction power in the non-linear multi-dimensional space of fuel consumption and emission generation. Hence, a machine learning modeling methodology relying on on-road data from a fleet of 35 vehicles is proposed. The accuracy of the proposed

instrument-independent models is tried to be improved by introducing estimates of influential engine variables to the feature set through a cascaded modeling procedure. As a result, the R-squared metric reached 83%, while score improvements as high as 37% are achieved depending on the vehicle class and the machine learning technique used.

Despite the considerable scores achieved by utilizing fully-connected neural networks architectures, use of techniques compatible with the serially-correlated nature of vehicular operation seems more promising in achieving higher accuracy and robustness. Moreover, generalizing the models developed for particular vehicles to more aggregate levels is a need for diversifying models' use cases. To this end, a two-stage ensemble learning methodology based on vehicle-specific *Recurrent Neural Network* (RNN) models is proposed.

Long Short-Term Memory (LSTM) cell architecture resulted in the best lag-specific modeling scores (compared to the other RNN cell types). Vehicle-specific ensemble models developed by combining predictions from lag-specific RNN models showed score improvement records of up to 28% compared to the best component model (4% on average). In addition, the category-specific ensembles developed on top of metamodels achieved score improvements of up to 32% compared to the best component metamodel (6% on average). Linear regression dominantly resulted in the best score improvements for NOx and PM rates at both forecast combination stages, while random forests and gradient boosting methods dominantly worked the best for fuel and CO2 rates.

Keywords**:** Vehicular Fuel Consumption, Vehicular Emissions, Eco-Driving, On-road Measurements, Model Validation, MOVES, Machine Learning, Deep Learning, Support Vector Regression, Artificial Neural Networks, Cascaded Modeling, Recurrent Neural Networks, Ensemble Learning.

Résumé

Le transport routier est un contributeur majeur à la consommation d'énergie et aux émissions mondiales. La validité des modèles développés pour l'évaluation environnementale des projets de transport lorsqu'ils sont utilisés hors de leurs origines est discutable car ils ne sont validés que pour les conditions prévalant à leur origine. Cette étude commence par la validation de l'un des modèles d'évaluation environnementale des transports les plus populaires, MOVES, pour une utilisation dans des régions non américaines comme le Canada en effectuant des mesures sur route. Des différences distinctes entre la vérité terrain et les prédictions MOVES sont révélées. MOVES sous-estime les taux de carburant et de CO2 de 17% et 35%, respectivement. Les prédictions sur les oxydes d'azote (NOx) et les particules (PM) établissent des records de surestimation allant jusqu'à + 420%. De plus, la sortie MOVES est biaisée pour les groupes de véhicules avec des attributs spécifiques.

Les résultats de la validation MOVES ont souligné la nécessité d'utiliser des modèles locaux de carburants et d'émissions alternatifs. Cependant, de nombreuses méthodologies existantes de modélisation des émissions et des carburants des véhicules sont critiquées pour des aspects tels que l'ignorance des données de formation du monde réel, la faible diversité du parc d'essai, l'impossibilité pratique dans les applications du monde réel (telles que l'éco-conduite indépendante des instruments ou l'utilisation en

parallèle avec microsimulation du trafic) et une faible puissance de prédiction dans l'espace multidimensionnel non linéaire de la consommation de carburant et de la production d'émissions. Par conséquent, une méthodologie de modélisation d'apprentissage automatique reposant sur des données routières d'une flotte de 35 véhicules est proposée. On essaie d'améliorer la précision des modèles proposés indépendants de l'instrument en introduisant des estimations des variables influentes du moteur dans l'ensemble de fonctionnalités grâce à une procédure de modélisation en cascade. En conséquence, la métrique R-carré a atteint 83%, tandis que des améliorations de score allant jusqu'à 37% sont obtenues en fonction de la classe de véhicule et de la technique d'apprentissage automatique utilisée.

Malgré les scores considérables obtenus en utilisant des architectures de réseaux de neurones entièrement connectés, l'utilisation de techniques compatibles avec la nature corrélée en série du fonctionnement des véhicules semble plus prometteuse pour obtenir une précision et une robustesse plus élevées. De plus, la généralisation des modèles développés pour des véhicules particuliers à des niveaux plus agrégés est nécessaire pour diversifier les cas d'utilisation des modèles. À cette fin, une méthodologie d'apprentissage d'ensemble en deux étapes basée sur des modèles de réseaux neuronaux récurrents (RNN) spécifiques au véhicule est proposée.

L'architecture de cellule de mémoire à long terme (LSTM) a abouti aux meilleurs scores de modélisation spécifiques au décalage (par rapport aux autres types de cellules

RNN). Les modèles d'ensemble spécifiques au véhicule développés en combinant les prédictions des modèles RNN spécifiques au décalage ont montré des records d'amélioration du score allant jusqu'à 28% par rapport au meilleur modèle de composants (4% en moyenne). En outre, les ensembles spécifiques à la catégorie développés en plus des métamodèles ont obtenu des améliorations de score allant jusqu'à 32% par rapport au meilleur métamodèle composant (6% en moyenne). La régression linéaire a principalement abouti aux meilleures améliorations de score pour les taux de NOx et de PM aux deux étapes de combinaison de prévisions, tandis que les forêts aléatoires et les méthodes de renforcement de gradient ont principalement fonctionné le mieux pour les taux de carburant et de CO2.

***Mots clés*:** consommation de carburant des véhicules, émissions des véhicules, éco-conduite, mesures sur route, validation de modèle, MOVES, apprentissage automatique, apprentissage en profondeur, régression vectorielle de soutien, réseaux de neurones artificiels, modélisation en cascade, réseaux de neurones récurrents, apprentissage d'ensemble.

Acknowledgements

I would like to convey my gratitude, first to my supervisor, Prof. Luis Miranda-Moreno, for his trust in me and consistent support throughout my Ph.D. studies. I should also thank Dr. Omid M. Rouhani, whom I started my Ph.D. studies with and was in charge of supervising me up to my comprehensive oral exam. His guidance and support at the early stages, oriented and boosted my research, especially, through the design of the experimental procedure and implementation of the field works.

I would also like to acknowledge the financial support provided by the *Natural Sciences and Engineering Research Council of Canada* (NSERC) as well as the complementary support provided by the *Mitacs* organization, without which completion of this research was impossible due to the high cost of equipment and other expenses related to field activities.

Lastly, I would like to thank my loved ones, my wife Atena and my daughter Rosha, who tolerated the difficulties of living in a foreign country for about three years to let me follow my academic ambitions.

Publications to Date

The candidate has been involved in the preparation and publication of 6 papers related to his PhD research topic through the course of his studies. He has been the first-author in the first 3 and co-author for the rest.

No.	Title	Journal / Conference	Year of Publication	Status
J1	Developing Category-specific Vehicular Fuel and Emission models; A Hybrid Machine Learning Methodology	Transportation Research Part D	-	Under Review
C1	On-road vs. Software-based Measurements: On Validity of Fuel, CO2, NOx, and PM Predictions by US EPA's MOVES	TRB Annual Conference	2021	Presented
J2	Vehicular Fuel Consumption Estimation using Real-World Measures through Cascaded Machine Learning Modeling	Transportation Research Part D	2020	Published
C2	A Comparative Analysis of the Vehicular Emissions Generated as a Results of Different Intersection Controls	TRB Annual Conference	2020	Presented
C3	Energy and Environmental Footprints of Urban Travel: The Case of the City of Montreal	TRB Annual Conference	2019	Presented
J3	Perceptions, preferences, and behavior regarding energy and environmental costs: the case of Montreal transport users	Sustainability	2018	Published

Contributions of Authors

This thesis is founded upon the content and contributions of the papers J1, J2, and C1 with minor references to paper C2. The papers J1, J2, and C1 are all the candidate's original work (co-authored by Prof. Luis Miranda-Moreno). In each of these 3 papers, the study conception and design and the interpretation of results have been the joint work of the candidate and Prof. Luis Miranda-Moreno, while the field data collection, data visualization, computer coding for model development, running simulations, and preparation of the manuscripts have been done by the candidate. A finding and a quote from the paper C2, authored by Molly Behan (and co-authored by the candidate and Prof. Luis Miranda-Moreno), is mentioned in Section 3.7.2 and Chapter 7, respectively.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The overall quality of life, household economy, and the environment are profoundly influenced by the transportation sector. Constituting to about 20% of the total *Greenhouse Gas* (GHG) emissions [1] and being a major contributor to fine *Particulate Matters* (PM) and other toxic compounds such as *Nitrogen Oxides* (NOx) in Canada [2], the transportation sector is at the epicenter of the environmental controversies. GHG emissions such as CO2 play a major role in global warming phenomenon, while particulate matters are identified as a main cause of respiratory and cardiovascular diseases. On the other hand, NOx are influential elements in forming the photochemical smog during summer and have negative impacts on plants growth.

Mobile sources contributed to about 54% of NOx emissions in 2014, 65% of which from on-road mobility [3]. Besides, Canadians paid an amount of $55 billion to fuel up their vehicles in 2017 [4]. This means that even a 1% reduction in annual vehicular energy consumption would lead to the saving of millions of dollars for the country as well as saving the environment and many lives due to less emission levels. Given such significant impacts and due to increasing worldwide environmental concerns and the growing attention to energy efficiency, transportation external costs have become a key priority at all government levels. Strategies and projects to improve road network efficiency, eco-driving behaviors, fleet efficiency, and maintenance routines towards environmentally-friendly directions are all being conducted as an answer to the aforementioned concerns.

The environmental impact assessments of projects and strategies requires accurate emission modeling and estimation tools. The current state of the practice relies on emission modeling software tools that can estimate emissions at different levels (link, city, or regions). The most commonly used tool by practitioners in the U.S. and Canada is the *Motor Vehicle Emission Simulator* (MOVES) developed by the *United States Environmental Protection Agency* (USEPA). The model is widely used for environmental assessment of projects at different scales, ranging from small-scale network treatment scenarios to large-scale policy implementations. MOVES covers a broad range of pollutants and estimates energy consumption at national/county (macro) and project (link-level) scales. Hence, it has the potential to be used for eco-driving purposes as well.

The major drawback of deploying MOVES in countries rather than where its core models are designed and calibrated (the U.S.) is the unreliability of estimations. The model provides adjustment options to adapt it to the prevailing conditions of the area under study (in terms of fleet distribution, vehicle types, fuel formulation, and even meteorology); however, it leaves the user limited to the range of conditions existing in the U.S. states and counties. But the U.S. fleet used for the estimation of MOVES core models is significantly different from the Canadian vehicle fleet [5,6]. Canadian fleet is composed of a higher proportion of smaller-size vehicles with lower *Fuel Consumption Rates* (FCR). Even with the increasing global demand for SUV vehicles in recent years, there has been a surge in Canadian's interest in crossovers over full-size SUVs considering their better fuel economy [7]. Another crucial difference lies in harsh weather during winter (and even fall) in Canada which is not comparable to the U.S. states. Furthermore, the average annual mileage varies significantly among different countries, which in combination with the meteorological conditions, leads to different depreciation rates even for cars of the same model and age and could result in different FCR and *Emission Rates* (ER) [8].

Additionally, the type of experiments conducted for collecting raw data to estimate MOVES core models is a matter of consideration. For this purpose, in-lab chassis dynamometer tests are performed by EPA through the *FTP-75* (Federal Test Procedure) driving cycles for urban driving simulation in addition to the supplementary US06 test procedure [9]. The latter is used to address the FTP-75 shortcomings in representing combined high-speed and/or high-acceleration driving behavior, rapid speed fluctuations, and driving behavior following start-up [10]. But controlled test environments could affect the quality of energy consumption and emissions estimates significantly [11]. Several factors including pavement quality, tire friction due to its type, age, and pressure, wind direction, rainfall, snowy/icy road conditions, and even complex driving patterns due to the topology of the transportation network are disregarded in lab experiments.

Whether the popular models such as MOVES are adequately reliable for the local needs in the environmental assessment of transport projects or if they could be trusted for incorporation in microsimulations or eco-driving services in Canada (and generally, regions other than the models' country of origin) is a matter of question which requires a thorough validation analysis.

An alternative approach to tackle the uncertainties corresponding to available models is developing region- and fleet-specific models. Nevertheless, due to the common perception that the similarities between the U.S. and Canadian transportation sectors outweigh the impact of differences, no significant effort has been put into developing alternatives so far.

Developing local vehicular fuel and emission models is itself a challenging task. To develop reliable and efficient real-time eco-driving assistance services, design smartphone-based energy and emissions quantification procedures, equip existing traffic simulation models with energy and emissions assessment modules, and assess the environmental impact of transport/urban development plans before or after the implementation, models for estimation of FCR and ER at the meso (link-level) and micro (second-by-second) scales are required. Meso-scale models provide an overall picture of eco-friendliness of the driving operation per link (or road segment), whilst the micro-scale models focus on understanding instantaneous correlations between the state of the vehicle and the FCR and ERs. Although much more difficult to estimate, micro-scale models provide richer information for analysis and they could be converted to lower-resolution models (even to macro levels) through aggregation and extrapolations over the whole fleet.

Using estimates of power as proxy variables, multi-variate linear/nonlinear regression, and basic *Machine Learning* (ML) solutions [12–19] are examples of techniques used for micro-/meso-scale modeling fuel consumption and emissions since the mid-1990s. However, there exist seven major issues with the available models usable for eco-driving. First, the majority of previously proposed models rely on simulated lab data instead of on-road measurements [14,17,20,21]. Earlier in this section, the drawbacks of dependence on lab simulations are explained. Second, there are many studies and models which have used a low-diversity test fleet [14,16,18,22–28]; as a result, the generalizability of their conclusions is questionable. Third, traditional statistical modeling techniques are incapable of capturing complex non-linear dependencies corresponding to fuel consumption and emissions generation mechanisms. Fourth, many of the existing models depend on *Internal Engine Variables* (IEVs) to achieve acceptable accuracy levels [17,18,21,27,29–34]. Such dependence eliminates the models' usefulness for developing universally-applicable and instrument-independent eco-driving services. It avoids their incorporation into traffic simulation software for conducting environmental assessment of different scenarios as well, as no knowledge regarding IEVs is retrievable from traffic microsimulation. In other words, there is a need for models founded upon smartphone sensors' readings (such as GPS, accelerometer, and gyroscope measurements) or simple kinematic instantaneous variables (such as speed, acceleration, and road grade) extracted from simulation software instead of IEVs recorded using specialized logger devices. Regarding eco-driving, drivers' smartphones are the only means of providing such service for many vehicles moving on roads which are still not equipped with on-board eco-driving systems while having a considerable share of current vehicle fleet. Fifth, some of the popular comprehensive energy and environmental assessment models such as MOVES [35] and *Comprehensive Modal Emission Model* (CMEM) [36] require heavy processing steps that do not allow real-time use. Sixth, the serially-correlated nature of instantaneous FCR and ER measurements is rarely addressed in the literature. The temporally extended impact of past driving events (a few seconds before time t) on current FCR and ERs is either left unobserved [14,27,29,32–34], acknowledged but disregarded and considered effectless [37], or at best is spread through time as an error using moving average techniques to improve instantaneous predictions [17,18]. Seventh and the last, majority of the existing microscale models are developed with a vehicle-oriented approach [14,16,17,25,37–39], while more general models (for instance, models that work acceptably for categories of vehicle rather than particular ones) are preferred to expand their domain of use cases.

This research work aims to provide some insights in each of these literature gaps. For this purpose, a large set of field experiments is conducted using vehicles representative of the most popular vehicles in Montreal (and other selected regions) to provide sufficient and proper data for evaluating existing popular models such as MOVES for use in non-U.S. regions as well as developing methodologies for estimation of alternative localized, fleet-specific, microscale fuel and emission models for Canada. Through the steps taken, the seven aforementioned issues with the existing models are considered. By using state-of-the-technology portable fuel and emission measurement sensors, on-road tests are conducted to collect as realistic as possible data instead of using lab simulations. Through testing on 35 passenger cars, increasing the diversity of data to boost the robustness of analyses and inferences is targeted. By utilizing powerful ML techniques, difficult-to-formulate and nonlinear dependencies between input variables and dependents are captured. Simple kinematic variables such as speed, acceleration, and road grade (or alternatively, GPS altitude) are used as input variables to eliminate models' dependency on IEVs. Straightforward, light and easily deployable procedures are defined to avoid computationally expensive execution processes of models. Special ML algorithms are used for addressing the serially-correlated nature of vehicular fuel consumption and emission generation. Finally, forecast combination methods are utilized to develop general category-specific models on top of vehicle-specific ones.

1.2 Motivations

Heedless use of MOVES in Canada despite the obvious signs of its unsuitability for use in non-U.S. regions (based on its core characteristics as well as validation studies conducted in other countries such as Mexico, India, and China [40–42]) is a matter of concern. Founding environmental assessment of transport projects or eco-driving guidance on erroneous forecasts causes irreparable damages to the environment in long term. Thus, it is important to make the transport consultants and authorities in Canada aware of the necessity of either taking additional adjustment steps or alternatively transitioning to locally-estimated energy and emissions models to come up with refined and more realistic interpretations and conclusions in their projects.

When rejecting an existing tool, it is crucial to propose alternative solutions. By making all fuel consumption and emission modeling efforts in this research, it is hoped that avenues be opened to the development of smartphone-based real-time eco-driving services usable by passenger car fleet (especially, the older vehicles which are the major energy consumers and pollutant emitters, while they are mostly not equipped with on-board eco-driving options), trucking industry, and even the public transit. In addition to eco-driving, microscale fuel and emission models could be incorporated into traffic microsimulations for environmental assessment of different transportation/urban development projects before or after the implementation.

1.3 Objectives

The main objectives of this study could be summarized as follows:

- Conducting a comprehensive literature review over the history of research about vehicular fuel and emissions modeling at different scales and identifying the gaps and major shortcomings of previously-proposed modeling methodologies.
- Introducing and practically evaluating on-road data-collection methodologies using state-of-the-technology portable activity, fuel, and emission sensors as an alternative to in-lab experiments.
- Evaluating the magnitude of error as well as the bias direction when using MOVES model for estimating project-level (link-level) energy and emissions rates in Canada (or other non-U.S. regions). In addition, assessing the sensitivity of MOVES output to fleet attributes as well as road network and environmental conditions to discover the impacting regional aspects which need to be addressed in developing adjustment procedures or even independent local models.
- Reducing the complexity of the vehicular fuel and emission models and increasing their practicality by limiting their feature set to instrument-independent and simply-retrievable variables. Moreover, proving that the ML techniques, if used in an appropriate combination, are capable of shouldering the burden of extracting the hidden impact of IEVs (which usually possess significant correlations with FCR and ERs) without the need to directly including IEVs in the models' feature sets.

- Introducing proper ML methodologies for addressing the serial correlation and the lagged impact of features on instantaneous FCR and ERs. Furthermore, building up a realistic foundation for developing category-specific models using well-recognized forecast combination techniques. The goals would be achieving an additional level of independence as well as expanding models' use cases (as with generalized models, there would be no need for detailed information about the technical specifications of vehicles for predicting FCR and ERs anymore).

1.4 Structure of the Thesis

This thesis is structured as follows: In Chapter 2, first, the most recent pieces of research conducted on evaluating the accuracy of MOVES output and the attempts on developing local models with the help of MOVES are reviewed. Then, the history of research about vehicular microscale fuel consumption and emissions modeling, from the use of traditional statistical methods to the rare recent attempts on using ML techniques, is briefly reviewed.

The experimental procedure and the proposed methodology of research for the MOVES validation as well as microscale fuel and emission modeling are explained in Chapter 3. To be more precise, this chapter elucidates data requirements of the research, data collection process and the equipment, preparation of field-measured data, execution of different MOVES scenarios, and the novel, simple, but efficient ML-based procedures introduced in this research for modeling local and fleet-specific fuel and emission models. In Chapter 4, the results of the validation study over MOVES predictions are presented and discussed. Then, in Chapter 5 and Chapter 6, outcomes of alternative modeling ideas (a cascaded ML modeling approach as well as a multi-stage forecast combination procedure for improving vehicle-specific *Recurrent Neural Network* (RNN) models and generalizing them to categories) will be visually and statistically discussed. Several sensitivity analyses are conducted in these chapters, mainly based on the fleet, region, and road attributes. Finally, in Chapter 7, conclusions are drawn and possible future research topics following this study are explained.

Chapter 2

Literature Review

2.1 On the Validity of MOVES predictions

MOVES is the U.S. *Environmental Protection Agency's* (EPA) Motor Vehicle Emission Simulator. It is used to create emission factors or emission inventories for both on-road motor vehicles and non-road equipment. The purpose of MOVES is to provide an accurate estimate of emissions from cars, trucks, and even non-highway mobile sources under a wide range of user-defined conditions.

In MOVES modeling process, the user specifies vehicle types, time periods, geographical areas, pollutants, vehicle operating characteristics, and road types to be modeled. The driving profile including second-by-second changes of speed and grade should be provided to the model as well. The model first calculates the second-by-second *Vehicle Specific Power* (VSP) based on the corresponding vehicle operating parameters. Using a combination of calculated VSPs and the vehicle speed, the calculation process then finds the appropriate operating modes from an operating mode bin table. This is followed by another table lookup with the vehicle emission rate table based on operating mode, vehicle type, and age. The emission and fuel consumption are accumulated and corrected with base cycle emission rates calibrated previously using MOVES [35,43].

A few studies have tried to validate MOVES output either at national/county levels or for the project-scale scenarios for use in non-U.S. regions or the conditions not included in MOVES development experiments. The scarcity of such studies, despite the passage of more than a decade since the public availability of MOVES, could be the result of over-trusting the model's generalizability and unawareness about the significant sensitivity of MOVES predictions to regional/case-specific conditions.

Findings of Chinese researchers, through a sensitivity analysis over MOVES outputs concerning model year, age group, and speed intervals showed that despite describing the relative changes between categories well, the absolute values of MOVES estimations are noticeably different from the ground-truth (such differences are linked to the regional regulations and fuel quality) [44].

The NOx emissions from vehicles in the Environmental Protection Agency's *National Emissions Inventory* (NEI), prepared using MOVES, are found to be overestimated by up to a factor of two [45,46]. Following the discovery of such error, EPA's Review Workgroup conducted a study focusing on NOx emission by *Light-Duty Vehicles* (LDV) using Inspection/Maintenance (I/M) test data which showed that MOVES estimates are higher than I/M data for pre-2000 LDVs while lower for 2010+ LDVs [3,47]. MOVES NOx estimation error of up to 24% is confirmed by EPA. The overpredictions are assumed to be linked to multiple compounding errors such as the time-space allocations of mobile emission sources.

Distrustful of solely using MOVES for local vehicular emission estimation, the study conducted in Beijing, China [48] used MOVES to fill the gaps that a locally developed model (using limited test data) was unable to cover, particularly, for ERs corresponding to high-power operating modes. Although still partially dependent on MOVES, this study was one of the first efforts in adjusting MOVES output for local use.

Near-road ambient air pollutant data collected in urban sites in Texas [49] revealed that despite the inclusion of local characteristics such as wind, MOVES consistently underestimates ambient CO and NOx ratios, a finding in total alignment with that of Fujita *et al.* [46]. Up to 24% difference in estimated rates was observed even when best available local data was used as MOVES input. Nevertheless, the use of stationary ambient emissions sensors for validating mobile-source emission predictions (in other words, conducting a non-homogenous comparison) introduces a level of error to the conclusions.

The dramatic difference observed between the true CO, *Hydro Carbons* (HC), and NOx rates and MOVES predictions for Indian cities, where MOVES underestimated the emissions around 9 times lower, encouraged researchers to revamp the model for Indian cities using alternative driving cycles reflecting the dominant driving conditions and traffic in India [41]. Their study confirmed that countries are not only different in terms of fleet distribution and general traffic and driving conditions, but also the deterioration rate of vehicles is different which disrupts MOVES assumptions and coefficients when including the impact of vehicle age on energy consumption and emissions.

A recent study conducted in mountainous parts of China questions the validity of MOVES predictions for use in similar regions [42]. It has been observed that within each of the MOVES operating modes, defined based on VSP ranges, statistically different emission levels are generated across different road grades. The study confirms the shortcomings of operating-mode methodology in correctly capturing influential factors that exclusively exist in non-U.S. regions.

The validity of MOVES output has been a matter of concern in literature for both light-duty as well as heavy-duty vehicles. An assessment of MOVES predictions for heavy trucks' FCR [50] showed that the model overestimates the rate during deceleration and cruising while producing inconsistent outputs for acceleration. Whether a similar condition applies to light-duty vehicles as well is not yet assessed in the literature.

Updated evaluations conducted by EPA researchers to particularly assess the impact of vehicles' age on accuracy of MOVES NOx estimates at the national level, report an increasing level of overestimation projected to reach 21% and 30% in 2028 and 2045, respectively [51]. Based on this report, no adjustment solution at any scale is still proposed by EPA.

In one of the most recent attempts on adjusting MOVES for special conditions, high-resolution GPS data is used to adapt the VSP formula for requirements of an intersection control study, where emissions need to be analyzed at the microscopic level and separately for acceleration, cruising, deceleration, and idling parts of turn/through movements [52]. Taking such an approach, MOVES error in estimating CO2 and NOx emissions reduced by around 40% and 45%, respectively.

Finally, on-road fuel consumption and emissions measurements, conducted using portable emission sensors in Montreal [53], as a part of a traffic-control impact study, showed that MOVES underestimates fuel consumption and CO2 rates by an average factor of two, while over-estimating NOx emissions significantly (predictions equal to one-fifth of true observations are reported). However, the test fleet of only three vehicles in this study is too small for concrete conclusions.

The interesting perception after reviewing the literature regarding the validity of MOVES predictions is the diversity of findings. Some studies have reported overestimation, while some others have observed opposite biases for particular emissions. This emphasizes the importance of evaluating MOVES output for use in new regions, in special conditions (such as microscopic traffic studies), and for vehicles with specific attributes.

2.2 A Transition from 0D/1D Simulation to Data-driven Modeling

In addition to the commercial models capable of estimating vehicular energy and emission rates at different scales, there have been several attempts in the literature targeting exclusively the development of microscale FCR and ER estimation modelling.

A major part of studies focus on 0D/1D simulation, which delve into detailed mechanical, thermodynamic, and chemical interactions inside the engine. Use of traditional statistical modeling techniques such as multivariate linear/nonlinear regression is another significant segment in the literature. Finally, in recent years, there has been a divergence towards use of emerging ML techniques such as *Support Vector Machines* (SVM) and *Artificial Neural Networks* (ANN). A well-known classification of models proposed in the literature is done based on their transparency by which there would be three categories of white-box, grey-box, and black-box models [54,55].

White-box models (known as first principle models or 0D/1D simulations) are highly deterministic and developing them requires a thorough understanding of underlying mechanical, thermodynamical, and chemical mechanisms that affect energy consumption and emission generation [32,56–64]. Even though the first-principle models are the most complete information source, they are too complex for real-time use. Besides, they require access to IEVs which makes them instrument-dependent. Such a characteristic limits the model deployment to offline eco-driving studies or at best, to onboard eco-driving services available on some vehicles. Thus, the model will not be applicable in third-party smartphone-based eco-driving services which solely rely on GPS data and user-provided information. Moreover, although providing accurate predictions, the sensitivity of the performance of white-box models to the accuracy of features is a matter of concern.

Combining the 0D/1D perspective with data-driven modeling is a trick to avoid the complexity of physics-based emission models for real-time use by correcting data-based estimations using engine readings as a complementary step. Such an approach is known as grey-box modeling as well. For instance, Falcone *et al*. [65] combined mechanical and thermodynamics attributes of torque generation inside the engine with data collected through the *New European Driving Cycle* (NEDC) test procedure to develop their combustion model for diesel-engine vehicles. In another study, Hirsch *et al*. [55] took a similar approach to estimate emissions. In their model, the *Zeldovich*-formation principle [66,67] takes the impact of temperature on NOx and PM formation into account, while the main emissions estimation is handled by a data-based model including other impacting factors.

Alternatively, cascaded techniques are utilized in the literature to develop less-complex models while still considering internal engine interactions in modeling. For instance, Frey *et al*. [18] calculated equilibrium concentrations of oxygen and nitrogen during NOx formation in the combustion chamber using the *Zeldovich* model based on IEVs and used the resulting estimates to predict the NOx rate as they were highly correlated with the target ERs.

Contrary to first-principle models are the black-box models (data-driven or input-output models) which lack physics or chemistry in their structure. They require large quantities of data to compute the optimal model parameters. As a drawback, they are only valid for the range of observations and have unknown extrapolation properties. Thus, to be valid for eco-driving, they require large training datasets covering all the dominant driving situations.

Multi-variate nonlinear regression has been the most popular traditional statistics method used in black-box modeling of vehicular energy consumption. However, capturing the nonlinearity while keeping the model simple is a challenge when training multi-variate regression models. Rakha *et al*. [17,60] and Saerens *et al*. [21] both focused on engine power as a proxy variable for estimating FCR. Relying on almost noise-free in-lab observations and incorporating RPM and other internal engine parameters, they used simple polynomial functional forms in the development of their models. Duarte *et al*. [68] used *Vehicle Specific Power* (VSP) as a proxy variable. Although providing a good fit using a piecewise polynomial functional form, the signs of biased experiment settings (in fleet diversity, selection of route and time of the experiment, etc.) are visible in their observations.

2.3 First Steps into Machine Learning

Stepping into ML modeling in the past few years, simple neural network architectures are widely used in literature to estimate FCR and ERs for the cold-start, hot-start, and hot-stabilized engine conditions [26,39,69,70]. Parlak *et al*. [71] used a single-layer perceptron to predict specific fuel consumption of a non-vehicular diesel engine.

However, the use of accurate readings of internal engine variables as input compensated for the simplicity of the ANN architecture they used. Amer *et al*. [19] used a wider single-layer ANN architecture (with more activation nodes in the hidden layer) to estimate fuel consumption for an eco-routing purpose. Nevertheless, their variable set included only two dynamic variables (speed and distance) and three constants which make their modeling results questionable. Logically, distance cannot affect engine operation and constants do not help the model training. Çapraz *et al*. [33] tried to evaluate ANN and *Support Vector Regression* (SVR) techniques to model FCR. They relied on data from on-road experiments conducted in highways using three test vehicles. Incautious selection of input variables (such as distance traveled, latitude, and longitude) is a common mistake in statistical modeling repeated by Wickramanayake and Bandara [72] who compared the prediction power of the random forest, gradient boosting, and neural networks methods in fuel consumption estimation.

A common drawback of these models is disregarding the assessment of deeper neural networks as well as more sophisticated ML techniques capable of capturing serially correlated and lagged effects. A popular method of dealing with lags and autocorrelation in literature has been time alignment of input data beforehand [25,73], a method utilized for preprocessing the data used for estimating sub-models of MOVES model. Nevertheless, the approach arises criticism as the lagged effects might not occur with a constant order.

The time-series forecasting of the vehicular FCR and ERs is a challenging problem due to the dynamic nature of data [74–76]. Additionally, volatility in features leads to increased forecasting error and when combined with lagged effect, the majority of the traditional statistical modeling techniques fail to perform acceptably. RNNs are gaining renewed interest among researchers as they provide promising results for modeling time-series and serially-correlated phenomena [77–79]. The study conducted by Kanarachos *et al*. [16] is the most recent and a rare example of using the RNN technique for modeling the instantaneous FCR in the literature. Although achieving outstanding prediction scores, their modeling procedure is founded upon data collected from a single vehicle prone to bias which makes the generalization of the developed model questionable. To the best of our knowledge, no research has been done yet in utilizing RNNs for ER modeling.

Chapter 3

Methodology

3.1 Data Requirements

Based on the defined objectives of the research (in Section 1.3), two categories of data from vehicles are required. First, data regarding real-world fuel consumption and emission generation in the finest possible units of time or distance should be collected. If time is selected as the basis, the measurements could be converted to the distance basis and vice versa. In this study, time (per-second basis) is selected; however, conversions will be conducted when needed. As the gasoline-engine vehicles are chosen to be studied, collecting data regarding their dominant emissions [80,81], including *Carbon Dioxide* (CO2), *Nitrogen Monoxide* (NO), *Nitrogen Dioxide* (NO2), and *Particulate Matters* (PM), is targeted.

To execute MOVES scenarios (for validation of its outputs) and to develop microscale fuel and emission models, complementary data, including important IEVs in addition to positional state of the vehicle, in a second-by-second manner is required as well. *Fuel Consumption Rate* (FCR) will be calculated with the help of engine variables, while positional variables provide information regarding instantaneous speed, acceleration, and road grade will be obtained.

3.2 On-road Experiments

Thirty-five different passenger cars in total were selected for on-road experiments in three cities of Montreal (Canada), Bucaramanga (Colombia), and Tehran (Iran). The selection criterion was the popularity of different makes and models in each region to keep the test fleet as a representative of the local fleet. Moreover, although Canadian cities such as Montreal were the main target, tests in Colombia and Iran were conducted due to having physical access to those regions as well as our desire to increase the diversity of experiment conditions (as a requirement for achieving the study goals) in terms of fleet, weather, network topology, etc. In total, 6493 minutes of on-road data (by traversing 2350 kilometers) was collected through the experiments conducted in three cities.

3.2.1 Set of Sensors and Equipment

On-Board Diagnostics (OBD-II) loggers, shown in Figure 1, are installed on all the vehicles to collect engine-state parameters. Instantaneous GPS coordinates and accelerometer measurements are added to the records in real-time using a tablet that wirelessly receives and saves the *Engine Control Unit* (ECU) data. The logged OBD-II parameter set includes *Engine Speed* (RPM), *Manifold Absolute Pressure* (MAP), *Mass Air Flow* rate (MAF), *Barometric Pressure* (P), *Fuel-Air Equivalence Ratio* (ϕ), and *Intake Air Temperature* (IAT). Engine parameters are monitored in real-time to better understand the microscopic reactions of engine operation to the driving behavior, road, and environmental conditions through modeling.

(a) (b)

Figure 1: (a) Wireless OBD Scanner and (b) OBD-II port under the steering wheel

In recent years, the advent of *Portable Emissions Measurements Systems* (PEMS) has dramatically changed the state of on-road emissions monitoring as more and more studies across the globe are being conducted using these light-weight, small-size, and ultra-portable sensors [24,82]. A state-of-the-technology PEMS is installed on the tailpipe of 17 vehicles under study in Montreal to monitor the instantaneous CO2, NO, NO2, and PM concentrations. CO2 measurement is done using *Non-Dispersive Infra-Red* (NDIR) absorption technology with a measurement range of 0-20% and an accuracy of ±70 ppm. For NOx, 3-electrode electrochemical sensors are incorporated in the unit capable of measuring up to 5000 ppm for NO and 300 ppm for NO2. The measurement resolution for NO and NO2 are 1-5ppm and 0.1ppm, respectively. Regarding PM, the unit measures undiluted emissions through the response of three dissimilar particulate sensors. Ionization is used for ultra-fine/fine particulates usually between 0.01 to 1 micron, while a combination of opacimeter and laser scattering is deployed for coarse particulates up to 10 microns. Figure 2 shows the details of the PEMS installation on one of the test vehicles.

An intake probe is clamped into the tailpipe which collects exhaust samples at a 2.5 $liters/minute$ rate with the help of an internal pump. As there is no dilution, no extrapolation of the sensor values to the full concentrations is required. A chiller unit condenses and removes the water vapor present in the exhaust. An additional water trap completes the water-removal process before sending the sample to the main unit.

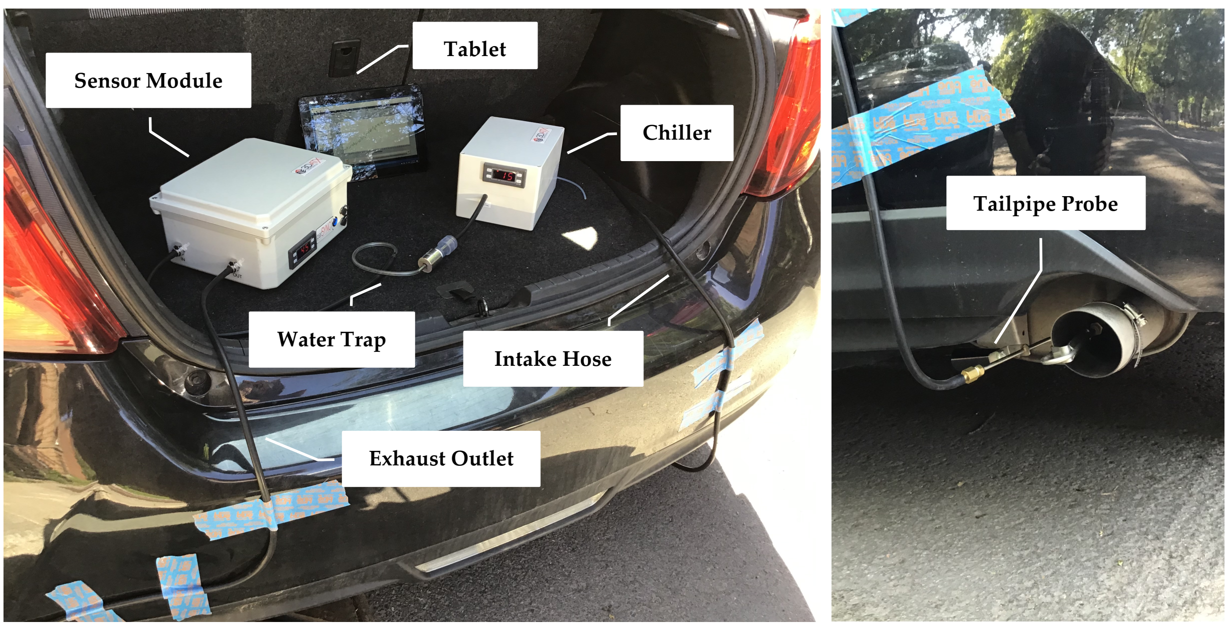

Figure 2: Details of the PEMS setup on a test vehicle

3.2.2 Planning for Experiments

Cold-start emissions are disregarded in this study as the sensing process was started after reaching the hot-stabilized engine operation. Pre-test and post-test *Zero-out* processes are conducted for all the field experiments to capture the ambient emission levels as a reference for calculating the net emission concentrations.

The maintenance quality of the vehicles is evaluated through short interviews with the volunteer car owners participating in the experiments and the vehicles with uncertain/unacceptable conditions are ignored and not included in the tests. A single person drove all the cars in each of the three cities, while the three drivers are coordinated in advance in terms of driving style to mitigate the chance of bias in collected data. The drivers are all asked to avoid aggressive driving and keep their speed coordinated to that of traffic flow. It is noteworthy that the impact of the traffic flow and traffic control systems would be implicitly captured in time-series logs of speed and acceleration.

A driving plan is set in advance to guide the drivers to drive approximately 30% of the time in highways, 30% in arterials, 30% in local roads, while dedicating the remaining 10% of the time to uphill and downhill driving. The equivalent distance for the above shares varies in each test. However, as the FCR and ERs are time-based, time is selected as the reference for scheduling the trip chains' plans. For the 10% uphill and downhill driving time-window, special road segments with grades beyond normal road design thresholds (higher than 7% or less than -7%) are targeted. Taking the aforementioned approach, the randomness of data in terms of speed, road grade, and diversity of acceleration/deceleration patterns is preserved.

3.2.3 A Visual and Statistical Look at the Experiments

Figure 3 represents the aggregated view of GPS trajectories for experiments conducted in Montreal and Bucaramanga, in addition to a part of experiments conducted in Tehran.

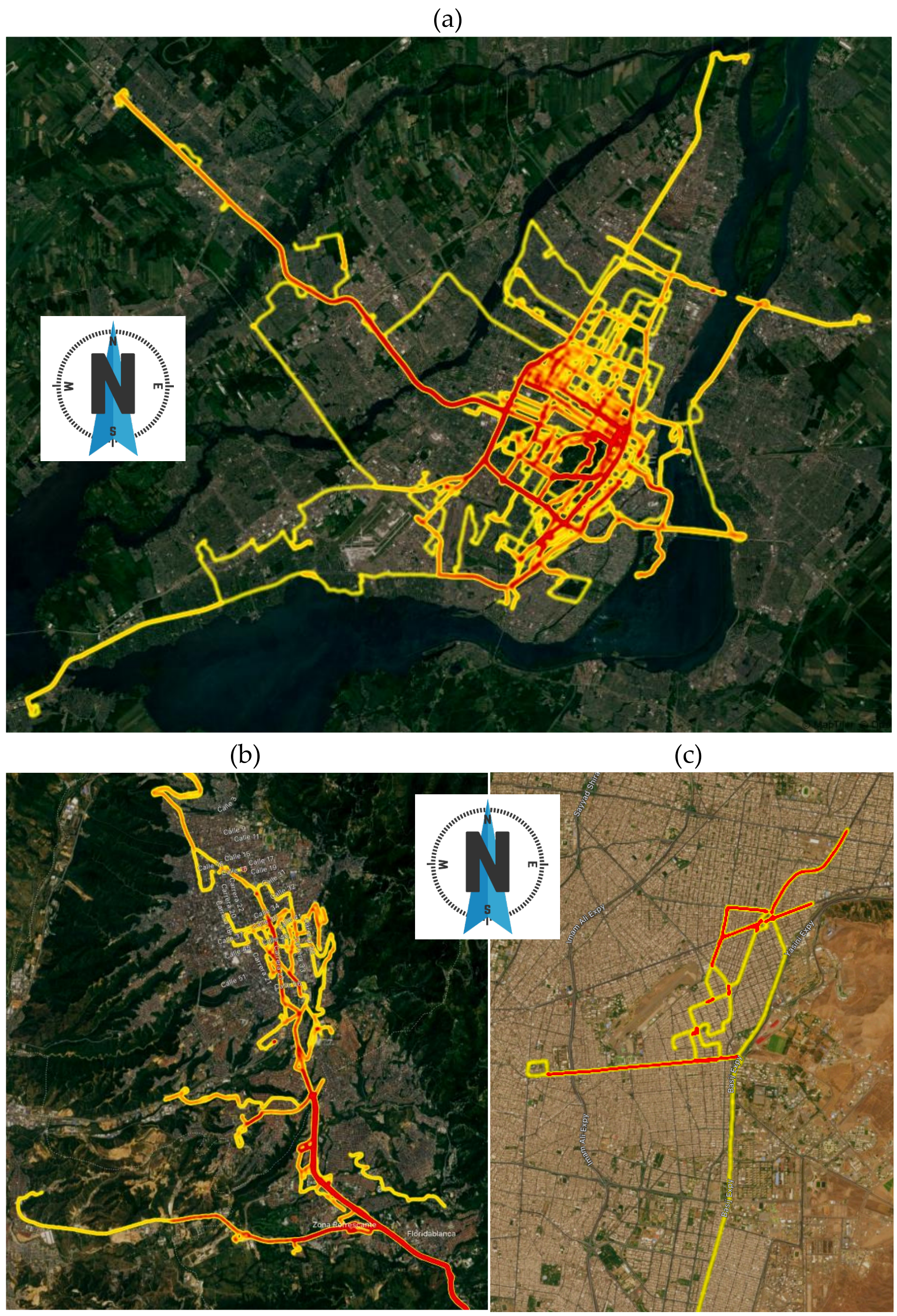

Figure 3: Trajectories of experiments in (a) Montreal, (b) Bucaramanga, and (c) Tehran

Figure 4 shows histograms of link average speeds and grades (slopes) for the road experiments performed in three cities. Moreover,

Table 1 and Table 2 provide descriptive statistics regarding experiments and the collected data.

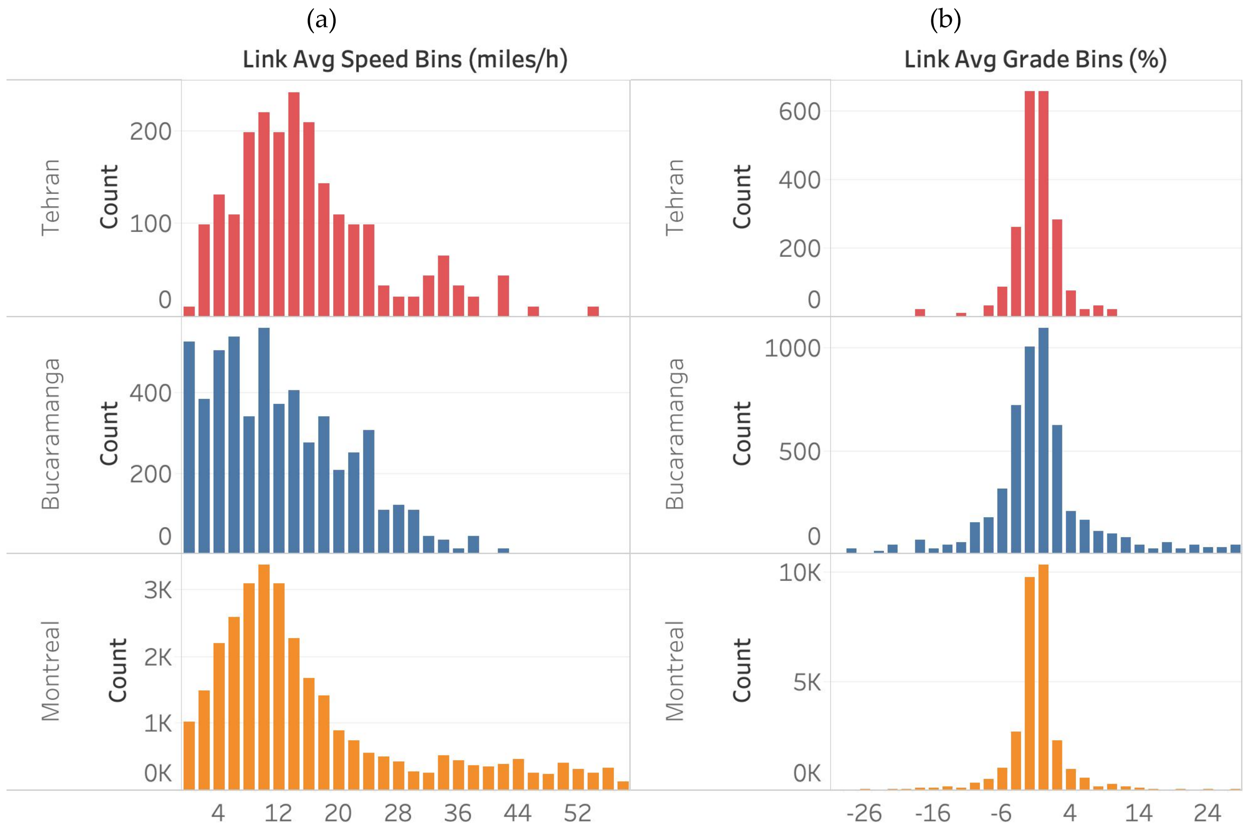

Figure 4: Histograms of (a) average link speeds and (b) average link grades

Table 1: Summary of the field experiments

Attribute	Numbers and Quantities		
	Montreal	**Bucaramanga**	**Tehran**
Total Trip Length (km)	1804	291	255
Total Trip Time (Minutes)	5224	825	444
Number of Test Vehicles	22	7	6

Table 2: Characteristics of the test fleet

Attribute	Category	Quantity
Vehicle Segments	Compact SUV	7
	Subcompact SUV	1
	Midsize Sedan	7
	Compact Sedan	9
	Subcompact Sedan	3
	Compact Van	1
	Compact Hatchback	1
	Subcompact Hatchback	6
Engine Types	Regular	31
	Turbocharged	4
Transmission Types	Manual	6
	Automatic	19
	Dual-Clutch (Auto)	2
	CVT * (Auto)	8
Vehicle's Age (years)	-	0 ~ 13

* Continuous Variable Transmission

3.3 Data Preparation

As the raw data logged by the set of sensing devices includes minor outliers (due to sensor errors) and some missing values (due to instantaneous malfunctioning of sensors or analog data-link interruptions), and more importantly, because there is a need for converting measurements to desired units and scales, a post-processing procedure is applied on raw measurements.

3.3.1 Outlier Filtering and Smoothing

The most significant number of outliers are observed in GPS altitude measurements, mainly caused by temporary loss of signals from GPS satellites when the vehicles passed through the tunnels or the underpasses (less than 1% of the duration of each experiment). To treat them, the Savitzky-Golay smoothing algorithm is used [83]. It generates more satisfying results compared to other available algorithms such as moving average, exponential, and convolutional smoothing methods. Note that wheel speed retrieved from ECU is prioritized over GPS speed due to relatively higher accuracy. Hence, no outlier filtering (and smoothing) is applied to instantaneous speed data. Figure 5 demonstrates the results of applying the Savitzky-Golay algorithm on raw GPS altitude data.

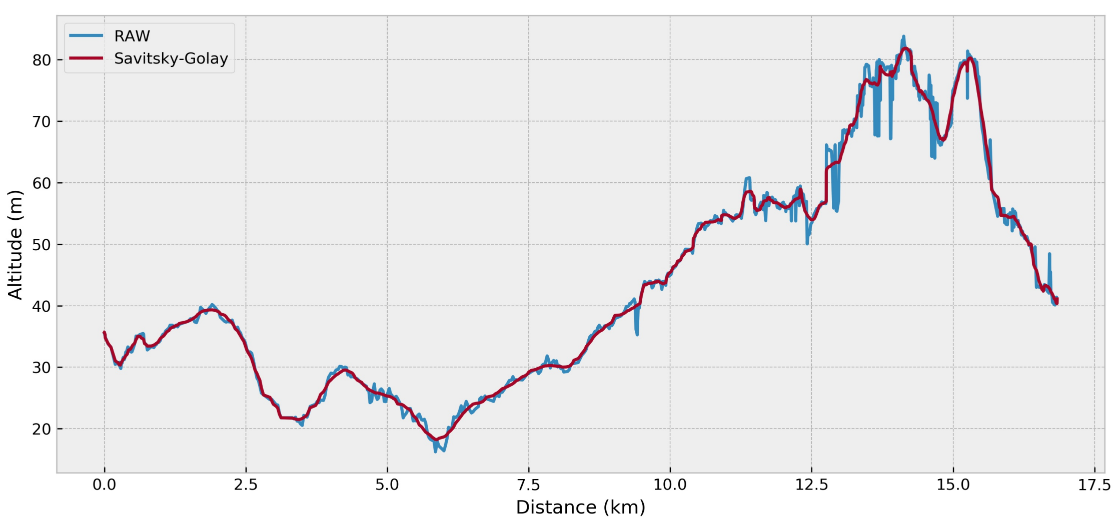

Figure 5: Efficiency of the Savitzky-Golay smoothing algorithm shown for a fraction of an experiment

Although rare, missing values should be fixed in data either by removal of the whole corresponding record or by substituting them with appropriate values. As in this study, the data is assumed to be time-series and one of the main goals of the study is addressing the serial correlation in data, no record is eliminated. Instead, missing values are filled out by the average of the neighboring data points. The process is applied to variables measured by both the OBD-II logger as well as the PEMS measurements.

3.3.2 Measure Extraction

Internet-synced timestamps were used to align and join data tables provided by the sensors. As almost no vehicle is equipped with direct fuel flow metering (using flow meters between the fuel pump and the engine), FCR has to be calculated indirectly using other internal engine observations. It could be done with the help of observed (Eq. 1) or estimated (Eq. 2) value of MAF, representing the flow of air entering a fuel-injected internal combustion engine. All modern vehicles are equipped with a MAF sensor, but not all of them report the MAF level through the OBD-II interface. In that case, MAF could be roughly estimated based on MAP as well.

$$FCR_t = \frac{MAF_t}{\lambda * AFR_{stoich}} \tag{1}$$

$$MAF_t = \frac{RPM}{120} * \frac{MAP_t}{IAT_t} * \frac{VE}{100} * ED * \frac{MM}{R} \tag{2}$$

In Equations 1 and 2, index t indicates the instantaneous nature of observations, FCR and MAF are both in g/s, and AFR_{stoich} denotes the air-to-fuel mixture ratio at the stoichiometric level. λ is the ratio of the actual air/fuel ratio (AFR) to its stoichiometric level [84], RPM is in revolutions per minute, MAP is the pressure at the intake air manifold in kPa, and VE is the volumetric efficiency, which is around 65% for regular gasoline engines and goes up to 85% and even higher for turbocharged models. ED denotes engine displacement in Liters, MM is the average molecular mass of air (28.97 grams/mole), IAT is the intake air temperature in Kelvin, and R is the ideal gas constant equal to 8.314 $J/°K/mole$.

Energy Consumption Rate (ECR) predicted by MOVES is reported in temporal or distance-based units of J/s or $J/mile$. Thus, Equation 3 is used to convert FCR observations to proper energy units. Having second-by-second temporal ECRs in hand, the distance-based rates are calculated by aggregating temporal rates and dividing the result by the corresponding distances.

$$ECR_t = FCR_t * 31{,}536{,}000 * Eff_{combustion} \tag{3}$$

$$Eff_{combustion} = \begin{cases} 100\%, & Regular\ Gasoline \\ 96.7\%, & E10\ Gasoline \end{cases}$$

Each liter of regular gasoline generates approximately 31,536,000 joules of energy. However, the combustion efficiency is different for other gas formulations. E10 gasoline (which has 10% ethanol content) has lower combustion efficiency, whereas it generates less pollution and is eco-friendlier. All the gas stations used for fueling up the test vehicles in Montreal provided E10 fuel. But, in Tehran and Bucaramanga, regular gasoline is used during the tests.

The PEMS setup reports instantaneous emission concentrations in percentage for CO2, ppm for NOx, and $\mu g/m^3$ for particulate matters. But the second-by-second concentrations should be converted to temporal rates. The challenge with this conversion is the absence of exhaust flow rate data. But with an all-in all-out assumption (ignoring the existence of minor leakage from the engine or the exhaust pipe), the MAF rate could be used as an alternative to the exhaust flow rate. The exhaust pipe lag, however, could introduce errors to the calculations. This lag refers to the time it takes for the exhaust to traverse the exhaust pipe from the engine to the tailpipe. Presence of resonator and catalytic converter in addition to the length of the pipe influence on this lag. Later in this research, different modeling approaches (by including lagged variables in the models or by use of sequence modeling techniques) are described and evaluated as solutions for capturing such lagged effects. To lessen the impact of such error only for use in the MOVES validation, a 6-second moving average of the MAF rate is considered instead. The 6-second threshold is selected based on the longest time lag observed when visually comparing engine RPM and emission fluctuations. For this purpose, peaks and valleys are traced for randomly selected time-windows from all 17 vehicles included in emission tests.

Equations 4 and 5 are used first to unify concentration units and adjust the concentrations for prevailing temperature and pressure. The concentrations are then converted to instantaneous emission rates using Equation 6.

$$For\ CO2\!:\ Conc_{ppm} = 10^{6} * Conc_{\%} \quad (4)$$

$$Conc_{mg/m^3} = Conc_{ppm} * \left(\frac{Molecular\ Weight\ of\ Gas}{22.4}\right) * \left(\frac{273}{273+T}\right) * \left(10 * \frac{P}{1013}\right) \quad (5)$$

$$ER_t = Conc_{mg/m^3} * 10^{-6} * MAF_t * \left(\frac{10^{-3}}{Air\ Density}\right) \quad (6)$$

In Equation 5, T is the intake air temperature in $°C$ and P is the ambient barometric pressure in kPa. In Equation 6, MAF_t is the mass air flow in g/s and the air density is equal to 1.2929 kg/m^3. The molecular weight of emissions is 44.01, 46.01, and 30.01 g/mol for CO2, NO2, and NO, respectively.

3.3.3 Feature Scaling

During the modeling stages, *Mean Normalization* is applied for feature scaling on five target features of speed, acceleration, grade, GPS altitude, and RPM as well as the dependent variable (either the FCR or one of the ERs). The features will be rescaled so that they will have the properties of a standard normal distribution. Feature scaling is recommended in ML to avoid attributes in greater numeric ranges (such as speed or RPM) dominating those in smaller numeric ranges (such as acceleration and grade).

Furthermore, feature scaling speeds up the gradient descent convergence during the training process of the ML models, especially when the data has high variance.

3.4 Executing MOVES Scenarios

MOVES is developed based on the MOBILE6 model to satisfy several use cases such as national and local inventory development, project-level analysis, and sensitivity analysis or policy evaluation in travel-demand or microscopic traffic studies. MOVES is founded upon several sub-models that are assigned to categories of source bins and operating mode bins of the fleet under study. Energy and emission rates are then assigned to each category by a direct lookup in predetermined databases.

In this study, a total of 214 hourly MOVES scenarios are set up and executed for all the 35 vehicles. The simulation is conducted at the project level domain. Regarding the geographical bounds, Oklahoma county, Los Angeles county, and Chittenden county in Vermont state are chosen to represent Tehran, Bucaramanga, and Montreal, respectively, as they had the most analogous meteorological conditions to the target cities. Hourly changes of ambient temperature and humidity are considered in scenarios based on weather records of the target cities [85].

A reverse-geocoding process using Google API is conducted for converting GPS coordinates to road names and define links. The speed and grade profiles of each link in addition to aggregate attributes of them are then injected to MOVES for having distance-based link-level energy and emission rates as a result.

The age of test vehicles is included in scenarios. It affects the choice of engine technology by MOVES. Furthermore, regular gasoline (and its corresponding chemical formulation) is input to the MOVES for tests conducted in Tehran and Bucaramanga while ethanol-based (E10) gasoline is considered for Montreal.

Regarding the emissions, cold-start, and crankcase emissions are disregarded and only the running exhaust (with warmed-up engine) is considered for MOVES simulation. As the equipment is designed for tailpipe emission testing, brake wear and tire wear particulate matters are ignored from the MOVES output options as well.

3.5 Cascaded Machine Learning Modeling

3.5.1 Proposed Feature Sets (Model Structures)

To initiate the modeling procedure in this study, the relationship of FCR and the parameters simply retrievable by a smartphone such as instantaneous speed (v), acceleration (a), and road grade (g) is in focus. Note that for the modeling, wheel speed reported by ECU and acceleration from the tablet/smartphone accelerometer sensor are used instead of GPS speed and GPS-based acceleration due to higher accuracy and reliability. At the model deployment time, GPS variables could be used, although they will add minor errors to the model prediction. For the instantaneous road grade calculation, the most precise available source of information is GPS altitude measurements, in combination with wheel speed, and wheel diameter of each car. Such a procedure is easily reproducible at the deployment stage of the model. Other sources such as Google's Elevation API are available for retrieving road elevation at any desired point as well; however, they are not as accurate as GPS-based calculations, based on an exclusive assessment conducted as a part of this study on around 300 random GPS points in Montreal as well as published articles in this regard [86].

Binary and triple combinations of features (including speed, acceleration, and grade), 6 in total, are examined as model inputs. The acceleration is instantaneous in nature; however, the changes in speed and road grade occur less abruptly. It is suspected that the changes in speed and road grade have an extended temporal impact on FCR. To evaluate this hypothesis, lagged speed and grade features (1^{st} order) are added to the feature set. It is noteworthy that the lagged speed and grade variables are calculated ahead of the random sub-setting of samples.

Two widely used ML techniques of *Support Vector Regression* (SVR) and *Artificial Neural Networks* (ANN) are evaluated and compared for modeling FCR in the following subsections. We focused on these two techniques as they have already shown acceptable results in different regression problems in other fields. In addition, alternative well-recognized ML methods such as *Random Forests* or *Gradient Boosting* are mainly designed for classification or forecast combination problems.

3.5.2 Support Vector Regression

When modeling a nonlinear feature space, taking the traditional statistical approaches usually lead to a time-consuming search for complex functional forms for the model. SVR is an extension of the widely deployed classification technique of *Support Vector Machines* (SVM) [87,88]. It is commonly used in ML as a powerful alternative to the traditional linear/nonlinear regression models. The ε-SVR variation is selected here, in which the observations with a distance further than ε from the nonlinear regression curve (or in higher-dimensional problems, from the regression plane or hyperplane) are penalized using a weight parameter C to preserve a soft margin and regularize the objective function.

The kernel trick in SVR helps us to implicitly map the lower-dimensional feature space into a higher dimensional space and address the nonlinearity and complexity of vehicular fuel consumption. Several kernel functions are introduced in the literature and many more could be defined through the composition of the base kernels (as long as the new kernel functions satisfy *Mercer* conditions [88]). However, the *Radial Basis Function* (RBF) kernel is mainly used when there is no prior knowledge about the data. Its stationarity attribute makes it invariant to translation. It looks at the differences of observations not their absolute values, which could help with improving model structures that include lagged speed and grade. Moreover, the smoothness of the RBF kernel adds a desirable level of smoothing to the data.

3.5.3 Artificial Neural Networks

Since the introduction of a single-layer perceptron [89] up until the recent advancements in recurrent network architectures [77,90,91], there have been several attempts at modeling time-series data using the concept of the neural networks. In contrast with the conventional statistical methods, ANNs are adaptive. Data is passed through the layers many times such that each pass of data results in a prediction that is compared to a corresponding observation. The adaptivity attribute, use of nonlinear activation units such as *Sigmoid* or *Rectified Linear Unit* (ReLU), and multi-layered architecture of ANNs make them great candidates for modeling complex unknown phenomena.

In addition to a comparison between the general quality of SVR and ANN in FCR estimation, it is desirable to understand how the wideness or the deepness of the neural network's architecture impact the prediction power. The hypothesis here is that deeper and narrower architectures work better for some vehicles (depending on the characteristics) compared to the shallower and wider architectures. Figure 6 presents the four candidate architectures evaluated.

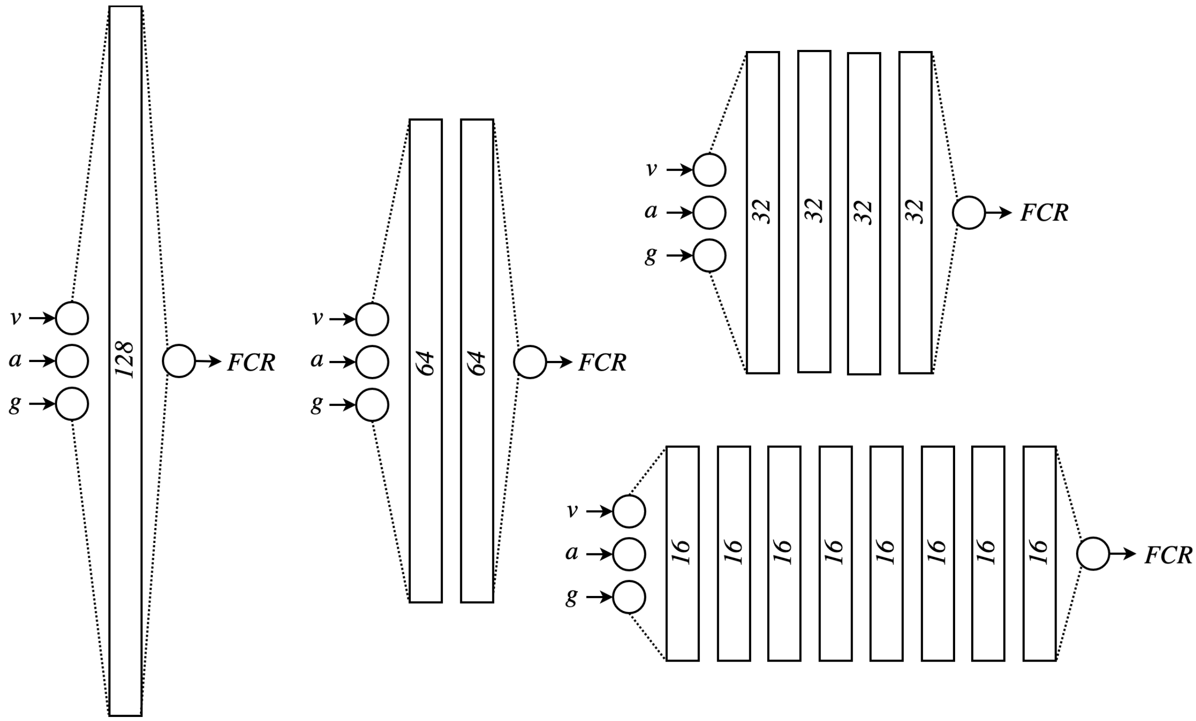

Figure 6: Graphical representation of the four candidate ANN architectures

To come up with an initial hidden-layer size for the base architecture, a trial and error is done using random sample sub-sets of data and by testing different single-layer architectures (starting from 4 neurons and increasing it in base-2 logarithmic scale up to 256). Based on the results, the base architecture is formed with 3 features as input layer (v, a, and g), 128 neurons for the hidden layer, and a single neuron for the output layer. Deeper and narrower architectures are generated using a similar base-2 logarithmic scale. ReLU activation function seems appropriate here as it boosts gradient descent convergence time (computationally efficient) and reduces the likelihood of vanishing gradient descent. However, as regression (not classification) is being performed, a linear activation function is used for the output neuron.

3.5.4 Estimates of IEVs for Improving the Models

The RPM is one of the internal engine variables that unlike other influential variables such as the throttle position, is available through the OBD-II interface of "all" the vehicles. As RPM represents the pace of the iterative combustion process inside the engine, one might expect that a linear correlation exists between FCR and RPM. However, the existence of the transmission system with various technologies and settings in the powertrain results in a complex and nonlinear relationship between these two.

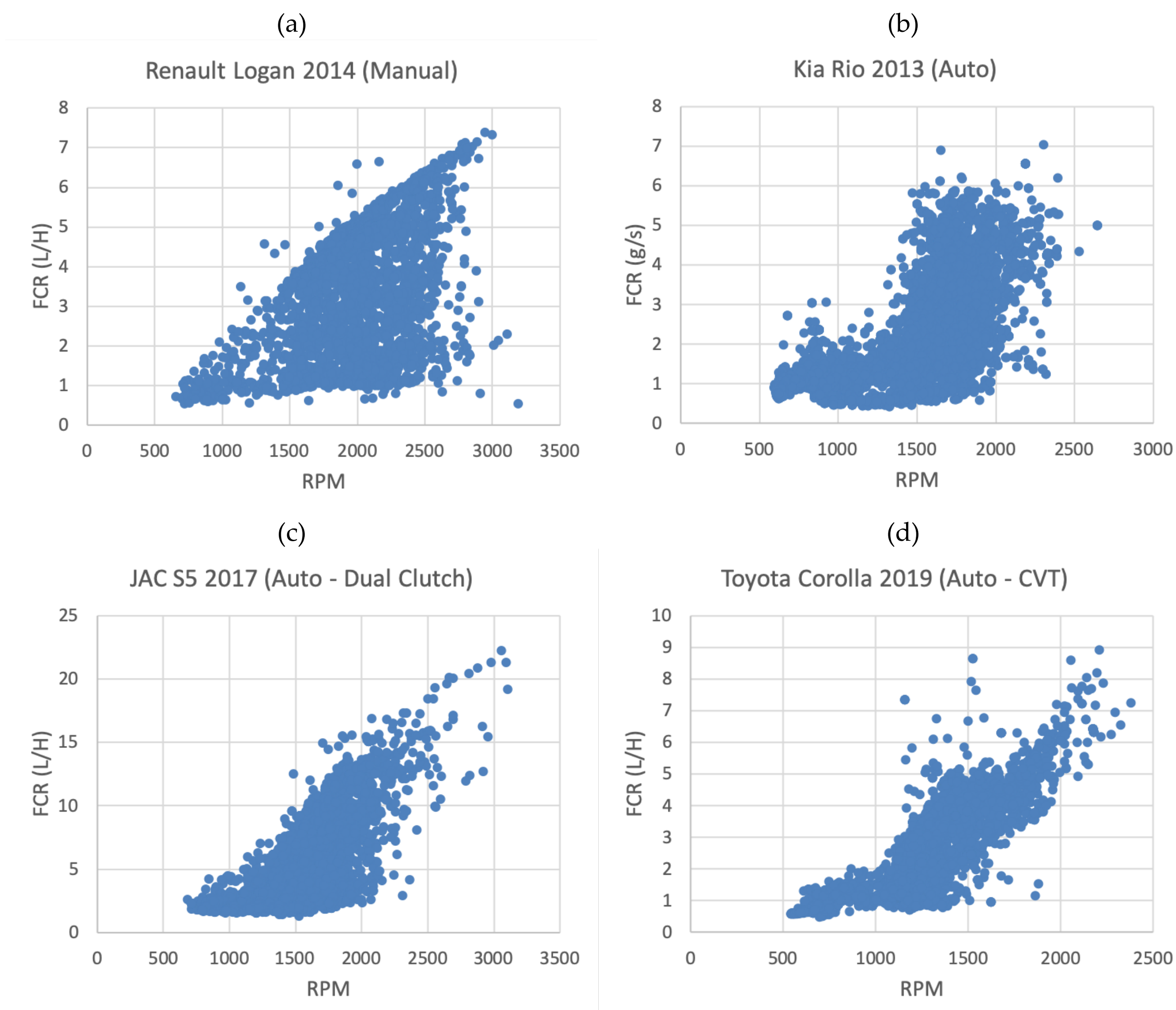

Figure 7: Diversity of FCR-RPM relationship concerning transmission technologies

As shown in Figure 7, the relationship becomes more linear (and less noisy) when moving from manual transmission (Figure 7-a) to CVT (Figure 7-d). Thus, a higher influence of RPM in modeling FCR for CVT-transmission vehicles could be expected.

To make use of RPM as a complementary feature to improve the estimation power of the ML models, two steps are taken. First, by adding RPM to the existing input feature set (v_t, v_{t-1}, a_t, g_t, g_{t-1}), new model structures are generated. Estimating SVR and ANN models is repeated for the new structures to understand the maximum possible improvement compared to the structures without RPM (108,000 SVR and 4,320 ANN additional training). Second, as RPM is inaccessible in practice without the use of specialized equipment, it should be estimated with an acceptable level of error. It is a requirement for developing smartphone-based eco-driving assistance services. Many vehicles currently on the roads (especially, the older ones which are the main fuel consumers and air pollutants), are not equipped with on-board eco-driving options. The drivers' smartphone is the only tool available and models are needed that are instrument-independent (by instrument, a tool that provides access to internal engine variables such as RPM, throttle position, etc. is meant). The traditional models for estimating the RPM require detailed knowledge of the transmission systems' operation mechanism and its technical specifications. Such models need the pre-programmed pattern of gear shifting at different speeds and throttle positions (transmission shift points) in addition to the internal gear ratios for each vehicle as input. In absence of such information, the base easily-observable features are considered for completing the cascaded modeling approach (demonstrated as a part of Figure 8). Eight ANN architectures are assessed and the two most promising structures (based on the results of the primary models for FCR estimation) are chosen to find the best RPM estimates. Four additional wider architectures (by doubling the number of neurons of each layer) are examined as well to provide flexibility for the ANN in learning the true underlying relationships.

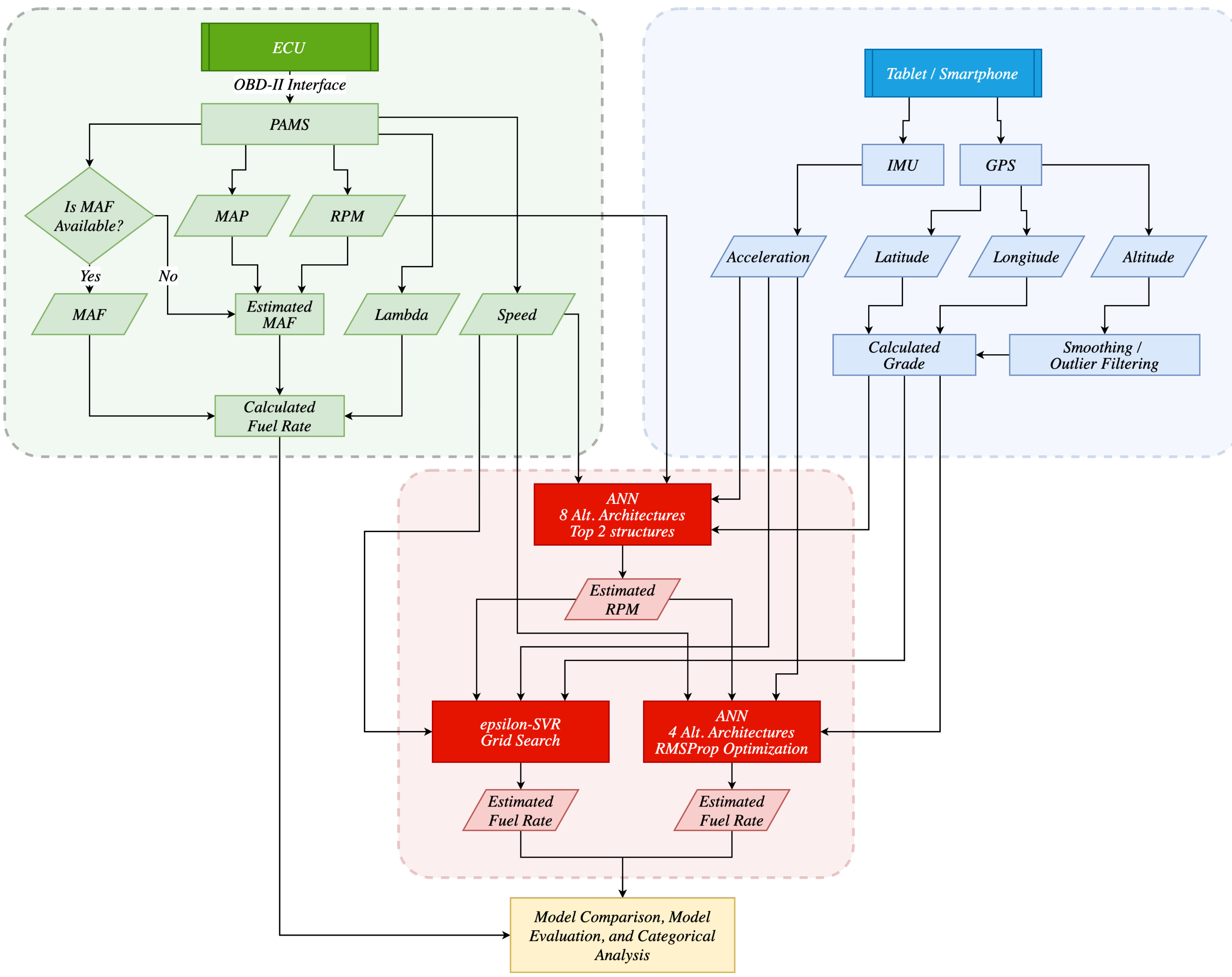

Figure 8: Representation of experimental and cascaded modeling procedures

ANN is used for the RPM estimation (not the SVR) because even with the same architecture, the result of training an ANN with a new dependent variable (and similar features) could lead to different weights representing different underlying interactions between features. While with SVR, this could only be achieved through switching the kernel function which is not easy to do in a data-driven modeling approach. Results of 30,240 training (including 2,160 ANN training for RPM prediction, and 1,080 ANN and 27,000 SVR training to finalize the cascaded modeling process), are presented and discussed in Chapter 5.

3.5.5 Model Training and Validation

For SVR, a 5-fold cross-validation approach is taken as a preventative measure against overfitting. For each model structure and each vehicle, the sample data is first shuffled and then split into five same-size segments. The training process is repeated five times, each time using four segments (80% of the data) as the training set and the remaining segment (20% of the data) as the test/validation set. The models' training and test scores would be the average of cross-validation results.

For each cross-validation process, a grid search is performed to find the best combinations of the three ε-SVR parameters, ε, C, and γ, which give us the best average cross-validation score. C is the regularization parameter to control the well-known bias-variance trade-off (to avoid overfitting). The γ parameter is the inverse of the standard deviation of the RBF kernel (Gaussian function), which is used as a similarity measure between two data points. A large γ value defines a Gaussian function with small variance and in this case, two points are considered similar just if they are close to each other (note that the distances in a multidimensional space is in focus here). When γ is very small, the model cannot capture the complexity or the shape of the data. The resulting model would behave like a linear model as two points can be considered similar even if they were far from each other. The value of ε defines a margin of tolerance where no penalty is given to errors. By tuning the epsilon value, the boundary of support vectors is defined and the regression curve/plane in n-dimensional space is guided towards the best fit. Note that no physical meaning, related to the vehicle operation or fuel consumption process, could be defined for these three parameters.

To minimize the computation time, first, valid ranges for each parameter are found through trial and error on a few sample sub-sets. Choosing a base-10 log scale, ε, C, and γ are selected from ranges of $[0.0001\colon 1]$, $[0.1\colon 100]$, and $[0.001\colon 10]$, respectively. A total number of 81,000 SVR training has to be performed to find the best parameter set corresponding to each of the vehicles and each model structure. The R-squared metric is used as the evaluation metric. The SVR computations are done in *Python* using the *scikit-learn* library. The results are discussed later in Chapter 5.

For ANN, a similar train/test splitting and cross-validation strategy is deployed. *RMSprop* is used as the optimizer. The stopping criterion for the iterative training process is based on the *Mean Squared Error* (MSE) metric. Regularization is done using the drop-out technique. Ten percent of the neurons of each layer is turned off randomly in each pass of the forward/backward propagation process. A total number of 3,240 neural network training is conducted at this step and the *TensorFlow* library [92] is used for training the ANN models.

3.6 Sequential Modeling to Address Lagged Dependencies

3.6.1 Different Sources of Lag

Earlier in Section 3.3.2, possible sources of lag in tailpipe emission measurements were described when the estimation of the exhaust flow rate based on MAF readings was done. It is noteworthy that the exhaust pipe lag is one of the several lag types which the modeling approach should be able to capture altogether. The other sources could be listed as follows:

- *Sensor Response Lag*, which is a fraction of a second for a pre-heated NDIR or electrochemical sensor (like the CO2 and NOx sensors included in the PEMS units).
- *Engine Response Lag*, which refers to the time between the moment a driver takes an action due to the power demand, road geometry, or network interruptions (for example, pushing the gas pedal, braking, shifting gears, performing a maneuver) and the time the engine responds. This time lag is impossible to be measured without additional specialized sensors.

- *Kinematic Distributed Lag*, which represents the delay between the moment a driver decides to increase speed and the time the change in speed to the desired level occurs. In other words, an amount of fuel is consumed (and consequently, a volume of emissions is generated) immediately after the driver's action to generate the required force for acceleration, but the corresponding impact on kinematic variables, which forms the feature set in modeling, is observed gradually and with delay.

3.6.2 Recurrent Neural Network Architectures

Recurrent neural networks are a special type of artificial neural network designed to recognize patterns and temporally-distributed effects on the dependent variable in sequences of data, such as time-series. The technique is widely used for *Natural Language Processing* (NLP), where it has shown its exceptional power in memorizing the past attributes of the sentences and reflecting their impacts on current predictions. Nevertheless, RNNs have been seldom used in vehicular fuel and emission rate modeling, and in the majority of studies in this field, data points are assumed random samples rather than time-series.

A fully-connected neural network takes in a fixed-size vector and gains no knowledge about temporal interactions of features and the dependent through the model training process, while an RNN model not only takes the features' vector at time t but also takes the measurements from the previous moments up to p lag steps ($t-1$, $t-2$,

$\dots, t-p$), assuming lagged (or distributively lagged) impact of features on the dependent variable exists. Figure 9 depicts the general architecture of a many-to-one RNN model which is applicable in vehicular fuel and emission rates modeling. Note that $\mathbf{X_t} = [X_{t-p}, \dots, X_{t-2}, X_{t-1}, X_t]$ is the input matrix corresponding to time t. Each X_i element defines a columnar vector holding the instantaneous values of the main model features. Hence, in our case, X_i would be equal to $[v_i, a_i, z_i]^T$, where v_i is speed in km/h, a_i is acceleration in m/s^2, and z_i is the GPS altitude in m all at time i for each vehicle selected for modelling. The measured instantaneous FCR or one of the ER values at time t could be used as output Y_t while training the model.

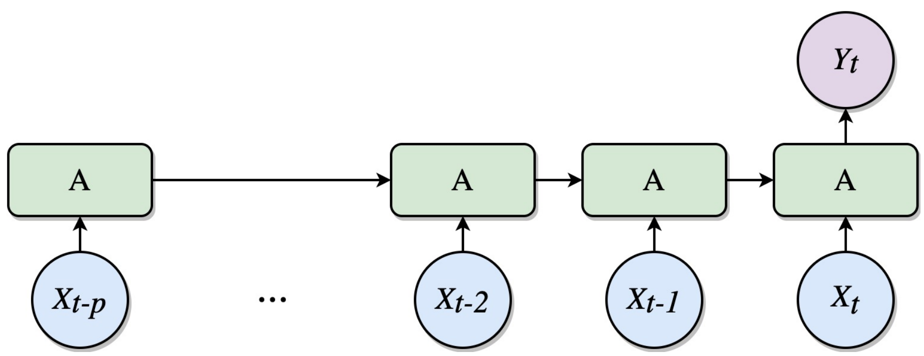

Figure 9: Many-to-one RNN architecture with lag order of p (RNN cells are in green)

Three different internal mechanisms are introduced in ML literature for RNN units. Although the non-gated simple RNN unit is powerful in including short-term dependencies, the mechanism is weak in capturing long-term effects due to the vanishing/exploding gradients phenomenon during the training (backpropagation) process. To deal with such problem, *Long Short-Term Memory* (LSTM) units are introduced, with proven effectiveness during the last two decades [76–78,90,93]. The complex internal gated structure of an LSTM unit allows the memorization of long-term dependencies in data. The gates manage the intensity as well as the temporal extent of past changes of features on the current output of the model. In recent years, a simplified version of LSTM, called *Gated Recurrent Unit* (GRU) has gained interest among ML modelers [94,95]. GRUs possess only two internal gates (compared to three gates of LSTM unit) and training/running a GRU-based sequence model is less computationally expensive, while providing comparably accurate results. Whereas, due to the existence of dedicated control over the memory in LSTM, rather than fully exposing the hidden content like what GRU does, LSTM is theoretically supposed to remember longer lagged effects. As the order of lag is unknown for us at first and is assumed to be dependent on the type of emission, type of vehicle, and even as it could change through time in different driving conditions, all three RNN unit types would be assessed in this study.

The prediction power of RNNs could be boosted by deepening them through stacking RNN layers over each other. In stacked many-to-one RNN architecture, each layer (except the last one) outputs a sequence of vectors which will be used as an input to a subsequent layer. The additional hidden layers are understood to recombine the learned representation from prior layers and create new representations at high levels of abstraction.

3.6.3 RNN Modeling Settings

As a foundation for developing category-based models, vehicle-specific RNN models for both fuel as well as emission rates are estimated. The input vectors consist of speed (v_t) in km/h, acceleration (a_t) in m/s^2, and GPS altitude (Z_t) in m. One hundred hidden RNN units are considered inside each cell of the network after performing a few warm-up training procedures and evaluation of the impact of the lower and higher number of hidden units. In addition, all three types of RNN units (Simple, LSTM, and GRU) as well as single-, double-, and triple-layer stacked architectures are assessed (deeper structures are disregarded due to exponentially increasing processing time). Depending on the lag order, data is converted into p-length sets of vectors, and 5-fold cross-validation (with 70% of data for training and 30% for validation) is used to achieve robust modeling results. Regularization is applied through the dropout technique with a drop probability of 50%. Besides, MSE is used as the loss function, and the *Adam* algorithm is considered for the neural network's optimization (as it resulted in acceptable accuracies much faster than other popular alternatives, *Momentum, RMSprop,* and *Stochastic Gradient Descent* (SGD), during the warm-up modeling attempts). Moreover, the iterative training process (forward and backward propagation through time) is continued until the validation error got stabilized. It is noteworthy that for all different types of modeling, *Python* programming language and two popular libraries of *TensorFlow* [92] and *Scikit-Learn* [96] are deployed.

3.7 Primary Forecast Combination of Models with Different Lag Orders

The capacity of the modeling methodology to allow generalization to more aggregate levels (such as categories of vehicles) is an important factor when developing vehicular fuel consumption and emission models either for eco-driving purposes, use in traffic simulations, or even for macro-scale transportation analyses. To achieve this goal, an appropriate forecast combination technique is required which results in ensemble models that perform at least as good as the component models. Consequently, the ensemble will represent the category and is assumed to perform acceptably for any vehicle with the same shared attributes.

3.7.1 Popular Forecast Combination Methods

Several forecast combination techniques are introduced in the traditional statistics literature. The most basic method, *Simple Averaging*, assumes the component models (also called *Weak Learners*) contribute at the same level to the ensemble forecast [97]. *Trimmed* and *Winsorized* means are samples of developments in averaging techniques [98]. But these methods work efficiently only when adequately accurate predictions by component models are available. Moreover, they underperform significantly when data with skewed distribution exists [74]. The more sophisticated method of *Ordinary Least Squares* (OLS) regression has been frequently used in the literature [99–101] for ensemble forecasting. But as the world is transitioning from the age of the traditional statistics to the machine learning era, ensemble learning algorithms are evolving accordingly. Methods founded upon the concepts of *Decision Trees* (DT), *Gradient Boosting* (GB) and its extensions such as *AdaBoost* (AB), *Random Forests* (RF), SVM, and even recently ANN have come into focus of ensemble modelers and have proved their capability in generating significantly improved ensemble predictions [102,103].

3.7.2 Finding the Extent of Lagged Effects

Due to the existence of different sources of lag, the true lag order in data is completely unknown and it is possibly dynamic. To deal with this uncertainty, first, vehicle-based RNN modeling is repeated for lag orders from 1 to 10 to find out the best range of lag orders to be considered for the next steps (the upper bound is selected based on the engineering judgment of the author after visually inspecting the overlapped curves of speed, FCR, and ER for different vehicles and the notion that 10 seconds is long enough for dissipation of temporally distributed effects in vehicles). The average normalized *Root Mean Squared Error* (RMSE) resulting from training RNN models is presented in Figure 10 for fuel, CO2, and PM rates. Note that the simple RNN modeling results are disregarded here as they showed weaker validation scores by a significant margin.

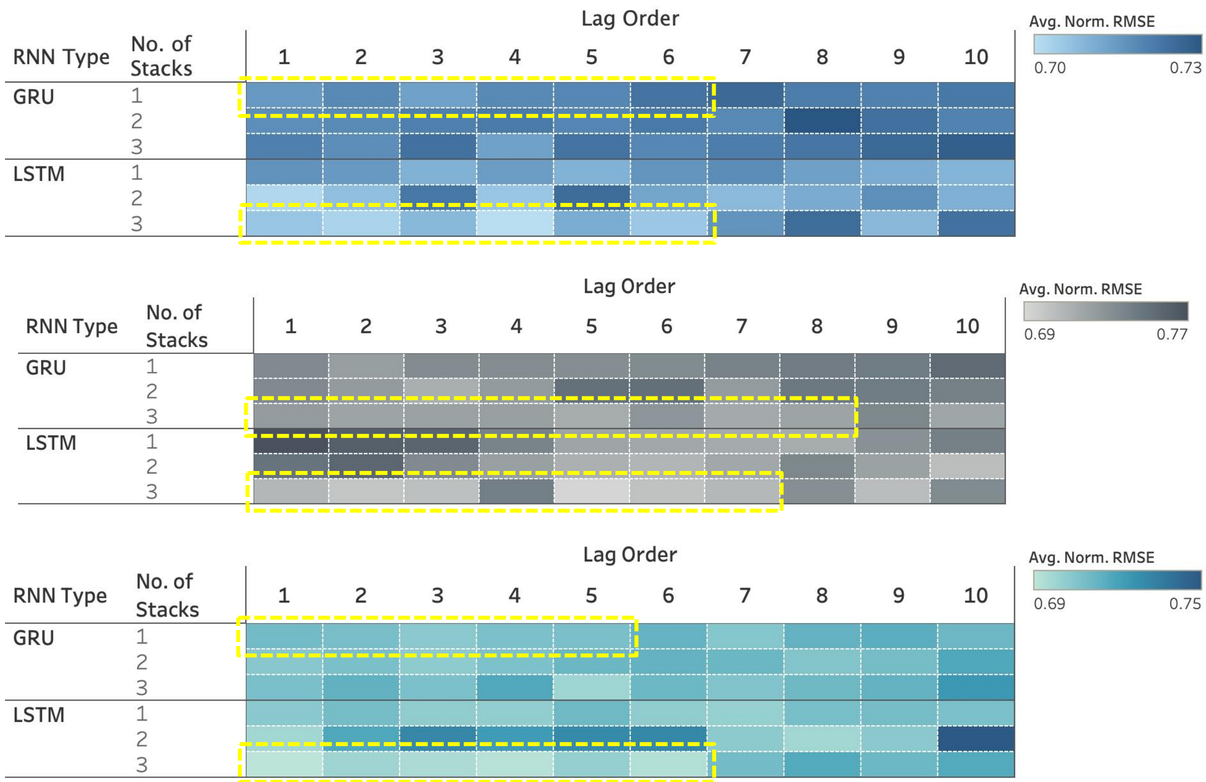

Figure 10: Average Norm. RMSE for RNNs regarding (top) FCR, (middle) CO2, and (bottom) PM

The best results obtained using each of the RNN types corresponds to a range of lag orders from 1 to an average of 6. Considering the proven capability of RNNs in capturing serially-correlated as well as lagged phenomena, it is concluded that the extent of lagged effects does not last for more than 6 seconds. This is compatible with the findings from the visual inspection of overlayed speed and fuel/emissions curves in literature [53,104].

The results of RNN modeling for lag orders of 1 and 6 for a randomly selected time-window for three of the vehicles under study are presented in Figure 11. The prediction curves clearly show that L1 and L6 models compete with each other at different ranges in terms of accuracy. An interesting understanding is that RNNs with lower lag order predict better the extreme and sudden peaks and valleys (see regions highlighted in magenta), while those with higher lag orders perform better at ranges with smaller/no variations (see regions highlighted in yellow). The observation brings the idea that combining forecasts conducted by RNN models of different lag orders might lead to a single but more accurate model.

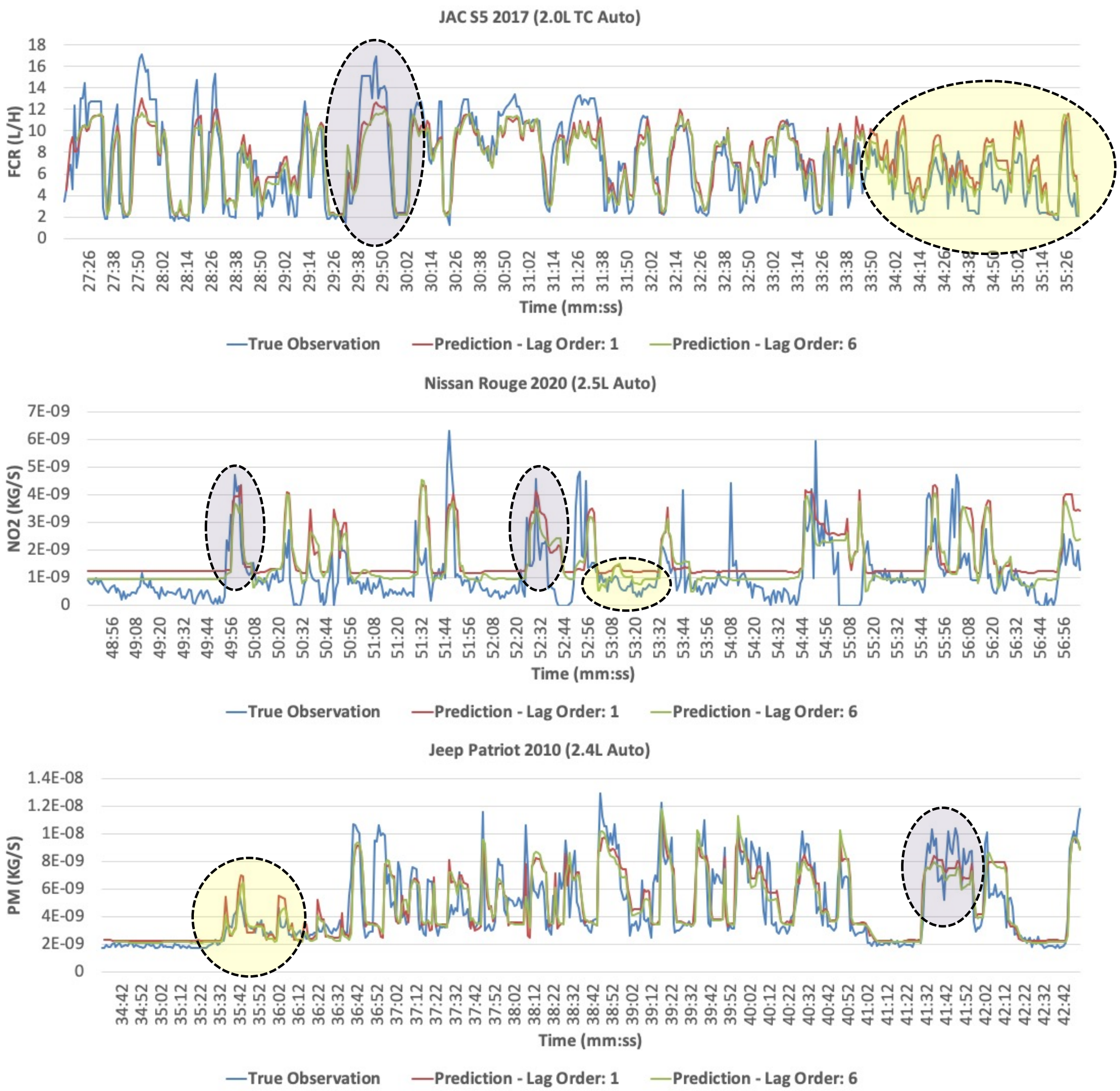

Figure 11: Best RNN predictions for lag orders of 1 and 6 on three vehicles

3.7.3 Developing the Ensemble of Lag-specific Models

As a complementary step to develop vehicle-specific RNN predictions, forecast combination techniques are utilized to combine predictions of best RNN models trained for each vehicle-dependent pair for the 6 lag orders (called *lag-specific* models). Taking such an approach, a *Metamodel* (or *Meta-Regressor*) is aimed to be trained for each vehicle-dependent pair which is expected to perform at least as good as the best lag-specific model, if not outperforms it. These primary metamodels will be used later for building higher-level ensembles for the categories of vehicles. The approach taken here is inspired by the *Stacking* method in the *Ensemble Learning* (EL) paradigm, as shown in Figure 12, where the component models are trained based on a complete dataset, and then their outputs are used as input features to train an ensemble function.

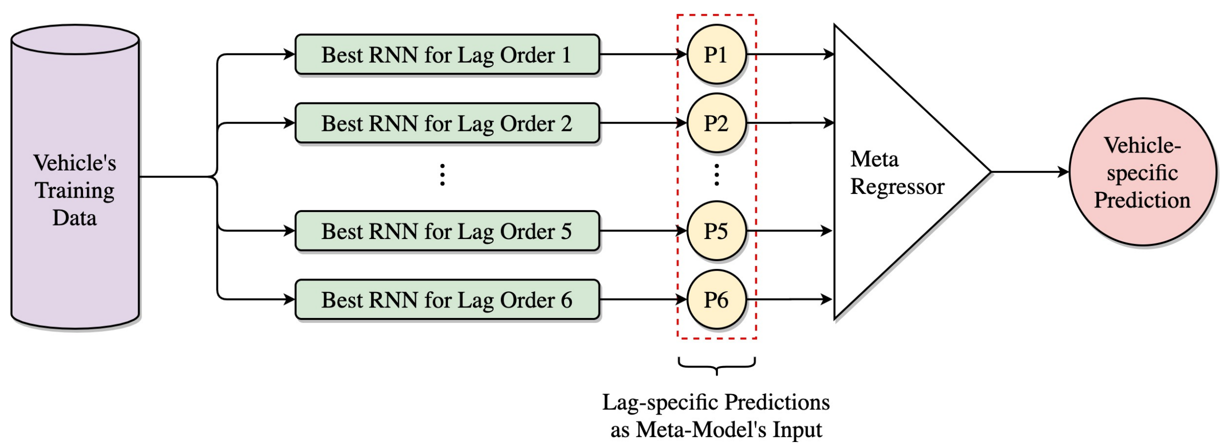

Figure 12: A Stacking EL architecture to develop vehicle-specific models

The performance of 8 widely used EL algorithms as the *Meta-Regressor* (two settings for each) to combine lag-specific RNN models is evaluated. Then, the best algorithmsetting combination for each vehicle is handpicked. The algorithms and the corresponding major settings are described in Table 3.

Table 3: Major details of EL algorithms evaluated for developing lag-specific models

Algorithm	**Settings**	
	Attribute	**Value**
Linear Regression	Feature Normalization	Active
Ridge Regression	Feature Normalization	Active
	Regularization Strength	α = {0.1, 1.0}
Support Vector Regression	Kernel	RBF
	Gamma	Scale
	Epsilon	0.1
	Regularization Parameter	C = {1.0, 10.0}
Decision Tree	Splitting Criterion	MSE
	Splitting Strategy at Nodes	{Best, Random}
	Maximum Tree Depth	Unbounded
Gradient Boosting	Loss Function	Least Squares Regression
	Splitting Criterion	MSE
	Learning Rate	0.1
	Number of Boosting Stages	{10, 100}
AdaBoost	Base Estimator	Decision Tree Regressor
	Loss Function	Linear
	Learning Rate	1.0
	Number of Boosting Stages	{10, 100}
Random Forest	Number of Trees	{10, 100}
	Splitting Criterion	MSE

Algorithm	Settings	
	Attribute	**Value**
	Maximum Forest Depth	Unbounded
Multi-Layer Perceptron (MLP) *or* Fully-Connected ANN	Number of Hidden Layers	{1, 2}
	Layer Size (No. of Neurons)	100
	Activation Function	ReLU
	Optimizer	Adam
	Learning Rate	0.001
	Maximum No. of Iterations	200

For all the ensemble models trained at this stage, the same train-test splitting strategy (70% of data for training and the rest for testing) is followed. A diverse spectrum of modeling techniques was tried. The sophisticated types (such as Random Forests or ANN) are not directly targeted because if acceptable results are achievable with simpler algorithms, at the time of deployment (for example, in an eco-driving assistance service), much lower processing power would be required. The results of the vehicle-specific forecast combination are visually presented and statistically discussed later in Chapter 6.

3.8 The Generalization to Category-specific Models

Having the vehicle-specific ensemble predictions in hand, the final step towards modeling category-specific models could be taken. Vehicle-specific modeling is naturally susceptible to bias; hence, generalizing and using such models for other vehicles could always be criticized. Furthermore, smartphone-based eco-driving services should be usable for different types of vehicles even with unknown detailed technical specifications. Appropriate microscale fuel and emission models should be easily adjustable to the new vehicles based on simple and general information retrieved from the driver or the car owner, such as vehicle class, age, engine type, engine size, transmission type, weight, payload, etc.

To fulfill the aforementioned requirement, higher-level metamodels (or *Supermodels*) are trained for categories by combining forecasts already conducted by vehicle-specific metamodels. However, making a heavy assumption is required that all the vehicles in a category possess important common attributes affecting their powertrain operation which result in similar fuel consumption and emission generation patterns. Furthermore, it is assumed that the data used for training lag-specific models and vehicle-specific ensembles for each category member is a subset of a hypothetical larger homogenous dataset dedicated to the category. Nevertheless, before training the category-specific models based on different categorization criteria, it is not clear which common attributes lead to categories with such homogenized members. The categorization criteria could be ranked only after comparing the modeling scores corresponding to category-specific supermodels. Figure 13 shows the comprehensive architecture of the two-stage EL approach for generalizing the basic RNN models to supermodels generalized to categories of vehicles.

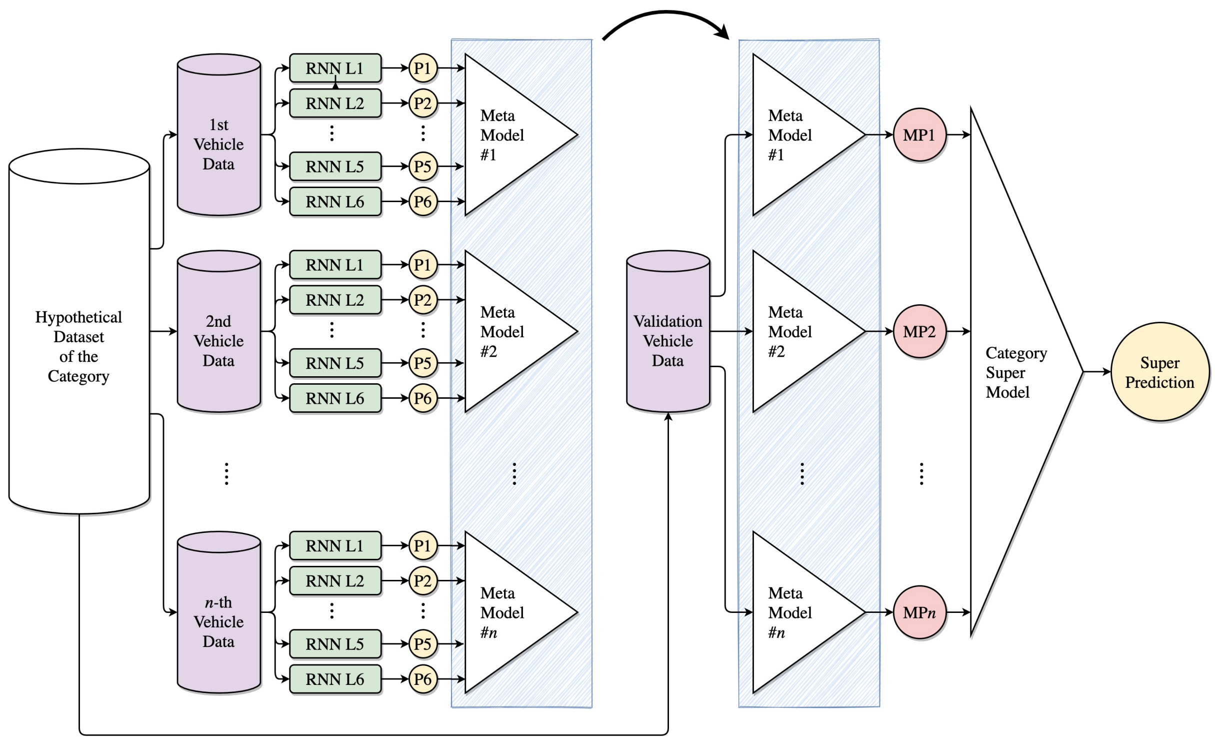

Figure 13: The architecture of the two-stage EL for category-specific modeling

Because the vehicle-specific metamodels are trained for each vehicle's dataset separately, the predictions regarding their test input data will not be valid for the second-level forecast combination. To deal with this issue and to avoid violating the heavy assumption regarding the homogeneity of category members, a special n-fold cross-validation approach (n is the category size) is taken in which after categorizing the vehicles based on the desired criterion, one member of each category as *Validation Vehicle* is pulled out. The validation vehicle's data is then injected into vehicle-specific metamodels (of the rest of the category members) and their predictions are used for developing the category-specific supermodel through EL. The average cross-validation scores are then used for evaluating the supermodels' performance.

The same set of EL algorithms and settings described in Table 3 are assessed for developing category-specific models as well. Like the previous steps, a similar train-test splitting strategy is deployed to prepare the metamodel predictions as input for the category-specific supermodels and keep a part of that data for validation (5-fold cross-validation is applied as well). Vehicles are categorized based on 6 criteria of *Age Range*, *Class*, *Engine Type, Engine Size Range*, *Transmission Type*, and *Weight Range*. Vehicle weight is calculated as the sum of curb weight and live/dead payload at the time of experimenting. All the modeling steps are repeated for FCR and each of the ERs. RMSE has been used in all three modeling steps (lag-, vehicle-, and category-specific modeling) as the evaluation metric. Results of EL modeling attempts (for both stages) are presented visually and discussed in Chapter 6.

Chapter 4

MOVES Validation using Real-World Measurements

4.1 Introduction

Following a major concern regarding the validity of using MOVES predictions for environmental assessment of transport projects or eco-driving purposes in non-U.S. regions (such as Canada), several MOVES scenarios are run based on the methodology described in Section 3.4 and the results are compared to the emission rates captured by the *Portable Emissions Measurement System* (PEMS) and the *Energy Consumption Rates* (ECR) calculated based on *On-Board Diagnostics* (OBD-II) readings in this chapter. Through this comparison, we target evaluating the magnitude of error as well as the bias direction when using MOVES model for estimating project-level (link-level) energy and emissions rates. In addition, the sensitivity of MOVES output to fleet attributes as well as local road network and environmental conditions can be assessed. Such sensitivity analysis helps us identify the regional aspects which should be addressed in developing independent local models or adjustment procedures.

4.2 Energy Consumption

The first comparison is for link-based ECRs (in Joules per mile). Figure 14 presents the ground-truth ECR observations versus MOVES predictions. The dashed red line in Figure 14 shows the expected (ideal) condition where ground-truth and predictions are in full compliance. The black solid line represents the existing correlation between observations and predictions. A linear regression model is applied to data to find the goodness-of-fit as a metric representing the accuracy of MOVES predictions.

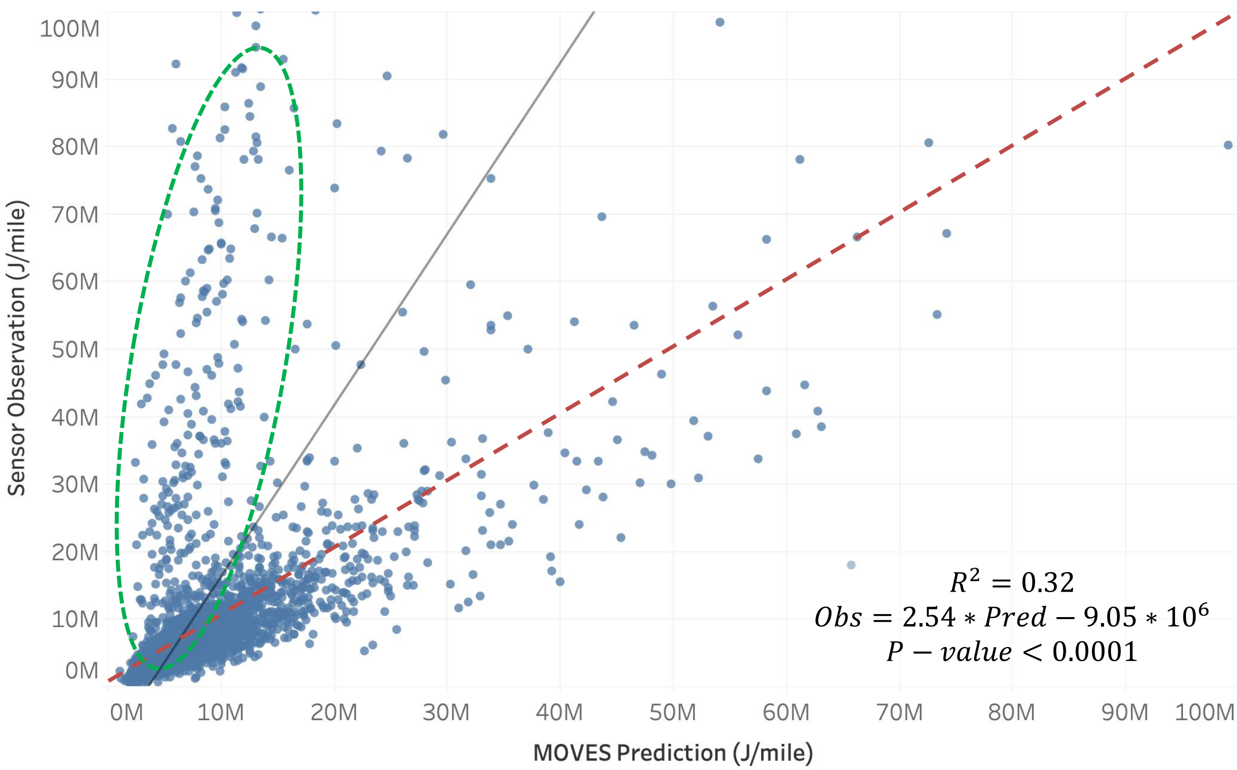

Figure 14: ECR observations vs. MOVES predictions

Thus, in general, MOVES shows low prediction power for energy consumption despite all the adjustments applied at the input stage to simulate the prevailing conditions regarding test fleet, fuel types, weather conditions, or road network attributes. The slope of 2.54 of the fitted line (while having a small negative intercept) shows that MOVES is underestimating energy consumption in general.

A significant unusual branch is visible in Figure 14 (highlighted with a green dashed line). The majority of data points in that branch is dedicated to three vehicles from the tests in Bucaramanga. All three vehicles have a manual transmission, regular engines, and used regular gasoline during the tests. But these are not the only manual-transmission vehicles among the test fleet in Bucaramanga (and the other two cities). Several tests in Bucaramanga as well as Tehran conducted on vehicles with manual transmission, which have shown normal behavior. Regular gasoline has been used in Tehran as well and several vehicles from the same age range exist in the test fleet; thus, the anomaly could not be linked to fuel type or age-related conditions. Moreover, the driving style has been regulated before the tests and the set of roads that drivers were supposed to traverse were almost similar for tests in each city. Nevertheless, none of the other vehicles have shown abnormal behavior. As the magnitude of the ground-truth ECRs for these three vehicles is significantly higher than observations regarding the rest of the test fleet, the anomaly could have roots in wrong vehicle engine settings, use of lower quality fuel, or extraordinary congestion at the time of experimenting on these three vehicles. Thus, those three vehicles are put aside from the sensitivity analysis of MOVES outputs conducted through the categorization of comparison results for different vehicle, trip, and location attributes (such as vehicle segment, age, weight, link average speed, link average grade, temperature, and humidity). It is important to note that by removing the data regarding three outlier vehicles, the R-squared score decreased to 0.24 as there is still high variance around the ideal line.

In Figure 15, the energy consumption data for manual-transmission vehicles is separated from those with automatic transmission (of any type) excluding the 3 outlier vehicles mentioned earlier. Although the slope of the fitted regression line is close to 1, the R-squared values are still low for both groups.

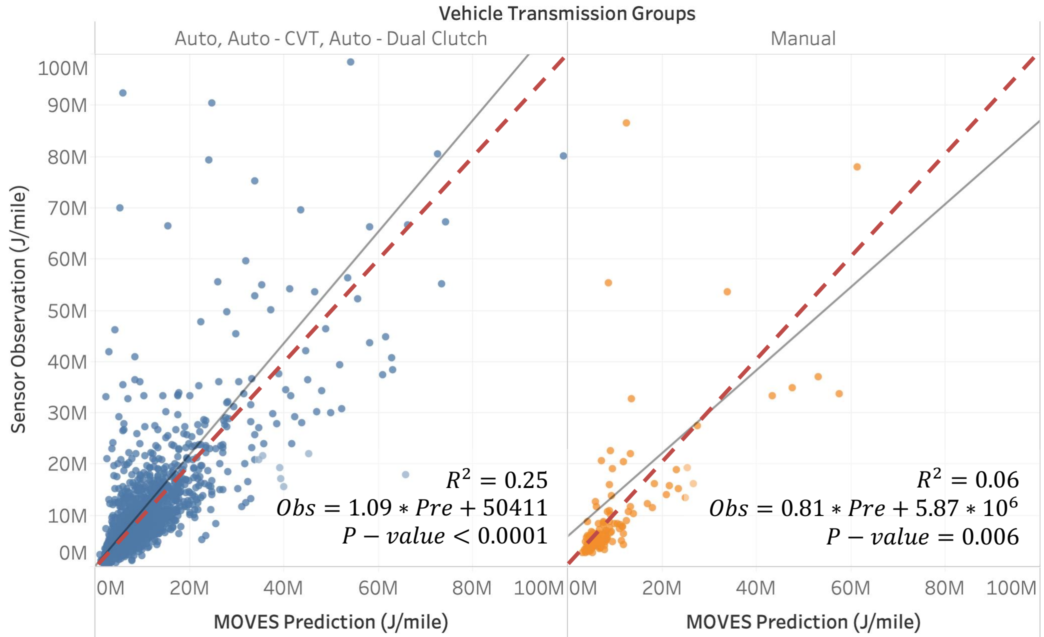

Figure 15: ECR observations vs. MOVES predictions for transmission-type categories

MOVES slightly underestimates ECR for automatic transmissions, whereas there is a minor over-estimation for manual transmissions. There is no option in MOVES to input information regarding the transmission type of vehicle fleet; thus, in general, it treats vehicles with different transmissions in the same way. However, as at the project level, second-by-second link drive schedules are used as part of the input, the impact of the transmission system (as means of converting generated power by the engine to the wheel speed) is considered implicitly in the model. But it is mixed with other influential factors such as road congestion or aggressive driving. Regarding fuel type, no significant difference between ground-truth ECR and predictions was observed for regular and E10 fuel.

Figure 16 takes the engine type as the categorization criterion. Compared to the regular-engine vehicles, MOVES performs exceptionally in the prediction of ECR for turbocharged ones with an R-squared of 0.74 and a slope of 0.91 for fitted lines. It is clear from the figure that the magnitude of ECR is considerably lower for turbocharged engines which was expectable. The significant difference between a turbocharged engine and a traditional naturally aspirated gasoline engine is that the air entering the engine is compressed before the fuel is injected. When air is compressed, the oxygen molecules are packed closer together. The increase in the air means that more fuel can be added for the same size naturally aspirated engine. This then generates increased mechanical power and leads to overall efficiency improvement of the combustion process. Therefore, the engine displacement can be reduced for a turbocharged engine leading to an overall improved fuel economy. The reason MOVES predicts much better for turbocharged vehicles could be linked to the fact that data collected for estimating MOVES are mainly results of in-lab simulations of the EPA’s standard driving cycles on chassis dynamometers in controlled environments on perfectly maintained vehicles.

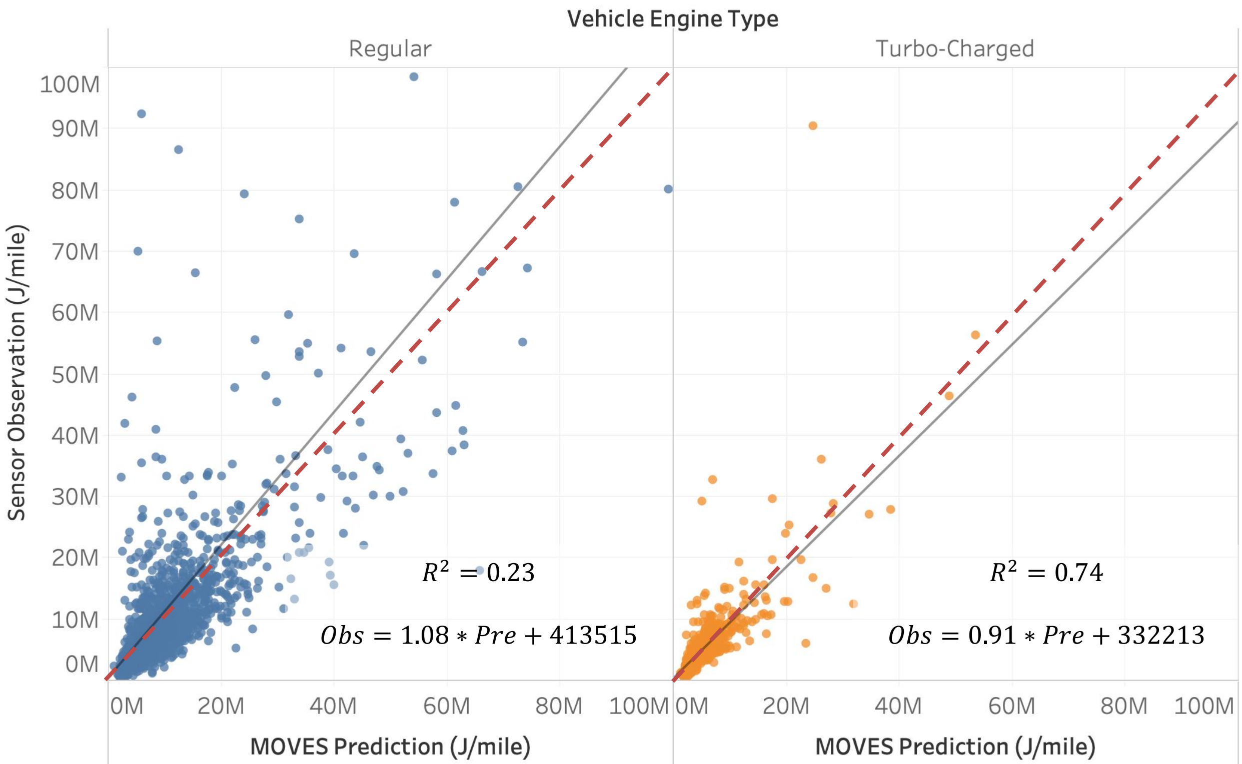

Figure 16: ECR observations vs. MOVES predictions for engine-type categories

In the lab, many external factors affecting the fuel economy of vehicles are ignored. Therefore, as the turbocharging pushes the engine towards working more ideally with less energy waste, the ground-truth for that type of engine would be more similar to the MOVES experiment conditions.

In Figure 17, age has been chosen as the grouping criterion. The significance of goodness-of-fit for vehicles aged between 5 to 8 years is interesting. Every year, vehicles with more advanced technologies and more efficient engines are introduced to the market. The last update to MOVES dates back to 2014. Thus, the majority of newly introduced vehicles are absent in MOVES' source fleet. MOVES has predicted ECR for 5-

8 years-old vehicles well as its core models are mainly estimated and calibrated for technologies of that era.

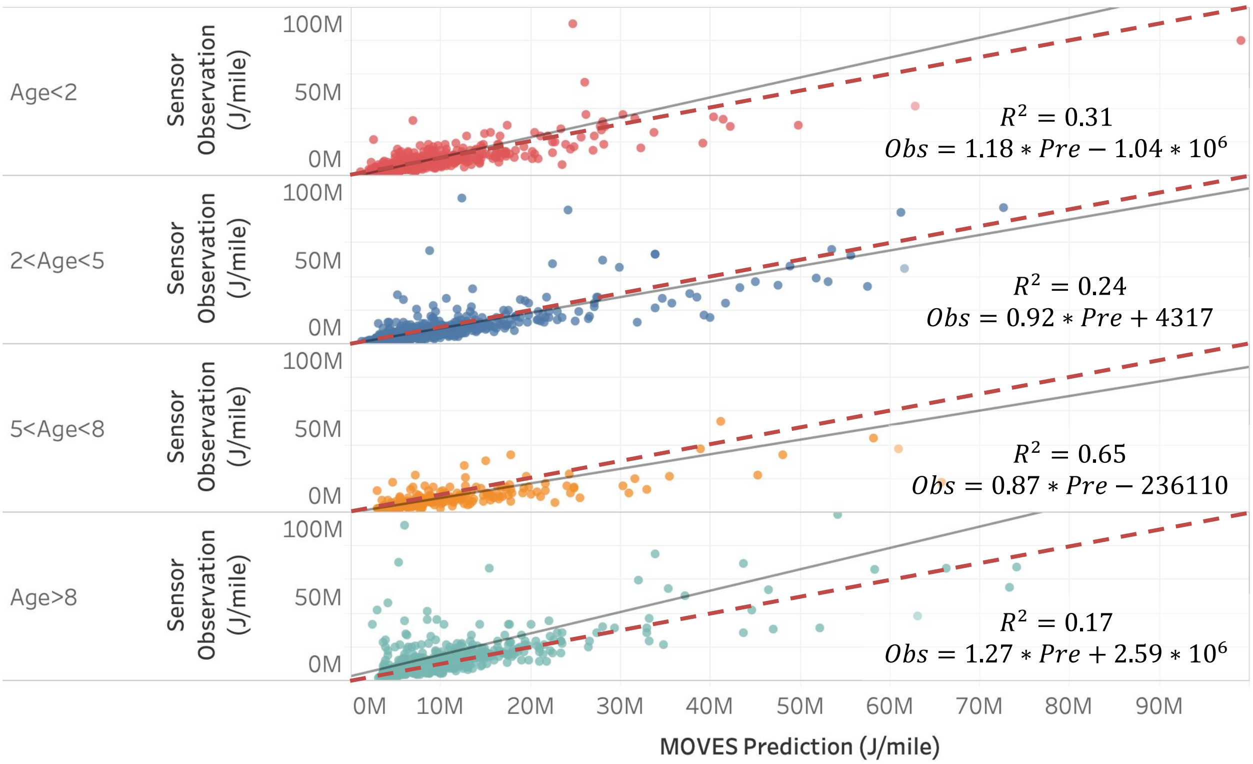

Figure 17: ECR observations vs. MOVES predictions for vehicle-age categories

Deepening the sensitivity analysis, the vehicle segment is targeted in Figure 18. The ECR for compact and subcompact hatchback vehicles are predicted with an R-squared score of 0.71. MOVES slightly over-estimates ECR for this category. For the rest of the segment groups, high variation is evident in comparison charts and the prediction accuracy is unacceptable. A concrete conclusion is not possible to be made unless the sensitivity analysis is performed based on vehicle weight as well (see Figure 19) because an average hatchback vehicle is smaller and lighter compared to vehicles from other segments.

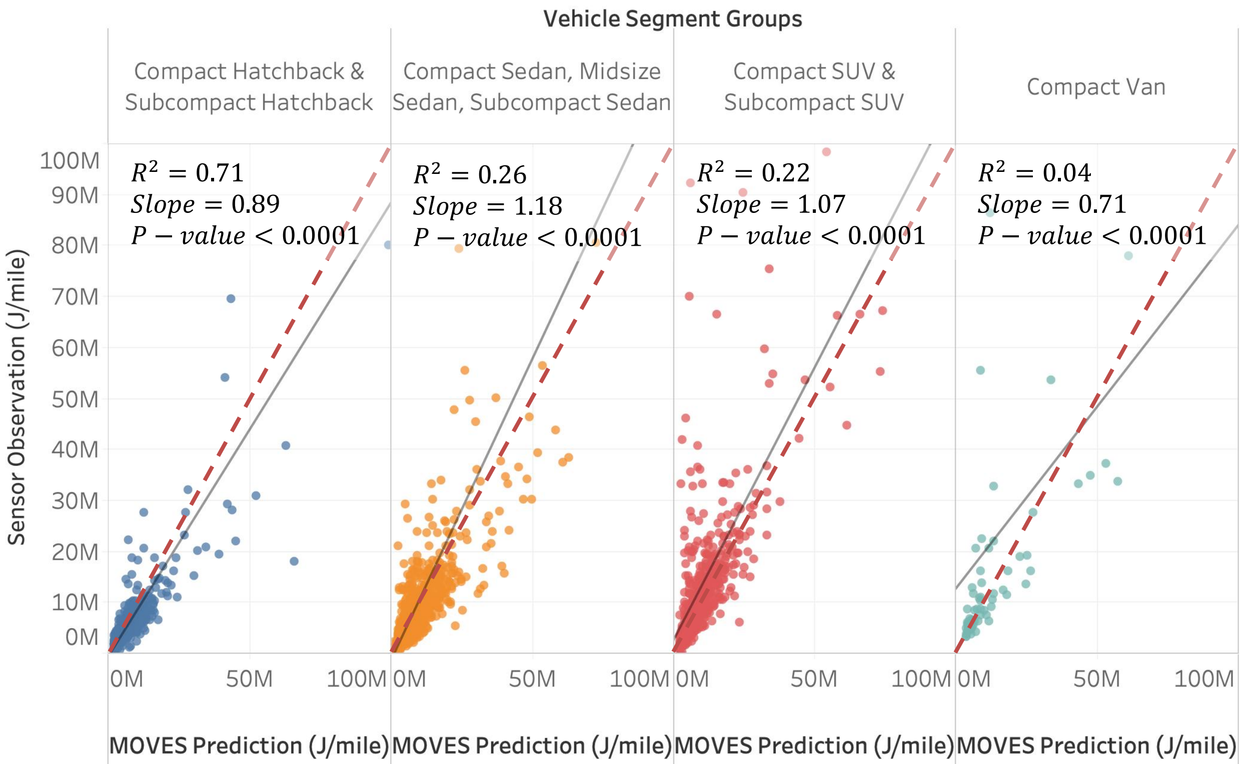

Figure 18: ECR observations vs. MOVES predictions for vehicle-segment categories

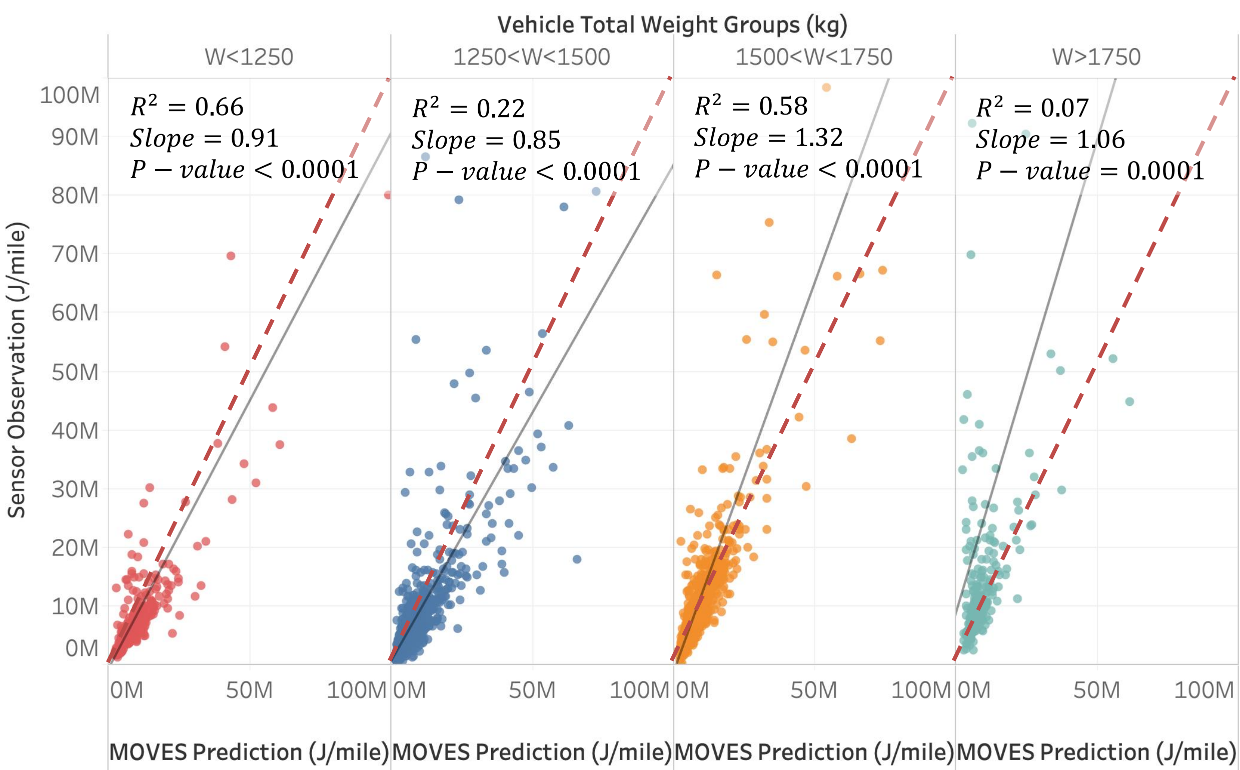

Figure 19: ECR observations vs. MOVES predictions for vehicle-weight categories

MOVES performs acceptably in predicting ECR for light (less than 1250kg) and medium-weight (1500 to 1750 kg) vehicles with R-squared values of 0.66 and 0.58, respectively. However, it noticeably underestimates ECR for the heavy-weight group. Therefore, it is confirmed that MOVES predicts much better for smaller/lighter vehicles like the hatchbacks. Usually, compact and subcompact hatchbacks are equipped with weaker engines less capable of high acceleration; therefore, lower *Vehicle Specific Power* (VSP) values are observed in their tests. It has already been proved (through studies by Jiménez-Palacios [105] and Wu *et al.* [44]) that the variation of observed emissions increases significantly for higher VSP values. Thus, it is expected that MOVES predictions include more errors for heavier vehicle segments.

No sensitivity was observed when investigating the results based on average-speed bins. However, a side-by-side comparison of observations/predictions against link average speeds in Figure 20 reveals interesting compliance of patterns between the ground-truth and MOVES output. It is visible that in the real world, there is a higher probability of deviation and noise in ECR which relates to influential factors ignored by MOVES.

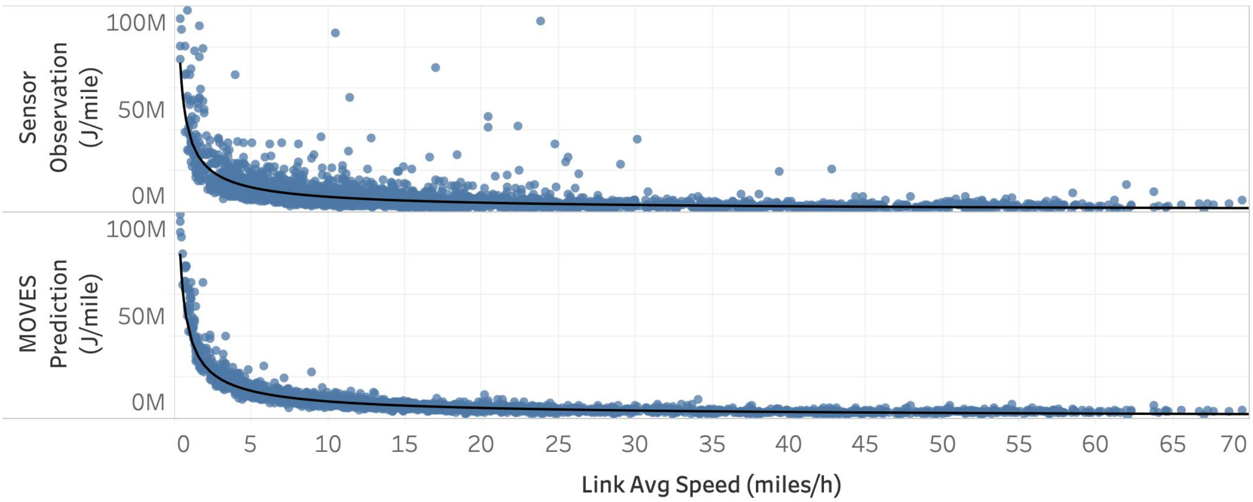

Figure 20: ECR observations and MOVES predictions against link average speed

Regarding grade, MOVES prediction power is exceptionally high for the positive grades from 5 to 20 percent (see Figure 21).

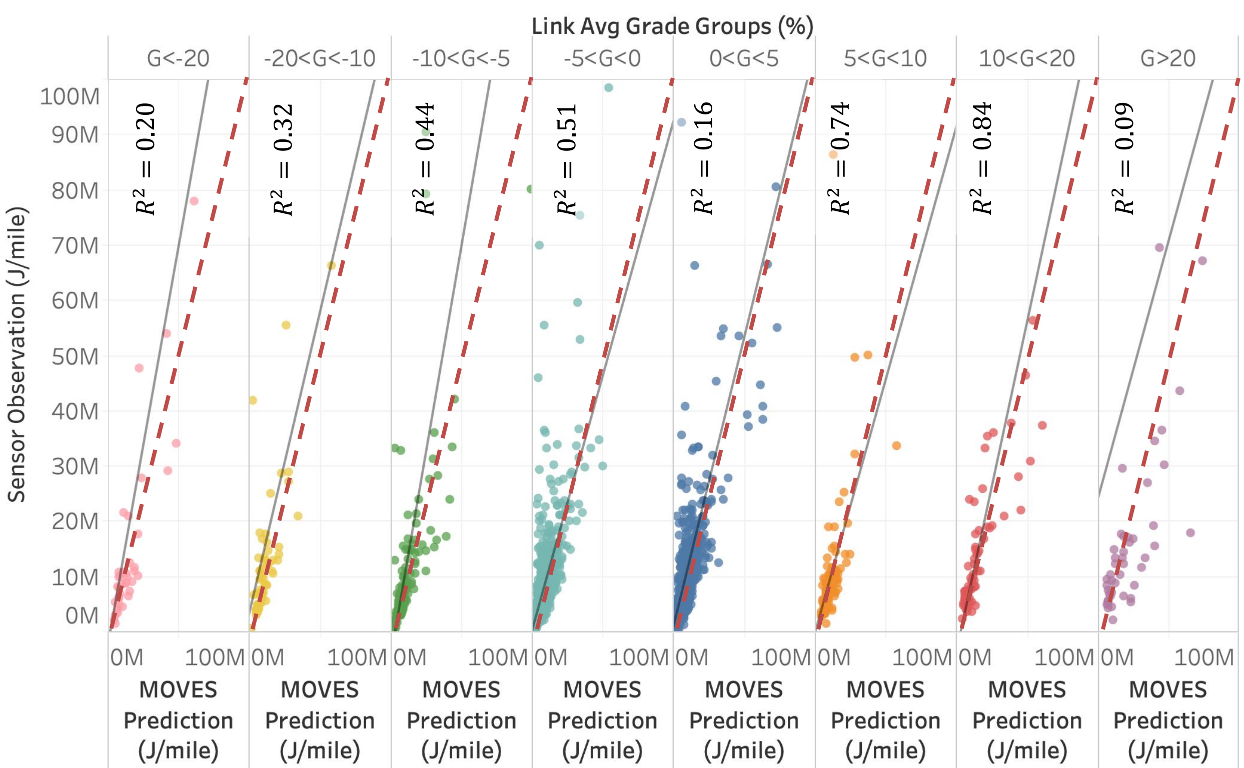

Figure 21: ECR observations vs. MOVES predictions for link-grade categories

In negative grades, the vehicle moves downhill while usually cruising or decelerating. As the need for power decreases in these conditions, lower VSP values (even negative) will happen. It is expected that MOVES predict with more accuracy as ECR at small/negative VSPs has less variation. But the current observation shows that the VSP formulation might not be capable of completely capturing the impact of grade on power demand.

MOVES ECR predictions are insensitive to ambient temperature changes. But, on the other hand, MOVES was quite accurate for ECR prediction in dry weather conditions (humidity lower than 50%). This finding is in accordance with the fact that MOVES core models are estimated and calibrated for the U.S. meteorological conditions.

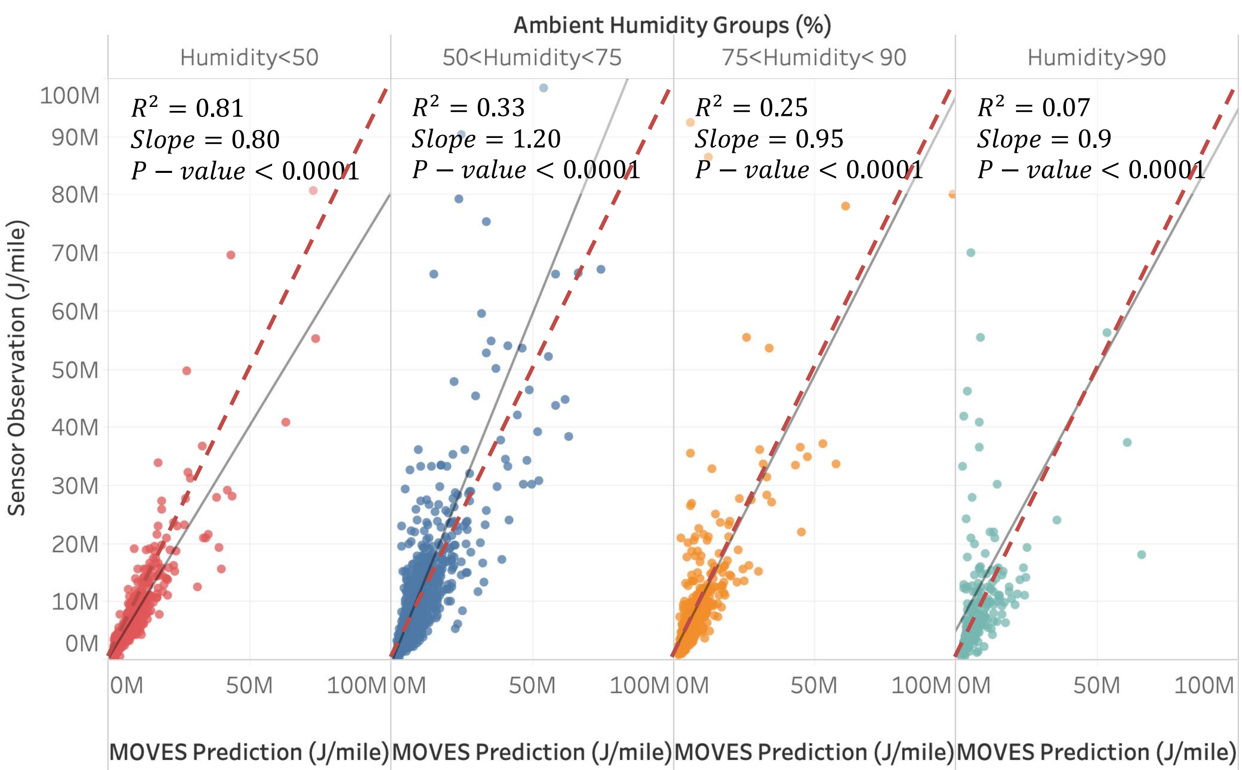

Figure 22: ECR observations vs. MOVES predictions for ambient-humidity categories

The U.S., on average, has much drier weather compared to Canada, Colombia, and even some parts of Iran. Thus, the reliability of MOVES is low for being used for All-year humid weather of Colombia or wet winters of Canada.

4.3 Carbon Dioxide (CO2)

Before starting to take a deeper look into MOVES capability in *Greenhouse Gas* (GHG) rate prediction, it is good to validate the PEMS measurements based on known facts from literature. Burning one liter of regular gasoline generates approximately 2.29kg of CO2 (the rate is 2.21 for E10 gasoline) [106]. Figure 23 represents a correlation analysis between the PEMS-based CO2 observations and the calculated CO2 rates based on the FCR.

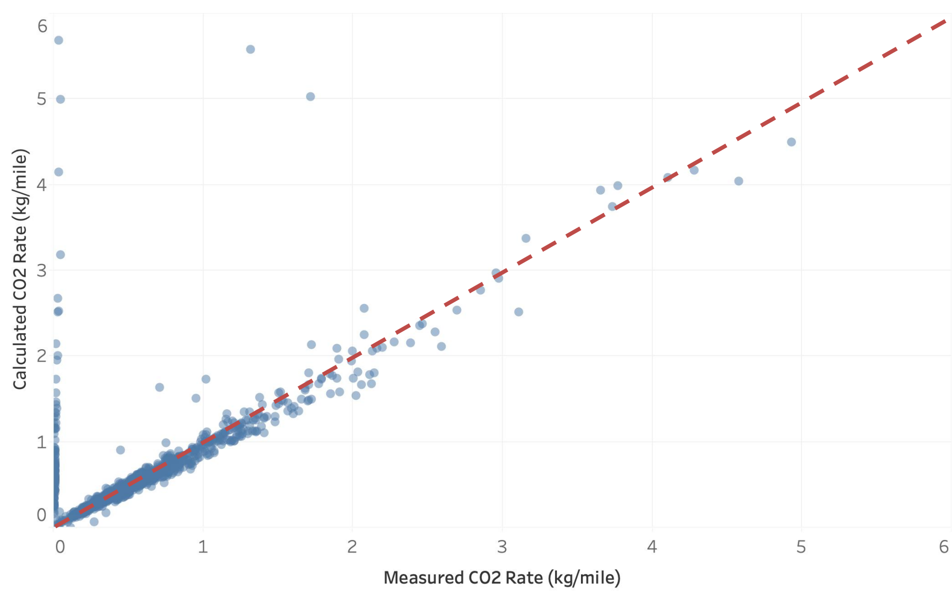

Figure 23: Directly-measured vs. indirectly-calculated CO2 rate

Except for a few outlier values where the PEMS unit has reported close to zero rates, the majority of the measurements show great agreement with the previously published finding regarding fuel-to-CO_2 conversion rates. The outliers could be the result of the instantaneous malfunction of the PEMS unit or instantaneous discontinuity of the exhaust stream which sometimes happens due to engine coughing phenomenon [107]. For the rest of the analysis, the above-mentioned few outliers are removed from the data.

As expected, all the CO_2-related comparisons conducted between MOVES predictions and PEMS observations are in agreement with that of ECR. However, in all the observations, the goodness-of-fit is considerably higher. It could be linked to the minor cumulative errors (approximations) produced while indirectly calculating the ECR based on engine attributes and *Mass Air Flow* (MAF) rate (refer to Section 3.3.2).

Moreover, less overall over/underestimation by MOVES is seen in the different categorizations (see examples in Figure 24). Nevertheless, an approximate 0.1 kg/mile initial bias (shift) in MOVES predictions is evident in the charts in low ranges of CO_2 rate. This could be linked to the fuel cut-off technology available in some vehicles in situations where there is no power demand (i.e., downhill cruising), the fuel stream to the engine stops to avoid waste of energy.

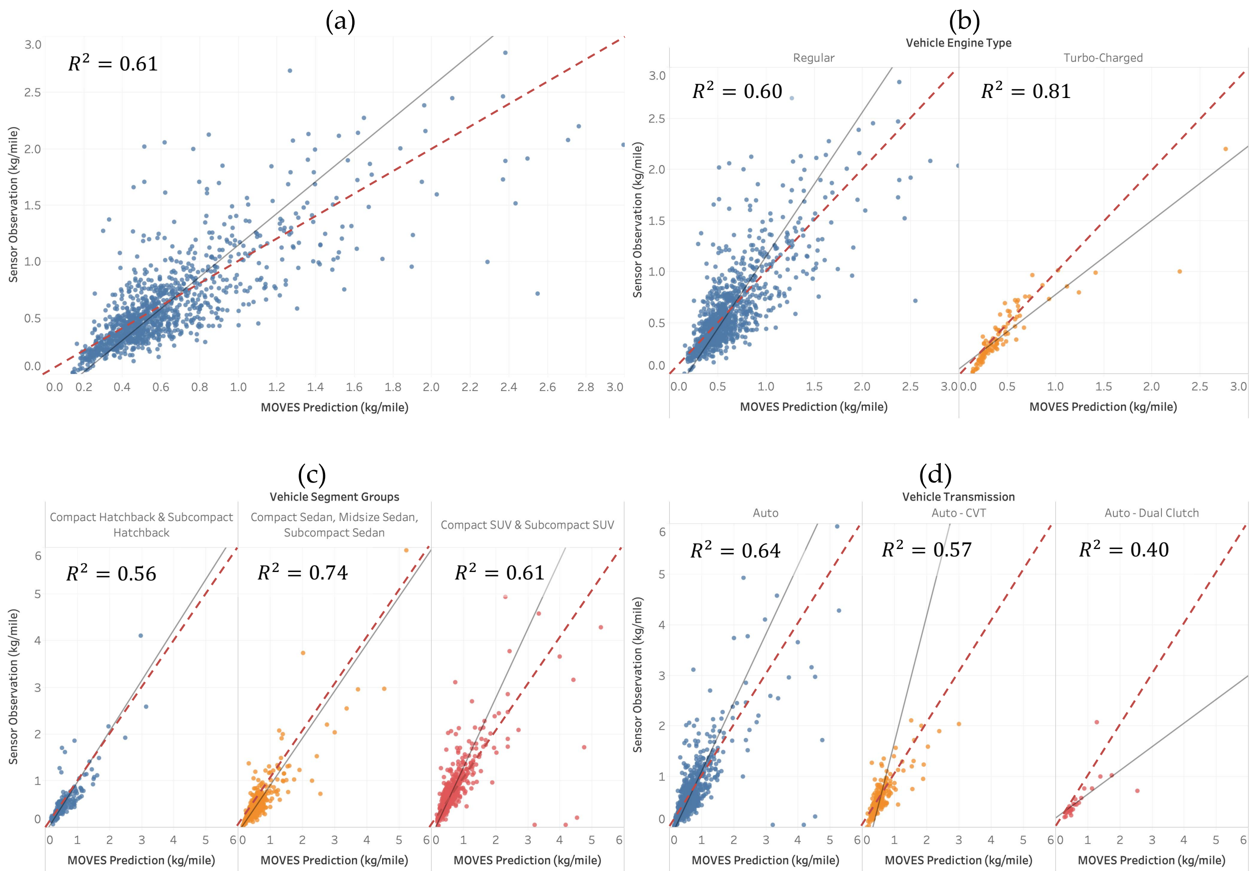

Figure 24: CO2 rate observations vs. MOVES predictions for (a) all links and categorized based on (b) engine type, (c) vehicle segment, and (c) vehicle transmission

VSP formulation used by MOVES (Equation 7) does not simulate such conditions as it is always positive while cruising (zero acceleration and positive speed). The VSP formula as presented first by Jiménez-Palacios [ref] is as follows:

$$VSP(kW/ton) = v(a(1+\varepsilon_i) + 9.81\, grade + 9.81\, C_R) + \frac{1}{2}\rho_a \frac{C_D . A_f}{m}(v+v_w)^2 . v \qquad (7)$$

Where v is speed, a is acceleration, the grade is in percentage, gravity is equal to $9.81\ m/s^2$, C_R is rolling resistance coefficient, C_D is the drag coefficient, A_f is the vehicle's frontal area, v_w is wind speed, ε_i addresses internal engine friction, and m represents the dead weight of the vehicle.

4.4 Nitrogen Oxides (NOx)

Around 95% of NOx in exhaust gas from a gasoline-engine vehicle is nitrogen monoxide (NO). The rest includes N2O and NO2 molecules. Figure 25 shows NOx-rate predictions against sensor observations for both NO and NO2 side by side.

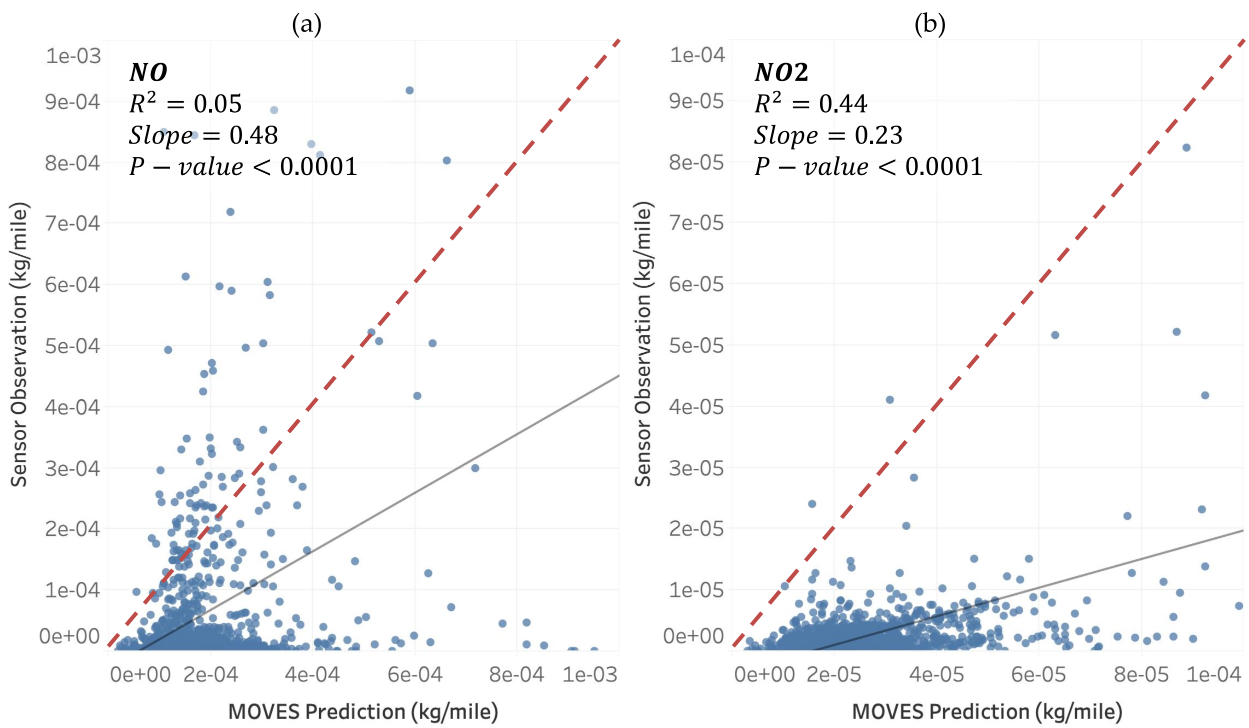

Figure 25: (a) NO- and (b) NO2-rate observations versus corresponding MOVES predictions

The most significant observation regarding MOVES NOx-rate predictions is the evident overestimation, with NO predictions possessing much higher variance. This could be due to the larger scale at which NO is emitted by gasoline-engine vehicles; especially, by the older ones. Evaluation studies by EPA workgroups [3,51] confirm this observation as the error in MOVES NOx estimates are projected to increase through time. NO2 is the dominant NOx emission in diesel-engine vehicles and a limited amount of it is observed in the measurements. A possible root cause of this overestimation lies in the catalytic converters' technology. Although the three-way catalytic converters are designed to reduce NOx levels (in addition to oxidizing carbon monoxide and unburnt hydrocarbons), their best efficiency only occurs in a narrow band around the stoichiometric air-to-fuel mixture level. Conversion efficiency falls very rapidly when the engine is operated outside of this band. If MOVES fails to capture the impact of catalytic converters, its predictions could become inaccurate like the observations above.

When categorizing NOx comparison charts by engine type, it is revealed that MOVES accuracy is lower for turbocharged engines. The increase in breathing capacity of the engine due to turbo-charging provides more oxygen for combustion, leading to more complete combustion. This effect leaves less oxygen in the exhaust, which results in a significant dropdown in oxidation/reduction reactions in catalytic converters. In general, it has been proved that turbocharged engines generate higher NOx, CO, and CO2 emissions compared to naturally-aspirated engines [108]. Thus, MOVES low prediction power, in this case, could be linked both to its weakness in simulating the effect of catalysts and the lower number of vehicles with turbocharged engines in its test fleet.

No other significant category-based finding for NOx emissions is achieved. The dominance of error levels in MOVES NOx-rate predictions does not allow extracting meaningful category-based insights. Nonetheless, a glance at Figure 26, presenting the sensor observations for NOx rates compared to MOVES predictions sorted by link average speed, reveals two important points. Although the pattern of true changes in NOx rates (by speed) is roughly similar to that of MOVES predictions, the MOVES overestimation is significant. In reality, the NOX rate converges to zero, whereas MOVES preserves a lower bound for its predictions.

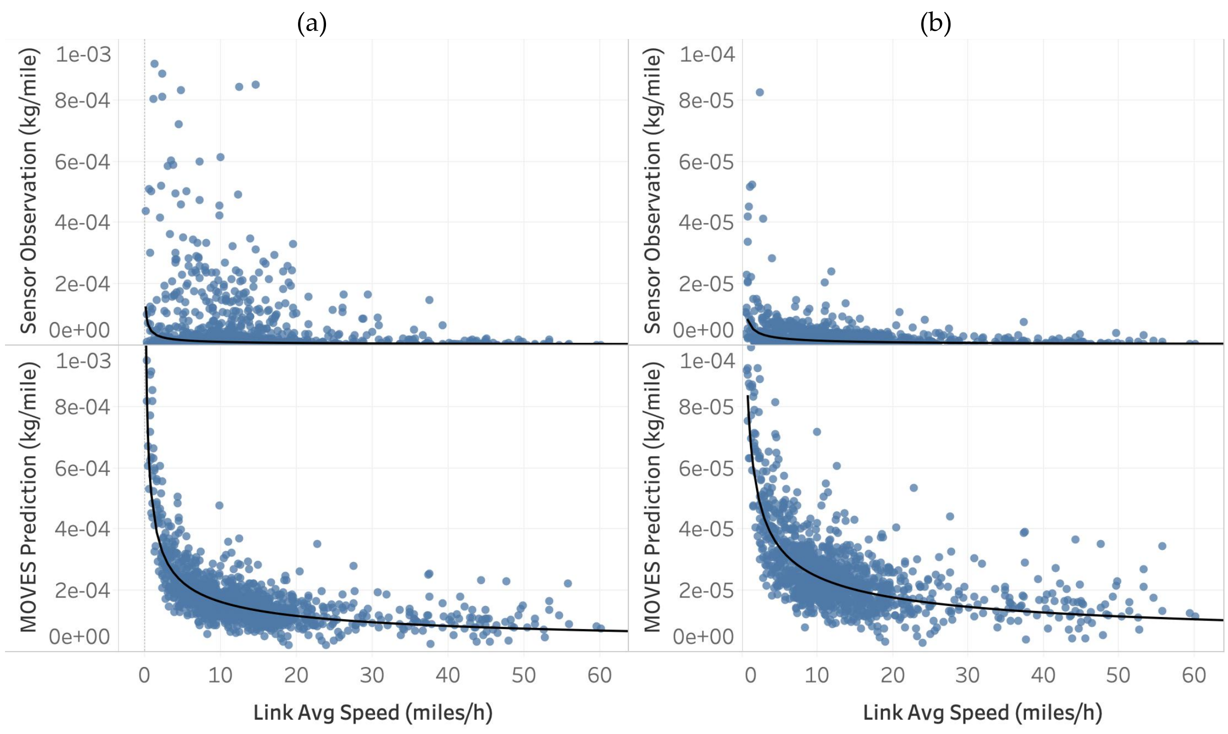

Figure 26: (a) NO- and (b) NO2-rate observations and MOVES predictions against link average speed

Moreover, even in MOVES predictions (which somehow simulate ideal conditions), much higher variance in NOx rates is observed for different link average speed values when compared to similar graphs for ECR or CO2. This indicates the nature of NOx generation in gasoline engine vehicles is relatively complex. Therefore, when modeling for the NOx rate, many more influential variables should be considered in addition to average speed.

4.5 Fine Particulate Matters (PM)

MOVES has the option to predict differently-sourced PM rates separately. However, our comparisons address the running-exhaust PM rates as the PEMS device we used is mainly designed for tailpipe measurement (hence, tire wear and brake wear particulate matters are disregarded in this study).

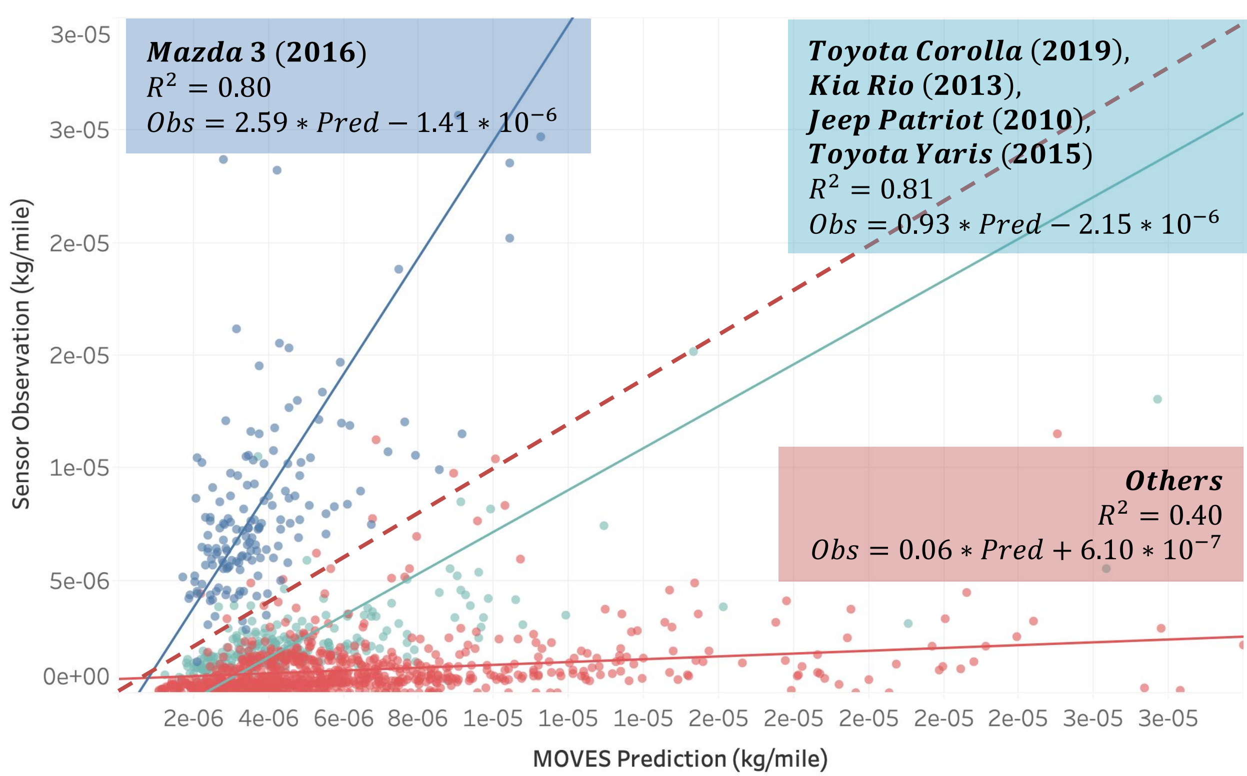

Figure 27: Tailpipe PM rate observations vs. MOVES predictions

Two anomalies are visible in Figure 27 both showing a high goodness-of-fit of around 0.8. However, for the blue part, corresponding to Mazda 3, underestimation is observed. For the cyan part, corresponding to 4 vehicles (Toyota Corolla, Kia Rio, Jeep Patriot, and Toyota Yaris), a good agreement with the bisector line and high R-squared score is evident. But for the rest of the vehicles, MOVES predicts with huge errors and generally overestimates PM rates. As all the PEMS measurements are conducted in Montreal, the observed segregation could not be linked to location differences. All four vehicles in the cyan area have regular engines with automatic transmission. But they are from different vehicle segments and have different total weights. They all used the same E10 fuel. They have had ages from zero to nine years and even the ambient humidity and temperature have been different while doing experiments on them. Therefore, the inconsistency in MOVES prediction accuracy could only be linked to the incapability of MOVES in simulating real-world conditions (especially, in regions rather than the U.S.) even though many adjustment options are incorporated in it.

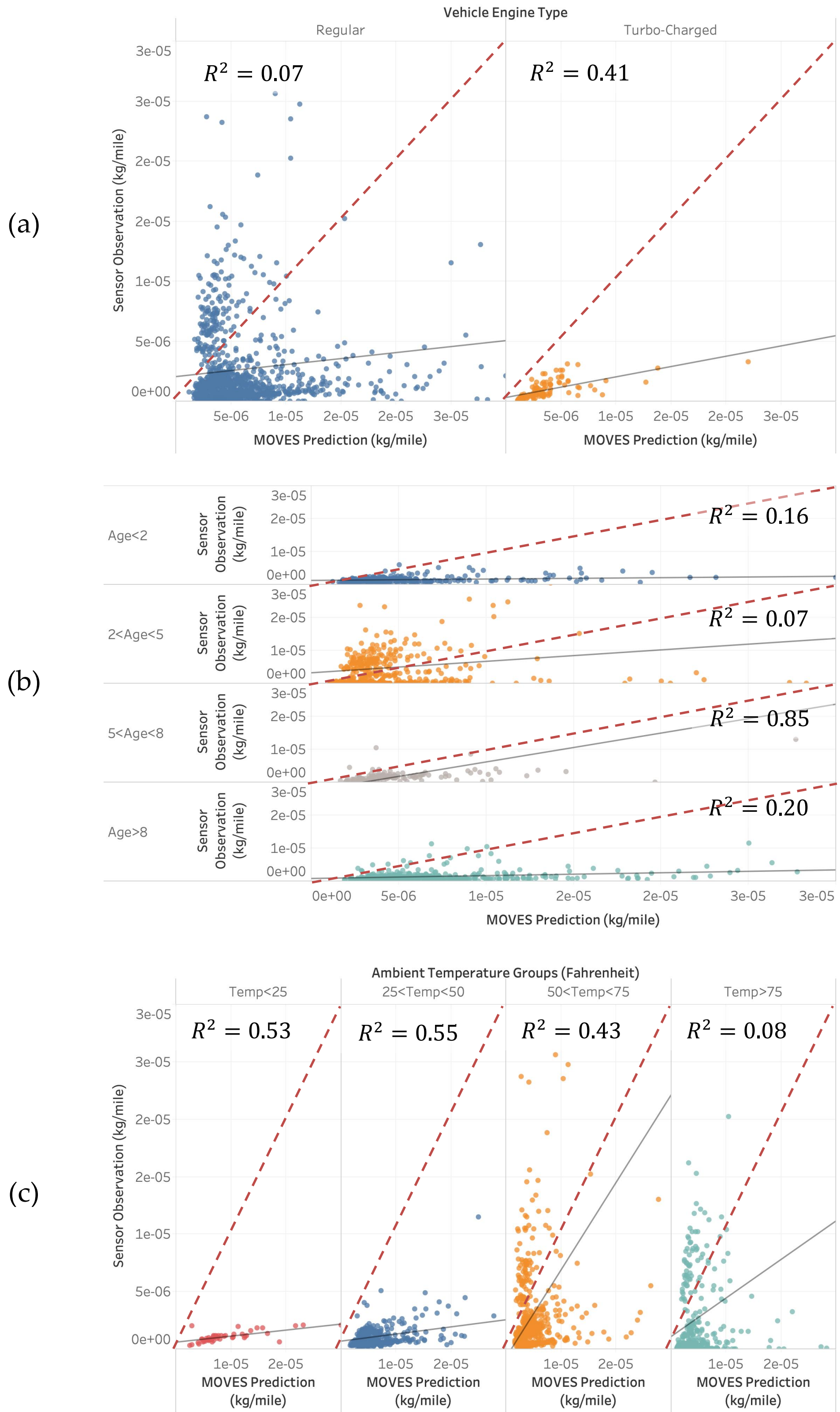

Figure 28: PM rate observations vs. MOVES predictions for (a) engine-type; (b) age; (c) ambient-temperature categories

The black regression lines show the over-estimation of PM rates by MOVES in different categorizations. Just like NOx, MOVES shows a slight initial bias of around 10^{-6} kg/mile in predictions; possibly, as a result of omitting fuel cut-off in models. Although predicting proportionally better for turbocharged engines (Figure 28-a), an R-squared of less than 0.5 is statistically insignificant. As shown in Figure 28-b, similar to the ECR and CO2 rates, MOVES has predicted acceptably for the age range of 5 to 8 years with an R-squared score of 0.85 while having negligible over-estimation. For the younger vehicles, MOVES core models are not adapted. Improvements in technology make the models estimated over older fleet less valid. Regarding the older vehicles, the physical depreciation of powertrain elements affects the normal behavior and efficiency of the vehicle. It is expected that predictions become unreliable as a result.

4.6 The Next Step

The knowledge regarding the level of MOVES predictions' validity for use in Canada (and other non-U.S. regions) sheds light on our way towards developing alternative local models. The sensitivity analysis conducted on MOVES output provided us with a valuable insight about how a model could lose its robustness due to changes in the fleet specifications, the road network characteristics, and the environmental conditions. Furthermore, the categorical analysis highlighted the need for category-based modeling of FCR and ERs, as such breakdown could dramatically simplify the modeling procedure by eliminating some of the impacting variables from the feature set.

In the following chapter, the results of a *Machine Learning* (ML) methodology for developing local microscale fuel models applied on data from a diverse fleet of 35 vehicles will be discussed. By limiting the feature set to the simple and easily retrievable kinematic variables (disregarding IEVs), improving the models' accuracy through a cascaded approach, and utilizing emerging ML algorithms capable of capturing complex and nonlinear interactions of variables, outperforming most popular commercial and academic models in terms of robustness, accuracy, and practicality is targeted.

Chapter 5

The Cascaded Approach to FCR Modeling

5.1 Introduction

All the modeling methodologies introduced in this study are oriented towards reducing the complexity of the vehicular fuel and emission models and increasing their practicality. Moreover, developing models capable of predicting *Fuel Consumption Rate* (FCR) with acceptable accuracy without the need to directly include *Internal Engine Variables* (IEV) in the feature sets is a major goal of the research. Hence, ML algorithms are assessed and a cascaded *Machine Learning* (ML) modeling procedure (explained in Section 3.5) is proposed which is expected to shoulder the burden of extracting the hidden impact of IEVs which usually possess significant correlations with FCR.

Choosing between *Support Vector Regression* (SVR) and *Artificial Neural Networks* (ANN) algorithms, primary evaluations are done first through the use of the combinations of the base features (speed, acceleration, and grade) as feature set. The impact of introducing engine speed directly to the feature set is then assessed. Finally, the engine speed is substituted by its estimations to shape and train a more accurate instrument-independent model.

Each modeling effort is followed by a category-based sensitivity analysis (similar to what was done in Chapter 4) to assess the possible limitations of the proposed methodology with respect to fleet, environment, or road-network attributes and to identify the combination of attributes which result in the highest prediction accuracies.

5.2 Primary SVR and ANN Modeling

Figure 29 demonstrates the magnitude and the variance of the best achievable scores for different vehicles, regarding each model structure evaluated (at different steps) and for both of the SVR and ANN methods. The first six rows (in green) correspond to the primary FCR modeling attempts. The next 8 rows (in red) correspond to finding the best achievable improvements in the model's prediction power when direct RPM observations are included in the feature set. The last 4 rows (in blue) show the results corresponding to the cascaded modeling step.

The experiment conditions (while collecting data) is kept as as constant as possible for all the vehicles through the use of the same measurement equipment, deployment of drivers with coordinated driving habits, and the same combination of road types traversed. Therefore, the significant variation of the scores most likely corresponds to the variety of the vehicles' technical specifications. This proves the necessity of avoiding the development of general models for FCR estimation and focusing on vehicle-/category-specific models.

The SVR scores for the primary FCR models range from 0 to 0.78, while the range is 0.08 to 0.79 for ANN. The combination of speed (v_t), acceleration (a_t), and grade (g_t) as input features result in the best SVR and ANN scores (comparing the median values for different structures). By including the lagged variables (v_{t-1} and g_{t-1}) different outputs from the two techniques are observed. SVR is not capable of capturing the lag effect. The best score improvement is only 0.01, whereas for most of the vehicles the scores drop dramatically. In contrast, the addition of lagged variables to the ANN feature set results in game-changing improvements up to 0.37 (i.e., from 0.32 to 0.69 for Hyundai Elantra GT 2019). Such improvement turns non-acceptable predictions of the model into much more reliable ones.

Comparing the median scores for the 1st, 2nd, 4th, and 6th structures (see Figure 29), the lagged grade (g_{t-1}) contributes much more than lagged speed (v_{t-1}) to the models. Despite all the exceptional cases of improvement, the median score for 5th and 6th structures are almost identical (1% difference). Nevertheless, the hypothesis regarding the extended temporal impact of speed and grade values on FCR could not be easily rejected.

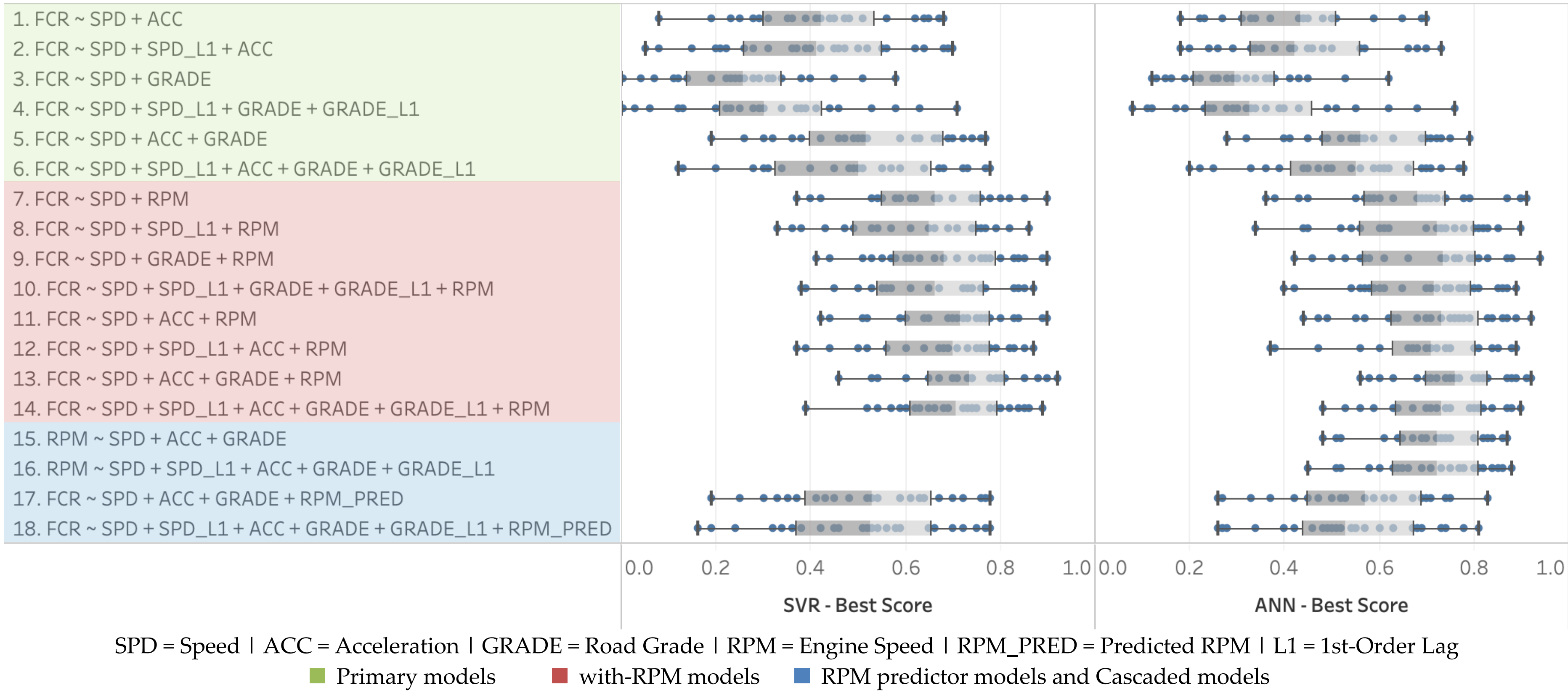

Figure 29: Score variation achieved for different model structures and vehicles

Comparing the scores corresponding to 1st, 3rd, and 5th structures (both for SVR and ANN), improvements of around 25% in the median score is observed, while the grade improves this score only by 9%. A part of this considerable difference could be the result of GPS measurement errors. The accuracy of the altitude measurements is highly affected by the number of satellites found by the GPS receiver [109]. Even though the outliers are filtered out and a smoothing algorithm on altitude data is applied prior to grade calculation and modeling, a residual error could still exist which makes the grade values less accurate and disrupts its true impact.

5.3 Optimum Model Settings

A majority of the best SVR results are achieved using ε value of 0.1, C values equal to 1 and 10, and γ values equal to 0.1, 1, and 10. Note that no penalty is associated in the training loss function with points predicted within a distance ε from the actual value. C is the regularization parameter to control overfitting and γ controls the level of nonlinearity (curvature) of the multidimensional hyperplane fitted on data. The optimum ε value of 0.1 guarantees a reasonable soft margin compared to observed FCR values, which are mainly ranging from ~1 L/H to 10 L/H in our experiments. More than 80% of the best results are achieved using average regularization (parameter C). This indicates that the choice of the *Radial Basis Function* (RBF) as the kernel has been appropriate, avoiding extreme overfitting. Nonetheless, diversity of optimal γ values (as a parameter controlling the way SVR treats the similarity of pairs of observations) is a sign of diversity in energy consumption procedures among different powertrains (as a result of technology differences).

On the ANN side, dominancy is neither with the wide and shallow nor with the deep and narrow neural network architectures. 58% of the best scores are associated with an ANN with two hidden layers each having 64 ReLU-activated neurons and only a negligible share of 0.2% are the results of our deepest alternative architecture. Thus, in general, there is a limit for internally estimating and using the unobserved influential variables by increasing the depth of ANNs.

5.4 Introduction of Engine Speed to the Feature Set

By adding RPM to the feature set a dramatic increase in model scores for all the vehicles (except one) is observed, either with SVR or ANN. As shown in Figure 30, the only vehicle showing a decrease in model score by the introduction of RPM to the feature set is IKCO Dena. Although it is a 2016 model, it is built on an old Peugeot platform from the early 1990s and many manufacturing flaws are reported about this model. Driving with full gas, heavy braking, and inappropriate gear selection with a manual-transmission vehicle or overall malfunctioning of the car could lead to such results. However, strict control was applied to the driving styles of drivers and the good maintenance of vehicles before tests was ensured. There were no tangible issues with the power transmission elements such as clutch plates, differential, axles, etc. Thus, this could only be linked to a technological flaw or faulty gear-ratio settings in the manual gearbox.

The levels of improvement for the rest of the vehicles are not in harmony for SVR and ANN. For example, for Kia Rio 2013, the SVR score increases only by 7%, while a significant increase of 47% is observed with ANN. On the contrary, for Honda Civic 2014, the improvements for SVR and ANN are 36% and 2%, respectively. As the two vehicles are almost the same age, the only observable differences between the vehicles (among the variables observed in this study) are vehicles' segment (compact sedan vs. compact hatchback), engine displacement (1.8L vs. 1.6L), and transmission type (*Continuous Variable Transmission* or CVT vs. Regular Automatic). While focusing on modeling score improvement due to the addition of RPM to the feature set, the most probable root cause would be the transmission system differences. Thus, a possible conclusion is that SVR performs better with CVT-equipped vehicles and ANN is the right choice for non-CVT vehicles. This is a reasonable judgment as CVT eliminates the sudden shifts between gear regimes occurring in manual or regular automatic gearboxes. CVT's pully-based mechanism changes seamlessly between a continuous range of effective gear ratios. It synchronizes the kinematic state of the vehicle and the RPM, and reduces the complexity of fuel consumption relationship with features. Moreover, this smoothness/coordination is compatible with what the RBF kernel virtually applies to feature interactions inside the

SVR model. Many of the top record holders are equipped with the CVT transmission. Later in this chapter, a categorical analysis is performed over the transmission type.

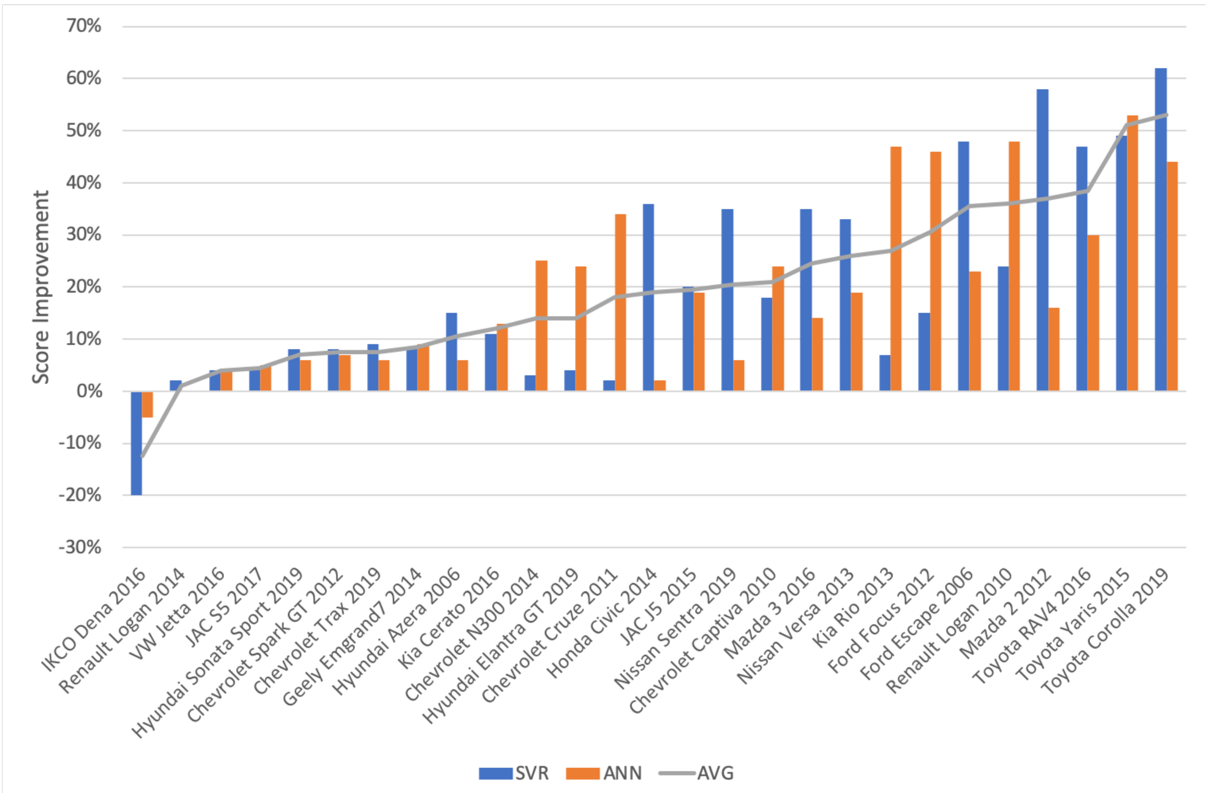

Figure 30: Best improvements by including original RPM to SVR/ANN models for the top model structure

To complete the cascaded modeling procedure, we estimate the RPM using separate models (with an acceptable level of error) and then use the estimations instead of real-world observations. The 15^{th} and 16^{th} rows in Figure 29 show the best RPM estimation scores using the 8 alternative ANN architectures used. The change in the range of the scores compared to that of FCR estimations (only using the base feature set) is noticeable. This proves the flexibility of ANN in capturing underlying multi-variate interactions. The scores top at 0.88, while the median score is about 0.73. Interestingly, all the best results are achieved by the original four ANN alternative architectures and not by any of the newly added alternatives (the wider ones).

The best results achieved through training SVR and ANN models using the new feature set (including the predicted RPM) are presented in Figure 31. Apparently, SVR shows extreme weakness in capturing complex interactions due to the introduction of the estimated variable. Improvements are either negligible or at most covering one-fourth of the score difference between non-RPM and with-RPM models. For some of the vehicles, the addition of the predicted RPM has led to a reduction of model score, which could be associated with SVR limitations. Observations regarding vehicles like VW Jetta 2016 or Hyundai Elantra GT 2019 with RPM prediction scores higher than 0.8 support this claim.

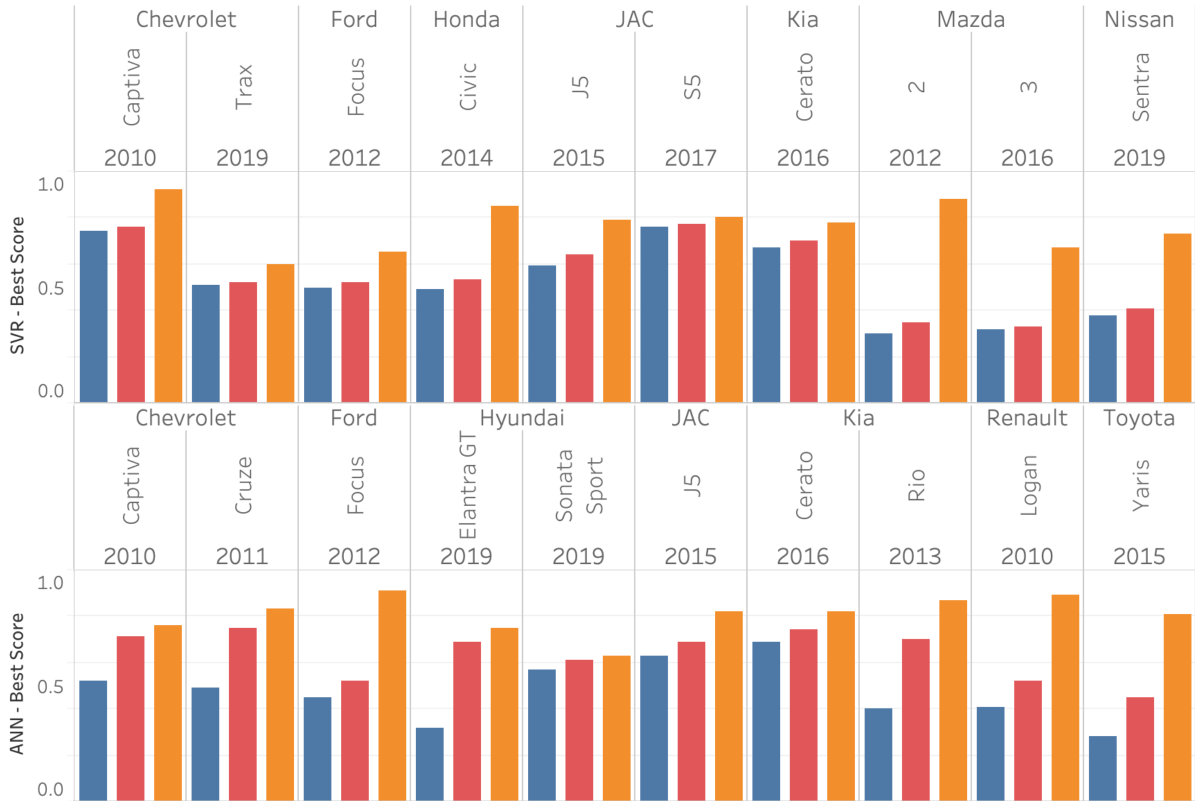

Figure 31: Score comparison between models with base feature set (blue), cascaded models (red), and ideal models including true RPM observations (orange) for 10 vehicles with best results

On the other hand, ANN shows more promising results than expected. For all the vehicles with a corresponding RPM prediction score of 0.75 or higher, a considerable improvement towards the ideal outputs has occurred. Chevrolet Cruze 2011 and Kia Rio 2013 have had the best improvement records, compatible with their corresponding RPM prediction accuracy which are among the highest. Nevertheless, there are still a few vehicles with RPM prediction accuracy of lower than 0.75. As a result, ANN has not been able to improve their corresponding model scores (and even the results have been worsened) compared to the base model.

Results of this limited attempt to separately model RPM indicates that neural networks have much higher potential to be used in the cascaded modeling procedure. However, lower degrees of prediction error (less than 20%) seem to be a prerequisite for guaranteed improved results for the majority of the vehicles. This could be achieved either by increasing the size of training data, testing more complex ANN architectures, or switching to more sophisticated neural networks techniques such as 1-D *Convolutional Neural Networks* (CNN) or *Recurrent Neural Networks* (RNN).

5.5 Comparison to the State-of-the-Practice Models

To compare and universally validate the cascaded model, two state-of-the-practice models, *Virginia Tech's Comprehensive Power-based Fuel Model* (VT-CPFM) [17] and USEPA's MOVES [35] are selected. The two benchmark models are applied to the same input data. VT-CPFM provides outputs at the same resolution of the cascaded model (second-by-second rates). However, MOVES output at its highest resolution is in form of energy consumption per unit of distance (or time) per link. In Figure 32, true FCR values are shown alongside the cascaded model and the VT-CPFM (Type I) output. This figure clearly shows how close the cascaded model predictions are to the true FCR values, how the real underlying trends and variations are being followed by the model, and how it outperforms the VT-CPFM Type I.

By doing aggregation and some post-processing, the ground-truth, the cascaded model output, and VT-CPFM (Type I) output are re-scaled to the same resolution of MOVES output (Joules per miles per link). Then, an additional comparison between model accuracies is conducted and results are visually presented in Figure 33. The superiority of the cascaded model is evident in this new comparison as well. The cascaded model still outperforms both MOVES and VT-CPFM (Type I) despite the aggregation being conducted over the instantaneous predictions.

It is important to note that to run the VT-CPFM model, vehicle mass, drag coefficient, frontal area, number of cylinders, engine size, number of gears, gear ratios, and final drive ratio were retrieved from the manufacturer online records. Regarding the rolling coefficient, c1, c2, driveline efficiency, and wheel slippage, the general information presented in Rakha *et al.* publication [17] is used. For the fuel-related parameters (idling fuel mean pressure and the fuel lower heating value), adjustments are done depending on the type of fuel, especially because the E10 gasoline has lower energy efficiency compared to regular gasoline. Moreover, some parameters such as the idling engine speed are extracted from the measurements of each vehicle.

Table 4 compares the R-squared score of VT-CPFM (Type I) with that of the cascaded model for second-by-second predictions. R-squared scores regarding the comparison between link-level distance-based ECR predicted by the cascaded model, MOVES, and VT-CPFM (Type I) are shown in Figure 33 next to each chart.

Table 4: Score comparison between VTCPFM and the cascaded model

	R-squared Score	
	VT-CPFM (Type I)	**Cascaded ANN**
Hyundai Elantra GT 2019	0.57	0.72
Chevrolet Captiva 2010	0.26	0.77
Chevrolet Cruze 2011	0.51	0.78

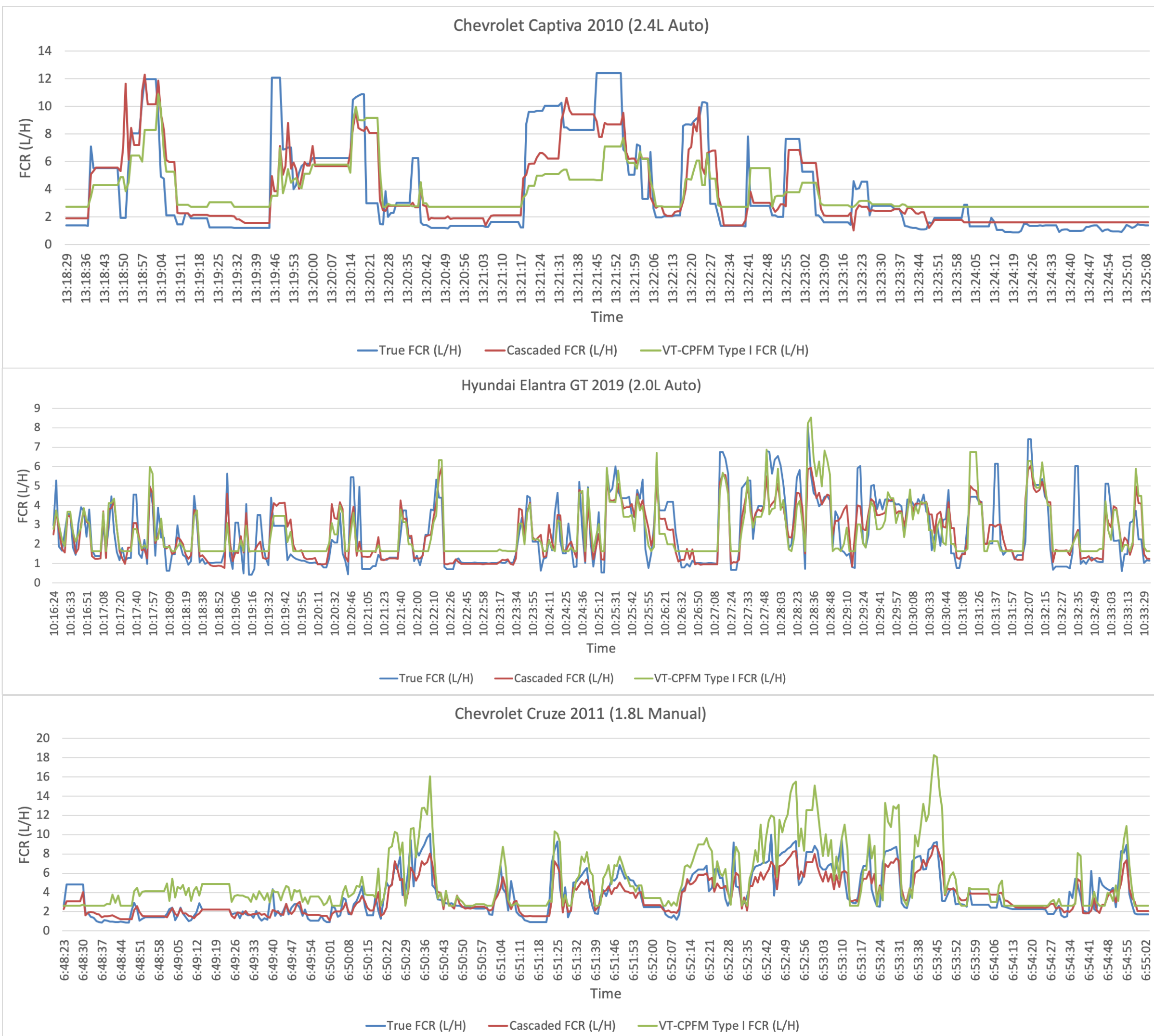

Figure 32: Time-series visualization of true FCR, cascaded, and VT-CPFM outputs

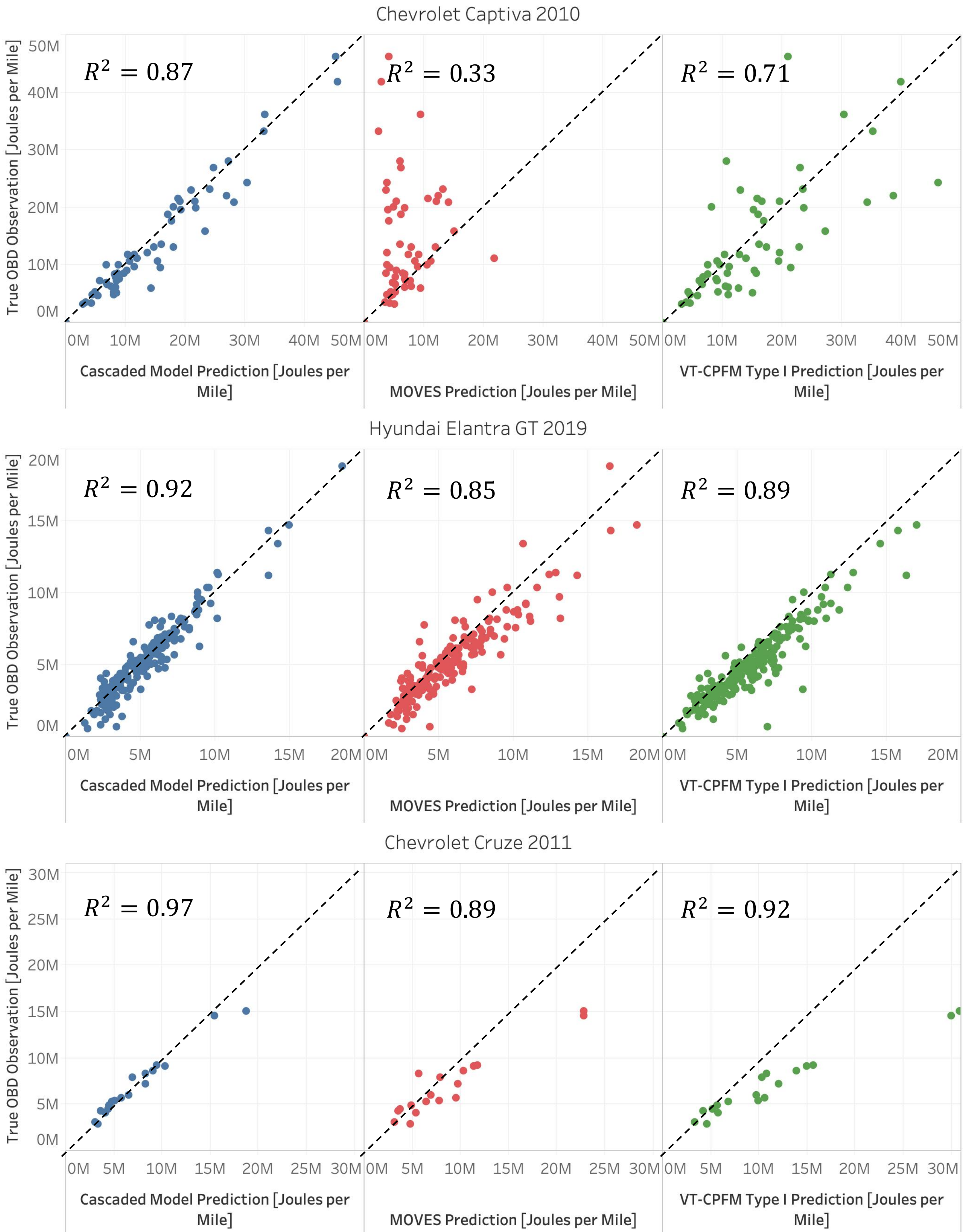

Figure 33: Link-level ECR predictions by cascaded model, MOVES, and VT-CPFM

MOVES has inconsistently predicted the ECR for different vehicles. As shown in Figure 33, although generating reasonable estimates for Hyundai Elantra GT 2019 and Chevrolet Cruze 2011, it has significantly underestimated ECR for Chevrolet Captiva 2010. Interestingly, even VT-CPFM predictions are more accurate than that of MOVES. Overall, a higher dispersion (from the ground-truth) exists in MOVES as well as VT-CPFM outputs when compared to the cascaded model.

5.6 Categorical Sensitivity Analysis of the Cascaded Model

In this section, an investigation is conducted on the modeling results by categorizing them based on car segment, transmission type, engine type, engine displacement, and age. It is hoped that such a categorical analysis helps developing category-specific models with less prediction error in the future. In Figure 34, an almost identical pattern of scores is observed between the two modeling methods. The only minor differences are that SVR results in higher scores for compact hatchbacks (compared to ANN), while ANN shows more power with compact vans.

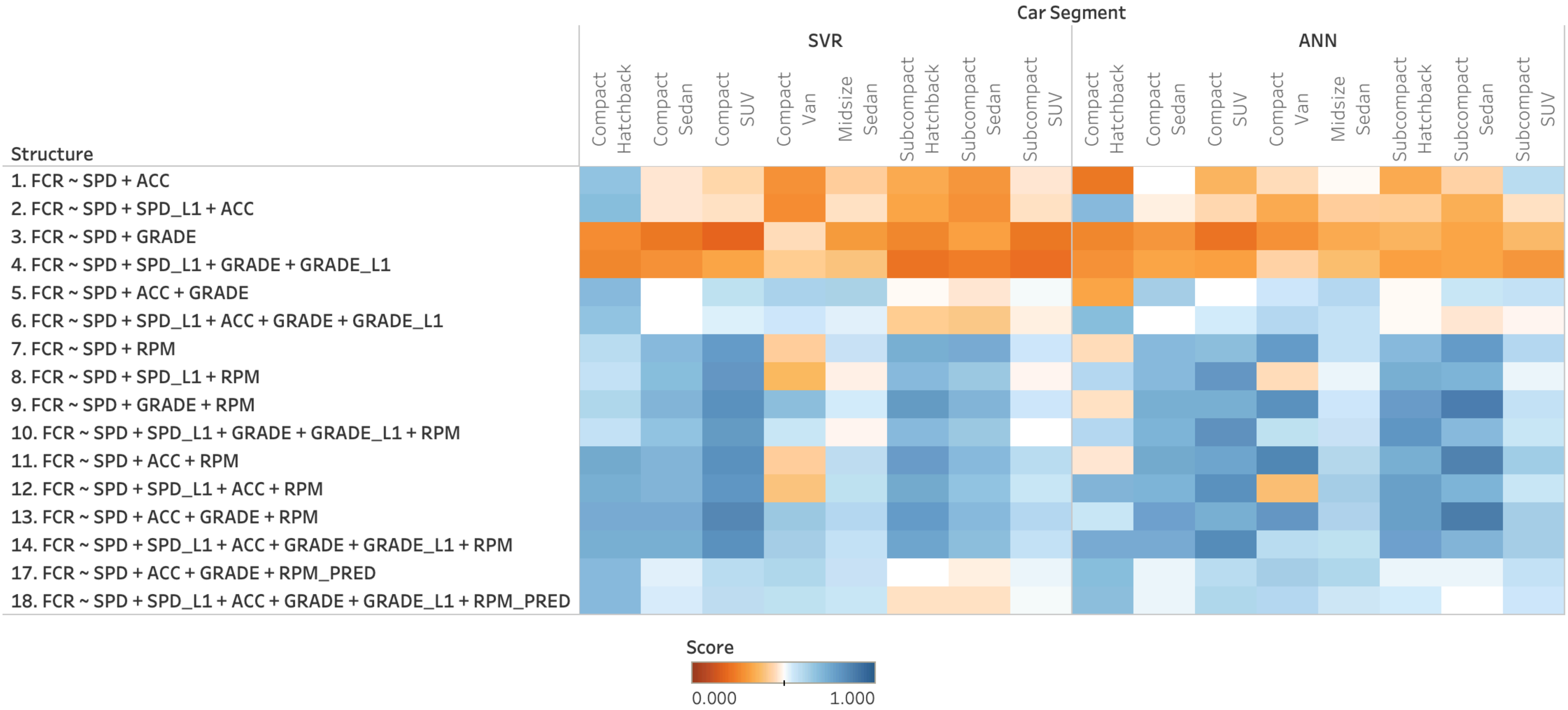

Figure 34: SVR/ANN median best score per car segment for different model structures

The superiority of with-RPM and cascaded models in comparison with the six base structures is visible in this figure. Compact SUVs (crossovers) and subcompact hatchbacks show the most dependence on RPM. However, compact hatchbacks have had the most significant cascaded modeling scores both with SVR and ANN. On the other hand, there is a significant difference between SVR and ANN scores for the subcompact hatchbacks. Compact and subcompact sedans, despite showing high dependency on RPM (in with-RPM models), have shown disappointing cascaded modeling results. In general, the cascaded modeling approach could be declared successful with the following three segments: compact hatchbacks, midsize sedans, and subcompact SUVs, as they have the least difference between average with-RPM and cascaded modeling results. Making a robust inference is impossible about the compact vans as there exists only one vehicle (Chevrolet N300) from that segment.

Vehicles with turbocharged engines show higher modeling scores compared to the vehicles of the same segment with regular engines. Compact SUVs have the best scores among the turbocharged vehicles examined (both with SVR and ANN), while compact sedans and subcompact hatchbacks show better results among regular-engine vehicles. Turbocharged engines have higher fuel efficiency as extra compressed air is fed into the combustion chamber avoiding unburnt or incompletely burnt fuel during the combustion process. That is exactly the reason why the trained models show more success with turbocharged engines. The FCR values used for modeling are indirectly calculated based on the intake air flow rate (logged using the MAF sensor). MAF variable does not account for losses due to incomplete combustion. Turbocharging technology reduces this unobservable natural error and helps the models gain a higher score and better prediction power.

A closer look at the modeling results for different engine displacements (engine volume) shows that vehicles with high engine displacement (>2.4L) have a relatively lower best score from both modeling techniques. Furthermore, the best results correspond to vehicles with average-size engines. It is important to note that two out of the three vehicles with the highest engine displacements are older than 10 years and the low scores could be associated with age-related efficiency losses rather than engine displacement.

FCR in vehicles with CVT transmission shows high dependence on RPM and this is apparent in with-RPM modeling results shown in Figure 35. Almost the same dependency is observed for the vehicles with dual-clutch automatic transmission. However, ANN seems more successful in capturing RPM's impact on FCR for dual-clutch ones. Rather than conventional automatic transmission systems, CVTs use a combination of chain/belt and pulley instead of gears for power transmission. There is no fixed number of gears in CVT and a pulley’s diameter could continuously change on demand. As a result, there would be no RPM surge/drop such as when conventional automatic transmissions shift gears. This leads to a much better fuel economy. The cascaded models show acceptable results for manual (with ANN only), automatic, and dual-clutch transmissions. Nonetheless, they show up weak for the CVT vehicles. As the RPM predictions score (using ANN) for the four vehicles with CVT has been on average 0.7, this weakness could be interpreted as the need for much higher RPM prediction accuracy for such vehicles. This is compatible with the RPM-dependent nature of FCR in these vehicles.

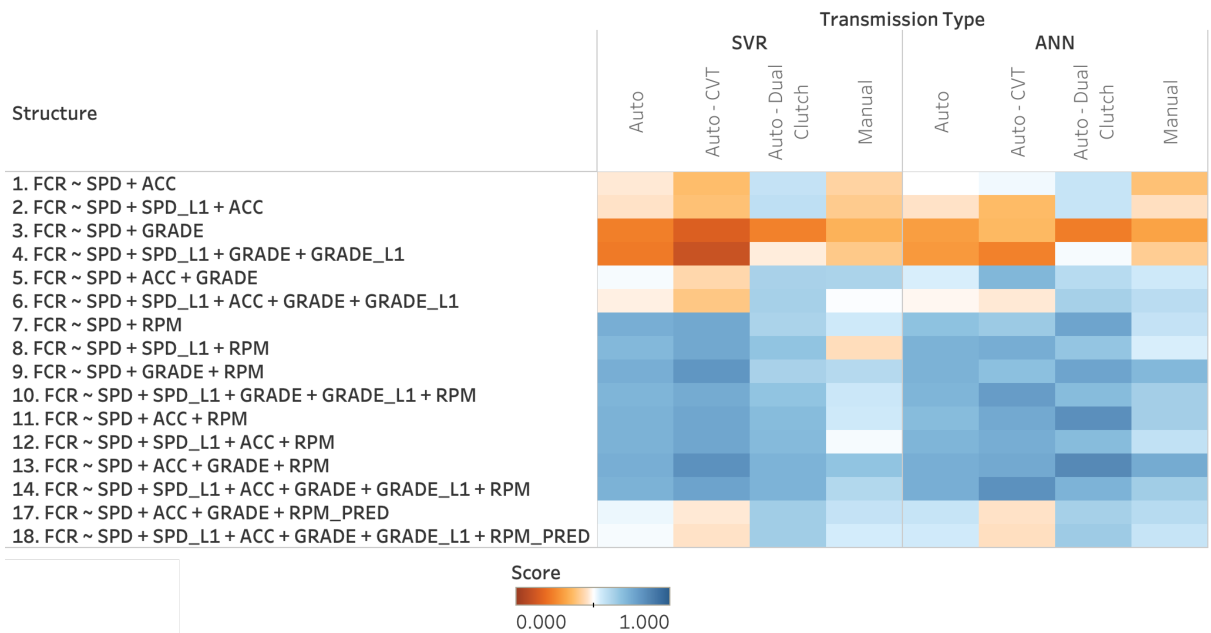

Figure 35: Median best score for different transmission technologies and different model structures using SVR and ANN

No significant trend is visible in model best scores as the vehicle age increases. This emphasizes the impact of maintenance quality and the need for quantifying it if including this attribute as a feature in the model is desired. The maintenance quality could be included in the model in the form of a correction factor as well. However, to extract such a measure, a thorough investigation of each car is required.

5.7 The Next Step

The promising results of the developed cascaded model confirmed the potential of ML not only in improving the prediction accuracies, but in simplifying practical use of the models. The RPM was only one of the influential IEVs and the prediction power of the cascaded procedure could be improved by inclusion of other variables such as *Throttle Position* and *Engine Torque*. Our basic effort in addressing the lagged effects of variables through direct inclusion of 1^{st}-order lagged variables in SVR and ANN architectures did not result in a significant outcome. Hence, evaluation of other ML algorithms developed specifically for time-series modeling should be done in future steps.

The categorical analysis results, especially the findings regarding the role of transmission type and vehicle class on the accuracy of model predictions, emphasized the necessity of category-based modeling. By having models calibrated for particular categories, it is expected that the variance in predictions would reduce and categorical models move towards the optimal trade-off between bias and variance.

In the next chapter, we first try to assess the possibility of simultaneously capturing the instantaneous as well as the lagged impact of variables on both FCR and ERs using more sophisticated ML algorithms. Then, forecast combination techniques are evaluated and compared for developing category-specific models using previously developed vehicle-specific models as building blocks.

Chapter 6

An RNN-based Two-stage Ensemble Learning

6.1 Introduction

Although achieving significant forecasting scores, having simple structures, and relying on instrument-independent variables as input, the *Fuel Consumption Estimation* (FCR) estimation models developed in Chapter 5 could at best be deployed in practice for vehicles with a similar make, model, and age. An alternative method, would be reverting to the use of complex parametric models which try to tune themselves based on vehicles' technical information. But such information might not be available for every vehicle when incorporating fuel and emission models into traffic micro-simulator software or if they were to be used for eco-driving purposes. Moreover, the compatibility of the models with the serially-correlated nature of the vehicular fuel consumption and emission generation and capturing the theoretically-proved lagged impact of variables (refer to Section 3.6.1), is not addressed yet.

In this chapter, the results of using *Recurrent Neural Network* (RNN) technique, a *Machine Learning* (ML) algorithm exclusively designed for time-series modeling and capable of capturing short and long dependencies in data, along with a forecast combination step for maximizing the prediction power of vehicle-specific FCR and *Emission Rate* (ER) models will be discussed. Then the outputs of a bottom-up procedure for combining the forecasts of RNN-based vehicle-specific models in search for generalized category-specific models is presented and discussed. Similar to the previous chapters, the discussion is followed by a complementary categorical analysis to identify the criteria which categorizing vehicles based on them lead to the highest improvements in modeling scores.

6.2 Lag-specific RNN Models

Figure 36 shows the share of different RNN settings (including RNN type and the depth of recurrent architecture) leading to the *best* modeling scores for the vehicles at each lag order.

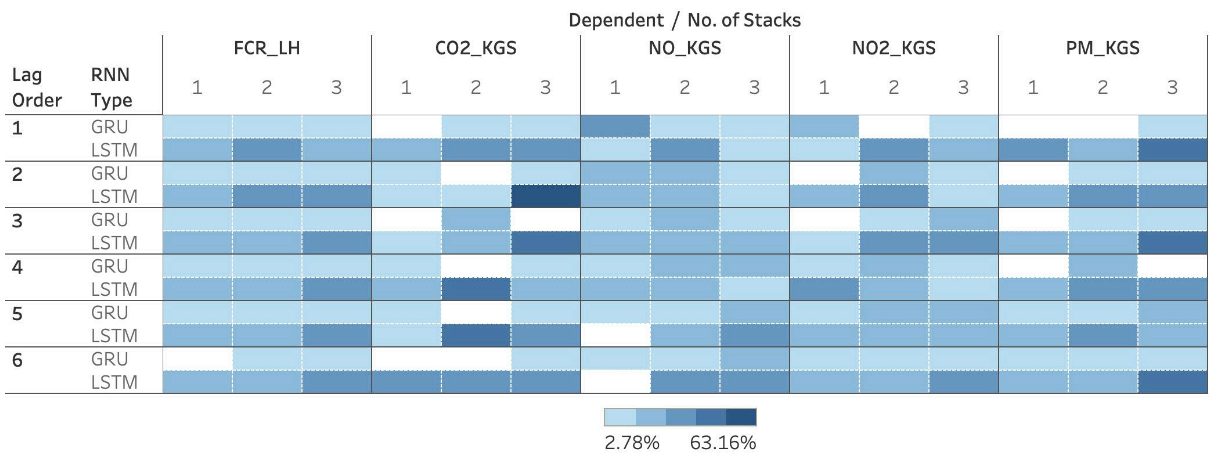

Figure 36: Share of different RNN settings leading to best RNN scores for each Lag-Dependent pair

Long Short-Term Memory (LSTM) mechanism worked the best for a significant number of the vehicles and almost at all the lag orders. Nevertheless, *Gated Recurrent Unit* (GRU) mechanism will not be disregarded from the next steps of modeling as it has outperformed LSTM for some vehicles. This observation emphasizes that for some powertrains, the extent of lasting dependencies is meaningfully shorter. Besides, keeping GRU models is reasonable as they execute faster than LSTM and are appropriate for smartphone-based eco-driving services due to the lower computation resources it requires.

Except for *Nitrogen Monoxide* (NO), having more than one layer of RNN (2 or 3) seems desirable. Based on the observations, the level of NO emissions has low volatility which lets less-complex modeling architectures predict its rate acceptably. Interestingly, the pattern of shares seems quite repetitive for FCR models, where at all lag orders the majority of best models are achieved using LSTM with 3 stacked layers. A possible inference is that the sources of lag influencing the FCR have complex, nonlinear, and lasting impacts, but are *not* from diverse roots. On the other hand, the absence of repeated patterns and the diversity in the distribution of best settings for the emissions could be linked to the combined impact of different sources of lag mentioned earlier in Sections 3.3.2 and 3.6.1. There is a proven significant correlation between FCR and CO2 rate [110]; thus, it seems reasonable that they possess similar lagged dependencies and have similarities between shares reported in Figure 36. But the differences confirm the disturbance which different sources of lag impose on CO2 rate observation patterns (especially, the impact of the exhaust pipe, resonators, and catalytic converter).

6.3 Vehicle-specific Metamodels

The exceptional power of forecast combination algorithms was revealed during our metamodel development stage. In only 10 out of 103 metamodels, no improvement was observed when comparing the *Root Mean Squared Error* (RMSE) scores with that of the best lag-specific models trained for each vehicle-dependent pair. For the rest of the metamodels, improvements up to 28% and on average 4% were obtained. Figure 37 shows the frequency distribution of improvements among the 103 metamodels. The notion that the ensembles always perform better than the component models is not guaranteed. However, it highly depends on the type of ensemble estimator as well as the level of the weakness of the component model.

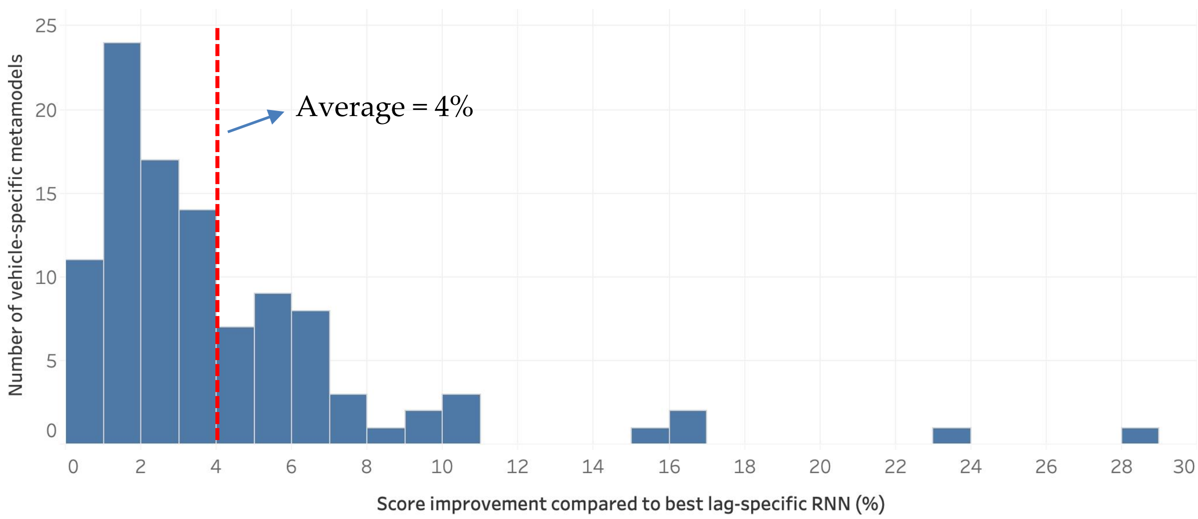

Figure 37: Score improvement rate histogram for best vehicle-specific metamodels

For example, none of the averaging methods evaluated in this study resulted in the best vehicle-specific metamodels. In addition, for the vehicle-dependent pairs where the lag-specific models performed quite well, the level of improvement by the ensemble estimator has been negligible.

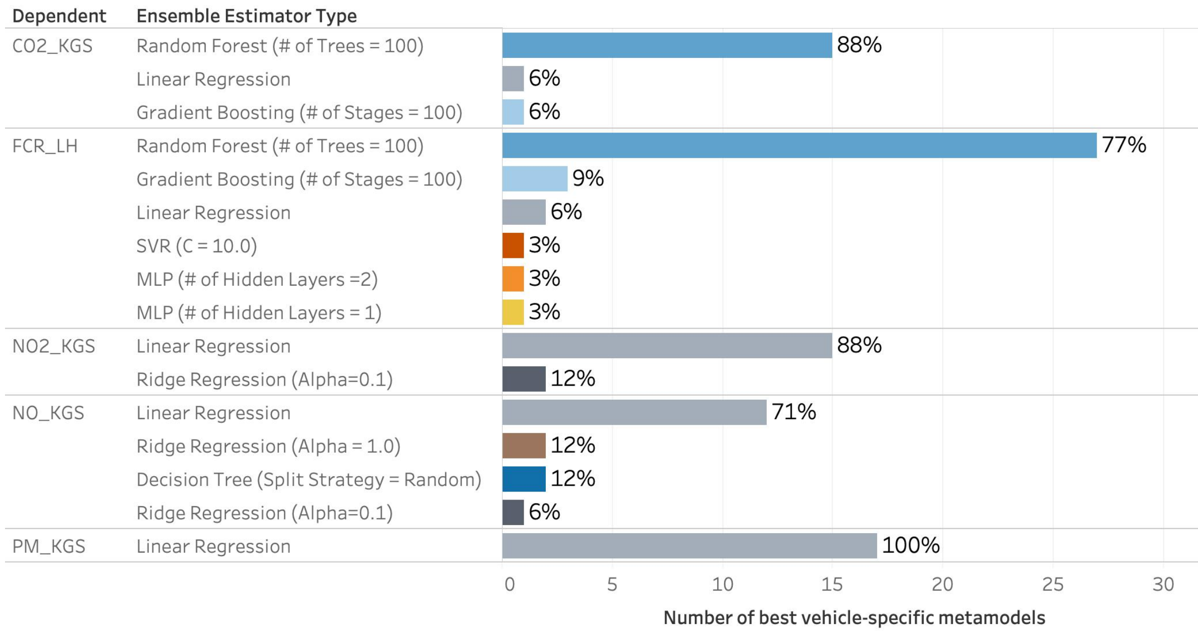

Figure 38: Share of best metamodel estimators for different dependents

Figure 38 shows the segregation between fuel consumption and CO2 generation mechanisms on one hand and that of other emissions on the other hand. The *Random Forest* algorithm has been the dominant metamodeling ensemble technique (leading to best scores) for FCR and CO2 rate, whereas the much simpler method of *Linear Regression* has resulted in the best results for *Nitrogen Dioxide* (NO2), NO, and *Particulate Matter* (PM) rates. Two facts could be deduced in this regard. First, small differences exist between the predictions of different lag-specific models (as inputs of the EL models) for FCR and CO2 rates. Hence, only sophisticated EL algorithms could extract underlying nonlinear dependencies and achieve considerable improvements. It is noteworthy that the 28%, 23%, and 16% improvement records are all dedicated to FCR and CO2 rate metamodels (an average of 6% improvement resulted in this group of metamodels). Such high improvements confirm the existence of higher-level nonlinear dependencies the lag-specific RNNs were incapable of capturing. Second, the lag-specific RNN predictions for NO, NO2, and PM are varied enough and as inputs to metamodels, they possess linear correlations with the dependent letting simple unregularized linear regression algorithm combine them and achieve improvements. Furthermore, it is important to note that the average improvement for these three emissions is equal to 2%. Interestingly, this could be due to the dominant effect of one lag-specific model on the metamodel performance. Therefore, a possible interpretation is that for NOx and PM emissions, the existence of a relatively constant lag order is feasible, while for fuel and CO2, distributively lagged effects exist.

Predictions of vehicle-specific metamodels regarding 3 sample vehicles are presented in Figure 39. The *Ensemble Learning* (EL) algorithms show undeniable effectiveness for FCR and CO2 rate. It is interesting how the EL algorithms have corrected some of the wrong local trends predicted by lag-specific RNN models (see regions highlighted in yellow). Moreover, metamodels have compensated component models' weakness in predicting sudden spikes (see regions highlighted in magenta). Even for the NO2 (as well as PM and NO), despite the higher level of prediction error, the metamodel outperforms the lag-specific component models. In Figure 40, the true observations are compared to the metamodel predictions for all datapoints corresponding to the same three vehicles discussed in Figure 39.

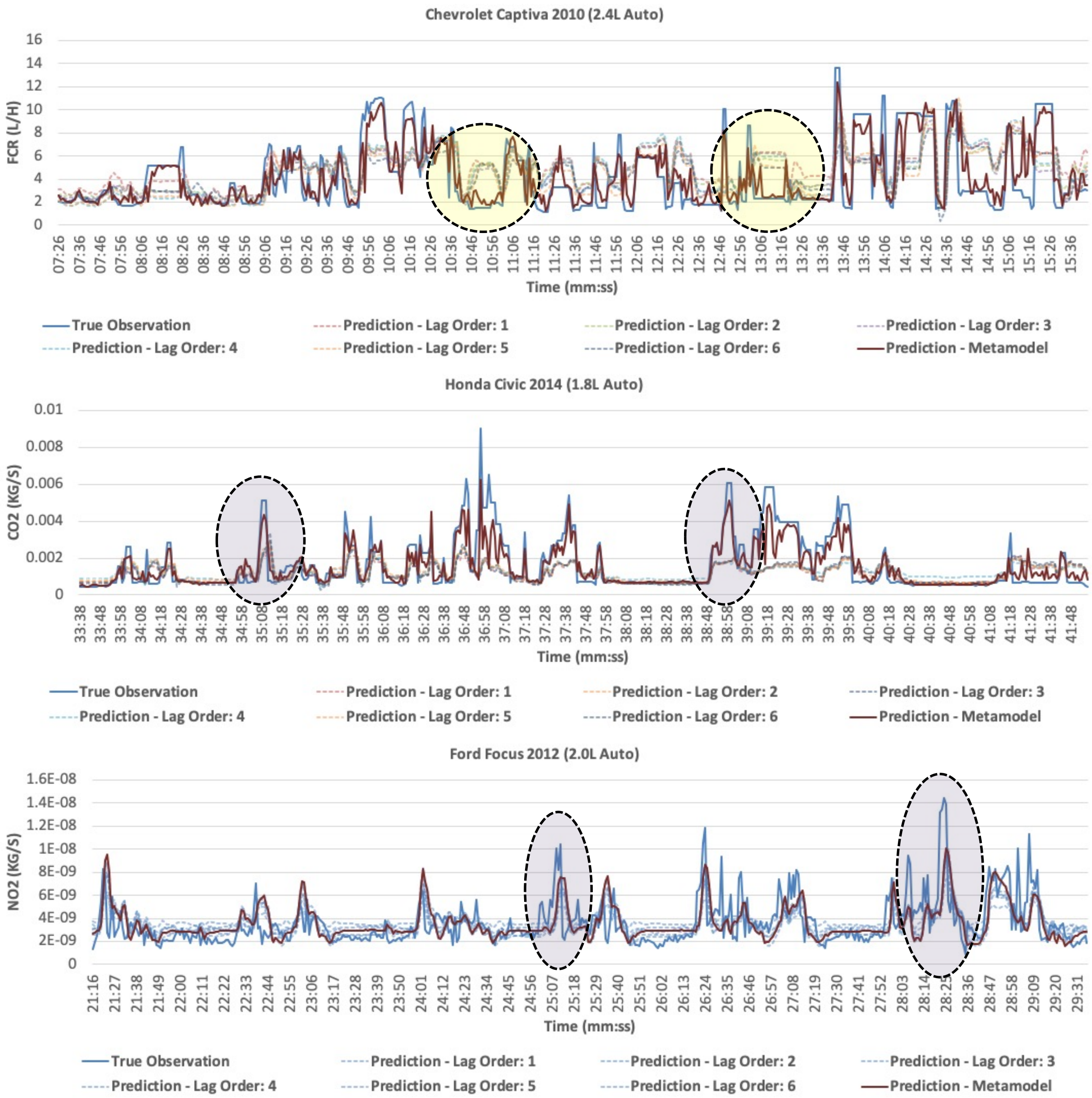

Figure 39: Random time-windows showing the prediction power of metamodels

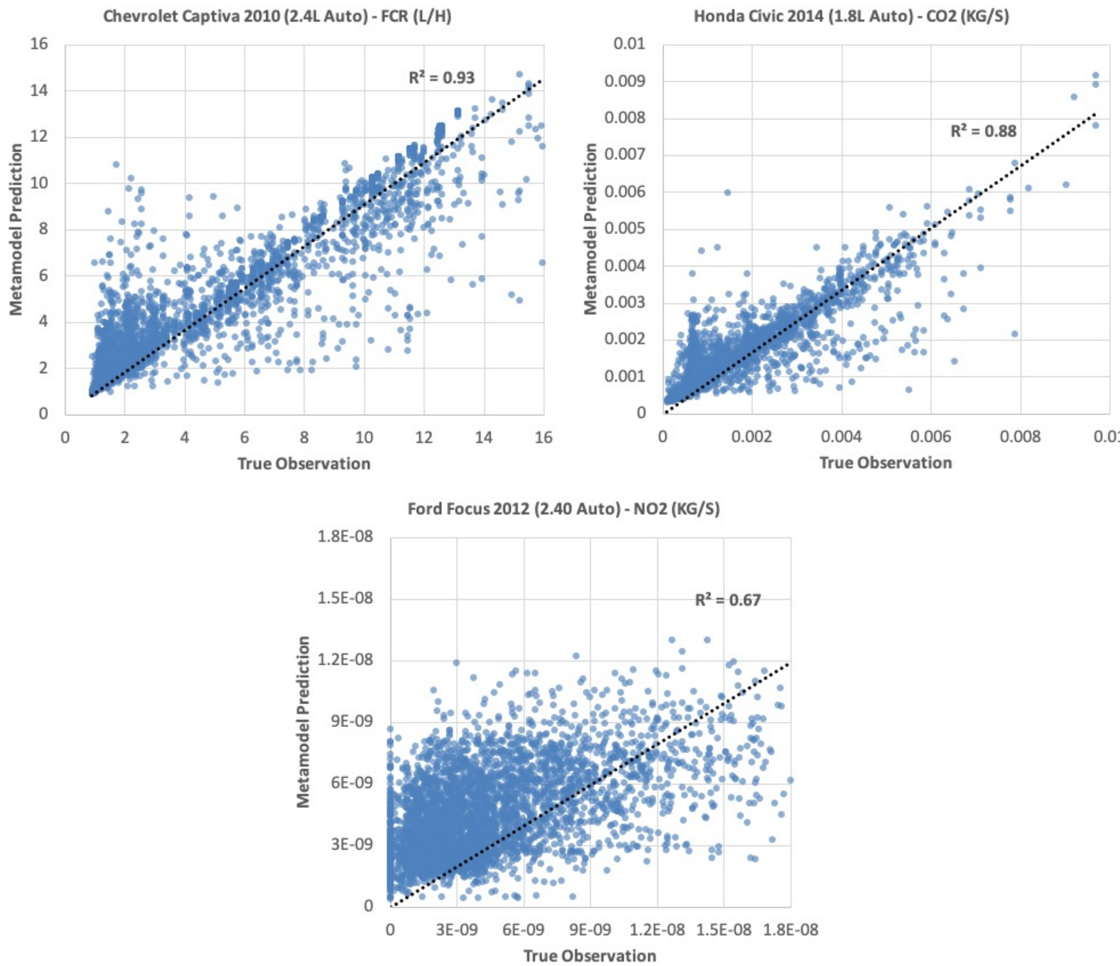

Figure 40: Comparison between true observation and metamodel predictions for the three vehicles

6.4 Category-specific Supermodels

The average score improvement of 6% is achieved as a result of developing category-specific supermodels, with a record of 32% improvement. The frequency distribution of improvement rates is shown in Figure 41.

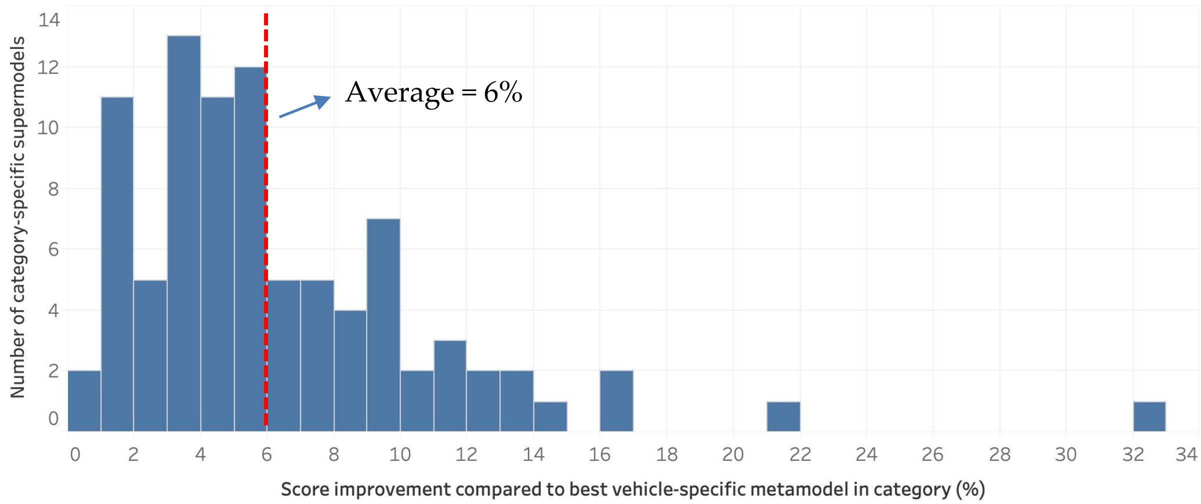

Figure 41: Score improvement rate histogram for best category-specific supermodels

Although the same set of EL algorithms has been used for developing the supermodels, relatively higher improvements have occurred. The diversity of the datasets corresponding to different category members, is one of the important root causes of this notable difference, despite the heavy assumption that was made about considering category member's datasets as homogenous subsets of a hypothetical large dataset.

In Figure 42, a deeper look is taken at the improvements achieved by ensemble category-specific supermodels. The percentages, all positive with rare zero values, support the idea that EL algorithms could work as a unifying medium for developing higher-level (aggregate) microscale fuel and emission models. Ranking the average of resulted improvements for each criterion, *Transmission Type* seems to be the most efficient aggregation measure for FCR supermodels. The transmission system directly deals with the quality of power transmission from the engine to the wheels and has a significant impact on the efficiency of the combustion process. Hence, its importance regarding fuel consumption and CO2 generation is expectable. However, this criterion does not seem appropriate for NO and PM emissions as limited improvements have been achieved for them.

Criterion	Category	Dependent FCR_LH
Age Range	Age < 2	8%
	3 < Age < 5	12%
	6 < Age < 9	10%
	10 < Age < 13	3%
Engine Size Range	1.0 < ED < 1.5	7%
	1.5 < ED < 2.0	11%
	2.0 < ED < 3.8	5%
Engine Type	Regular	7%
	Turbo-Charged	6%
Segment	Compact Sedan	11%
	Compact SUV	3%
	Midsize Sedan	16%
	Subcompact Hatchback	13%
	Subcompact Sedan	13%
Total Weight Range	1000 < W < 1200	3%
	1200 < W < 1400	4%
	1400 < W < 1600	16%
	1600 < W < 2000	14%
Transmission Type	Auto	32%
	Auto - CVT	4%
	Manual	8%

Criterion	Category	Dependent CO2_KGS	NO2_KGS	NO_KGS	PM_KGS
Age Range	Age < 2	3%	5%	3%	2%
	3 < Age < 5	6%	3%	0%	21%
	6 < Age < 9	9%	11%	5%	5%
	10 < Age < 13	2%	5%	1%	3%
Engine Size Range	1.0 < ED < 1.5	7%	9%	4%	2%
	1.5 < ED < 2.0	4%	5%	1%	5%
	2.0 < ED < 3.8	9%	8%	4%	1%
Engine Type	Regular	9%	2%	1%	6%
Segment	Compact Sedan	4%	3%	1%	6%
	Compact SUV	3%	1%	1%	4%
	Midsize Sedan	9%	5%	5%	1%
	Subcompact Hatchback	3%	8%	1%	3%
Total Weight Range	1000 < W < 1200	4%	7%	1%	5%
	1400 < W < 1600	9%	7%	3%	2%
	1600 < W < 2000	5%	5%	1%	4%
Transmission Type	Auto	9%	10%	0%	3%
	Auto - CVT	12%	6%	4%	4%

Figure 42: Score improvement concerning best metamodel for different criteria and corresponding categories

For the criteria in which a large difference is observed between the improvements achieved by category-specific supermodels, an inference is that possibly modifying the categorization thresholds or combining some of the categories could lead to more coordinated improvements among categories. For example, the low improvement achieved for the *Compact SUV* class brings the idea of merging this class with another one. Note that such low improvement (compared to other classes) is obtained despite the presence of 8 vehicles in the category which eliminates the chance of low diversity of data leading to inefficient forecast combination.

The *Age Range* criterion seems to work best for the PM rate. The finding was expectable as the increase in vehicle age leads to physical degradation of the engine and adds to the inefficiencies of the powertrain. Moreover, in an aged vehicle, usually the catalytic converter and the particulate filters lose their effectiveness resulting in higher PM rates. Like some other criteria, the age thresholds considered in this study need to be tweaked to achieve better improvements for all the ranges. Although there have been small but positive improvements for all NO-related supermodels, the two-stage EL approach does not seem to be the perfect match for this emission. As mentioned earlier, the low volatility of NO rate observations makes predictions of much simpler nonlinear modeling algorithms (even single-stage and without EL) acceptable enough. Such weak results (with respect to other emissions and FCR), could be linked to the sensor measurement errors as well. Although the state-of-the-technology PEMS units utilized provides unbeatable accuracies, as NOx emission rates are generally so low in gasoline-engine vehicles, even minor sensor errors affect the readings considerably.

A glance at Figure 43 shows that even for the supermodels, sophisticated algorithms (*Gradient Boosting* as well as *Random Forest*) outperform others for the majority of criteria/categories for FCR and CO2 rate. Similar to the case of metamodels, *Linear* and *Ridge Regressions* shoulder the forecast combination burden of NOx and PM supermodels better.

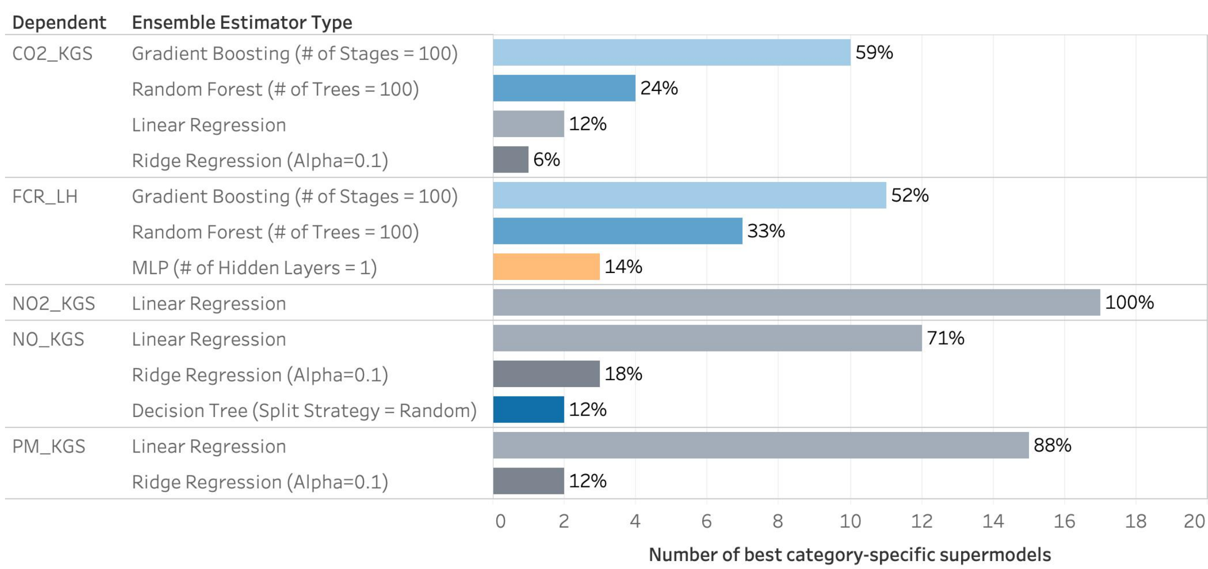

Figure 43: Share of the best supermodel estimators for different dependents

Gradient Boosting and *Random Forest* algorithms possess similar nature of using multiple *Decision Trees* and combining their forecasts to achieve better results, however, the former builds trees one at a time, where each new tree helps to correct errors made by the previously trained tree. Their difference could be better explained using the concept of bias and variance in the ML paradigm. *Boosting* is based on weak learners which basically have high bias and low variance (like the vehicle-specific metamodels in each category). Boosting reduces errors mainly by reducing bias. On the other hand, *Random Forest* uses fully grown decision trees with low bias and high variance (similar to lag-specific RNNs). It tackles the error reduction task by reducing variance. This explanation clarifies why *Gradient Boosting* and *Random Forest* algorithms have been the dominant best estimators for supermodel and metamodel development, respectively.

Finally, in Figure 44, sample time-windows are randomly selected for 3 category-dependent combinations to visually evaluate the performance of trained supermodels. Figure 45 depicts the impressive accuracy of our proposed two-stage EL approach is perceivable for FCR and CO2 rates with R-squared of 0.95 and 0.88, respectively. Although relatively lower, our method has still produced acceptable predictions for NO2, although it appears to be weak in capturing peaks. Nevertheless, as our categorization process still requires refinement, an R-squared score of 0.7 seems a satisfying score at this stage.

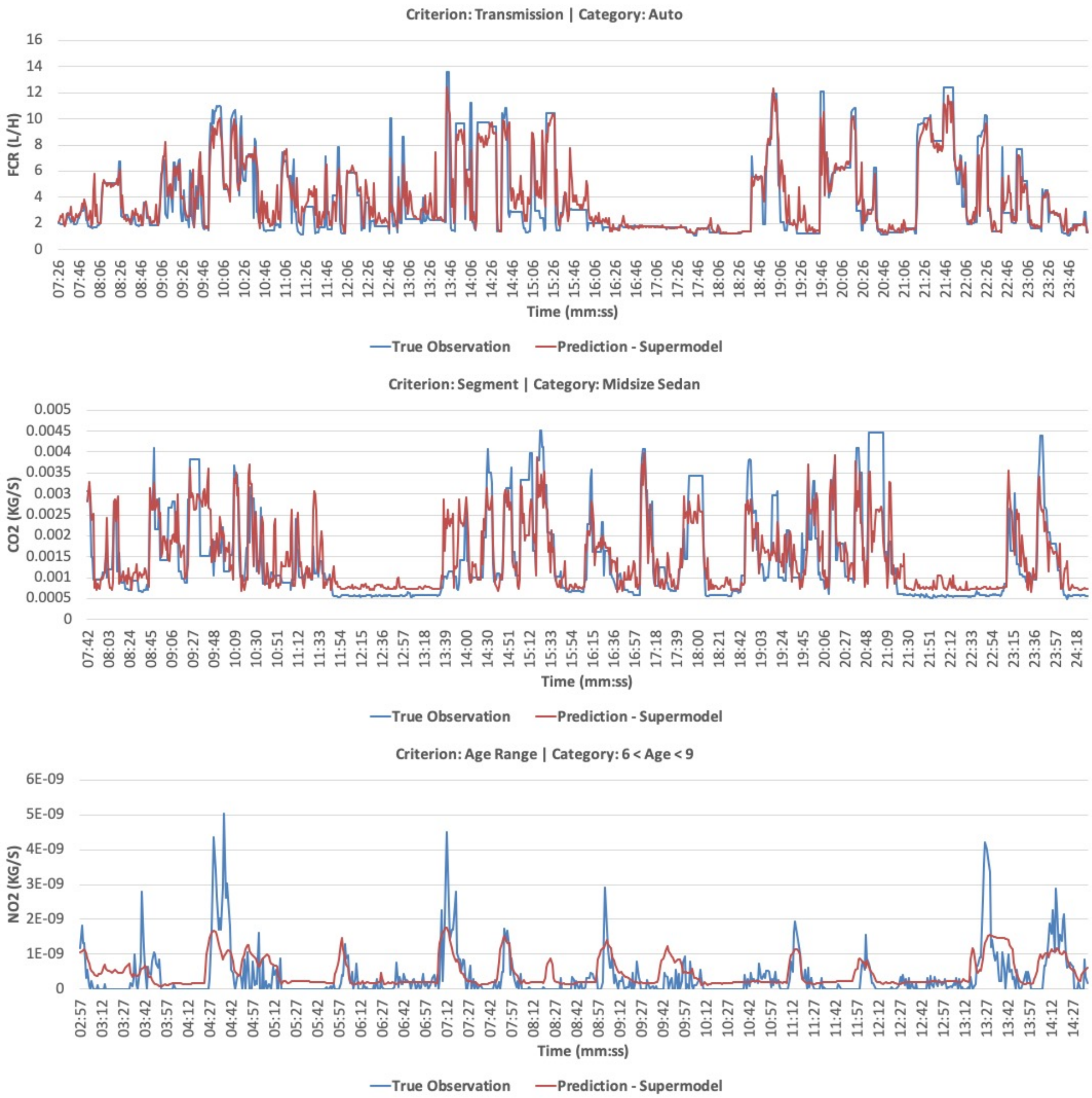

Figure 44: Random sample time-windows showing the prediction power of supermodels for 3 criteria/categories

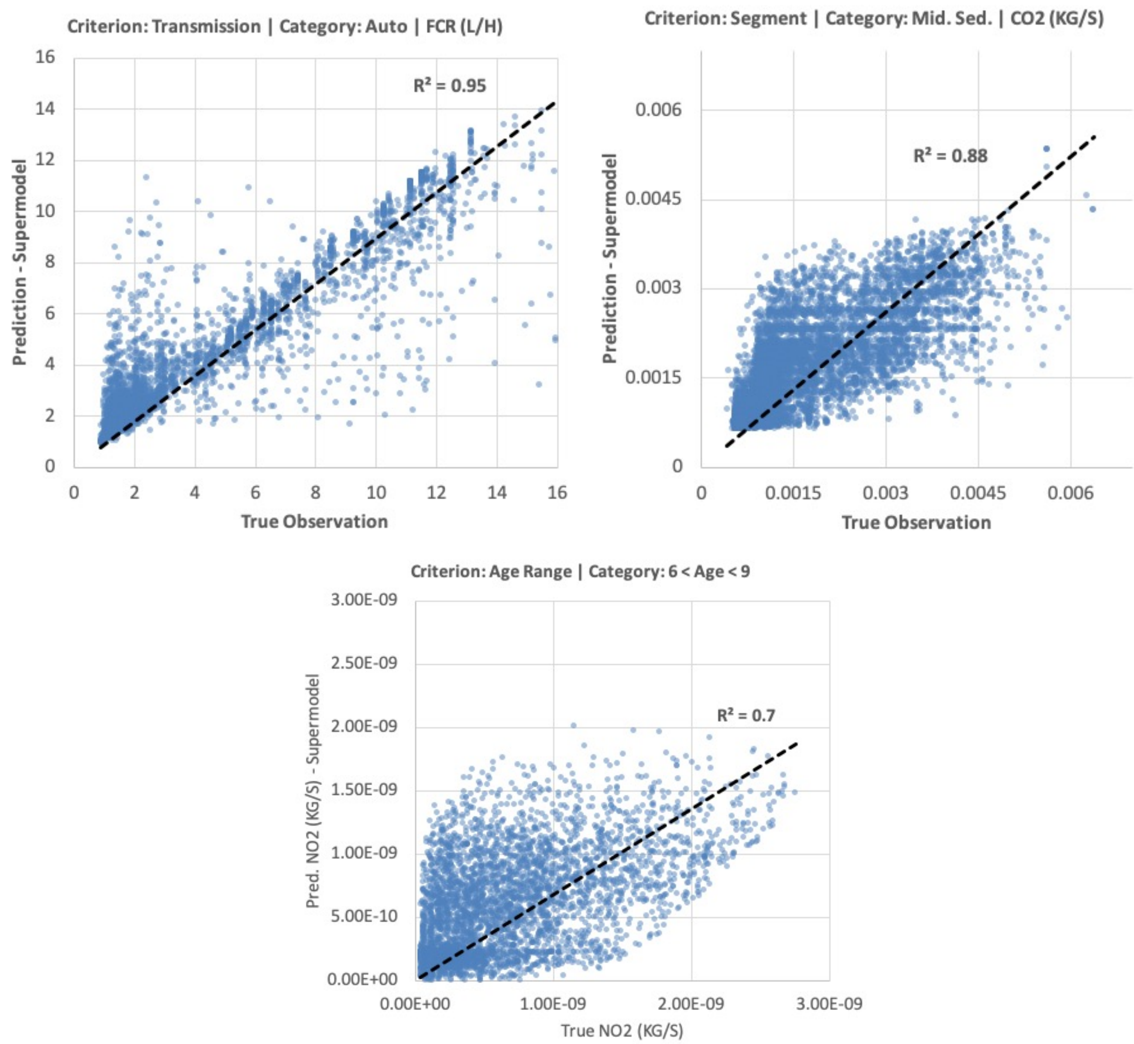

Figure 45: Comparison between true observation and metamodel predictions for three selected criterion-category pairs

Chapter 7

Conclusions and Future Research

This research contributed in various ways to the field of microscale vehicular fuel and emission modeling. In terms of data collection, we practiced transitioning from the in-lab simulation-based experiments to the realistic on-road activity, fuel, and emissions measurements. As a result, a replicable procedure including elaborate details about planning the field experiments, use of different sensors, and adjusting the sensor outputs for the needs of modeling was introduced. Such a transition provides the opportunity for the modelers to capture the impact of several influential factors in their measurements that simulating them in the lab is impossible or costly (such as rainfall and snow, icy road, low-quality pavements, real-world and local stop-and-go movement patterns in congested roads or grid networks).

Questioning the validity of using existing commercial/academic models in regions rather than their country of origin, a methodology for evaluating MOVES output, one of the most popular environmental assessment models in North America, was proposed. The evaluation results would avoid consulting companies and the official transport authorities in countries like Canada from relying on erroneous energy and emissions assessment in projects. This could dramatically affect the transport policies, change the development strategies, and hopefully in long-term, save millions of dollars for fixing the environmental damages due to making wrong estimates.

Several specific insights are concluded from the evaluation of MOVES. The model's inaccuracies in predicting the fuel consumption and emission rates do not come as a surprise. MOVES is created for use in the US and although it has been adjusted to the prevailing conditions of the test scenarios, the runs are still outside of the model's intended use. Furthermore, MOVES is designed primarily for use in *State Implementation Plan* (SIP) conformity analysis and using it at the microscopic level (project level in MOVES) also likely decreases its accuracy.

Regarding *Energy Consumption Rate* (ECR) and CO2 rate, an underestimation was observed with *Mean Percentage Error* (MPE) of -17% and -35%, respectively. The high variation which existed in predictions resulted in low goodness-of-fit scores (R-squared of 0.32 and 0.6, respectively). The deficiency of MOVES in simulating non-U.S. scenarios was revealed through detailed categorization of MOVES outputs based on vehicle-, road-, and location-specific attributes. MOVES predicted ECR and CO2 better for automatic transmission vehicles compared to manual ones, and significantly better for turbocharged-engine vehicles and the light-weight/small vehicle segments (by R-squared margins of 19%, 51%, and 45%, respectively). Such attributes are either omitted in MOVES or are not directly included. The fact that MOVES predicts much better for turbocharged engines is in accordance with lab- and simulation-based (close-to-ideal) testing conditions for estimation of MOVES core models. Furthermore, the categorization based on age showed ECR and CO2 rate for vehicles with older ages are better predicted by MOVES (with 34% higher R-squared score). The finding was linked to MOVES core models not being updated for newer engine technologies and vehicle design factors.

For *Nitrogen Oxides* (NOx) and *Particulate Matters* (PM), the dominant finding was over-estimation of rates by MOVES (The MPE metric goes up to +420%), most of which could be linked to poor addressing of the impact of catalytic converters on emissions and disregarding fuel cut-off events. Moreover, age-based categorization for NOx and PM rate comparisons, confirmed the previous finding for ECR and CO2 rate that MOVES requires regular updating as it loses validity year by year due to fast-paced changes in vehicular engine and general design technologies.

Due to the significant influential differences of the regions (in terms of fleet, road network, meteorology, etc.), a wise alternative for use in transportation studies or incorporation in simulations and eco-driving services seems to be developing local models. The advent and availability of ultra-portable emission measurement technologies for on-road monitoring of vehicle activities has now opened avenues for all the countries/provinces to achieve such goals with low investments.

Independence from *Internal Engine Variables* (IEV) is a requirement for an eco-driving system to work with equal quality either incorporated into the vehicle's on-board information system or installed as a service on smartphone platforms. Besides, such systems should depend on a limited set of input variables as it provides a hassle-free and smooth experience for the users at the deployment time. The same requirement exists for fuel and emission models that work on top of traffic simulation models and are used for environmental assessment of transport projects or policies. Many of the state-of-the-practice models such as *Virginia Tech's Comprehensive Power-based Fuel Model* (VT-CPFM) or *Comprehensive Modal Emissions Model* (CMEM) work oppositely and rely on many input variables. For example, VT-CPFM requires 14 technical parameters to perform the proxy estimations of resistance force and engine power beforehand, and finally get ready to estimate *Fuel Consumption Rate* (FCR) using nonlinear regression. As a part of this study, developing vehicle-specific models that only rely on speed, acceleration, road grade, and general attributes of the vehicle such as vehicle segment, engine type, engine size, transmission type, and age as their feature set was targeted. A methodology based on *Machine Learning* (ML) techniques, capable of capturing the complex underlying patterns of interactions between features and the dependent variable, was introduced. Using ML techniques for vehicular FCR and ER estimation, the models could automatically adjust themselves to the performance parameters of the powertrain.

The kernel trick in *Support Vector Regression* (SVR) and the layered nature of *Artigical Neural Networks* (ANN) translate the lower-dimensional feature space into estimated impacts of unobserved but influential variables. An empowerment of the base SVR/ANN models was attempted through a cascaded approach and introduction of the estimate of RPM to their feature set. Similar median scores were achieved with SVR and ANN when modeling using base (no-RPM) structures. However, the superiority of ANN was apparent in cascaded modeling as we experienced improvements of up to 116% (for Hyundai Elantra GT 2019). Whereas the SVR received extremely weak benefits from the addition of RPM estimates to the feature set (with 8% as the best improvement for Honda Civic 2014 and Jac J5 2015). Considering the modeling results, the flexibility of ANN due to the possibility of training over various architectures, and the requirements of a data-driven modeling approach, ANN is found to be the more promising option for FCR modeling.

Neither deep and narrow nor wide and shallow ANN architectures, but some settings in-between led to the best training/test scores with ANN (performing on average 21% better than shallowest and 25% better than deepest architectures). We found that with a limited number of input features, there is an upper bound for increasing the ANN depth to achieve improved scores. Concerning the ANN architectures, although for most of the vehicles the best scores were achieved with an intermediary architecture, for some vehicles, deep and narrow, and for some others wide and shallow architectures worked the best. This observation suggests that modeling FCR for different vehicles might not be possible with a single functional form (which is the case for regression models such as VT-CPFM), even if they are designed in parametric form and some flexibility is added to them by adjusting parameters.

The categorical sensitivity analysis conducted on cascaded modeling results was a first step towards improving the cascaded modeling results and generalizing the models from vehicle-scale to category- or even region-scales in future steps. It is necessary to understand the biases of the best-achieved results towards categories of vehicles or vehicles with specific attributes. We found that the cascaded modeling approach could be applied to compact hatchbacks, midsize sedans, and subcompact SUVs with acceptable results (maximum cascaded modeling score of 0.83 was achieved for these car segments). However, an adjustment methodology is required for other car segments. Vehicles with turbocharged engines are perfect options for the cascaded modeling (with a maximum-achieved score of 0.75), while those with conventional engines need further elaborate investigation. Regarding the vehicle's age, including a quantified measure for maintenance quality is a more promising approach. Moreover, the mileage could be a more reasonable alternative to the age.

By taking the cascaded modeling approach instead of traditional nonlinear regression, the process of updating models with data collected from the newly introduced vehicles (in a way that the model stays valid for both the older vehicles as well as the new ones) would be straightforward and simple. Utilizing the *Transfer Learning* techniques enables researchers to build up models by taking a base model already trained on a fleet and continue training it for the vehicles which specifically exist in the new fleet.

In addition to the goals we sought when developing the cascaded modeling methodology, the microscale FCR and *Emission Rate* (ER) models need to keep in accordance with the hybrid instantaneous/distributively-lagged nature of dependencies and interactions between features and the dependent variable. Moreover, they need to gain generalizability to more aggregate levels as well (for instance, category-specific models instead of vehicle-specific ones). Only in such circumstances, microscale FCR and ER models reliable for incorporation in eco-driving services, traffic simulation procedures, and transportation impact studies could be developed. The trained ML-based category-specific models would have less variance (as they are trained for a subset of data with a significant common attribute) and closer to the optimal bias-variance trade-off point.

By putting the burden of extracting complex and combined effects on sophisticated machine learning algorithms such as *Recurrent Neural Networks* (RNN) and popular forecast combination methodologies, the first abovementioned goal is tackled. *Long Short-Term Memory* (LSTM) cell architecture was the dominant type of RNN leading to best lag-specific modeling scores, while the impact of deepening RNN architectures (through stacking layers on top of each other) was found significant for fuel consumption and PM rates, almost negative for NOx rates, and negligible for CO2 rate. The vehicle-specific metamodels, trained by combining forecasts made by lag-specific RNN models, showed improvement records of up to 28% concerning RMSE score (with an average improvement of 4% among different vehicles and dependent types). The proposed solution for targeting the second goal, generalizing the vehicle-specific metamodels to more aggregate levels, was to apply another EL layer on top of previous layers. The category-specific supermodels developed not only could be used as representative models for all the existing/new vehicles dedicated to the corresponding categories, but they achieved score improvements of up to 32% (with an average improvement of 6% among the criteria/categories). Besides, linear regression dominantly resulted in the best score improvements while developing both metamodels as well as supermodels for NOx and PM rates, while sophisticated methods of random forests and gradient boosting were the dominant algorithms for FCR and CO2 rate.

This study opens avenues to use of ML techniques for rapid development of light, generalized, and localized microscale fuel and emission models. But it lacks in a few aspects which could be addressed in future works. Although our test fleet (including 35 vehicles of different makes, models, and ages) is much larger compared to that of many other studies in this field, simultaneously increasing the size and diversity of the fleet could lead to more robust conclusions especially in the categorical analysis. A similar methodology could be taken for diesel vehicles, considering that NOx and soot (particulate matters) are generated in more significant amounts from diesel engines. Moreover, a question worth answering through further research is whether *Vehicle Specific Power* (VSP) as a proxy variable (used in MOVES core models) is capable of estimating correct power demand for manual transmission vehicles or not. The presence of so many manual-transmission vehicles, most of which older cars prone to generate significantly higher emissions and consume more fuel, on the roads around the globe emphasizes the importance of such investigation.

Expanding the size of the training dataset can definitely improve the modeling scores. Moreover, we had to limit the search ranges for the grid search in SVR modeling or the number of iterations in training ANN models because of time and computational limitations. Relaxing some of these constraints could result in improved model accuracy as well.

In the categorical sensitivity analysis, the focus has been on general and available-to-public attributes of the vehicle such as vehicle segment, engine size and type, transmission, weight, etc. Other factors such as frontal area, drag coefficient, tire characteristics, use of air conditioner, and vehicle mileage could be used for further categorical analysis in future research. If a dramatic increase happens to the number and diversity of test fleet, mixed categories could be defined as well which leads to richer sensitivity analysis as well as an overall improvement in the accuracy of the supermodels when deployed in practice (Ensemble supermodels developed for the mixed categories would be more reliable representatives of corresponding vehicles).

A possible extension to the current methodology is to train ensemble models that could predict FCR and different emission rates simultaneously (by adding a soft-max layer to the models). Such an approach would make separate models unified and eases their incorporation process in eco-driving services or even traffic simulation software.

As transmission technologies play an important role in fuel consumption and emission generation patterns, a detailed look should be taken at existing transmission technologies and expand the number of categories corresponding to this criterion in hope of achieving improved results, especially, for the emission rates.

Finally, a combination of the cascaded modeling and the two-stage RNN-based EL approach is expected to result in even more promising accuracies. Estimates of engine speed (or other influential IEVs) could be introduced to the feature set of the lag-specific models to give more flexibility to the modeling algorithms for accurately predicting FCR and ERs.

Bibliography

1. Natural Resources Canada GHG Spotlight on Transportation Available online: https://www.nrcan.gc.ca/energy-and-greenhouse-gas-emissions-ghgs/20063#L4 (accessed on Aug 30, 2019).
2. Crouse, D.L.; Peters, P.A.; van Donkelaar, A.; Goldberg, M.S.; Villeneuve, P.J.; Brion, O.; Khan, S.; Atari, D.O.; Jerrett, M.; Pope, C.A.; et al. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A canadian national-level cohort study. *Environ. Health Perspect.* **2012**, doi:10.1289/ehp.1104049.
3. Sonntag, D.; Choi, D.; Warila, J.; Beardsley, M.; Toro, C.; Simon, H.; Timin, B.; Dolwick, P.; Eyth, A.; Henderson, B.; et al. *Update on MOVES Model Evaluation; NOx*; 2017;
4. Woodruff, M. Gas Spending Around the World Available online: https://www.businessinsider.com/gas-prices-around-the-world-2012-5 (accessed on Aug 30, 2019).
5. Thiel, W. 2020 Canada Vehicle Sales Figures By Model Available online: https://www.goodcarbadcar.net/2020-canada-vehicle-sales-figures-by-model/ (accessed on Jul 30, 2020).
6. Thiel, W. 2020 US Vehicle Sales Figures By Model Available online: https://www.goodcarbadcar.net/2020-us-vehicle-sales-figures-by-model/ (accessed on Jul 30, 2020).
7. Consumer Reports The Rise of the Crossover: The segment that’s really driving the auto industry’s sales Available online: https://advocacy.consumerreports.org/research/the-rise-of-the-crossover-the-segment-thats-really-driving-the-auto-industrys-sales/ (accessed on Jul 30, 2020).
8. Comparisons, I. Transportation statistics by country Available online: http://internationalcomparisons.org/environment/transportation.html (accessed on Jul 30, 2020).
9. Bielaczyc, P.; Woodburn, J.; Szczotka, A. The WLTP as a new tool for the evaluation of CO 2 emissions. In Proceedings of the FISITA 2014 World Automotive Congress - Proceedings; 2014.
10. EPA fuel economy ratings - what’s coming in 2008 Available online: https://web.archive.org/web/20071008015835/http://www.earthcars.com/articles/article.htm?articleId=6 (accessed on Jun 7, 2020).

11. Pelkmans, L.; Debal, P. Comparison of on-road emissions with emissions measured on chassis dynamometer test cycles. *Transp. Res. Part D Transp. Environ.* **2006**, doi:10.1016/j.trd.2006.04.001.

12. Exploring Trip Fuel Consumption by Machine Learning from GPS and CAN Bus Data. *J. East. Asia Soc. Transp. Stud.* **2015**, doi:10.11175/easts.11.906.

13. Ahn, K.; Rakha, H.; Trani, A.; Van Aerde, M. Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. *J. Transp. Eng.* **2002**, doi:10.1061/(ASCE)0733-947X(2002)128:2(182).

14. Duarte, G.O.; Gonçalves, G.A.; Baptista, P.C.; Farias, T.L. Establishing bonds between vehicle certification data and real-world vehicle fuel consumption - A Vehicle Specific Power approach. *Energy Convers. Manag.* **2015**, doi:10.1016/j.enconman.2014.12.042.

15. Wu, J. Da; Liu, J.C. Development of a predictive system for car fuel consumption using an artificial neural network. *Expert Syst. Appl.* **2011**, doi:10.1016/j.eswa.2010.09.155.

16. Kanarachos, S.; Mathew, J.; Fitzpatrick, M.E. Instantaneous vehicle fuel consumption estimation using smartphones and recurrent neural networks. *Expert Syst. Appl.* **2019**, doi:10.1016/j.eswa.2018.12.006.

17. Rakha, H.A.; Ahn, K.; Moran, K.; Saerens, B.; Bulck, E. Van den Virginia Tech Comprehensive Power-Based Fuel Consumption Model: Model development and testing. *Transp. Res. Part D Transp. Environ.* **2011**, doi:10.1016/j.trd.2011.05.008.

18. Frey, H.C.; Zhang, K.; Rouphail, N.M. Vehicle-specific emissions modeling based upon on-road measurements. *Environ. Sci. Technol.* **2010**, doi:10.1021/es902835h.

19. Amer, A.; Abdalla, A.; Noraziah, A.; Fauzi, A.A.C. Prediction of vehicle fuel consumption model based on artificial neural network. In Proceedings of the Applied Mechanics and Materials; 2014.

20. Rakha, H.; Ahn, K.; Trani, A. Development of VT-Micro model for estimating hot stabilized light duty vehicle and truck emissions. *Transp. Res. Part D Transp. Environ.* **2004**, doi:10.1016/S1361-9209(03)00054-3.

21. Saerens, B.; Rakha, H.; Ahn, K.; Van Den Bulck, E. Assessment of alternative polynomial fuel consumption models for use in intelligent transportation systems applications. *J. Intell. Transp. Syst. Technol. Planning, Oper.* **2013**, doi:10.1080/15472450.2013.764801.

22. Weng, J.; Liang, Q.; Qiao, G.; Chen, Z.; Rong, J. Taxi fuel consumption and emissions estimation model based on the reconstruction of driving trajectory. *Adv.*

Mech. Eng. **2017**, doi:10.1177/1687814017708708.

23. Saboohi, Y.; Farzaneh, H. Model for developing an eco-driving strategy of a passenger vehicle based on the least fuel consumption. *Appl. Energy* **2009**, doi:10.1016/j.apenergy.2008.12.017.

24. Kousoulidou, M.; Fontaras, G.; Ntziachristos, L.; Bonnel, P.; Samaras, Z.; Dilara, P. Use of portable emissions measurement system (PEMS) for the development and validation of passenger car emission factors. *Atmos. Environ.* **2013**, doi:10.1016/j.atmosenv.2012.09.062.

25. Jaikumar, R.; Shiva Nagendra, S.M.; Sivanandan, R. Modeling of real time exhaust emissions of passenger cars under heterogeneous traffic conditions. *Atmos. Pollut. Res.* **2017**, doi:10.1016/j.apr.2016.07.011.

26. Li, Q.; Qiao, F.; Yu, L. A Machine Learning Approach for Light-Duty Vehicle Idling Emission Estimation Based on Real Driving and Environmental Information. *Environ. Pollut. Clim. Chang.* **2017**, doi:10.4172/2573-458x.1000106.

27. Bifulco, G.N.; Galante, F.; Pariota, L.; Spena, M.R. A linear model for the estimation of fuel consumption and the impact evaluation of advanced driving assistance systems. *Sustain.* **2015**, doi:10.3390/su71014326.

28. Wang, H.; Fu, L.; Zhou, Y.; Li, H. Modelling of the fuel consumption for passenger cars regarding driving characteristics. *Transp. Res. Part D Transp. Environ.* **2008**, doi:10.1016/j.trd.2008.09.002.

29. Zhou, Q.; Gullitti, A.; Xiao, J.; Huang, Y. Neural network-based modeling and optimization for effective vehicle emission testing and engine calibration. *Chem. Eng. Commun.* **2008**, doi:10.1080/00986440701568830.

30. De Lucas, A.; Durán, A.; Carmona, M.; Lapuerta, M. Modeling diesel particulate emissions with neural networks. *Fuel* **2001**, doi:10.1016/S0016-2361(00)00111-3.

31. Thompson, G.J. Neural network modelling of the emissions and performance of a heavy-duty diesel engine. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2000**, doi:10.1243/0954407001527277.

32. Kayes, D.; Hochgreb, S. Mechanisms of particulate matter formation in spark-ignition engines. 3. Model of PM formation. *Environ. Sci. Technol.* **1999**, doi:10.1021/es981101o.

33. Çapraz, A.G.; Özel, P.; Şevkli, M.; Beyca, Ö.F. Fuel Consumption Models Applied to Automobiles Using Real-time Data: A Comparison of Statistical Models. In Proceedings of the Procedia Computer Science; 2016.

34. Nie, Y.; Li, Q. An eco-routing model considering microscopic vehicle operating

conditions. *Transp. Res. Part B Methodol.* **2013**, doi:10.1016/j.trb.2013.06.004.

35. Koupal, J.; Cumberworth, M.; Michaels, H.; Beardsley, M.; Brzezinski, D. Design and Implementation of MOVES: EPA's New Generation Mobile Source Emission Model. *Int. Emiss. Invent. Conf.* **2003**.

36. Scora, G.; Barth, M. Comprehensive modal emissions model (cmem), version 3.01. *User Guid. Cent. Environ. Res. Technol. Univ. California, Riverside* **2006**, *1070*.

37. Zhai, H.; Frey, H.C.; Rouphail, N.M. A vehicle-specific power approach to speed- and facility-specific emissions estimates for diesel transit buses. *Environ. Sci. Technol.* **2008**, doi:10.1021/es800208d.

38. Bishop, J.D.K.; Stettler, M.E.J.; Molden, N.; Boies, A.M. Engine maps of fuel use and emissions from transient driving cycles. *Appl. Energy* **2016**, doi:10.1016/j.apenergy.2016.08.175.

39. Du, Y.; Wu, J.; Yang, S.; Zhou, L. Predicting vehicle fuel consumption patterns using floating vehicle data. *J. Environ. Sci.* **2017**, doi:10.1016/j.jes.2017.03.008.

40. Guevara, M.; Tena, C.; Soret, A.; Serradell, K.; Guzmán, D.; Retama, A.; Camacho, P.; Jaimes-Palomera, M.; Mediavilla, A. An emission processing system for air quality modelling in the Mexico City metropolitan area: Evaluation and comparison of the MOBILE6.2-Mexico and MOVES-Mexico traffic emissions. *Sci. Total Environ.* **2017**, doi:10.1016/j.scitotenv.2017.01.135.

41. Perugu, H. Emission modelling of light-duty vehicles in India using the revamped VSP-based MOVES model: The case study of Hyderabad. *Transp. Res. Part D Transp. Environ.* **2019**, doi:10.1016/j.trd.2018.01.031.

42. Zhang, L.; Hu, X.; Qiu, R.; Lin, J. Comparison of real-world emissions of LDGVs of different vehicle emission standards on both mountainous and level roads in China. *Transp. Res. Part D Transp. Environ.* **2019**, doi:10.1016/j.trd.2019.01.020.

43. Koupal, J.; Michaels, H.; Cumberworth, M.; Bailey, C.; Brzezinski, D. EPA's plan for MOVES: a comprehensive mobile source emissions model. In Proceedings of the Proceedings of the 12th CRC On-Road Vehicle Emissions Workshop, San Diego, CA; 2002; pp. 15–17.

44. Wu, Y.; Song, G.; Yu, L. Sensitivity analysis of emission rates in MOVES for developing site-specific emission database. *Transp. Res. Part D Transp. Environ.* **2014**, doi:10.1016/j.trd.2014.07.009.

45. Anderson, D.C.; Loughner, C.P.; Diskin, G.; Weinheimer, A.; Canty, T.P.; Salawitch, R.J.; Worden, H.M.; Fried, A.; Mikoviny, T.; Wisthaler, A.; et al. Measured and modeled CO and NOy in DISCOVER-AQ: An evaluation of emissions and

chemistry over the eastern US. *Atmos. Environ.* **2014**, doi:10.1016/j.atmosenv.2014.07.004.

46. Fujita, E.M.; Campbell, D.E.; Zielinska, B.; Chow, J.C.; Lindhjem, C.E.; DenBleyker, A.; Bishop, G.A.; Schuchmann, B.G.; Stedman, D.H.; Lawson, D.R. Comparison of the MOVES2010a, MOBILE6.2, and EMFAC2007 mobile source emission models with on-road traffic tunnel and remote sensing measurements. *J. Air Waste Manag. Assoc.* **2012**, doi:10.1080/10962247.2012.699016.
47. Sonntag, D.; Choi, D.; Warila, J.; Beardsley, M. *Comparison of light-duty NOx emission rates estimated from MOVES with real-world measurements*; 2017;
48. Li, C.; Yu, L.; He, W.; Cheng, Y.; Song, G. Development of local emissions rate model for light-duty gasoline vehicles: Beijing field data and patterns of emissions rates in EPA simulator. *Transp. Res. Rec.* **2017**, doi:10.3141/2627-08.
49. Bai, S.; Du, Y.; Seagram, A.; Craig, K. MOVES-based NOx analyses for urban case studies in Texas. In Proceedings of the International Emission Inventory Conference, Baltimore, MD, August; 2017; Vol. 18.
50. Ramezani, H.; Lu, X.-Y.; Shladover, S.E. *Calibration of motor vehicle emission simulator (MOVES) using real heavy-duty truck data*; 2019;
51. Toro, C.; Warila, J.; Sonntag, D.; Beardsley, M.; Choi, D. MOVES Light-Duty Emission Rate Evaluation in the Context of Reconciling Modeled and Ambient NOx. **2019**.
52. Lin, C.; Zhou, X.; Wu, D.; Gong, B. Estimation of emissions at signalized intersections using an improved MOVES model with GPS data. *Int. J. Environ. Res. Public Health* **2019**, doi:10.3390/ijerph16193647.
53. Behan, M.; Moradi, E.; Miranda-Moreno, L. A Comparative Analysis of the Vehicular Emissions Generated as a Results of Different Intersection Controls. In Proceedings of the Transportation Research Board 99th Annual Meeting; 2020.
54. Ehmer, M.; Khan, F. A Comparative Study of White Box, Black Box and Grey Box Testing Techniques. *Int. J. Adv. Comput. Sci. Appl.* **2012**, doi:10.14569/ijacsa.2012.030603.
55. Hirsch, M.; Alberer, D.; Del Re, L. Grey-box control oriented engine emissions models. In Proceedings of the IFAC Proceedings Volumes (IFAC-PapersOnline); 2008.
56. Eriksson, L.; Wahlström, J.; Klein, M. Physical modeling of turbocharged engines and parameter identification. *Lect. Notes Control Inf. Sci.* **2010**, doi:10.1007/978-1-84996-071-7_4.

57. Saerens, B.; Diehl, M.; Van Den Bulck, E. Optimal control using pontryagin's maximum principle and dynamic programming. *Lect. Notes Control Inf. Sci.* **2010**, doi:10.1007/978-1-84996-071-7_8.

58. Tunestal, P.; Lewander, M. Model predictive control of partially premixed combustion. *Lect. Notes Control Inf. Sci.* **2010**, doi:10.1007/978-1-84996-071-7_11.

59. Cachón, L.; Pucher, E. Fuel Consumption Simulation Model of a CNG Vehicle based on Real-world Emission Measurement. In Proceedings of the SAE Technical Papers; 2007.

60. Rakha, H.A.; Ahn, K.; Faris, W.; Moran, K.S. Simple vehicle powertrain model for modeling intelligent vehicle applications. *IEEE Trans. Intell. Transp. Syst.* **2012**, doi:10.1109/TITS.2012.2188517.

61. Arrègle, J.; Javier López, J.; Guardiola, C.; Monin, C. Sensitivity study of a NOx estimation model for on-board applications. In Proceedings of the SAE Technical Papers; 2008.

62. Payri, F.; Arrègle, J.; Javier López, J.; Mocholí, E. Diesel NOx modeling with a reduction mechanism for the initial NOx coming from EGR or re-entrained burned gases. In Proceedings of the SAE Technical Papers; 2008.

63. Tauzia, X.; Karaky, H.; Maiboom, A. Evaluation of a semi-physical model to predict NOx and soot emissions of a CI automotive engine under warm-up like conditions. *Appl. Therm. Eng.* **2018**, doi:10.1016/j.applthermaleng.2018.04.005.

64. Demesoukas, S. 0D/1D combustion modeling for the combustion systems optimization of spark ignition engines 2015.

65. Falcone, P.; De Gennaro, M.C.; Fiengony, G.; Glielmo, L.; Santini, S.; Langthaler, P. Torque generation model for Diesel Engine. In Proceedings of the Proceedings of the IEEE Conference on Decision and Control; 2003.

66. Anetor, L.; Odetunde, C.; Osakue, E.E. Computational Analysis of the Extended Zeldovich Mechanism. *Arab. J. Sci. Eng.* **2014**, doi:10.1007/s13369-014-1398-7.

67. Blauwens, J.; Smets, B.; Peeters, J. Mechanism of "prompt" no formation in hydrocarbon flames. *Symp. Combust.* **1977**, doi:10.1016/S0082-0784(77)80395-0.

68. Duarte, G.O.; Gonçalves, G.A.; Farias, T.L. A Methodology to Estimate Real-world Vehicle Fuel Use and Emissions based on Certification Cycle Data. *Procedia - Soc. Behav. Sci.* **2014**, doi:10.1016/j.sbspro.2014.01.104.

69. Kim, D.; Lee, J. Application of neural network model to vehicle emissions. *Int. J. Urban Sci.* **2010**, *14*, 264–275.

70. Wu, J. Da; Liu, J.C. A forecasting system for car fuel consumption using a radial

basis function neural network. In Proceedings of the Expert Systems with Applications; 2012.

71. Parlak, A.; Islamoglu, Y.; Yasar, H.; Egrisogut, A. Application of artificial neural network to predict specific fuel consumption and exhaust temperature for a Diesel engine. *Appl. Therm. Eng.* **2006**, doi:10.1016/j.applthermaleng.2005.10.006.
72. Wickramanayake, S.; Bandara, D.H.M.N. Fuel consumption prediction of fleet vehicles using Machine Learning: A comparative study. In Proceedings of the 2nd International Moratuwa Engineering Research Conference, MERCon 2016; 2016.
73. Ajtay, D.; Weilenmann, M. Static and dynamic instantaneous emission modelling. In Proceedings of the International Journal of Environment and Pollution; 2004.
74. Adhikari, R. A neural network based linear ensemble framework for time series forecasting. *Neurocomputing* **2015**, doi:10.1016/j.neucom.2015.01.012.
75. Kourentzes, N.; Barrow, D.K.; Crone, S.F. Neural network ensemble operators for time series forecasting. *Expert Syst. Appl.* **2014**, *41*, 4235–4244.
76. Bianchi, F.M.; Maiorino, E.; Kampffmeyer, M.C.; Rizzi, A.; Jenssen, R. An overview and comparative analysis of recurrent neural networks for short term load forecasting. *arXiv Prepr. arXiv1705.04378* **2017**.
77. Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci. Rep.* **2018**, doi:10.1038/s41598-018-24271-9.
78. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv Prepr. arXiv1506.00019* **2015**.
79. Petneházi, G. Recurrent neural networks for time series forecasting. *arXiv Prepr. arXiv1901.00069* **2019**.
80. Patterson, D.J.; Henein, N.A. Emissions from combustion engines and their control. **1981**.
81. Marotta, A.; Pavlovic, J.; Ciuffo, B.; Serra, S.; Fontaras, G. Gaseous Emissions from Light-Duty Vehicles: Moving from NEDC to the New WLTP Test Procedure. *Environ. Sci. Technol.* **2015**, doi:10.1021/acs.est.5b01364.
82. Varella, R.A.; Giechaskiel, B.; Sousa, L.; Duarte, G. Comparison of portable emissions measurement systems (PEMS) with laboratory grade equipment. *Appl. Sci.* **2018**, *8*, 1633.
83. Schafer, R.W. What is a savitzky-golay filter? *IEEE Signal Process. Mag.* **2011**, doi:10.1109/MSP.2011.941097.

84. X-engineer.org Air-fuel ratio, lambda and engine performance Available online: https://x-engineer.org/automotive-engineering/internal-combustion-engines/performance/air-fuel-ratio-lambda-engine-performance/ (accessed on Sep 12, 2019).

85. Past weather in Montreal, Quebec, Canada Available online: https://www.timeanddate.com/weather/canada/montreal/historic (accessed on Jul 31, 2020).

86. Wang, Y.; Zou, Y.; Henrickson, K.; Wang, Y.; Tang, J.; Park, B.J. Google Earth elevation data extraction and accuracy assessment for transportation applications. *PLoS One* **2017**, doi:10.1371/journal.pone.0175756.

87. Drucker, H.; Surges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In Proceedings of the Advances in Neural Information Processing Systems; 1997.

88. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* 2004.

89. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, doi:10.1037/h0042519.

90. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, doi:10.1162/089976600300015015.

91. Gers, F.A.; Eck, D.; Schmidhuber, J. Applying LSTM to time series predictable through time-window approaches. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2001.

92. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016; 2016.

93. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, doi:10.1162/neco.1997.9.8.1735.

94. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Gated feedback recurrent neural networks. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015; 2015.

95. Alcan, G.; Yilmaz, E.; Unel, M.; Aran, V.; Yilmaz, M.; Gurel, C.; Koprubasi, K. Estimating Soot Emission in Diesel Engines Using Gated Recurrent Unit Networks. In Proceedings of the IFAC-PapersOnLine; 2019.

96. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

97. Jose, V.R.R.; Winkler, R.L. Simple robust averages of forecasts: Some empirical results. *Int. J. Forecast.* **2008**, doi:10.1016/j.ijforecast.2007.06.001.

98. Wu, M. *Trimmed and winsorized estimators*; Michigan State University, 2006; ISBN 054268795X.

99. Chan, L.W. Weighted least square ensemble networks. In Proceedings of the Proceedings of the International Joint Conference on Neural Networks; 1999.

100. Ferreira, W.G.; Serpa, A.L. Ensemble of metamodels: the augmented least squares approach. *Struct. Multidiscip. Optim.* **2016**, doi:10.1007/s00158-015-1366-1.

101. Hansen, B.E. Least-squares forecast averaging. *J. Econom.* **2008**, doi:10.1016/j.jeconom.2008.08.022.

102. Ren, Y.; Zhang, L.; Suganthan, P.N. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]. *IEEE Comput. Intell. Mag.* 2016.

103. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249.

104. Moradi, E.; Miranda-Moreno, L. On-road vs. Software-based Measurements: On Validity of Fuel, CO2, NOx, and PM Predictions by US EPA's MOVES. In Proceedings of the Transportation Research Board 100th Annual Meeting; 2021.

105. Jiménez-Palacios, J.L. Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing, Massachusetts Institute of Technology, 1999.

106. Natural Resources Canada *Learn the facts: Horsepower's effect on fuel consumption*; 2014;

107. Okuno, I.; Oda, H.; Kaneko, T. Engine control system 1988.

108. Mahmoudi, A.R.; Khazaee, I.; Ghazikhani, M. Simulating the effects of turbocharging on the emission levels of a gasoline engine. *Alexandria Eng. J.* **2017**, doi:10.1016/j.aej.2017.03.005.

109. Dussault, C.; Courtois, R.; Ouellet, J.-P.; Huot, J. Influence of Satellite Geometry and Differential Correction on GPS Location Accuracy. *Source Wildl. Soc. Bull.* **2001**.

110. Gao, Y.; Checkel, M.D. Experimental measurement of on-road CO2 emission and fuel consumption functions. In Proceedings of the SAE Technical Papers; 2007.