Doctoral Dissertation
Doctoral Program in Energetic Engineering (35^{th}cycle)

Data Driven Techniques for On-board Performance Estimation and Prediction in Vehicular Applications

By

Alessandro Falai

******

Supervisor(s):
Prof. Daniela Anna Misul
Prof. Ezio Spessa

Doctoral Examination Committee:
Prof. Alfredo Gimelli , Referee, Università degli studi di Napoli Federico II
Prof. Manfredi Villani, Referee, The Ohio State University
Prof. Stefano D'Ambrosio, Politecnico di Torino
Prof. Phillip Kollmeyer, McMaster University
Prof. James Marco, University of Warwick

Politecnico di Torino
20-06-2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Alessandro Falai
20-06-2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

I would like to dedicate this thesis to my family, which has always supported me throughout my career.

Acknowledgements

First and foremost I am extremely grateful to my supervisors, Prof. Daniela Anna Misul for her invaluable advice, continuous support during my PhD study.

A debt of gratitude is also owed to experienced engineers of the Addfor SPA and Danisi Engineering companies with whom I had the opportunity to collaborate and from whom I learned a lot during my PhD carrier.

Abstract

The primary objective of this doctoral dissertation is to devise data-driven models for the purpose of performance evaluation within the domain of terrestrial transportation. The contemporary automotive industry is witnessing an increasing need for precise estimation methodologies, and data-driven models have surfaced as a viable solution. This dissertation puts forth three significant contributions to fulfil this requirement. Initially, a virtual sensor is suggested for the purpose of promptly forecasting and supervising NOx discharges in diesel engine contexts, particularly during variable on-road driving conditions. The utilization of AI algorithms, specifically the XGBoost machine learning model, has demonstrated exceptional suitability and reliability in the execution of this task. The model has exhibited remarkable flexibility, robustness, and outstanding performance in predicting NOx engine-out. The implementation of this virtual sensor can be carried out on the engine control unit (ECU), thereby facilitating the uninterrupted monitoring and regulation of emissions.

Subsequently, a simulated environment has been created to emulate the functioning of electric vehicles, with a specific focus on evaluating the performance of a two-wheeled electric vehicles. The global model employs the battery discharge current as an input variable and forecasts the instantaneous velocity of the vehicle. The battery model and vehicle dynamic model parameters were established via a calibration procedure utilizing data obtained from on-road experimental measurements. Through the integration of a battery model and a dynamic vehicle model, this environment facilitates a thorough evaluation of the overall performance, thereby assisting in the optimization and design decisions.

Finally, a tool for estimating the optimal state-of-health (SOH) is presented for the purpose of managing battery systems (BMS). The development of the estimator involved the utilization of multiple Bi-LSTM neural networks. These networks were employed to leverage various datasets that contained time series of charge data with varying lengths across the entire SOC domain, creating a predictive tool that might

be smoothly incorporated into the Battery Management System (BMS). This tool accurately predicts battery cell lifetime by determining the optimal state of charge (SOC) window, thereby enhancing battery management efficiency and reliability. The utilization of data-driven models presents notable benefits and enhancements within the realm of land vehicles. The provision of precise and up-to-date evaluations of diverse automobile parameters facilitates the implementation of improved control strategies and optimization methodologies. Moreover, these models serve to facilitate the practice of predictive maintenance, thereby enabling the timely identification of potential faults or degradation. In addition, they assume an essential role in the advancement of eco-friendly and energy-efficient vehicles through the facilitation of emission monitoring and the optimization of battery performance. To summarize, this dissertation presents innovative data-driven models for evaluating the performance of land-based vehicles. These models provide valuable insights and advancements in emission prediction, electric vehicle performance analysis, and battery health estimation. The applications of these models are crucial for enhancing vehicle efficiency, reducing emissions, and improving overall performance and reliability in the realm of land vehicles.

Contents

List of Figures

List of Tables

List of Symbols

Abbreviations

R^2	R-squared, coefficient of determination
AI	Artificial Intelligence
BMS	Battery Management System
CCCV	Constant Current Constant Voltage protocol
CNN	Convolutional Neural Network
CRA	Customized Regression Accuracy
ECM	Equivalent Circuit Model
ECU	Electronic Control Unit
EGR	Exhaust Gas Recirculation
EOL	End Of Life
EV	Electric Vehicle
FNN	Feed-forward Neural Network
ICE	Internal Combustion Engine
LIB	Lithium-Ion Battery
LSTM	Long Short–Term Memory neural network
LUT	Look–UP Table
ML	Machine Learning
NMC	Nickel–Manganese–Cobalt
OBD	On-Board Diagnostic
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RUL	Remaining Useful Life
SIL	Software–In–the–Loop

Physical Variables

$\dot{\theta}$	Motor rotational acceleration

η_{b2r}	Battery–to–road efficiency
λ	air–fuel ratio
ω_{eng}	Engine Speed
ρ_{air}	Air density
θ	Motor rotational speed
A, B, C	Coast–down coefficients
A_f	Vehicle frontal area
C_1	Capacitance of first pair RC
C_2	Capacitance of second pair RC
C_x	Drag coefficient
f	Rolling resistance coefficient
I	Current
$IMAP$	Intake manifold pressure
$IMAT$	Intake manifold temperature
J_{motor}	Motor rotating inertia
L_{opt}	Optimal data length
m	mass
NOx	Nitrogen Oxide
O_2	Oxygen concentration
OCV	Open–Circuit Voltage
P_{batt}	Battery power
P_{diss}	Dissipative power
P_{mot}	Driving power
P_{rail}	Rail pressure
Q_{actual}	Cell actual capacity
Q_{air}	Air quantity
Q_{cell}	cell nominal capacity
Q_{main}	Fuel quantity of main injection
Q_{pil}	Fuel quantity of pilot injection
Q_{rated}	Cell nominal capacity
Q_{tot}	Fuel injected quantity
R_0	Internal resistance
R_1	Resistance of first pair RC
R_2	Resistance of second pair RC
SOC	State Of Charge

SOH	State of health
SOI_{main}	Start of injection of main injection
SOI_{pil}	Start of injection of pilot injection
V	Voltage
V_{veh}	Vehicle speed

Chapter 1

Introduction & Background

1.1 Motivation: technical and historical factors

[1] The automotive sector has undergone a substantial shift in recent times, primarily due to the progress made in data-driven methodologies and the growing accessibility of data from diverse vehicular applications. The capacity to gather, manipulate, and scrutinize extensive quantities of data has introduced novel prospects for augmenting vehicular performance and security. The utilization of data-driven techniques has become imperative in tackling issues pertaining to the estimation and prediction of on-board performance in vehicular applications. The accurate estimation and prediction of performance in vehicles has been a persistent concern within the automotive industry. Historically, physical models and sensor measurements have been utilized to establish performance metrics. Nevertheless, these techniques frequently encounter constraints, such as intricate modeling efforts, elevated computational demands, and uncertainties linked to real-world operational circumstances. The utilization of data-driven methodologies has resulted in a significant change in approach towards the development of performance estimation and prediction models, with a focus on utilizing vast amounts of data to enhance accuracy and reliability.

[1] Part of this chapter has been published in the form of papers as: **(1)** Falai, A.; Misul, D.A. Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine. *Energies*, vol. 16, 2125, 2023; **(2)** Falai, A.; Giuliacci, T.A.; Misul, D.; Paolieri, G.; Anselma, P.G. Modeling and On-Road Testing of an Electric Two-Wheeler towards Range Prediction and BMS Integration. *Energies* 2022, 15, 2431; **(3)** Falai, A.; Giuliacci, T.A.; Misul, D.A.; Anselma, P.G. Reducing the Computational Cost for Artificial Intelligence-Based Battery State-of-Health Estimation in Charging Events. *Batteries* 2022, 8, 209.

The integration of data-driven models in automotive applications has a rich history that dates back several decades. Initially, vehicle performance estimation and prediction were based on empirical models and simplified mathematical formulations. However, with the proliferation of sensors, embedded systems, and connectivity, the automotive industry has transitioned towards data-driven approaches. The incorporation of data-driven models in automotive applications has a substantial historical background that spans several decades and can be attributed to the nascent stages of computing and the advent of machine learning (ML) algorithms. In the beginning, data-driven methodologies were predominantly employed for the purpose of identifying faults and regulating systems in automobiles. The estimation and forecasting of vehicle performance, however, relied heavily on empirical models and rudimentary mathematical expressions. With the advent of enhanced computing capabilities and the availability of vast amounts of data, coupled with the widespread use of sensors, embedded systems, and connectivity, the automotive sector has shifted towards data-centric methodologies, encompassing a diverse array of applications.

The emergence of data-driven techniques in the automotive industry can be attributed to a confluence of technical and historical factors. Technical factors refer to the various aspects of technology that can impact a system or process. These factors can include hardware, software, networks, and other technological components that are necessary for the functioning of a system. Understanding technical factors is important in order to ensure that systems are designed, implemented, and maintained in a way that maximizes their efficiency and performance. The main points can be summarized below.

- The complexity of vehicles has been on the rise, with modern models featuring a multitude of interconnected subsystems and components. The intricate relationships and interactions among various parameters present challenges for conventional analytical models, as their complexity makes it arduous to capture them all. The article depicted in reference [2] illustrates how the proliferation of digitalization, along with its disruptive processes, is poised to dissolve numerous traditional approaches.

- The proliferation of data has been observed as a consequence of the development of sensors, embedded systems, and connectivity in automobiles. Vehicles produce copious amounts of data pertaining to their functionality, usage, ecological circumstances, and user conduct. Conventional analytical methodologies encounter difficulties in managing such a vast amount of data and deriving significant insights from it. For instance, [3] examine currently available car data, including sensor data for autonomous driving, and underscore the potential financial gains that can be realized through the development of services that leverage this data.

- The field of data analysis has witnessed significant advancements in recent times, particularly in the area of machine learning [4]. These developments have introduced novel methodologies for extracting meaningful insights from voluminous and intricate datasets. Machine learning algorithms have the capability to detect patterns, correlations, and anomalies in data that may not be immediately discernible through conventional analytical techniques.

- Real-time decision making is a critical factor in optimizing performance, efficiency, and safety within the automotive industry. In order to adhere to rigorous global standards pertaining to safety, quality, sustainability, and efficiency, automotive suppliers are obligated to conform to novel product and requirements [5]. The utilization of data-driven techniques allows for the ongoing observation and examination of data, which in turn enables the ability to promptly adjust and make informed decisions in a proactive manner.

Factors pertaining to the past events and circumstances that have influenced the current state of the art.

- The triumph and extensive implementation of data-driven methodologies in sectors such as finance, healthcare, and e-commerce have motivated the automotive industry to investigate their prospective benefits. Automotive manufacturers and researchers have demonstrated an increased interest in the development of data-driven models and algorithms for a variety of applications, such as performance estimation and predictive maintenance, due to the observed benefits in other domains.

- The incorporation of sophisticated sensors and embedded systems in automobiles has transformed the process of data acquisition and application. In the initial stages of vehicle development, the sensory functionalities were restricted to measuring fundamental parameters such as engine RPM and speed. With the progression of technology, a wider array of sensors, including but not limited to GPS, accelerometers, pressure sensors, and temperature sensors, have become ubiquitous. The sensors offer a plethora of information that can be utilized for evaluating and forecasting performance.

- The emergence of big data and cloud computing has had a notable impact on the integration of data-centric methodologies within the automotive industry [6]. The advent of cloud-based computing resources and scalable storage solutions has enabled the efficient processing and analysis of vast quantities of data. The establishment of a direct virtual connection among human, vehicle, and infrastructure has facilitated our entry into the Digital Twin epoch [7].

It is clear how the proliferation of connected vehicles and telematics systems has intensified the demand for data-driven methodologies in the realm of connectivity and telematics. The ability of connected vehicles to transmit real-time data to central servers or cloud platforms facilitates uninterrupted monitoring and analysis. The interconnectivity has facilitated prospects for remote diagnostic, prognostic maintenance, and enhanced performance optimization.

In terms of theoretical explication, it is pertinent to commence with the definition provided by Toshika Srivastava, an accomplished Artificial Intelligence (AI) technical team lead at Audi. Srivastava characterizes data-driven development as the process of creating features, tools, methods, and services based on the insights derived from the vast amount of data that an organization has accumulated. This involves acquiring knowledge of features and patterns from databases to facilitate specific functionalities [8]. Here is explained how in contemporary times, a prevalent approach to extracting insights from data is through the utilization of artificial intelligence and machine learning techniques, which have demonstrated high levels of efficacy. The methodologies being utilized in the development of the described software are also encompassed within the product development cycle, which is a multi-step process, as shown and referred in Srivastava's work [8].

This study provides a comprehensive account of the product development cycle, focusing on the in-vehicle acquisition utilized for the purposes of development, testing, and implementation of a data-driven approach. The workflow delineates the systematic approach of data acquisition, refinement, and augmentation, with a primary emphasis on data preparation, labeling, and subsequent training and testing in a simulated environment. Subsequently, the tool that has been developed may undergo code generation and hardware integration processes, enabling it to be subjected to Hardware–In–Loop (HIL) testing and ultimately facilitating its actual vehicle on-board implementation.

The inquiry remains unresolved: What are the benefits of developing and employing data-driven models for the prediction of vehicle performance? Let us proceed to furnish specific particulars regarding the potential implications of this phenomenon within the realm of the automotive sector.

1.2 Evolution of land vehicle performance: history pills

The remarkable advancements in ground vehicle performance witnessed today are the result of a lengthy evolutionary process and significant milestones attained through comprehensive research. The advent of the internal combustion engine during the latter part of the 1800s represented a significant turning point in the realm of terrestrial vehicle capabilities. The utilization of petroleum-based fuels enabled the propulsion of vehicles, thereby facilitating the advancement of automobiles and motorcycles. In the 1885, Karl Friedrich Benz is known to have made the first true automobile. It was a gasoline powered automobile with an internal combustion engine and it had three wheels [9].During the early 20th century, the launch of the first "Performance Cars" occurred. The primary objective of a performance automobile was to possess maneuverability, reduced weight, and an aerodynamic configuration optimized for competitive racing. The Tatra Rennzweier is considered to be among the pioneering automobiles that were exclusively engineered for motor sport purposes [10]. The emergence of mass production techniques, which were pioneered by Henry Ford, marked a significant development in the early 20th century. In 1908, the Ford Model T was unveiled as the inaugural mass-produced automobile that was economically accessible to the general public. This event denoted a noteworthy change in the proficiency of terrestrial vehicles by rendering automobiles attainable to a broader demographic, thereby revolutionizing the domain of transportation and the community at large [11]. During the 1920s and 1930s, scholars directed their attention towards optimizing aerodynamics in order to augment the performance of land vehicles. The implementation of streamlining techniques, such as the utilization of sloping designs and enclosed bodies, resulted in a decrease in drag and an enhancement of fuel efficiency. Streamlined vehicles were utilized to establish land speed records, thereby expanding the limits of vehicular performance. The enhancement of land vehicle performance was significantly influenced by the evolution of suspension systems. Advanced suspension systems offer improved maneuverability, stability, and comfort. The implementation of advanced features such as independent suspension, shock absorbers, and anti-roll bars has significantly augmented the handling and steering capabilities of automobiles. Advancements in engine technology have resulted in notable enhancements in power output over the course of several decades. The implementation of advanced fuel injection systems, turbocharging mechanisms, and variable valve timing has resulted in a significant enhancement of both performance and efficiency. Engines with high performance, such as those commonly found in sports cars, exhibit remarkable capabilities in terms of acceleration and maximum velocity. The incorporation of electronic systems and controls has significantly influenced the operational efficiency of land vehicles. The implementation of electronic fuel injection systems, anti-lock braking systems (ABS), traction control, stability control, and advanced driver-assistance systems (ADAS) has resulted in improved safety, performance, and handling capabilities. The prominence of alternative fuel and electric vehicles has increased due to mounting environmental concerns. Hybrid automobiles, which utilize both internal combustion engines and electric motors, provide enhanced fuel economy. Electric vehicles that operate solely on electricity are capable of eliminating emissions and delivering immediate torque, thereby transforming the performance of land vehicles. Rimac and Tesla are prominent corporations that prioritize electric technologies and the advancement of hypercars. Hence, within the power systems of an automobile, the energy storage system, which presently comprises lithium-ion batteries, assumes a pivotal role with regard to safety, efficacy, ecological footprint, and performance.

The utilization of software modeling has played a significant role in the development and production of high performance vehicles over the past century. This can be attributed to the emergence of new technologies, the imperative to mitigate environmental impact, and the pursuit of optimal levels of safety and efficiency. Historically, the foundation of scientific and engineering knowledge was reliant on limited data that was often acquired through experimental methods designed to test a particular hypothesis (physics-based models). The experimental results were constrained to a narrow scope, producing a restricted amount of data. Currently, there is a copious amount of data that is readily available and can be easily collected in every individual experiment at a relatively low expense. The utilization of data-driven modeling and scientific discovery represents a paradigm shift in the approach to problem-solving across various fields of science and engineering.

1.3 Physics-based models vs data-driven models in automotive sector: literature review

In recent years, the engineering industry, particularly in the automotive sector, has witnessed a distinct divergence between the two primary methodologies employed for modeling vehicle systems and their constituent parts. These methodologies are characterized as physics-based models and data-driven models. This main classification of the two methodologies adopted in practice could be found in [12]. A simplified representation of such classification is shown in the Figure 1.1 which depicts the categorization of the two methodologies involved. Occasionally, it is possible to develop a methodology that combines both approaches into a hybrid blended format.

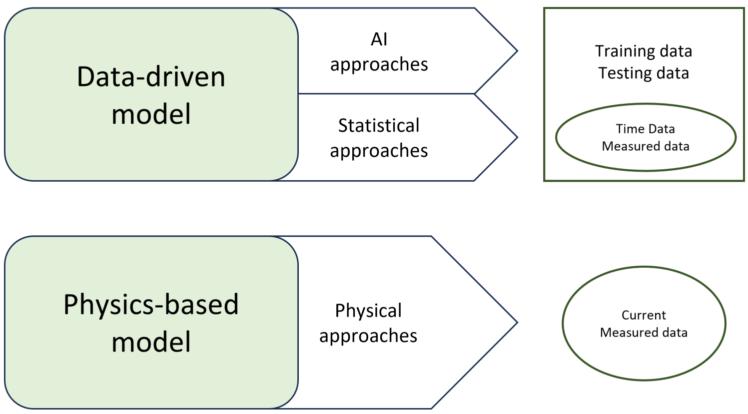


Fig. 1.1 Classification of modeling approaches: data-driven, physics-based.

Physical models are propelled by an understanding of specific mechanisms. The automotive industry has been utilizing mathematical equations to advance their product development and conceptualization based on the physical properties of vehicles for a considerable duration. During the design phase, the physical behavior of the vehicle system and its individual components, such as the engine [13–15], suspension [16–18], tire [19], powertrain [20], and after-treatment system [21] for burnt gases, are modeled using advanced software. The utilization of physics-based modeling necessitates numerous experimental methodologies to procure the parameters of said models, leading to increased development costs and time [22]. As previously mentioned, data-driven models are constructed through the utilization of algorithms that acquire knowledge of patterns and correlations from the data itself, without any explicit comprehension of the fundamental physical principles that govern the system. The aforementioned models employ machine learning methodologies to extract valuable insights and generate forecasts by analyzing patterns derived from either historical or real-time data.

Numerous studies in the literature have focused on the development of AI/ML models for the purpose of estimating and predicting performance in the vehicular application domain. For instance, the process of estimating an advanced model of vehicle dynamics is a complex task, as it involves dealing with uncertainties and nonlinearities [23]. Moreover, the computational expense associated with resolving the dynamics of intricate multi-body systems, such as automobiles, is considerable. To adequately tackle this matter, the utilization of data-driven modeling through deep learning methodologies offers a proficient approach for instantaneous simulation of the multi-body systems of vehicles [24]. The second chapter of this work focuses on a key topic, namely the reduction of polluting emissions in diesel vehicles. Within this field, a number of studies in the literature have explored the use of neural networks for various applications. The study presented in reference [25] compares the effectiveness of an Artificial Neural Network (ANN) and a Neuro-Fuzzy approach in predicting the volumetric oxygen concentration at the engine intake during transient operational conditions. In [26] An ANN was used for NOX evaluation in real driving emission (RDE) tests conducted using vehicles equipped with portable emissions measurements systems.

The present dissertation's main points of interest are identified and accompanied by relevant literature works as follows. Numerous studies have been devoted to the design and development of battery systems in response to the widespread commercialization and increasing market penetration of electric vehicles (EVs). The primary emphasis of their efforts has been directed towards achieving enhanced energy efficiency, superior thermal performance, and optimized designs for battery enclosures that incorporate multiple materials. The implementation of simulation-based design optimization for the battery pack and Battery Management System (BMS) has undergone development and now encompasses advancements such as artificial intelligence/machine learning (AI/ML) to enhance efficiency in design, manufacturing, and operations for their utilization in electric vehicles and energy storage systems. Regarding battery management systems (BMS), these sophisticated concepts facilitate a more precise estimation of battery functionality, including the State of Health (SOH) and State of Charge (SOC) estimations.

Accurate estimation of the state of charge (SOC) is of paramount importance in prolonging cell lifespan and guaranteeing safe operation in electric vehicle applications. In [27] a deep learning-based transformer model is developed without the requirements of feature engineering. In addition, research involving comparative analyses or the implementation of hybrid artificial intelligence models has yielded favorable outcomes in the prediction of State of Charge (SOC), as evidenced by references [28–31].

The evaluation and control of cells deterioration, particularly in the context of electrified vehicular operations, is a critical concern and a topic of significant discourse in contemporary times. The monitoring of the state-of-health (SOH) is of utmost importance in this regard. Over the past year, a number of novel works have been produced in the context of SOH estimation [32–34]. During this period, researchers have explored the potential of artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNN) for monitoring the remaining useful life of batteries. This investigation was made possible by the availability online of open-source data pertaining to cell aging tests conducted in specialized experimental laboratories.

Battery modeling has demonstrated to be a valuable tool for the on-board implementation of battery management systems (BMS) in predicting and optimizing fundamental battery parameters, including state of charge (SOC), state of health (SOH), and charge/discharge characteristics, as well as the overall performance along its lifespan. Various individual models exhibit variations in their complexity, computational expenses, and overall predictive accuracy, rendering them appropriate for diverse domains of application [35]. In the literature, models utilizing either Equivalent circuit models (ECM) [36, 37] or ML [38–40] are commonly employed for the purpose of monitoring, controlling, and managing batteries through BMS in a software-in-the-loop (SIL) environment. Battery modeling, in-vehicle diagnostic tools and data-driven approaches can be combined in a framework to create a battery digital twin [41]. The utilization of battery digital twins enables researchers and engineers to enhance battery designs, refine battery management systems, and formulate sophisticated control strategies. The utilization of virtual replicas allows for the evaluation of crucial parameters such as state-of-charge and state-of-health without the necessity of physical testing, ultimately leading to a reduction in both time and costs [42].

1.4 Objectives and original contributions

The present dissertation aims to address the current challenges in the automotive industry, focusing on three key objectives: the development of a real-time estimator of NOX emissions in diesel engines that can be incorporated into the engine control unit, or more broadly, the electronic control unit (ECU), the modeling of batteries for their implementation in batteries BMS and the development of a lithium cell state of health (SOH) estimator for implementation on BMS. Through this research, it aims to contribute to the development of advanced technologies and improvements in the automotive sector by exploiting data-driven techniques. Within this framework, there exist a number of unresolved matters that continue to serve as the foundation for the current study.

- Original Equipment Manufacturers (OEMs) in the automotive industry are encountering several obstacles in the contemporary landscape. The automotive industry is distinguished by a rise in ecological regulations and consumer demands for enhanced performance, efficiency, and reduced emissions. OEMs must therefore constantly seek to improve their vehicles to meet these ever-increasing demands.

- Moreover, the shift towards electric mobility poses a significant opportunity and challenge for OEMs. The incorporation of lithium battery technologies and the enhancement of battery management systems (BMS) are crucial factors for the advancement of dependable and effective electric vehicles.

This dissertation aims to address the aforementioned needs and make a contribution to the research domain of developing performance estimators and predictors that can be integrated on-board in ground vehicles. To achieve this, the following research objectives have been identified:

- Development of virtual sensing techniques based on machine learning for the control and monitoring of engine-out NOx emissions under transient on-road vehicle operating conditions. The model that has been formulated utilizes the predictive potential of the extreme gradient boosting (XGB) algorithm. The research presents a potential methodology for designing a virtual sensor as an alternative to conventional in-vehicle physical sensors, thereby mitigating their limitations.

- Development of a data-driven identification approach for modeling a lithium-ion battery-powered two-wheeler vehicle that can be used to analyze its performance under real-world driving conditions. The objective of this study is to create a computational model that can verify and describe the behavior of batteries by means of experimental tests conducted on city routes.

- Development of a deep learning model-based estimator that can accurately predict the state of health (SOH) in lithium-ion batteries in real-time, while also maintaining a low computational cost. The objective of this study is to examine a methodology that has the potential to enhance the accuracy of the estimation of the battery's remaining useful life. This feature can be integrated into the BMS.

The present dissertation has been designed as a papers collection, therefore the format of the manuscript was chosen based on the type of structure.

1.5 Outline of the dissertation

The dissertation comprises three distinct research chapters, namely Chapter 2 through Chapter 4.Finally, Chapter 5 presents the primary findings and suggestions for future research.
In Chapter 2, the development of a virtual sensor for NOx monitoring in diesel engines was carried out using a machine learning approach, specifically the Extreme Gradient Boosting (XGBoost) algorithm. A campaign of experimentation was conducted to gather data from the engine test bench and the engine electronic control unit (ECU) in order to develop and calibrate a virtual sensor under steady-state conditions. Subsequently, the virtual sensor underwent comprehensive testing during an on-road driving mission to assess its prediction performance under dynamic conditions. The

calibration procedure underwent an optimization process that enabled the assessment of the efficacy of this methodology in the real-time estimation of NOx.
In Chapter 3, the development and subsequent assessment of a two-wheeled electric scooter model, with a focus on its performance under real-world driving cycle conditions. The proposed model is based on the energy based-longitudinal dynamic approach. Additionally, it is integrated with a second-order RC equivalent circuit model for Li-Ion Battery, which enables accurate prediction of the electric range. The model's validity was confirmed through experimental tests carried out in urban streets, which were also utilized to retrieve the principal parameters of the model. Subsequently, an assessment was conducted to gauge the effectiveness and electrical power of the two-wheeled pure electric vehicle. The development and assessment of electric vehicle models serve as a foundation for the design of BMS. This approach offers an efficient and cost-effective method to examine the optimal control logics of batteries within a Software-in-the-Loop setting.
Chapter 4 focuses on the development of a computationally lightweight approach for estimating the SOH of lithium-ion cells in electric vehicles during partial charging procedures. The utilization of artificial intelligence (AI) algorithms has demonstrated significant potential as a data-driven modeling technique for predicting the SOH of cells. This is attributed to their high level of suitability and low computational requirements. The attainment of a precise on-board SOH estimation is accomplished by identifying an optimal State of Charge (SOC) window during the cell charging procedure. A random-search algorithm has been utilized to train multiple Bi-LSTM networks using data from a constant current constant voltage (CCCV) test protocol. Finally, a potential integration within the BMS control unit could enable it to guarantee optimal performance and extend the lifespan of the cells.
The last chapter in this study will provide a summary of the scientific discoveries and outcomes of each research theme, while also suggesting potential future advancements to enhance the practicality of the models presented.

Chapter 2

Real-time pollutant emissions estimation in heavy-duty CI engines using data-driven approach

[1] Pollution from heavy-duty diesel engines is a significant concern due to its negative impact on human health and the environment. Heavy-duty diesel engines are a significant contributor to air pollution and have been associated with various health problems. Nitrogen oxide (NOx) emissions from these engines are particularly problematic, as they contribute to air pollution, climate change, and respiratory problems. Traditional approaches to estimate these emissions using mathematical models based on engine performance parameters have limitations and can be inaccurate due to the complex nature of diesel engine combustion. Additionally, the current emission regulations impose strict limits on NOx emissions, and compliance with these regulations is crucial for engine manufacturers. Real-time monitoring and control of these emissions are necessary to ensure compliance and improve engine efficiency. However, current monitoring techniques are often costly, time-consuming, and require complex equipment, making them unsuitable for real-time monitoring.

Recently, data-driven approaches, such as machine learning and deep learning algorithms, have shown promise in improving the accuracy of real-time pollutant emissions estimation. Machine learning algorithms can learn patterns from large

[1] Part of this chapter has been published in the form of a paper as: Falai, A.; Misul, D.A. Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine. *Energies*, vol. 16, 2125, 2023.

datasets and make predictions based on the learned patterns. On the other hand, deep learning algorithms, a subset of machine learning, use artificial neural networks to extract features from raw data, allowing for more accurate predictions. For instance, these algorithms can be trained using engine sensor data such as pressure, temperature, and fuel injection rate, to estimate pollutant emissions in real-time. The advantages of using data-driven approaches for real-time pollutant emissions estimation in heavy-duty CI engines are:

- Improved accuracy: Data-driven approaches can learn patterns from large datasets and make predictions based on the learned patterns. This can lead to more accurate estimations of pollutant emissions compared to traditional mathematical models.
- Real-time monitoring: Data-driven approaches can provide real-time estimations of pollutant emissions, allowing for timely and effective control of emissions.
- Cost-effective: Data-driven approaches can be implemented using existing engine sensors and data acquisition systems, making them a cost-effective solution for real-time pollutant emissions estimation.
- Reduced complexity: Data-driven approaches eliminate the need for complex mathematical models and can be easily integrated into existing engine control systems.

In this chapter, we will discuss the application of data-driven approaches for real-time NOx emissions estimation in heavy-duty CI engines. We will review the literature on the topic and provide an overview of the methodology used for data collection, processing, model training and validation.

2.1 Introduction

The emission of pollutants from vehicles, particularly nitrogen oxide (NOx), is a significant issue of both environmental and public health significance. The implementation of stringent regulations on a global scale has been aimed at mitigating NOx emissions from vehicles. In response, car manufacturers are proactively exploring avenues to enhance the efficiency of their internal combustion engines (ICEs) to comply with these regulations [43].

The predominant methods for mitigating and regulating engine-out NOx primarily encompass the subsequent approaches: the implementation of Exhaust Gas Recirculation (EGR) [44–48] and advanced combustion techniques, such as homogeneous charge compression ignition (HCCI) [49, 50] and premixed charge compression ignition (PCCI) [51, 52], have been shown to effectively reduce NOx emissions by regulating the air–fuel mixture and combustion process. Additionally, diverse injection strategies targeting pressure and timing [53, 54] have been explored for further NOx reduction. Concerning vehicle tailpipe emissions, the primary techniques for reducing NOx emissions include the utilization of filters and catalysts. One such catalyst is Selective Catalytic Reduction (SCR) [55], which involves the injection of a reducing agent, such as urea, into the exhaust stream to reduce NOx emissions through a catalytic converter. Another technique is the Lean NOx Trap (LNT) [56], which temporarily stores NOx emissions in a catalytic converter for reduction at specific engine operating conditions. Additionally, the Diesel Particulate Filter (DPF) [57] utilizes a filter to capture particulate matter from the exhaust stream and can also assist in reducing NOx emissions. These techniques are commonly employed to mitigate the negative impact of vehicle emissions on the environment. However, achieving the desired level of NOx reduction performance is contingent upon the utilization of dependable NOx sensing technology.

There exist various models that can be employed to predict NOx emissions in engines. One established method for calibrating the model is through the exploitation of engine maps. During transient conditions, the model can be tested on-road [58–60]. Engine maps are graphical representations that illustrate the correlation between different engine operating variables, such as load, speed, and fuel flow rate, and the resultant engine performance attributes, such as power output and emissions. The aforementioned maps possess the capability to forecast the NOx emissions of an engine based on a specific set of operating conditions [61, 62]. In the study referenced in [63], a technique was devised to utilize engine maps for the purpose of predicting NOx emissions. The electronic control unit (ECU) was equipped with an engine map, which was expressed through mathematical models. This map was utilized to calculate NOx emissions based on the engine's observed operating conditions. The calibration of the model was conducted through the utilization of engine test bench data, followed by an on-road assessment to evaluate its predictive capacity in real-world scenarios. The literature contains various applications pertaining to virtual NOx sensing that utilize engine map operating points during steady-state conditions [64]. Additionally, applications derived from experimental test bench acquisition have also been documented [65]. Insufficient scholarly inquiry has been carried out regarding the virtual sensing and prediction of NOx emissions in the context of authentic on-road driving conditions, while also factoring in the impact of transient events.

The reduction of NOx emissions continues to be a significant challenge in contemporary times. To circumvent the limitations of physical solid-state sensors in obtaining engine-out NOx information in diesel engine applications, virtual sensing techniques have gained widespread usage, as evidenced by sources such as [66] and [67]. In recent times, there have been various endeavors to develop virtual sensors that utilize machine learning algorithms [68, 69]. The objective of these sensors is to forecast the engine-out NOx levels by considering the engine operating conditions, as obtained from the ECU. The potential of machine learning algorithms as virtual measurement tools is noteworthy, owing to their capacity to identify highly non-linear behavior in the examined physical system [70, 71]. Various machine learning algorithms have been utilized for virtual sensor applications in predicting engine NOx. These algorithms include linear regression [72], support vector machines (SVM) [73], and artificial neural networks (ANNs) [74]. According to a study published in [75], the XGBoost model has demonstrated a notable level of efficacy in comparison to other machine learning models, including gradient boosting (GBT) and random forest (RF). XGBoost's capability to manage voluminous datasets and handle missing values makes it a suitable choice for virtual sensor applications, where data may be incomplete or contain noise. Furthermore, it has been determined that XGBoost is a powerful and efficient machine learning instrument for a diverse range of applications, such as engine-out NOx levels prediction[76]. Recent research has investigated the utilization of ensemble models, including random forest and adaptive boosting, in conjunction with XGBoost [77, 78]. Ensemble models amalgamate the forecasts of several individual models to attain enhanced precision, while maintaining a level of resilience comparable to that of the individual models. Furthermore, ensemble methods primarily depend on randomization techniques, resulting in the generation of numerous distinct solutions for the given problem. Within this particular framework, it has been observed that the XGBoost algorithm has demonstrated superior predictive capabilities in comparison to other ensemble models, including GBT and RF, [79]. Therefore XGBoost turned out to be a model particularly suitable for the use of real-time applications. In addition, several Extreme Gradient Boosting Regression Tree models have been developed to examine the precision of estimating physical parameters, including tailpipe NOx emissions [80, 81]. The existing body of literature is constrained in terms of investigating the regulation and monitoring of NOx emissions from engines during transient conditions. This study aimed to investigate the application of the XGBoost algorithm in real-time engine-out NOx sensing under real-world driving conditions. The virtual sensor utilizing XGBoost was calibrated under stationary conditions and subsequently validated during dynamic on-road traces, as part of experimental efforts carried out both on the test bench and on the road.

2.2 Materials and methods

In recent years, virtual sensing techniques have emerged as a promising alternative to physical measurement instruments, particularly in situations where the latter are economically or practically challenging to use. In this context, the present study focuses on the development of a machine learning-based virtual sensor for predicting NOx emissions in a diesel engine application. To ensure the robustness and reliability of the developed model, a thorough analysis was conducted on the steady-state data acquired from the engine. The performance evaluation of the machine learning algorithms was performed using various metrics such as root mean squared error (RMSE), and R-squared value. The results of this analysis helped to identify the most suitable architectures of the machine learning algorithm for the specific problem task. Following this, the study employed XGBoost, a state-of-the-art machine learning algorithm, to develop the virtual sensor for predicting NOx emissions. The performance of the developed model was verified and validated through an experimental campaign conducted on-road. The real-world driving mission data was collected using an on-board diagnostic (OBD) system and the developed virtual sensor was tested against it. The evaluation of the prediction accuracy was performed using the same metrics as before. The case study vehicle and the involved datasets, as well as the proposed method for model training and validation, were thoroughly investigated in this study. The data used in this study were collected under controlled laboratory conditions, i.e steady-state tests, and in the real-world driving scenarios, i.e transient tests. The data processing and cleaning techniques were applied to ensure the accuracy and reliability of the data.

The present study followed a consistent and well-defined methodology which is composed by sequential steps for the robustness analysis and validation process, and shown in Figure 2.1.

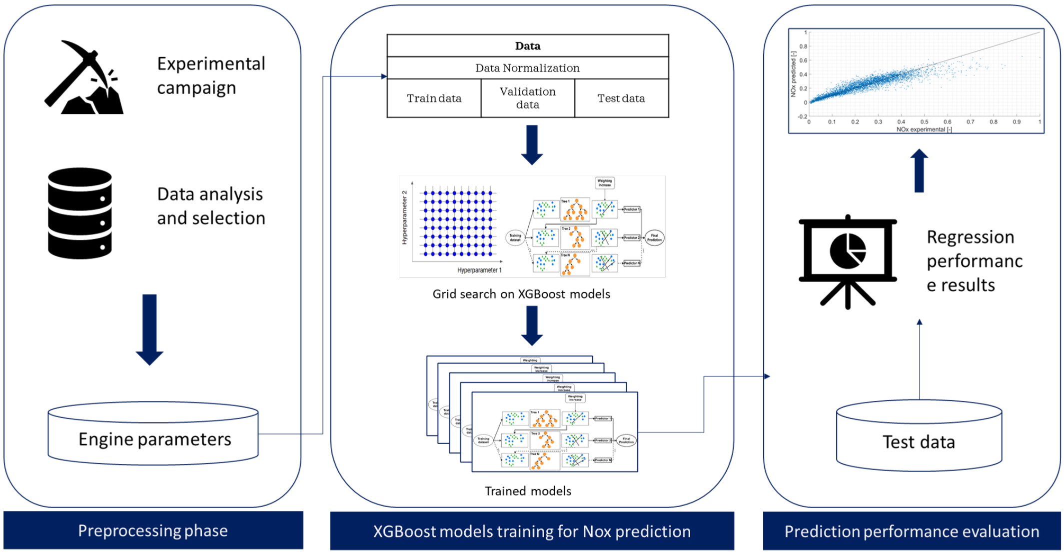


Fig. 2.1 The sequential steps describing the proposed methodology of the performance model evaluation.

As shown in figure, the main sections are described as composed by several sequential steps, filling each represented box. Specifically, the method follows a recognized approach in the field of data science and model development. The three main phases and relative sub-steps are described as follows:

- preprocessing phase: the engine data values acquired from the experimental campaigns were analyzed, handled, and cleaned to ensure their accuracy and reliability. The data were checked for outliers, missing values, and inconsistencies, and appropriate measures were taken to address any issues found. The cleaned data were then used for the subsequent phases of the study;
- model training: the data were normalized and split into training, validation, and test sets as necessary. The XGBoost algorithm was chosen as the machine learning model of choice for this study due to its ability to handle complex non-linear relationships and its high prediction accuracy. Several XGBoost

architectures were trained using the training set to find the most accurate model for the specific case study. To achieve this, the grid search algorithm was employed as a powerful hyperparameters tuning technique to identify the optimal combination of hyperparameters for each XGBoost model architecture. The performance of an AI model is heavily dependent on the hyperparameter values, so a combination of grid search and cross-validation approaches was used to identify the optimal values for a specific model. The hyperparameters of the XGBoost model were specified to tailor the model for the specific application.;

- Performance evaluation: the best XGBoost architecture was evaluated by considering a test dataset according to different metrics, i.e., RMSE and the coefficient of determination (R-squared). RMSE is a measure of the difference between the predicted and actual values, while R-squared is a measure of how well the model fits the data. The XGBoost model's performance was evaluated by comparing the predicted values with the actual values of the test dataset.

To evaluate the development approach and better emphasize the practical details that were engaged throughout the current work, it is necessary to provide a detailed description of the major steps briefly described above.

2.2.1 Preprocessing phase

In this study, the case study for the deployed AI-based virtual sensor was an 11 L diesel engine for heavy-duty application. The engine was tested both at steady-state and under transient conditions to capture a comprehensive range of operating conditions. The details of the experimental campaign, together with a deep insight into the experimental set-up and the engine behavior, can be found in [82]. The engine's primary parameters during steady-state conditions were obtained via an engine test bench that was equipped with physical sensors integrated into the acquisition system. Additionally, the variables obtained from the ECU sensors and maps were also utilized. These parameters correspond to the main descriptors of the combustion process inside the cylinders, as well as the formation of the emission at the engine outlet. On the other hand, in the transient case, a prototype heavy-duty (HD) vehicle equipping the same CI engine was tested on-road to capture different driving conditions under real-world scenarios, and the main engine parameters were acquired by the on-board system acquisition. The acquired engine data values were analyzed and handled in the preprocessing phase to ensure their accuracy and reliability. Any missing or erroneous data values were identified and corrected or removed as appropriate. The data were checked for outliers and inconsistencies, and appropriate measures were taken to address any issues found. The cleaned engine data values were then used for the subsequent phases of the study, including the XGBoost models training phase and the performance evaluation phase. The preprocessing phase was critical for a data driven-based virtual sensor design and assessment, especially for real-time applications. Table 2.1 provides an overview of the main characteristics of the tested vehicle, including the engine type and test conditions. The table serves as a useful reference for the reader and provides context for the subsequent phases of the study.

Table 2.1 Main engine parameters and test characteristics.

Segment of Application	**Displacement**	**Turbocharger**
Heavy-duty vehicles & trucks	11 L	VGT type
Fuel injection system	**Engine params**	**n° engine points**
High pressure common rail	14	4711 [1]

[1] The operating engine points are related to the steady state tests over the engine map.

During the experimental campaign, the main engine parameters acquired in the steady-state and on-road scenarios follow:

- ω_{eng}: engine speed in [RPM];
- Q_{tot}: total amount of injected fuel inside the cylinders in [mm^3/(cycle $\times$ cylinder)];
- P_{rail}: pressure in the rail system in bar;
- Q_{main}: amount of injected fuel during the main injection in [mm^3/(cycle $\times$ cylinder)];
- Q_{pil}: amount of injected fuel during the pilot injection in [mm^3/(cycle $\times$ cylinder)];
- SOI_{main}: start of the injection of the main injection in degree;

- SOI_{pil}: start of the injection of the pilot injection in degree;
- $IMAP$: pressure value in the intake manifold in [bar];
- $IMAT$: temperature value in the intake manifold in [K];
- O_2: oxygen concentration in the chamber in %;
- Q_{air}: amount of air introduced in the chamber in [kg/(cycle $\times$ cylinder)];
- λ: ratio between air and fuel quantities.
- EGR: amount of exhaust gas recirculated in [kg/(cycle $\times$ cylinder)];
- NOx: amount of nitrogen oxides emitted in [PPM].

The experimental campaign was conducted at Politecnico di Torino utilizing an engine test bench that was outfitted with an ELIN APA 100 AC dynamometer. The gaseous emissions that were produced by the engine without any treatment were quantified using an AVL AMAi60 device, which was equipped with two sets of instruments for measuring the concentrations of the primary gaseous species. These measurements were taken at both the intake and exhaust manifolds simultaneously. The testing equipment was configured with thermocouples and piezoresistive pressure transducers to collect temperature and pressure data at multiple locations, encompassing both upstream and downstream positions relative to the turbine, compressor, and intercooler. Additionally, measurements were taken in the intake manifold and EGR circuit [83]. The experimental acquisitions were executed utilizing diverse engine strategies, which are not disclosed due to confidentiality concerns. The information is consistently presented in a normalized form. The data obtained from the engine experiment was utilized in the development of a machine learning model and the creation of a real-time virtual NOx sensor. Figure 2.2 displays the engine's steady state operating points as a function of engine speed and total injected fuel quantity.

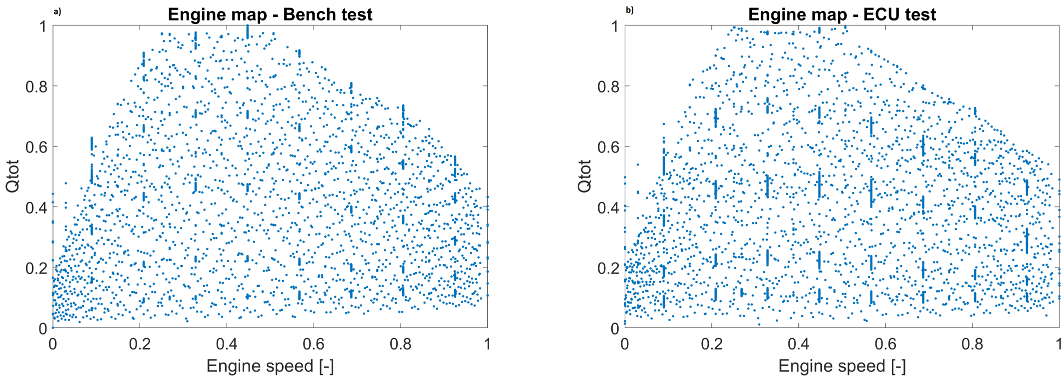


Fig. 2.2 Engine operating points in stationary conditions through (**a**) test bench and (**b**) ECU acquisition systems.

The graphs depict the typical configuration of the engine map shape and the full load (FL) curve, along with the implementation of various experimental strategies that have resulted in an atypical grouping phenomenon. The experimental acquisition was conducted under steady-state testing conditions, resulting in the elimination of transient events' characteristic time for each engine point represented on the map. In contrast, the road tests were characterized by transient events exhibiting a high degree of variability. Consequently, the sampling frequency utilized for acquiring the engine variables was established at 100 Hz.

2.2.2 Model training approach

The development and assessment of the NOx predictor involved the use of the XGBoost algorithm, which is known for its high accuracy and efficiency in regression problems [84]. In this study, the XGBoost algorithm was used to develop virtual sensors for predicting nitrogen oxides in the compression ignition engine-equipped vehicle. The mathematical problem belongs to the class of supervised learning, due to the presence of desired output corresponding to the data labels for the learning process.The steady state analysis involved defining and training the machine learning models on engine operating points detected using two different acquisition systems: a test bench equipped with physical sensors and the ECU sensors and maps. As required by machine learning algorithms for training and testing purposes of models, the whole dataset was split in training, validation and test sets, before fed into the process phase corresponding to the learning tasks. The training set is the set of data that is used to train model in learning the hidden patterns in the data and should have a diversified set of inputs so that the model is trained in all scenarios. The validation set is a set of data used to validate the model performance during training and it gives information for hyperparameters tuning process. Finally, the test set is a set of data used to test the model after completing the training and it provides an unbiased final model performance metric. To ensure the robustness of the selected models, a sensitivity analysis was performed on the size of the datasets used for the training and testing phases. As machine learning algorithms are driven by data, the amount of data supplied to the training phase can have a significant impact on the learning performance. However, increasing the number of training points can lead to overfitting, where the model learns the training data too well and fails to generalize to new data. To avoid this, the size of the training dataset was reduced while increasing the number of test datasets, allowing for the assessment of the models' robustness despite a potential loss in accuracy. The amount of data considered in the validation set was set to be the 10% of the training set, and for the sensitivity analysis the datasets' size is computed defining the test set dimension through $train\_test\_split$ parameter, which was set to $[10, 50, 95]$ values as the percentage of total amount of data. On the other hand, the training set is the complement to one of the validation and test set. The three different values employed in the sensitivity analysis is shown in Figure 2.3.

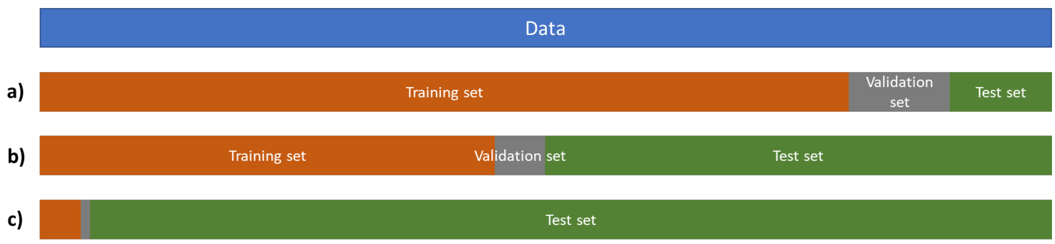


Fig. 2.3 Splits between train, test and validation over sensitivity analysis. The test set size is **(a)** 10%, **(b)** 50% and **(c)** 95% of the whole dataset, rispectively.

Since machine learning models are data-driven techniques, there is no optimal percentage of the train-test split. Generally, the split percentage is highly dependent on the project goals and taking into consideration the following factors:

- computational cost in the model training,
- computational cost in the model evaluation,

- training set representativeness,
- test set representativeness.

As seen in Figure 2.1, the second phase of the data split and learning process takes data normalization into consideration during the preprocessing phase. Typically, machine learning models, particularly those belonging to the deep learning branch, need a step consisting of data normalization prior to the learning phase in the order to increase the model stability and accelerate training. Nevertheless, not every models' performance is impacted by this strategy. Similar to the current research, a decision tree-based algorithm provides no benefit in terms of the prediction accuracy through its normalized data, the learning phase may be accelerated. In addition, for purposes of confidentiality, the numerical values of the data have been hidden by using the Min-Max normalization described in Equation 2.1

$$X_{i,norm} = \frac{X_i - min(X_i)}{max(X_i) - min(X_i)} \tag{2.1}$$

To identify the most influential and weighted variables, a feature extraction approach was employed during the creation of the predictors. This approach, commonly known as feature importance, is considered a critical preprocessing step in creating ML models. The feature extraction process involves analyzing the input data and selecting the most relevant features for the model. This not only helps to improve the accuracy of the model but also reduces the computational burden, making it more suitable for practical applications. Moreover, feature extraction is particularly useful if the number of system variables is large and finding the most relevant ones is an objective of the analysis, or if is important to know which features are driving the model predictions. In the case of the XGBoost algorithm, feature importance is relevant for several reasons:

- Model interpretability: by identifying the most important features, a better understanding of how the model is producing the predictions, and how different features are interacting with each other, can be gained. This can be helpful to interpret the results of the models and to make more informed decisions based on its predictions.
- Feature selection: identifying the most important features can also be useful for feature selection, which is the process of selecting a subset of features to use

in the models. By selecting the most important features, the performance of the models can be potentially improved by eliminating less important features that may be adding noise or reducing the model's ability to generalize.

- Model debugging: if poor model performance is experienced, identifying the most important features can be helpful to debug the model and identify potential issues. For example, if a particular feature is not statistically very important, it can be removed from the model.

Basically, the feature importance algorithm identifies, for each feature, a corresponding weight that represents the total gain of splitting data along the feature. It is calculated as the sum of the improvement in the loss function for all splits that use the selected feature. Features with higher weights are generally considered to be more important. The feature extraction technique was applied to the engine parameters listed in Section 2.2.1. The output of the algorithm was the computation of the relative importance of each parameter for NOx formation, and the results are presented in the Result section. In literature there are also other methodologies for investigating and interpreting the relationships of the input features with the model performance. A SHAP [85, 86] and feature importance techniques were investigated in [87] to study the correlation between the features and the output of the model, resulting in a faithful evaluation between the two methods.

As far as the mathematical definition of the feature importance technique is concerned, there are different ways to compute the relative importance of each input variable. In the present work, the Gini importance was employed for the considered algorithm [88]. The Gini importance was exploited to compute the node impurity and the feature importance was basically through the impurity reduction metric computed in a node weighted by the number of samples reaching that node from the total number of samples [89]. From the mathematical point of view, the importance of a node which is used to calculate the feature importance for every decision tree is described in Equation 2.2 [90]:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{2.2}$$

where ni_j is the importance of node j, w_j is the weigthed number of samples reaching the node, C_j is the impurity value of the node, $left(j)$ and $right(j)$ are respectively the child node on the left and on the right of the node j. Each single feature was used in the different branches of the tree and, thus, its feature importance can be computed through the Equation 2.3 [90].

$$fi_i = \frac{\sum_{j:NodejSplitsOnFeaturei} ni_j}{\sum_{j \in AllNodes} ni_j} \tag{2.3}$$

The values are first normalized relative to the entire number of feature values that are represented in the tree, and then the result is divided by the total number of trees reaching the estimate of the overall feature significance.

XGBoost algorithm & hyperparameters tuning

The field of data science makes extensive use of the algorithm known as XGBoost, which is classified under the umbrella term of ensemble models [91]. This algorithm takes multiple weak learners and combines them into a single strong learner, which then produces results that are significantly more accurate [92]. The decision tree is the fundamental component of XGBoost, and each of them is trained on a specific subset of the available data. As a consequence, the final prediction of the model is the result of a combination of the predictions produced by each of the component building blocks. The real strength of this approach lies in the fact that it is a gradient-boosting decision tree technique that is both scalable and distributed, and that it offers concurrent boosting of the tree [93, 94]. Gradient boosting decision tree (GBDT) is a supervised learning which uses the gradient descent method to generate progressively weaker models based on a particular loss function. Iteratively shallow decision trees are trained, and at each iteration, the residual error of the model that came before it is used to construct the model that comes after it. The final prediction is found to be a weighted prediction sum of the individual decision trees. This idea is where XGBoost originates; yet, it pushes the boundaries of what can be accomplished with computer power. In point of fact, the trees in this scenario are constructed in parallel as opposed to sequentially as they are in GBDT, and in addition, second order gradients of the loss function are used. Moreover, the loss function is an advanced objective function that also includes regularization terms. These terms impose penalties on the expansion of the algorithm by adding more decision trees. This enables the development of the ensemble model to be constrained, which helps avoid overfitting [91]. As far as the objective function is concerned, the expression to be minimized at each iteration as a combination of loss function and regularization term is given in the Equation 2.4 [95]:

$$L^t = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \tag{2.4}$$

where the first term is the loss function which measures the difference between the target y_i and the prediction $\hat{y}_i$, while the second term represents the penalty for controlling overfitting and shown in Equation 2.5 [95]:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{2.5}$$

where T is the number of leaves in the tree, w is the score of each leaf and γ, λ are the regularization degrees. The Equations 2.4, 2.5 can be found in [91].

Due to a large number of hyperparameters to configure and a wide range of values that each can assume in the XGBoost model, it has been found necessary to apply optimization strategies to increase its performance. For each of the hyperparameters a definition space is created, after which the GridSearchCV method has been applied to find the optimal combination across the entire space. Grid search and k-fold cross-validation [96] techniques were combined to perform hyperparameter tuning in order to determine optimal values for the given model. All different parameters were fed into a parameter grid, and based on a scoring metric (accuracy), the best combination was identified. The k-fold cross-validation process was allowed to perform the models' learning, evaluating each parameter combination over different datasets, i.e., validation set. The procedure involves partitioning the original training data into k subsets, where k is a positive integer, typically 5 or 10. The model is then trained on k-1 folds before being evaluated on the remaining fold. This procedure is repeated k times, with each test set fold being utilized once. The model's performance is measured by its average performance throughout all k iterations. In this investigation, the k parameter was assigned a value of 10. Using k-fold cross-validation, the model is trained and assessed on multiple different subsets of data, thus yielding a more accurate performance estimation [97]. The model hyperparameters involved in the study are reported in Table 2.2,

Table 2.2 XGBoost hyperparameters defining the model architecture.

Hyperparameter	Description	Domain space
n_estimator	Number of gradient boosted trees	[500 - 1500]
learning_rate	Shrinkage factor	[0.1 - 0.01]
max_depth	Maximum number of tree levels	[3 - 5]
min_child_weight	Maximum weight to create a new node	[3 - 7]
colsample_bytree	Subsample ratio of features for each tree	[0.5 - 1]

where *n_estimator* is the number of trees which build the entire XGBoost model architecture; *learning_rate* is the step size shrinkage used in update to prevent over-fitting; *max_depth* is the maximum depth of a tree and it was exploited to control over-fitting as higher depth allows model to learn specific relations and pattern to a particular sample; *min_child_weight* is the minimum sum of weights of all observations required in a child and it was used to control over-fitting; *colsample_bytree* specifies the subsample ratio of columns when constructing each tree and subsampling occurs once for every tree constructed;

2.2.3 Performance evaluation

The chosen models were utilized as virtual predictors of NOx pollutant, and their prediction performance was evaluated using various metrics. The evaluation of the XGBoost models for each case study was conducted by analyzing their performance using specific metrics in conjunction with the test data. Hence, the involved metrics are as follows:

- the coefficient of determination R^2, defined by the Equation 2.6;

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2} \tag{2.6}$$

- the real RMSE in *ppm* considering the test dataset and defined by the Equation 2.7;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \tag{2.7}$$

- the normalized RMSE considering the test dataset, defined by the Equation 2.8;

$$RMSE_{norm} = \sqrt{\frac{\sum_{i=1}^{n}(x_{n,i} - \hat{x_{n,i}})^2}{n}} \tag{2.8}$$

where x_i is the experimental data, $\hat{x}_i$ is the model estimated value, $\bar{x}_i$ is the mean of experimental measured data, $x_{n,i}$ are the normalized experimental data, $\hat{x_{n,i}}$ are the normalized estimated value and n is dataset sample size. The performance evaluation of the machine learning models and the selection of the best hyperparameter values are widely discussed and analysed in the Results and Discussion section.

As far as the model tuning of hyperparameters is concerned, a combination of a large amount of XGBoost architectures was investigated through the GridSearchCV technique which has been defined in the Section 2.2.2, combining the hyperparameter space definition with the grid search and the optimization procedure by the k-fold cross-validation technique. Therefore in this context, the selected optimized model architecture was chosen basing on the performance metric defined by the mean squared error shown in Equation 2.9

$$MSE = \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n} \tag{2.9}$$

The computational efforts in the design of the virtual sensor based on machine learning and the whole preprocessing phase were covered exploiting the high-level Python programming language and a PC worker with an Intel(R) Core(TM) i7-8700 processor ar 3.20 GHz and 64GB RAM architecture.

2.2.4 Virtual NOx Sensing in Steady-State Conditions

In the automobile industry, virtual sensing is a strategy that is frequently adopted to assure sensorless solutions that are capable of correctly calibrating the primary engine parameter, hence allowing for consumption and emissions reduction. Despite this, the virtual NOx prediction sensor need to be calibrated under stationary conditions over the whole of the engine's operating range in order to account for the fluctuating dynamic scenarios that an engine is subjected to during the course of regular on-road driving missions. These kinds of operations call for a significant number of experimental measurements, but they would ultimately make it possible for the ECU to provide accurate engine control during actual driving scenarios. As a result, the focus of this research moved from the development and calibration of an AI-based virtual sensor through the steady-state data collection. It is important to keep in mind that the experimental campaign resulted in the collection of two distinct sets of acquisition data: one set of data was collected at the test bench (Bench), and the other set of data was directly acquired from the engine control unit. As a result, the ML virtual sensor was developed according to various combinations of the available data sets for the training and testing phase. This was done in order to further study how the algorithm reacts when given certain inputs.

Since this work is based on the development of a virtual NOx sensor calibrated in steady-state conditions and validated in transient scenarios, the first real step in determining whether the data collected by the bench sensors and the control unit are sufficiently representative of the engine's behavior in real-world operating conditions is to investigate the model's steady-state behavior. How it is mentioned in the Section 2.2.2, a sensitivity analysis was carried out on the training and test sub-set splits, defining the test dataset size as a percentage amount of whole data, in order to assess the robustness of the methodology. The different cases analyzed are reported in Table 2.3, whereas the XGBoost model performance results are widely discussed in the Results and Discussion section.

Table 2.3 Steady-state case analysis based on datasets and train–test split sensitivity.

Case Study	Train–Test	Test Dataset Size [%]
#1	Bench–Bench	[10; 50; 95]
#2	ECU–ECU	[10; 50; 95]
#3	Bench–ECU	NO [1]

[1] The datasets involved in the training and test phases were derived from the two different acquisition systems. Hence, a train–test split technique was not adopted.

In both the Bench–Bench and ECU–ECU examples, the datasets that were used for model learning were distinct, although they originated from the same experiments.

More specifically, the Bench–Bench case relied on engine bench test data, while the ECU–ECU case relied on engine data received from the ECU. In the last investigated scenario, Bench—ECU, the training dataset and the test dataset were both received from the respective measures system applications. As a direct consequence of this, there was no train-test split, and all of the dataset samples were used for the different learning and testing phases. The evaluation of model performance in Case Study 3 was dependent on the extent to which the virtual sensor, which serves as a potential real-time estimator on ECU, was able to detect physical phenomena accurately. This was due to the fact that the bench acquisition system utilizes precise physical sensors to identify real events, while the engine map employs both data tables and sensors.

2.2.5 Virtual NOx Sensing in Transient Conditions

Following the calibration and testing of the virtual sensor in a steady-state environment to evaluate the model's robustness and its ability to predict NOx emissions, validation was conducted under dynamic conditions. To ensure precise forecasting of NOx emissions across diverse operational scenarios, it was imperative to train the virtual sensor model utilizing engine map data and subsequently validate it through real-world driving missions. The engine map data, encompassing details of engine speed and load, serves to simulate a range of stable conditions that the engine could potentially encounter. Nevertheless, it is possible that these conditions do not accurately reflect the dynamic and realistic operating conditions that an engine would encounter while driving on a roadway. Through the process of testing the sensor during an on-road driving mission, it becomes feasible to evaluate the efficacy of the virtual sensor in accurately predicting real-time emissions and its performance under actual driving scenarios. Identification of potential issues or limitations that require attention is possible when integrating a virtual sensor into a control system. Hence, in this study, the algorithm was trained using stationary experimental data obtained from the ECU and bench, and validated using on-road-based measurements to assess the accuracy of NOx prediction during real-world driving missions. In the course of empirical road trials, the vehicle propelled by a compression ignition engine was outfitted with a physical measuring equipment for the purpose of managing and monitoring engine-out operations. The creation and evaluation of a virtual NOx sensor suitable for immediate implementation would facilitate the replacement of real sensors, potentially leading to enhancements in cost-effectiveness and reliability.

The experimental campaign involved the vehicle performing a real driving mission in Piedmont, encompassing a range of operating conditions and maneuvers that were pertinent to the constraints and demands of both urban and extra-urban scenarios. To evaluate and establish the virtual NOx sensor model, the analysis incorporated data obtained from both steady-state measurement systems.

Table 2.4 Dynamic on-road case analysis based on training phase over different acquisition systems.

Case Study	**Train–Test**
#1	Bench–On-road mission
#2	ECU–On-road mission

Bench-On road relates to training done on a stationary engine test bench, whereas ECU-On road examines the learning process on stationary engine ECU data and validation in transient scenarios. Given the heterogeneous nature of the variables and the fact that the creation of the virtual sensor was geared toward a future real-time implementation and might thus be implemented in the ECU, it was essential to evaluate the two data gathering systems separately. Hence, the performance prediction of the engine map-based model may be compared to the actual collection of physical sensors.

2.3 Results and Discussions

The present section reports the scientific findings obtained through the analyzes conducted for the development of the virtual sensor. In this case, prediction performances are shown in steady-state conditions through engine bench acquisitions. Consequently, the XGBoost model for the real-time estimation of NOx in on-road applications was built by training the model under steady-state conditions, from which the entire engine map domain might be completely covered during experiments, and testing it on real-world driving data.

2.3.1 NOx Predictions in Steady-state Conditions

The case study results concerning the training and test phases on bench test conditions (Bench–Bench), and taking into account the sensitivity analysis conducted on the train–test split, are summarized in Table 2.5 and the regression results are shown in Figure 2.4.

Table 2.5 Summary of performance prediction results for test case Bench–Bench. The test size is here reported for the sensitivity analysis, considering the different percentage values 10%, 50% and 95%.

Test Size [%]	**10**	**50**	**95**
$\mathbf{R^2}$	0.98	0.97	0.85
RMSEnorm	0.017	0.022	0.052
$\mathbf{RMSE_{ppm}}$	60.0	77.4	186.5
q	0.0034	0.0062	0.0224
m	0.985	0.976	0.912
n_estimator	1500	1500	1200
learning_rate	0.08	0.08	0.08
max_depth	4	3	3
min_child_weight	3	7	7
colsample_bytree	1	0.7	1

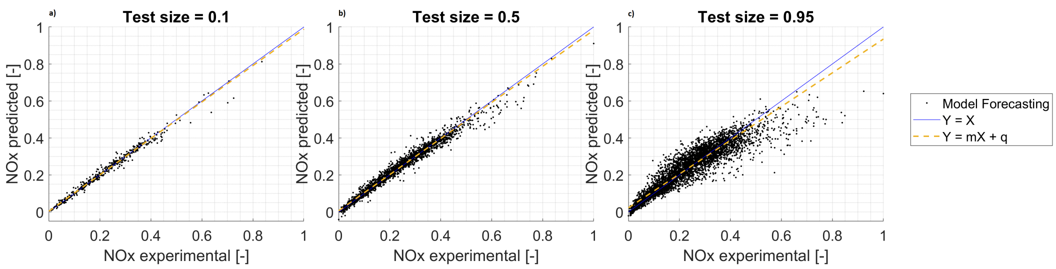


Fig. 2.4 Regression prediction results on test dataset for bench test conditions. Sensitivity analysis on the three different sizes of test dataset: (**a**) 10%, (**b**) 50% and (**c**) 95%.

The decrease in engine points allocated to the model learning phase, as the test dataset was increased from 0.1 to 0.95, resulted in a significant reduction in prediction performance. The results indicate that when 95% of the data was allocated for testing, notable errors were observed, despite a moderate rise in RMSE as the partition shifted from 10% test size (60.0ppm) to 50% test size (77.4ppm). It is noteworthy to emphasize that the model's predictive ability remained robust even after reducing the training dataset by half, indicating a remarkable capacity for accurately capturing physical phenomena. Additionally, the regression line in Figure 2.4 demonstrates the rise in noise brought on by the test size-appropriate sensitivity analysis.

The case study results concerning the training and test phases on ECU test conditions (ECU–ECU) are summarized in Table 2.6 and the regression results are shown in Figure 2.5.

Table 2.6 Summary of performance prediction results for test case ECU–ECU. The test size is here reported for the sensitivity analysis, considering the different percentage values 10%, 50% and 95%.

Test Size [%]	**10**	**50**	**95**
$\mathbf{R^2}$	0.98	0.97	0.76
RMSEnorm	0.021	0.024	0.067
$\mathbf{RMSE_{ppm}}$	73.3	84.9	237.7
q	0.0038	0.0047	0.0309
m	0.985	0.9802	0.862
n_estimator	1500	1500	1200
learning_rate	0.08	0.08	0.08
max_depth	4	4	3
min_child_weight	7	7	5
colsample_bytree	1	0.7	1

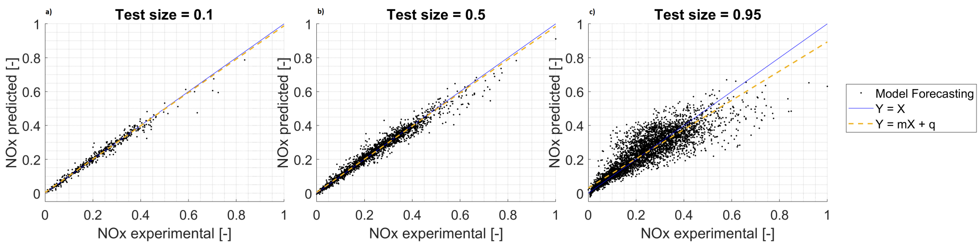


Fig. 2.5 Regression prediction results test dataset for ECU test conditions. Sensitivity analysis on the three different size of test dataset: (**a**) 10%, (**b**) 50% and (**c**) 95%.

Like the previous scenario, it was observed that the precision of NOx estimation exhibited a tendency to deteriorate (and the discrepancy to augment) with an increase in the test size and a decrease in the training size. The marginal reduction in R2, specifically 0.98 for 90% and 0.97 for 50% of the full dataset, due to the utilization of almost 50% less data for model training, is a significant observation. The observed phenomenon can be attributed to the considerable efficacy of the employed machine learning techniques, particularly in light of the fact that the data utilized in this specific context were obtained in a state of equilibrium. Hence, commencing with stable engine operating conditions was found to be a more effective approach for capturing NOx generation phenomena.

As a summary, Figure 2.6 shows the trend of R^2, RMSE and computational time as the train-test split varies in order to evaluate the performance of the models in different data configurations.

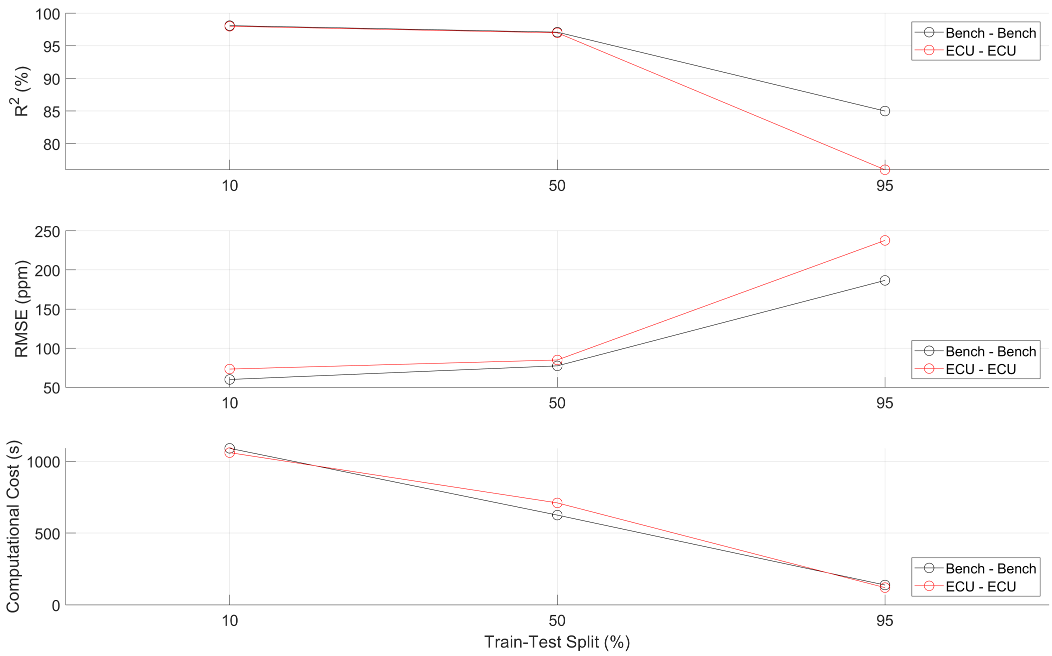


Fig. 2.6 R^2, RMSE and computational cost trend as functions of train-test split parameter.

The trends of R2 and RMSE indicate that the model's fit and performance are satisfactory. Moreover, the behavior exhibited during the process of model acquisition aligns with the computational duration, as a reduction in the quantity of data required for training results in an almost-linear decrease in processing time.

Finally, the case study results concerning the training conducted on Bench test conditions and test conducted on ECU test conditions (Bench–ECU) are summarized in Table 2.7 and the regression results are shown in Figure 2.7.

Table 2.7 Summary of performance prediction results for test case Bench–ECU.

R^2	0.97
RMSEnorm	0.024
$RMSE_{ppm}$	86.7
q	0.0037
m	0.972
n_estimator	1500
learning_rate	0.08
max_depth	4
min_child_weight	3
colsample_bytree	1

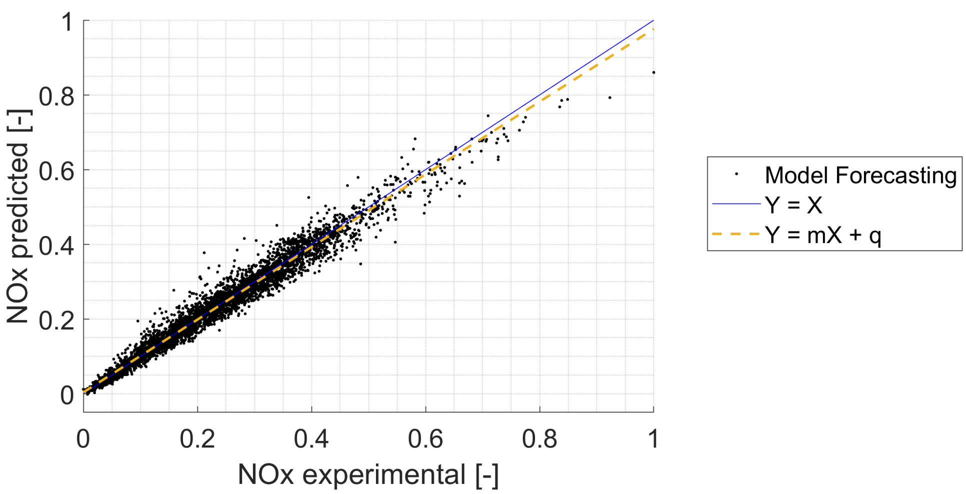


Fig. 2.7 Regression prediction results of training conducted on bench test conditions and test carried out under ECU test conditions.

The results obtained in this example demonstrate consistency with the findings of previous cases, particularly in relation to the robustness of the virtual sensor. Furthermore, the Bench-ECU illustration exemplifies the precise estimation (R^2 of about 98%) of the formation phenomenology of nitrogen oxides through the utilization of the developed virtual sensor, which was evaluated using ECU engine data.

During the learning process of the training phase, the XGBoost based predictive model provided the relative importance in percentage that each engine variable had in the predicting the instantaneous NOx value. Hence the feature importance algorithm was applied to identify the most influential from mathematical point of view, defining the close correlations between input and output variables. The complexity of a model increases with the number of features it possesses, resulting in sparser data and greater susceptibility to errors induced by variance. This, in turn, affects the model's prediction performance. The feature importance approach is a crucial aspect of machine learning, specifically within the feature engineering process. Its primary objective is to select the minimum number of features necessary to generate a feasible model, thereby avoiding any potential bias in results that may arise from the inclusion of irrelevant data [89]. The production of nitrogen oxides in diesel engines is significantly impacted by the engine's particular operating conditions. To streamline the models and evaluate their effectiveness, the feature extraction method is utilized. The verification of prediction robustness can be achieved through a comparison of the statistically significant engine variables utilized in the models with the empirical understanding of the relevant physical phenomenon. Given the complete availability of data during the training phase of the models, where the test size is equivalent to 10% for both the *Bench–Bench* and *ECU–ECU* case studies, the results of the feature extraction method are presented in Table 2.8. Table 2.9 presents the hyperparameters of the optimal XGBoost architecture for each analysis considered, using the same train-test split value.

Table 2.8 The feature importance outputs for the three steady-state cases study which highlight the relative importance of each variable. The percentage values identify the most influential engine variables. Considering a threshold value of 90%, the variables beyond this value are labeled in red and they were discarded as they were considered non-influential.

Variable	Bench–Bench	ECU–ECU	Bench–ECU
1	$SOI_{main}(13.9\%)$	$SOI_{main}(13.5\%)$	$SOI_{main}(14.2\%)$
2	$P_{rail}(13.4\%)$	$P_{rail}(13.4\%)$	$\omega_{eng}(13.3\%)$
3	$\omega_{eng}(13.3\%)$	$\omega_{eng}(13.0\%)$	$P_{rail}(13.3\%)$
4	$\lambda(9.5\%)$	$\lambda(10.7\%)$	$\lambda(9.7\%)$
5	$O_2(9.3\%)$	$IMAT(7.9\%)$	$O_2(9.6\%)$
6	$IMAT(7.5\%)$	$Q_{tot}(7.9\%)$	$IMAT(7.7\%)$
7	$Q_{tot}(7.4\%)$	$O_2(6.3\%)$	$EGR(7.4\%)$
8	$EGR(6.8\%)$	$Q_{main}(6.3\%)$	$Q_{tot}(6.5\%)$
9	$IMAP(5.6\%)$	$IMAP(6.1\%)$	$Q_{air}(5.9\%)$
10	$Q_{air}(5.5\%)$	$SOI_{pil}(5.4\%)$	$IMAP(5.1\%)$
11	$SOI_{pil}(4.5\%)$	$Q_{air}(4.6\%)$	$SOI_{pil}(3.8\%)$
12	$Q_{main}(2.6\%)$	$EGR(4.1\%)$	$Q_{main}(2.7\%)$
13	$Q_{pil}(0.7\%)$	$Q_{pil}(0.8\%)$	$Q_{pil}(0.8\%)$

According to the feature importance definition, the variables SOI_{pil}, Q_{main} and Q_{pil} were discarded as they proved not to be significantly affected by large variations. The algorithm, therefore, did not consider SOI_{pil} and Q_{pil} to be particularly significant in the evolution of NOx emissions. Furthermore, since the Q_{main} was strictly proportional to Q_{tot}, only one of them was taken into account.

Table 2.9 Best XGBoost architectures and grid values setup for the three case studies under investigation. The mean squared error (MSE) evaluation metric was used to obtain the optimal value by the GridsearchCV algorithm.

Hyperparams	Grid Values	Bench–Bench	ECU–ECU	Bench–ECU
learning rate	[0.1, 0.05, 0.08, 0.1, 0.15]	0.08	0.08	0.08
max_depth	[3, 4, 5]	4	4	4
n° estimators	[500, 1000, 1200, 1500]	1500	1500	1500
min_child_weight	[3, 5, 7]	3	7	7
colsample_bytree	[0.5, 0.7, 1.0]	1	1	1

It is evident that the feature extraction process identified nearly identical engine variables as the most pertinent in all three cases examined. It is probable that the variables with the highest relevance for forecasting NOx emissions remain consistent across various data sources and testing environments. Variables such as engine load, fuel flow, and exhaust gas temperature are crucial in predicting NOx emissions, regardless of the source of data collection, be it from a test bench or the engine control unit of the vehicle. Furthermore, the XGBoost algorithm identified comparable engine variables as the most significant across the three distinct scenarios, due to the similarity of the training and testing datasets. The dataset utilized to train the model on the test bench exhibited similarities to the dataset employed for testing the model on the vehicle's ECU. Thus, the pertinent variables exhibited similarity in both instances. The algorithm's ability to detect descriptive patterns in the data would yield similar results in predicting the occurrence of NOx formation, for identical reasons. Hence, the architectures of XGBoost in all three instances exhibited a high degree of similarity.

2.3.2 NOx Predictions in Transient On-road Conditions

After evaluating the model's ability to accurately predict and capture phenomenological events during steady-state conditions, a virtual NOx sensor was developed for real-time use by utilizing on-road experimental data. The model demonstrated high predictive performance and robustness. In contrast to bench tests, which entail collecting data while the engine is operating under controlled conditions to ensure consistent and replicable testing conditions, on-road data is obtained while the vehicle is being driven on real roads in realistic scenarios. The aforementioned data accurately portrayed the working conditions of the engine, encompassing environmental factors such as temperature, humidity, and altitude, in addition to the fluctuating dynamic loads that transpired throughout the driving mission. The measurements taken during on-road testing encompassed the vehicle's velocity and rate of change of velocity, which might impact engine performance. The virtual NOx sensor's design was established through calibration in stationary conditions, such as bench or ECU, and subsequently verified through real-world experimental tests. The engine parameters under consideration were identified through the feature importance process in the preceding stages of the analysis. Thus, the acquisition stage was encompassed by the feature importance outcome derived from the steady-state conditions. The predictive performance results for the two case studies are listed in Table 2.10 outlining the same statistical metrics.

Table 2.10 Performance prediction results over on-road test driving mission.

Case Study	**#1 Bench—On-Road Mission**	**#2 ECU—On-Road Mission**
$\mathbf{R^2}$	0.76	0.75
RMSEnorm	0.054	0.055
RMSE$_{[}$**ppm**$]$	192.0	194.9
m	0.92	0.97
q	−0.007	−0.017

The analysis of the metrics revealed a significant decline in predictive performance, with a reduction of up to 75% in accuracy. Additionally, a substantial shift in the regression line was observed, with a significantly higher q value compared to the stationary cases. Furthermore, a considerable increase in the error was also noted. The observed phenomenon can be attributed to the dissimilarity between the operating conditions experienced by the engine while performing a road driving task and those encountered during the model learning phase, which was characterized by stationary conditions. The predicted levels of NOx in the two case studies are presented in conjunction with the empirical traces in Figure 2.8. This figure highlights three distinct time intervals [50 s 250 s], [500 s 700 s], [850 s 1050 s] during the missions, each spanning 200 s. The model's forecasting capability was evaluated across various real-world conditions that the vehicle may encounter, by presenting three distinct time windows that corresponded to different operating conditions. These time windows were of considerable duration.

Evidently, the estimated NOx signals precisely depicted both the qualitative and quantitative pattern of the experimental unit, along with its transient trend throughout the driving mission. During the time interval of [50,250], certain mission points resulted in negative NOx values due to overestimation or underestimation of experimental outcomes. The observed phenomenon can be attributed to the prevalence of operating points where the lambda values exceeded the defined domain of the learning phase, coupled with the sudden surges in the values of fuel injected quantity. Furthermore, the signal exhibited significant spikes owing to the instantaneous nature of the measurement sensor utilized in the model. Each point of the mission was characterized by distinctive engine variables, which differed substantially from those of the subsequent instants. The aforementioned factor, coupled with the inadequacy of the learning phase to account for the NOx formation dynamics and their corresponding characteristic times, contributed to the differences between the anticipated and observed signals.

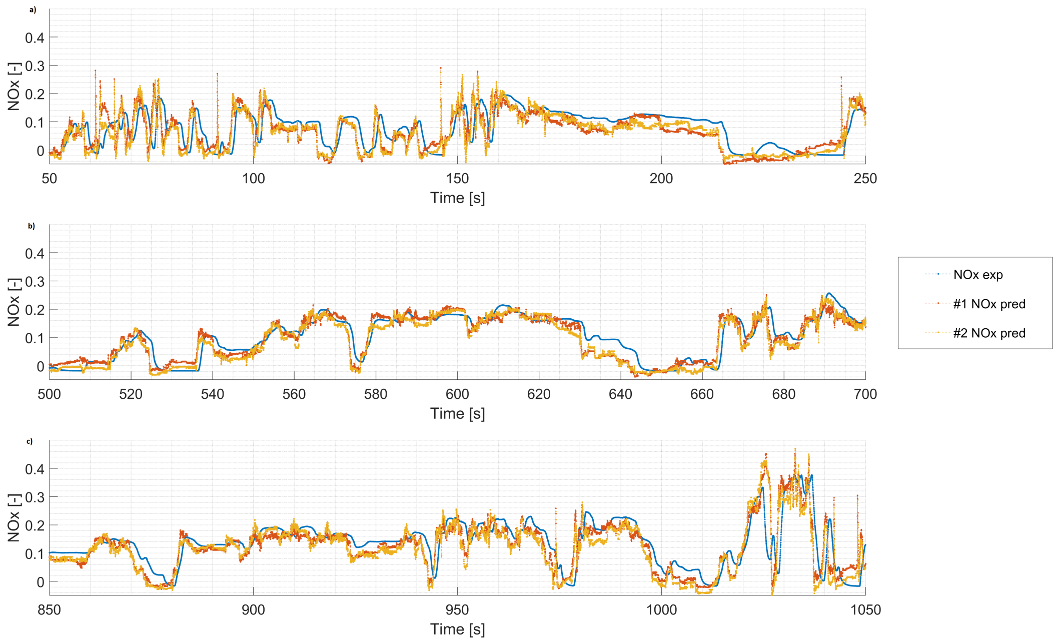


Fig. 2.8 The NOx predicted signals for cases study #1 and #2. (**a**)–(**c**) depict three distinct 200 s-length windows of driving mission.

The lower R2 values observed in Table 2.10 and Figure 2.8 may be attributed to the anticipation of the predicted signals in relation to the observed signal. Real-time virtual sensing applications may experience a significant decrease in performance due to the timing of predicted signals, either in anticipation or delay. This can result in incorrect emission control. This behavior is associated with the characteristic time delay of the measurement system and the line. The latter phenomenon arises due to a time lag between the measured level of NOx and the corresponding time frame to which the measurement pertains. As a matter of fact, the predicted NOx were referred to the implemented engine parameters at the given time-step, whereas the measured NOx would have been produced by the engine setting at a previous time frame. The latter phenomenon is correlated with the duration required for the discharge of exhaust gases from the engine to reach the acquisition probes, in addition to the sensor's related time lag. In conditions of steady-state, the delay is effectively eliminated due to the stationary nature of the test, in contrast to its origin from transient conditions. The model's NOx signal output was delayed by 1 second [98]. As a result, the expected signal closely approximated the experimental signal, as depicted in Figure 2.9, which pertains to a specific time interval.

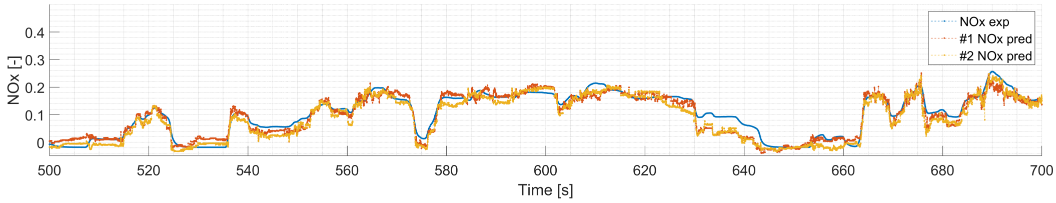


Fig. 2.9 The NOx predicted signals for cases studies #1 and #2, taking into account the line delay of measurement system.

The R2 value for both case studies was observed to be 85%, indicating the virtual sensor's ability to predict NOx emissions with a high degree of accuracy in real-time, as reported in [99].

Chapter 3

Li-ion Battery and Vehicle Modeling for On-Road Testing of an Electric Two-Wheeler

[1] Electric vehicles (EVs) are gaining popularity worldwide due to their low carbon footprint, reduced operating costs, and improved performance. In 2020, the sales of electric vehicles totaled over 400,000 units and exhibited exponential growth of 157% over the preceding year [100]. These statistics represent a significant shift in the market trajectory of electric automobiles. However, one of the main challenges associated with EVs is their limited driving range, which causes range anxiety among consumers. To address this challenge, it is crucial to develop accurate models of EVs that can predict their range and optimize their performance.

Traditional modeling techniques for EVs have relied on physics-based models, which can be complex and time-consuming to develop. Physics-based models require detailed knowledge of the components of the vehicle and their interactions, making them difficult to develop for newer technologies such as lithium-ion batteries. Moreover, they may not capture the variability in real-world driving conditions, such as traffic, weather, and driver behavior. Therefore, there is a need for alternative modeling techniques that can capture the complex interplay between the various components of EVs and provide accurate predictions of their performance. From

[1]Part of this chapter has been published in the form of papers as: Falai, A.; Giuliacci, T.A.; Misul, D.; Paolieri, G.; Anselma, P.G. Modeling and On-Road Testing of an Electric Two-Wheeler towards Range Prediction and BMS Integration. *Energies* 2022, 15, 2431

the perspective of the accumulators, a reliable and cost-effective management of the energy storage system is an essential component for the development of this device, as well as for their longevity and the optimization of vehicle performance. Data-driven modeling techniques offer a promising approach to modeling EVs, where data collected from real-world driving can be used to build models that accurately capture the vehicle's performance or, generally speaking, the on-road behaviour.

In this chapter, we present a data-driven approach for the electric two-wheeler parameters identification, which includes both the li-ion battery and the vehicle. The objective of this study is to develop a model that accurately predicts the vehicle's distance driven and optimize its performance. The model is developed using data collected from on-road testing, and it is developed towards an online integration in a battery management system (BMS) to estimate the battery states and adjust the vehicle's performance accordingly.

The contributions of this chapter are two-fold.

- first, we present a data-driven approach to identify the parameters of an electric two-wheeler model that can accurately predict the vehicle's range and optimize its performance during on-road driving missions. The data-driven modeling is focused combining a standalone energy storage system model with a dynamic vehicle model for a reliable and manageable approach in the global vehicle performance perspective;
- second, we demonstrate the feasibility of modeling techniques for EVs, combining batteries and vehicle dynamics, where the model parameters were calibrated through real-world driving mission data.

This study provides a foundation for future research on data-driven modeling techniques for EVs, which can help to improve the accuracy of range predictions and optimize the performance of these vehicles.

3.1 Introduction

The worldwide proliferation of electric vehicles is currently facing a number of constraints. The autonomy of e-vehicles is expected to be limited, ranging from 100 to 300 km, due to the battery size [101]. Increasing the capacity of the battery pack results in elevated expenses, augmented vehicle weight, and subsequently, amplified energy consumption. Therefore, it is imperative to focus on determining an appropriate balance between the electric range and battery capacity. One of the primary benefits of e-powertrain pertains to the superior efficiency of electric motors at lower vehicle velocities in comparison to internal combustion engines (ICEs). Consequently, electric mobility is predominantly aimed at urban areas. Given the uncommon and limited availability of electric charging stations [102], coupled with the absence of extensive fast charging infrastructure [103], it is imperative to effectively manage and mitigate range anxiety through precise predictive models. Electric two-wheelers have the potential to be a significant advancement in electric mobility, particularly in urban traffic, where two-wheelers have historically been a popular transportation option. Reference [104] presents a comprehensive vehicle model that elucidates the dynamic characteristics of electric scooters, with the aim of enhancing their design and production capabilities. Conversely, the implementation of electric two-wheelers necessitates the creation of specialized BMS in order to guarantee adequate performance and electric range. The successful deployment and implementation of BMS necessitates a comprehensive understanding of both the vehicle and battery behaviors. Regarding the battery, the utilization of lithium has become extensively prevalent. Lithium-ion batteries (LIBs) are renowned for their intricate yet efficient electrochemical mechanisms, which result in exceptional energy storage capacity, high-power density, extended lifespan, minimal self-discharge, low maintenance expenses, and negligible environmental footprint. Therefore, the investigation of LIBs and their modeling are crucial aspects for evaluating their performance and application constraints, as well as for designing a tailored, adaptable, and dependable Battery Management System [105].

The performance models of LIBs are categorized into three primary groups, namely electrical models [106, 107], analytical models [108, 109], and electrochemical models [110, 111]. The reduction in computational cost is accompanied by a decrease in the degree of complexity and an increase in the levels of accuracy. Equivalent circuit models (ECM) have been found to be appropriate for analyzing BMS controllers from a system-level perspective due to their high precision and low computational expense, making them a popular choice among electrical models [112, 113]. The current study is based on the utilization of Thevenin equivalent circuit representation. The second-order equivalent circuit model is utilized to describe each battery cell, which comprises of a voltage source, resistors, and capacitances.

The present research work did not take into account the battery thermal management issues, despite their extensive investigation in the literature [114, 115]. The proposed model is intended to be integrated into a software-in-the-loop (SIL) environment, serving as a foundation for the development of battery management systems. Future research endeavors may explore various strategies and control logics to optimize battery management. Furthermore, the implemented instrument will be advantageous in examining and contrasting diverse battery architectures and configurations.

3.2 Material & Methods

Data-driven modeling is an approach to modeling complex systems, such as electric vehicles, that relies on data rather than explicit knowledge of the underlying physical principles. In data-driven modeling, a large amount of data is collected from the system of interest, and this data is used to develop a mathematical model that can accurately predict the system's behavior. Moreover, the experimental data can be exploited for parameters identification involving physical modeling of a complex system. In order to complete this assignment, it is necessary to specify the data generation system, the collection system, and the experimental environment, in addition to the required amount of data to be collected. In the present work, an extensive experimental campaign was carried out on a prototype electric two-wheeled vehicle. The on-road tests were conducted in Turin (Italy). A physical-based model was constructed and then connected with a battery model in order to evaluate the vehicle's behavior in terms of energy consumption. Using the experimental data gathered from many road testing, model identification and validation were performed following an optimization process.

The case under consideration was an electric two-wheeler equipped with a prototype lithium-ion battery pack consisting of 180 cells grouped in 9 parallel modules, each of which included 20 cells in series. According to the classification terminology, the investigated battery layout is a 20s9p configuration. The basic component of the battery pack was Li-ion cylindrical cell of Samsung INR21700-50E type. The cell physical dimensions nad technical specifications are shown and listed in Figure 3.1 and Table 3.1. Finally, the composition of the cells in series and parallel to form the entire battery pack is described in the Table 3.2, showing the technical information.

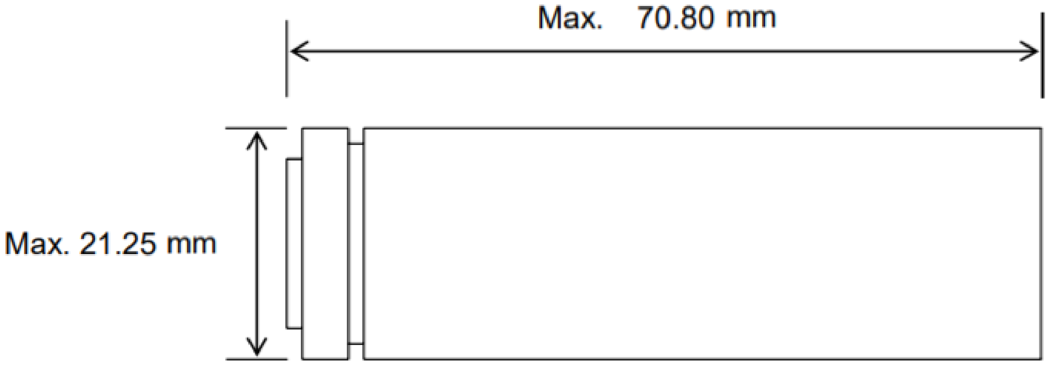


Fig. 3.1 Outline dimensions of INR21700-50E.

Table 3.1 Technical specifications of SAMSUNG INR21700-50E cell as reported on Samsung company data sheet [1].

Parameters	Values	Measurement Unit
Cell Format	Cylindrical	-
Technology	Li-Ion	-
Nominal Voltage	3.6	V
Nominal Capacity	4.9	Ah
Maximum continuous discharge current	9.8	A
Maximum non continuous discharge current	14.7	A
Recharge maximum current	4.9	A
Discharge Cut-off Voltage	2.5	V

Table 3.2 Technical specifications of the prototype battery pack.

Parameters	Values	Measurement Unit
Pack configuration	20s9p	-
Nominal Voltage	72	V
Nominal Capacity	44.1	Ah
Total Energy	3.17	kWh
Maximum continuous discharge current	88.2	A
Maximum non continuous discharge current	117.6	A
weight	12.5	kg

The electric two-wheeler was a vehicle on the market, therefore due to industrial and project secrecy it is not possible to provide information concerning the brand. However, some technical characteristic regarding the vehicle's performance and powertrain is shown in the Table 3.3. It is worth specifying that the standard battery of the vehicle has been replaced by the prototype one described in the Table 3.2 according to a previously existing industrial project. Therefore, the technical and performance specifications of the vehicle related to the original battery on board (such as electric range, voltage, charging time) are no more meaningful.

Table 3.3 Technical information of the electric two-wheeled vehicle.

Parameters	Values	Measurement Unit
Maximum velocity	90	km/h
Weight	98	kg
Powertrain - engine	Brushless	-
Powertrain - power	4	kW

3.2.1 Global Modeling approach

The adopted modeling method in this activity is based on an energy-based approach, which takes into account the electrical energy produced by the battery flowing via the vehicle drivetrain components to the driven wheels. Each drivetrain component under consideration can be represented as stated in the following sections. Matlab© and Simulink© were utilized for data post-processing and model building. The DSP System and Model Identification toolboxes were the most useful for this goal. The physical differential equations are solved using Simulink, and the overall vehicle modeling work is divided into two sections: battery and vehicle dynamic modeling, as indicated in the global model overview in Figure 3.2. The battery model takes as input the driver current request, which can also be computed by opening the accelerating knob, and returns the output battery pack electrical power. The electrical power from the battery pack is then the input of the vehicle model in order to estimate the vehicle longitudinal speed. From a global point of view, the developed model takes battery current as input and forecasts the real two-wheeler longitudinal speed.

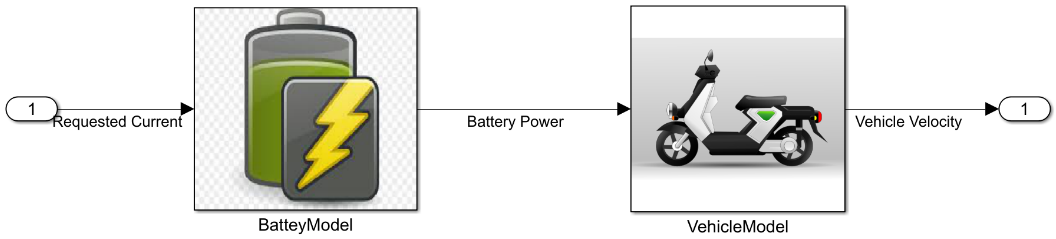


Fig. 3.2 Global model: battery and dynamic vehicle blocks set implemented on Simulink©.

The detailed analysis of the models, with the motivated choice of approaches is shown in the following sections.

Modeling approach for Li-ion battery pack

The modeling techniques of energy storage systems, such as lithium-ion batteries, belong to a topic widely discussed nowadays. As previously described in the 3.1, in the literature there are different approaches based on different levels of detail, with consequent different levels of computational power required and achievable accuracy in the estimation of parameters and in the prediction of operating variables. Depending on the application and the objectives to be achieved, it is essential to select a battery model that fits the design requirements. Since the present work was born as a precursor to the development of a customized and flexible BMS, the need arose to adopt computationally efficient and simplified techniques, but in which accuracy would not be too negatively affected. A common model technique which is widely employed in several different applications [116–118] is the equivalent circuit model due to the low obtained relative errors which can generally satisfy the precision requirements for the practical engineering calculation. In order to capture both fast dynamic events, e.g., resistance and charge transfer effect, and slow dynamic events, e.g., diffusion effects, a second-order RC model was exploited to characterize each single cell as the basic element of the battery pack. The behavior of each cell can be analyzed independently, which lays the groundwork for BMS control logic development. Usually, first-order RC and second-order RC are used interchangeably as the modeling levels are similar in terms of computational effort, but with results slightly in favor of the second order. In the [119] is shown how the first-order RC model could be the preferred choice for portable consumer electronics, while the second-order RC model could could be more suitable for stringent applications such as automotive. In the present study the same initial comparative analysis was conducted, obtaining only slightly higher accuracies in the second-order RC model case. An example of a result is shown in the Table 3.4, where the two models are compared in terms of error and precision in estimating the output voltage.

Table 3.4 First and second-order RC models compared in an initial step of the project.

	First-order RC model	**Second-order RC model**
RMSE[V]	0.048	0.041
$\mathbf{R^2}$	0.942	0.965

The cell model adopted, therefore, is shown in the Figure 3.3, where V and I are, respectively, the cell terminal voltage and current, V_j with $j = 0,1,2$ are voltages drops across resistors due to the flowing current. R_0 and R_j and C_j with $j = 1,2$ are the time-varying model parameters corresponding to resistors and capacitors which model both the static and dynamic behaviors of the cell.

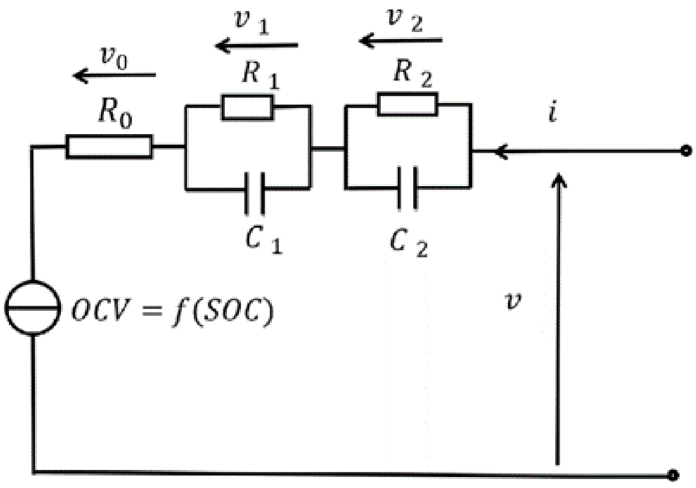


Fig. 3.3 The second-order RC cell model with operating parameters.

Specifically, R_0 determines the cell static resistance, and each RC pair determines different time constant dynamic behaviors: the first R_1C_1 models charger transfers and the double layer effect, while the last pair of R_2C_2 are employed to capture the cell diffusion phenomena occurring at a much higher time scale. The OCV is the cell Open-Circuit Voltage, which is modeled as an ideal voltage source and corresponding to the theoretical voltage that the cell can ideally supply. All these parameters are function of the cell's SOC as internal state of the system. In some advanced modeling approaches, hysteresis for the OCV [120] and parasite current phenomena are included. These two aspects were not considered in this work in order to preserve the simplicity of the model while obtaining a satisfactory accuracy.

The RC pairs is directly linked to the model order, which is a trade-off between model accuracy and complexity. For L-ion batteries, an ECM with two RC branches is commonly employed and it achieves good level of accuracy [121]. The operating conditions of cells were described through the ECM model according to the differential equation in Equation 3.1 [122]

$$\begin{cases} \frac{dV_1}{dt} = -\frac{V_1}{R_1C_1} + \frac{I}{C_1} \\ \frac{dV_2}{dt} = -\frac{V_2}{R_2C_2} + \frac{I}{C_2} \\ \frac{SOC}{dt} = -\frac{I}{Q_{cell}} \end{cases} \tag{3.1}$$

where Q_{cell} is the cell nominal capacity in ampere-seconds. The only variable in (3.1) is the current amplitude I. Once all parameters have been defined, the model takes the current I as input and solves a first-order differential equation which is able to simulate the total voltage delivered by the cell according to the Second Kirchhoff's Law, applied to the circuit model in Equation 3.2 [122]:

$$v = OCV - R_0 I - V_1 - V_2 \tag{3.2}$$

This modeling approach was implemented as a Simulink block, and it constitutes the initial component of the global vehicle model: for a given input current, it returns an output voltage. Using the model of a single cell to connect multiple cell models in accordance with the battery pack's 20s9p configuration, a general battery pack model was generated. The 20 ECMs were connected in series to form a single module, and then 9 modules were connected in parallel to replicate the pack layout, as shown in Figure 3.4.

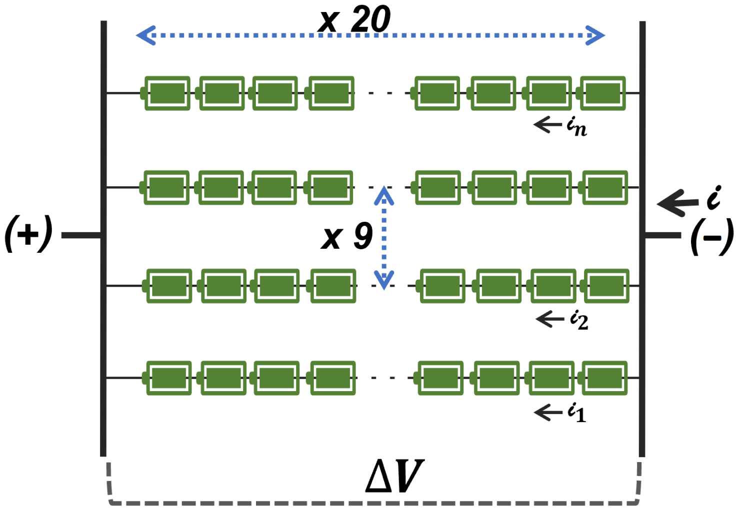


Fig. 3.4 Battery pack modeling: 9 modules are connected in parallel and each module has 20 cells connected in series, each single one modeled as an independent ECM: 'i' is the battery pack current, and ΔV is the voltage of the battery pack.

The Simulink implementation was performed by exploiting the Simscape Electrical libraries: The entire battery pack model is illustrated in Figure 3.5. The first layer of the battery pack model composed by 9 modules in pair is represented inside the black box. Then, each of the 9 modules are composed by 20 cells connected in series (green box). Finally, the single cell structure can be appreciate in the last red box: a SOC block estimation is presented, and it was useful to evaluate the SOC-dependency of the cell parameters in the look-up table. This allows each cell to have a different behavior. According to the Simulink libraries, the blue, red and black lines represent respectively the electrical, the physical and the signal connections.

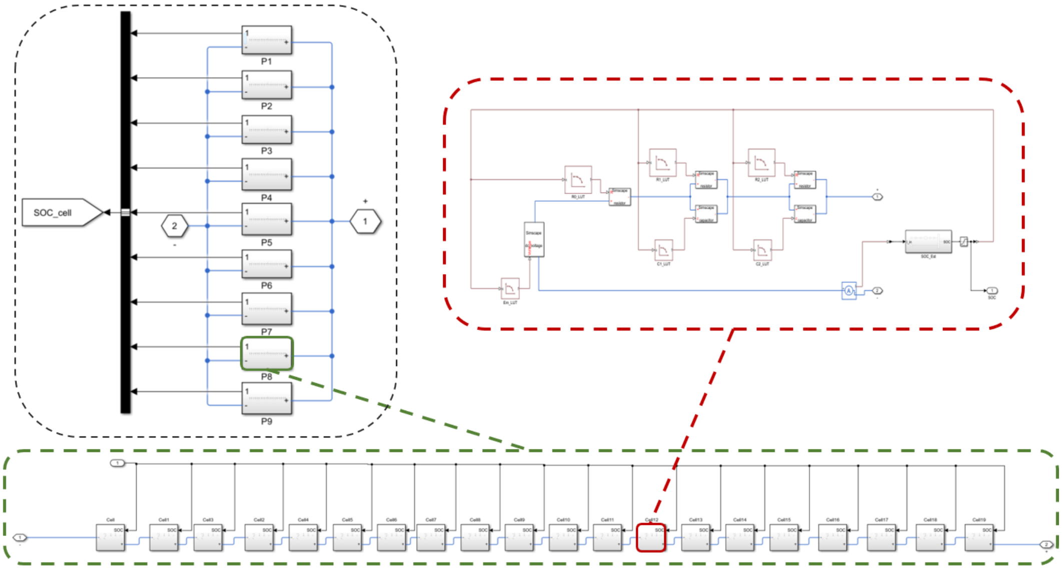


Fig. 3.5 Model of the battery pack configuration on the Simulink virtual environment.

As already discussed for Figure 3.3, each cell was modeled as a *2RC-ECM*, and the unknown parameters which characterize the cell model properties are:

- OCV [V];
- R1 [Ω];
- C1 [F];
- R0 [Ω];
- R2 [Ω];
- C2 [F].

These are implemented in appropriately calibrated one-dimensional look-up tables as a function of SOC, as explained in the next sections of the present work. The SOC for each cell is in turn evaluated according to the Coulomb Counting equation. This is obtained by integrating the third equation of (3.1) as show in the Equation 3.3:

$$SOC_t = SOC_{T0} - \int_{T0}^{t} \frac{I}{Q_{cell}} dt \tag{3.3}$$

This approach enables the modeling independence of each cell, and this is necessary for future BMS integration and the development of some future optimization strategies, such as cell equalization. However, this makes the model more expensive from the computational load point of view. In this phase of the activity, all the cells were modeled assuming the same values for all the parameters, which makes them identical to each others. This assumption is a good approximation of the reality considering that the cells in the tested battery pack are fresh and they should therefore be identical [123], assuming appropriate manufacturing tolerances [124, 125]. After the present work, a dispersion regarding values of the parameters will be investigated in order to analyze the battery pack behavior more realistically.

Vehicle modeling approach: longitudinal dynamic model

The physical modeling of the longitudinal vehicle dynamics can be defined by considering the instantaneous energy balance for the two-wheeler's body in the Equation 3.4:

$$m_{tot}\, a\, v = P_{mot} - P_{diss}, \tag{3.4}$$

where m_{tot} is the vehicle equivalent mass, v and a are the velocity and the acceleration, respectively, of the two-wheeler's center of gravity at a specific time instant t, P_{mot} is the power generated by the power train system at the wheels level, and P_{diss} stands for the power of all the dissipative phenomena occurring. The equations are derived from the longitudinal vehicle dynamic modeling theory. By integrating Equation (3.4) over a time interval, the energy associated to the body can be calculated following the Equation 3.5:

$$\int m_{tot}\, a\, v\, dt = \frac{1}{2} m_{tot}\, v^2 \tag{3.5}$$

The total energy of the vehicle body is related to the contributions of both the longitudinal speed and the rotational speed of the drive train components. Among the latter, the most important one is the rotatory inertia of the electric motor, which absorbs part of the traction power delivered by the battery, and it dissipates energy in deceleration phases. The same holds for the mechanical transmission, yet its contribution can be assumed as negligible. Hence, just the inertia of the electric motor was considered in this work. Equation (3.4) can thus be written as Equation 3.6:

$$m_{tot}\, a\, v = m\, a\, v + J_{motor}\, \theta\, \dot{\theta}. \tag{3.6}$$

where m is the translating mass of the vehicle, consisting of the sum of vehicle and driver masses, J_{motor} is the motor rotating inertia, while θ and $\dot{\theta}$ are the motor rotational velocity and its acceleration, respectively. By considering the kinematic of the vehicle transmission, we obtain the Equation 3.7:

$$J_{motor} \cdot \theta \cdot \dot{\theta} = J_{motor} \cdot \frac{\tau^2}{R^2} \cdot v \cdot a. \tag{3.7}$$

where R is the wheel dynamic radius, and τ is the speed ratio between the motor and the driven wheel rotational velocity. Two-wheelers equipped with ICE generally embed a CVT (continuously variable transmission). However, nowadays, CVTs are solutions generally not exploited for electric vehicles [126]. The drivetrain of a two-wheeler is an "in-body" transmission. Hence, it is reasonable to assume τ as a constant parameter. So, it can be assumed by the Equation 3.8:

$$J_{motor} \cdot \frac{\tau^2}{R^2} = m_r; \tag{3.8}$$

Therefore, the left-side term of Equation (3.4) becomes Equation 3.9:

$$m_{tot}\, a\, v = (m + m_r)\, a\, v \tag{3.9}$$

where m_{tot} is the equivalent mass.

As far as the motor power P_{mot} is concerned, it is considered corresponding to the battery power model output $P_{battery}$, V and I being the voltage and the current delivered by the overall battery pack, respectively, and assuming a battery-to-road efficiency η_{b2r} due to electrical and mechanical energy conversion (e.g., inverter, power electronics, bearings, tire friction and other transmission loss terms). The P_{mot} is, then, expressed in the Equation 3.10:

$$P_{mot} = \eta_{b2r}\, V\, I \tag{3.10}$$

As far as the dissipation terms are concerned, different contributions could be included into the equation term P_{diss} depending on the level of accuracy required. P_{diss} is equal to the product between the overall dissipative force acting on the vehicle F_{diss} and its velocity v, shown in Equation 3.11:

$$P_{diss} = F_{diss}\, v. \tag{3.11}$$

The main phenomena considered in F_{diss} are:

- aerodynamic forces, which represent the air resistance to the vehicle motion and dependent on the square of vehicle velocity, air proprieties and geometrical vehicle shape. The mathematical formulation is expressed in Equation 3.12:

  $$F_{Aerodynamic} = \frac{1}{2} \cdot \rho_{air} \cdot A_f \cdot C_x \cdot v^2, \tag{3.12}$$

  and more precisely, ρ_{air} is the air density, A_f is the frontal area of the vehicle and C_x is a dimensionless parameter lower than 1 called the drag coefficient, describing the aerodynamics of the body. According to the mathematical formulation, the lower this coefficient's value, the better the aerodynamic performances of the considered vehicle.

- Tire rolling resistance, which represents the resistance produced by the contact between tires and the road surface. The energy dissipation of this term is due to the elastic micro-deformation in the tire body. Its magnitude is modeled by the Equation 3.13:

  $$F_{Tire} = m \cdot g \cdot f \cdot cos(\alpha), \tag{3.13}$$

  The α parameter indicates the road grade, g is the gravity acceleration and m is the vehicle mass. Particular importance is give to the value of the f coefficient, which is called *rolling resistance coefficient*, and it depends on the tire composition and the road surface material. Concerning passenger cars, its value can range from 8 $\frac{kg}{t}$ considering homogeneous asphalt [127] and normal tires up to 45 $\frac{kg}{t}$ considering an off-road path [128]. The F_{Tire} formulation can also include a factor which depends on the square of vehicle velocity and an f_2 coefficient, but because this values is quite small, considering the application of a two-wheeler, these were neglected in this work.

- Gravity resistance consisting in resistance forces produced by the road slope. Its value can be calculated following Equation 3.14:

$$F_{Slope} = m \cdot g \cdot sin(\alpha); \tag{3.14}$$

  Since all the experimental tests were driven on approximately flat road, the values of both α and F_{slope} are neglected.

All the other elements which cause the dissipation of energy and which are not included above (for instance the internal friction between the mechanical elements) will be afterwards considered in the generic factor η_{b2r}, standing for the overall efficiency of the system. In general, the sum of the dissipative forces F_{diss} can thus be broken down according to the Equation 3.15:

$$F_{diss}(v) = F_{Aerodynamic} + F_{Tire} + F_{Slope} = \frac{1}{2}\rho_{air}A_f C_x v^2 + mgfcos(\alpha) + mgsin(\alpha) \tag{3.15}$$

which is rewritten in the form of Equation 3.16:

$$F_{diss} = A + B \cdot v + C \cdot v^2, \tag{3.16}$$

with:

$$\begin{cases} A = mgf; \\ B = 0; \\ C = \frac{1}{2}\rho_{air}A_f C_x; \end{cases} \tag{3.17}$$

Finally, the vehicle velocity can be predicted from the all previous equations and the battery electrical power, solving and substituting all terms of the (3.4), obtaining Equation 3.18:

$$\left(m + m_r\right) v \frac{dv}{dt} = \eta_{b2r} \cdot VI - v\left(A + Cv^2\right). \tag{3.18}$$

where the acceleration a is written as $\frac{dv}{dt}$ exploiting the differential relationship.

The equation (3.18) is implemented in the second block of the global model and shown in Figure 3.6

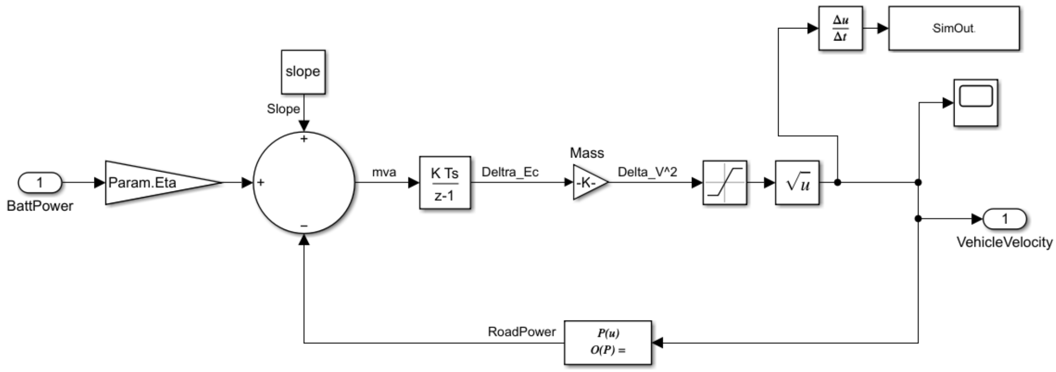


Fig. 3.6 Longitudinal dynamic equation of the electrified two-wheeler vehicle model implemented on Simulink.

3.2.2 Data acquisition system and on-road experimental campaign

The data of the two-wheeled prototype vehicle provided by the tests of the conducted experimental campaign were acquired on various driving missions. This was made possible by equipping the vehicle with an acquisition board capable of recording the signals from the vehicle's standard sensors, used for internal communication between the ECU, BMS, and driver monitor, as well as the signals from additional sensors, such as a Hall's effect current (LEM) sensor capable of enhancing the quality of the acquired current signal. Finally the raw signals of data are driven to an external data logger through the CAN bus protocol system and handled by a commercial software DEWEsoft® on an user device and memorized on a dedicated memory. The simplified scheme of the acquisition system is shown in the Figure 3.7.

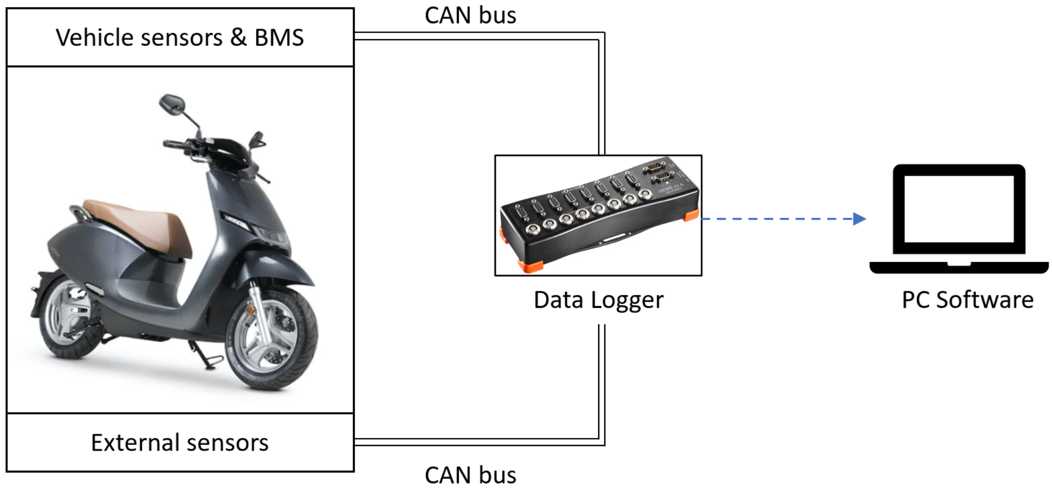


Fig. 3.7 A simplified scheme of the acquisition system used in the experimental campaign.

The main signals of interest collected through the BMS channels and external sensor were:

- battery pack voltage [V],
- battery pack current delivered [A],
- lowest and highest voltage among all internal cells [V],
- lowest and highest temperature among all internal cells [V],
- state of charge (SOC) of the entire battery pack,
- state of health (SOH),
- vehicle speed [km/h],
- opening percentage of throttle knob.

However, not all of these acquired signals were useful for the analyzes developed as they were not exploitable for the modeling approaches adopted. Several different driving missions were performed in order to define the parameters of battery and dynamic vehicle models and for the validation process among different operating on-road conditions. The different performed tests of the experimental campaign are listed in Table 3.5 and the relative routes performed in Turin are shown in Figure 3.8 and 3.9.

Table 3.5 Experimental tests available for the models identification.

Test	**Test Type/Session**	**Test Specifications**	**Repetitions**
1	Discharging Test	SOC in range [1, 0.3]	1
2.1	Coastdown 1	from 22 to 8 m/s	4
3.1	Coastdown 2	from 22 to 8 m/s	4
3.2	Coastdown 2	from 22 to 8 m/s	4
4.1	Coastdown 3	from 22 to 8 m/s	4
4.2	Coastdown 3	from 22 to 8 m/s	4
5.1	Constant Speed 1	5.5 m/s	2
5.2	Constant Speed 1	11 m/s	2
5.3	Constant Speed 1	16.5 m/s	2
5.4	Constant Speed 1	22 m/s	2
6.1	Constant Speed 2	5.5 m/s	2
6.2	Constant Speed 2	11 m/s	2
6.3	Constant Speed 2	16.5 m/s	2
6.4	Constant Speed 2	22 m/s	2
7	Partial Knob	50 %	1

The selected route for test #1 is shown in Figure 3.8, while the route for the remaining tests is shown in Figure 3.9. The same 'Test Type/Session' name refers to different tests performed in the same operating conditions. More details about tests conditions are provided below.

Fig. 3.8 Selected route in Turin, Italy, for the battery model calibration procedure.

Fig. 3.9 Selected route in Turin, Italy, for the calibration of the dynamic vehicle model.

Regarding the battery characterization tests, a driving mission was repeated until the BMS disconnected the battery voltage and the vehicle was turned down. This particulate test consisted of acceleration, constant speed, and deceleration phases, taking into account all possible operating conditions. The battery installed on the vehicle is a prototype pack unrelated to the manufacturer's on-board BMS, with a capacity approximately 50 percent greater than the battery for which the on-board BMS is designed. When the on-board BMS detected that the capacity delivered by the battery was close to the limit imposed by the manufacturer (which is calibrated on the original, non-over sized battery), it disabled the vehicle and turned it off to prevent critical issues arising from the under-voltage condition. The BMS system is a closed system and then the internal parameters cannot be altered. Therefore, it was not possible to completely discharge the prototype battery pack. As a result, the prototype battery was only discharged to approximately 70% of its entire capacity, whereas the calibration procedure was only feasible above this threshold value. In addition, temperature data were unavailable during the acquisitions phases due to the absence of specific sensors. In the present investigation, the battery has been partially modeled in the entire SOC domain, while temperature dependence has been neglected. Thus, the model's performance only referred to the standard environmental condition. The experimental signals of the variables exploited in the parameters identification of each single cell numerical model is graphed in Figure 3.10 showing the dynamic operation of the battery pack during the on-road trace and highlighting the transient, corresponding to the test #1 of "Discharging Test" name of the Table 3.5.

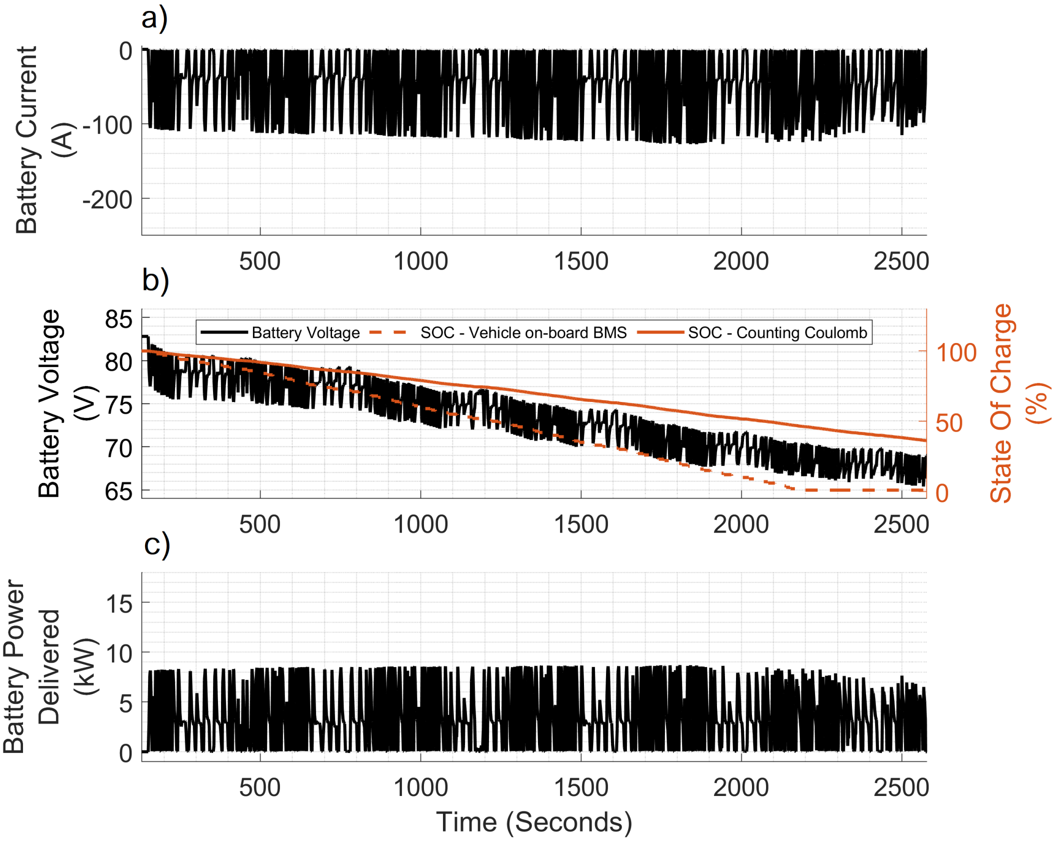


Fig. 3.10 Operating battery pack (**a**) voltage, (**b**) current and (**c**) power delivered during the tests #1.

A better highlight can be appreciate looking at the details represented in the Figure A.1 of the Appendix A, where the typical voltage response is shown under transient conditions and the total voltage drop due to the electrochemical phenomena, such as charge transfer, diffusion and static resistance. As can be seen, the SOC estimation reported in the figure had dual natures, as already explained. The dashed orange line corresponds to the SOC computed by the on-board BMS which is calibrated on the original battery pack, while continuous orange line is the post-processing SOC estimation performed and considered in this study through the Equation 3.3. Literature reveals that a number of SOC estimation methodologies take battery capacity into consideration; this may explain why the original BMS of the two-wheeler conducts an incorrect SOC estimation when monitoring the prototype pack. For the current analyses, however, only the estimated SOC after post-processing was considered, as determined by the Coulomb Counting method.

Once the initial condition SOC_{T0} is known [129, 130], this method has a powerful and reliable computation, which is sufficient for the actual needs in this work.

As far as the vehicle dynamic model identification, the experimental data exploited were those belonging to the tests from #2.1 to #7 of the Table 3.5 and carried out in different part of Turin (Italy). These tests were used to estimate the parameters of the dynamic vehicle equation. Specifically in the test sessions, different types of conditions among the driving mission were covered by the driver and explained as follows:

- the coastdown test consists of a phase of vehicle acceleration up to a certain speed, followed by a phase of deceleration down to zero speed, on an approximately flat road without braking and under stationary environmental conditions. In general, it is crucial to avoid selecting any gear during the deceleration phase of this type of test in order to avoid the inertia of driveline rotating elements [131, 132]. The two-wheeled electric vehicle, however, lacks a manual transmission selector. The vehicle is driven up to 80 km/h before the coastdown test begins, and numerous experimental records of this test are performed.
- In the constant speed test the drive had the objective of keeping the vehicle speed constant for several seconds and with several repetitions. The constant speed levels at which the vehicle was held were 20, 40, 60 and 80 km/h.
- In the partial knob tests, the driver held the acceleration knob at a constant partial opening of 50% for a few seconds.

In the Figure 3.11 some repetitions are shown concerning the three different driving conditions. How to use the different tests to identify the model parameters is explained below.

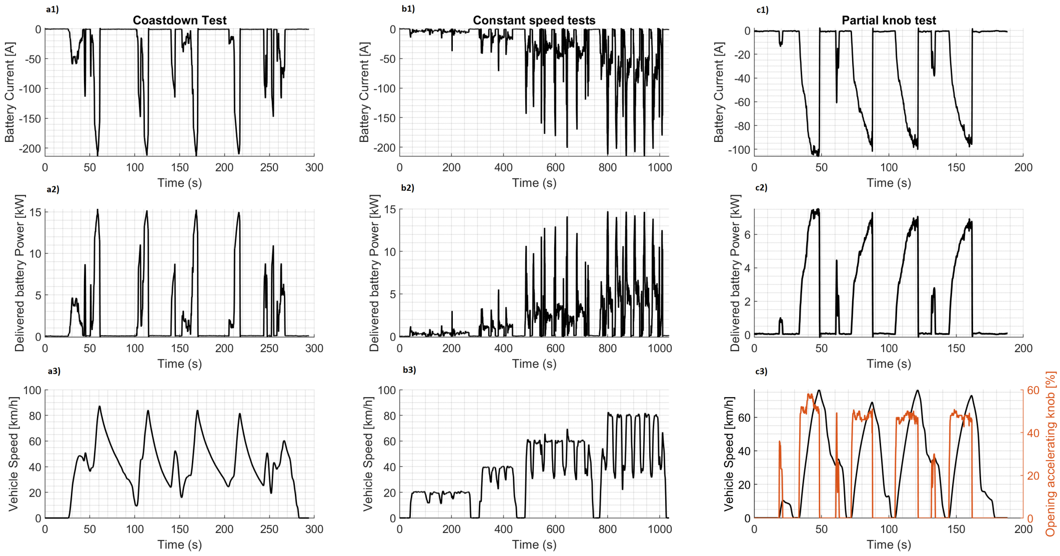


Fig. 3.11 Examples of some repetitions of the on-road experimental tests conducted over the three different driving conditions: (**a1**) Current, (**a2**) power and (**a3**) vehicle speed of the coastdown test; (**b1**) Current, (**b2**) power and (**b3**) vehicle speed of the constant speed test at 20, 40, 60, and 80 km/h; (**c1**) Current, (**c2**) power and (**c3**) vehicle speed of the partial knob test.

Each of the distinct subfigures depicts the complete time series obtained throughout the corresponding on-road test. The appropriate driving traces were appropriately truncated for the purpose of parameter calibration procedures, enabling the application of the modeling procedure. In the context of identifying coastdown parameters using the coastdown model, it is pertinent to note that only the segments pertaining to the vehicle's free deceleration, wherein the speed reduces and the battery's power output is zero, are considered. Similarly, with respect to the constant speed tests, the parameter identification model was applied by segmenting the data to exclusively analyze the vehicle's traces at the predetermined vehicle speed levels where it maintained a constant speed.

3.2.3 Parameter identification of the battery cells and vehicle models

Unknown parameters were found through an identification procedure in order to fully and uniquely define battery pack and vehicle models. Using multi-experiment data, parameters were estimated and validated in accordance with the tests defined in the table 3.5. As a result, the procedure for identifying the associated parameters has been established differently based on the subsystem taken into account via multiple datasets. As displayed, dataset #1 was used to calibrate the battery model's parameters, whereas datasets #2.1, #3.1, #4.1, and #5 were used to calibrate the dynamic vehicle model's parameters. On the same dataset consisting of datasets #3.2, #4.2, #6 and #7, both the standalone models and the global model have been tested and validated. For a better visualization of the subdivision between the datasets used for calibration and those used for validation, the values are reported in the Table 3.6.

Table 3.6 Datasets used for model identification and validation, with reference to column "Test" of the Table 3.5.

	Model Identification	**Model Testing**
Battery Model	1	3.2 – 4.2 – 6 – 7
Dynamic Vehicle Model	2.1 – 3.1 – 4.1 – 5	3.2 – 4.2 – 6 – 7
Global Model	- - -	3.2 – 4.2 – 6 – 7

Since the two models and their parameters were independently optimized, the global model was built by coupling the standalone models. Consequently, the identification of the global model has already been completed, and the relative box of the table contains no reference datasets.

In the present study, the model identification process is the technique according to which the values of the parameters that describe the behavior of the system are defined. This approach was based on an optimization procedure carried out through the Matlab© and Simulink© tool called *ParameterEstimator* [133, 134], which is based on a non-linear least squares (NLS) optimization method. This technique is based on the trust-region-reflective algorithm [135, 136] which is widely employed in the literature [137].

Regarding the calibration of the battery model's parameters, every single parameter was a function of the discretized state variable SOC, which ranged from 0 to 1 in 0.1-step increments. The parameters were represented in a 1D look-up table (LUT), as a function of SOC values. Due to the implemented modeling methodology, each single cell model would have had six 10x1 LUTs, one for each identifying parameter. Taking into account that the complete battery pack consisted of 180 cells, the optimization procedure should have identified a 1080 LUT, which corresponds to a total of 10800 values in the discretization values. To simplify the optimization procedure and save computation time, the hypothesis of cell equality has been used, limiting the task to determining the parameters of just one sample cell. This was accomplished by using a basic Simulink model with just one cell and taking into account the cell current $\bar{I}_{cell}$, which can be calculated by dividing the battery experimental current I by the number of battery pack modules which was equal to nine. The single cell voltage $\bar{V}_{cell}$ was then similarly computed taking into account the number of cells in series in a single module. The $\bar{I}_{cell}$ and $\bar{V}_{cell}$ are computed through the Equation 3.19.

$$\begin{cases} \bar{I}_{cell} = \dfrac{I}{\text{number of modules}} \\ \bar{V}_{cell} = \dfrac{V}{\text{number of series}} \end{cases} \tag{3.19}$$

In this calibration procedure, a graphic user interface was used in the optimization tool in order to set the NLS method as the objective function to be minimized, the acquired data were specified as input and output and the parameters were set to be controlled for the minimization process. Finally evaluating and minimizing the error function of the mathematical problem between model output and experimental measurements, the parameters' values of the second-order RC model of lithium-ion cell has been determined. The Table 3.7 shows a summary of the variables involved in the procedure.

Table 3.7 Parameters involved in the battery model identification procedure.

Input	**Output**	**Control variable**	**State variable**
$\bar{I}_{cell}(A)$	$\bar{V}_{cell}(V)$	$R_0, R_1, R_2, C_1, C_2,$ OCV	SOC

The terminology used to describe the types of variables, such as state and control, is a reference to what is known in the literature as dynamic optimization. The reference has only the purpose of underlining the role that these variables had in the problem of the present work . Thus, the state variable SOC is an independent variable that determines the system's state, while control variables are parameters whose values must be selected (and hence optimized) to minimize an objective function.

On the other hand, as regards the definition of the calibration problem of the vehicle model parameters, with reference to Equation 3.18 the aim was to determine the values of A, C related to the dissipative forces of the road load, and η_{b2r} and m_r related to the driveline. To carry out this activity, the identification process has been divided into two different phases characterized by different physical operating conditions. The datasets relating to the coastdown tests were therefore used to determine the coastdown coefficients A and C. In a theoretical coastdown test application, the vehicle is decelerated without any mechanical coupling between the transmission and the wheel and thus not providing any power to the motion through the battery. Hence, in this case, the reference equation was reduced to the Equation 3.20.

$$\left(m+\cancel{m_r}\right)v\frac{dv}{dt}=\cancel{\eta_{b2r}\cdot VI}-v\left(A+Cv^2\right). \tag{3.20}$$

Then, the objective function to be minimized in the optimization process was defined and it is reported in the Equation 3.21.

$$min\{\left(m\right)v\frac{dv}{dt}+v\left(A+Cv^2\right)\}; \tag{3.21}$$

After the coastdown coefficients characterizing the road load were obtained, the constant speed test datasets were applied to compute the battery-to-road efficiency η_{b2r} and the mass associated with the rotating inertia m_r. Given the nature of constant speed testing, vehicle acceleration should be theoretically zero and the equation obtained was that of Equation 3.22

$$\cancel{\left(m+m_r\right)v\frac{dv}{dt}}=\eta_{b2r}\cdot VI-v\left(A+Cv^2\right). \tag{3.22}$$

and then, the new objective function to minimize in the identification process of the two vehicle parameter is shown in Equation 3.23

$$min\{\eta_{b2r}\cdot VI-v\left(A+Cv^2\right)\} \tag{3.23}$$

Nevertheless, in the current investigation, the electric two-wheeler had a single-speed transmission, and the electric motor could not be disconnected from the wheels, resulting in the resistive torque created by the spinning components' inertia persisting. As a result, the term m_r could not be neglected during the coastdown calibration phase, and the parameters were determined iteratively. Initially the value of m_r was set equal to 0 and the numerical values of the parameters A and C was determined. Knowing the coastdown coefficients, through test maneuvers and tests available under constant speed conditions, the parameters m_r and η_{b2r} have been identified. After this, we proceeded with the next iteration until the convergence of the process was reached, i.e. until the values of the parameters were in a narrow neighborhood of their values at the previous steps. Then, for the optimization problem the considered objective functions to be iteratively minimized at each single step are show by the Equation 3.24.

$$\begin{cases} min\{\left(m+m_r\right)v\frac{dv}{dt}+v\left(A+Cv^2\right)\} \\ min\{\left(m+m_r\right)v\frac{dv}{dt}-\eta_{b2r}\cdot VI+v\left(A+Cv^2\right)\} \end{cases} \tag{3.24}$$

It is worth mentioning that the electric two-wheeler had not energy recovery mechanism, such as the regenerative braking. Thus, all vehicle deceleration where a braking event was expected were carried out by the mechanical brakes. For the vehicle model parameters identification, the Table 3.8 shows a summary of the involved variables

Table 3.8 Parameters involved in the vehicle model identification procedure.

Input	**Output**	**Parameter Identified**
$P_{batt}(W)$	$V_{veh}(m/s)$	A, C, η_{b2r}

A global model has been developed, which combines the Li-ion battery and vehicle body models. This model has been fine-tuned and validated using experimental tests conducted under real-world driving conditions. Therefore, during the testing phase, the model's predictions were evaluated against the experimental data, and the discrepancies between the datasets were analyzed using statistical methods to assess the accuracy of the regression. The performance analysis involved the implementation of the following metrics.

- The *Root-Mean-Square Error* (*RMSE*):

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}x_i-\hat{x}_i^{\,2}}{n}} \tag{3.25}$$

- The determination coefficient, *R squared* (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2} \tag{3.26}$$

where x_i is the experimental measured data, $\hat{x}_i$ is the model estimated value, $\bar{x}_i$ is the mean of experimental measured data in the considered test session and n is dataset sample size. The primary physical parameters employed for assessing performance were the power and voltage of the battery, the velocity of the vehicle, the distance covered, and the total electrical energy dispensed during the entire test. The battery model and vehicle model outcomes are initially presented and evaluated independently. Subsequently, an analysis is conducted on the overall performance of the global model, with a focus on benchmarking, to facilitate a comparison between the distance traversed by the vehicle in experimental settings and the distance predicted by the models. The aforementioned parameter serves as the primary metric in evaluating the model when the objective is to estimate a range. The evaluation of the global battery-vehicle coupled model is based on two parameters, namely,

$$\text{global distance traveled } = \sum_{k=1}^{n} v_k \cdot Ts \tag{3.27}$$

$$\text{total energy delivered} = \sum_{k=1}^{n} V_k \cdot I_k \cdot Ts \tag{3.28}$$

where v_k is the vehicle speed, T_s is sampling time or test time and equal to 0.1 s, V_k is battery voltage, I_k is battery current and k is the time index on data acquisition.

3.2.4 The data handling during a pre-processing phase

In the post-acquisition stage, the acquired data underwent a rigorous data handling process to facilitate their utilization in the development of a data-driven model in a subsequent phase. The data obtained from on-road test acquisitions underwent pre-processing procedures to ensure its consistency, robustness, and quality. The process entailed scrutinizing the data for any instances of missing values, spikes, outliers, or other undesirable occurrences that could have arisen due to transient malfunctions in certain sensors. The CAN data transmission-based acquisition system comprised distinct channels that corresponded to various sensors. Consequently, it was imperative to verify the synchronization of the physical signals to ensure accurate management and advancement of the battery and vehicle models. The process of cleaning the data was of utmost importance in guaranteeing the robustness and reliability of the results derived from the subsequent analysis. Furthermore, in order to ensure accurate analysis and manipulation of the data, it was necessary to apply signal filtering techniques to eliminate extraneous sensor noise in frequency bands that fall outside of the operational range. Additionally, consistent frequency resampling was required to facilitate the proper functioning and development of the algorithms.

In the context of coast-down acquisitions, the deceleration phases were segregated from the remaining tests and consolidated to determine the coefficients of the resistance forces. An identical procedure was executed for acquisitions at a consistent velocity. Through the implementation of pre-processing procedures, the dataset was appropriately formatted for utilization in subsequent analytical processes, which encompassed the creation of a model driven by data.

3.3 Results & Discussions

The vehicle's global model was constructed through the integration of the battery model and the dynamic behavior modeling of the vehicle. Consequently, this section presents the evaluations of the individual models as well as the comprehensive model. Subsequently, the parameterization and response of the battery, vehicle, and global models are presented as follows.

3.3.1 Battery model

The efficacy of the battery pack model was evaluated through a comparative analysis of the simulated Voltage and correlated Simulated Power against the corresponding real-world measurements. The present study maintained the Experimental Current, as depicted in Figure 3.12, which corresponds to the current supplied by the battery. The present study examines the power signal to assess the efficacy of the battery pack model and to investigate the error propagation in the sequential model chain.

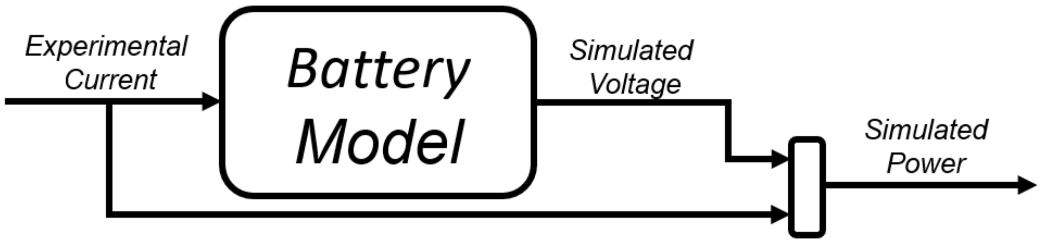


Fig. 3.12 Battery model scheme with Input/Output representation.

The model under consideration employs an ECM for every individual cell. All individual Li-ion cells were modeled using identical parameter values that were determined during the identification procedure. It is widely acknowledged in practical settings that Li-ion cells contained within a battery pack exhibit slight variations from one another as a result of manufacturing tolerances and temperature gradients. Hence, it is imperative to have an appropriate battery management system (BMS) to regulate the state of the battery due to the varying behaviors exhibited by cells during operational phases. Nonetheless, it is plausible to consider the hypothesis of cell similarity [138, 139] in this instance, as the effect on the prediction of the two-wheeler's range is relatively minor. As previously stated, the determination of cell parameters is based on data from experiments and remains significant until the state of charge (SOC) diminishes to 30% of the overall charge capacity, as shown in Figure 3.13.

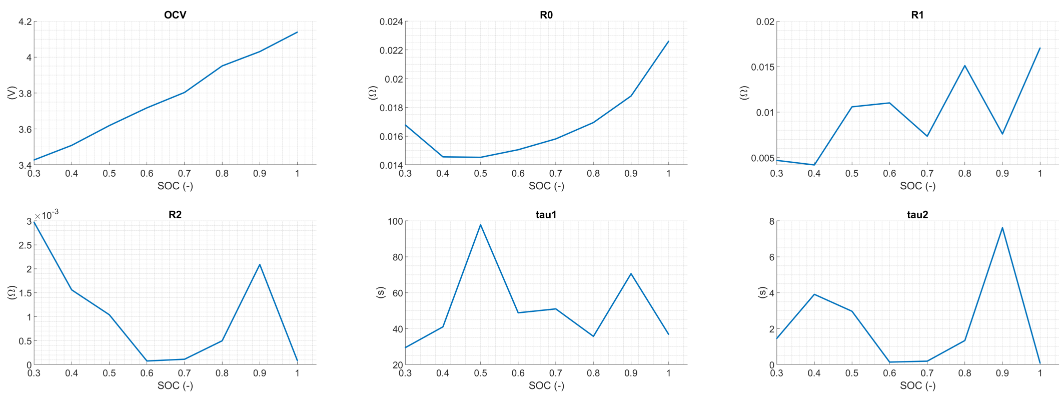


Fig. 3.13 Second-order cell model parameters estimated through the identification process.

The BMS constraint outlined in the Material & Methods section restricts our ability to fully utilize the battery's complete operational range. Nevertheless, the methodology may be expanded to encompass the value of SOC equivalent to 0% subsequent to further experimental assessments aimed at attaining total discharge. The assessment of model performance can be conducted through a comparative analysis between the voltage signal obtained from the model output and the experimental data. Subsequent to the calibration phase, the simulated voltage is graphed in relation to the target voltage across a period of time, as illustrated in Figure 3.14. Figure A.5 depicts an enlarged representation of the simulated voltage in comparison to the measured voltage, accompanied by the corresponding point-wise error.

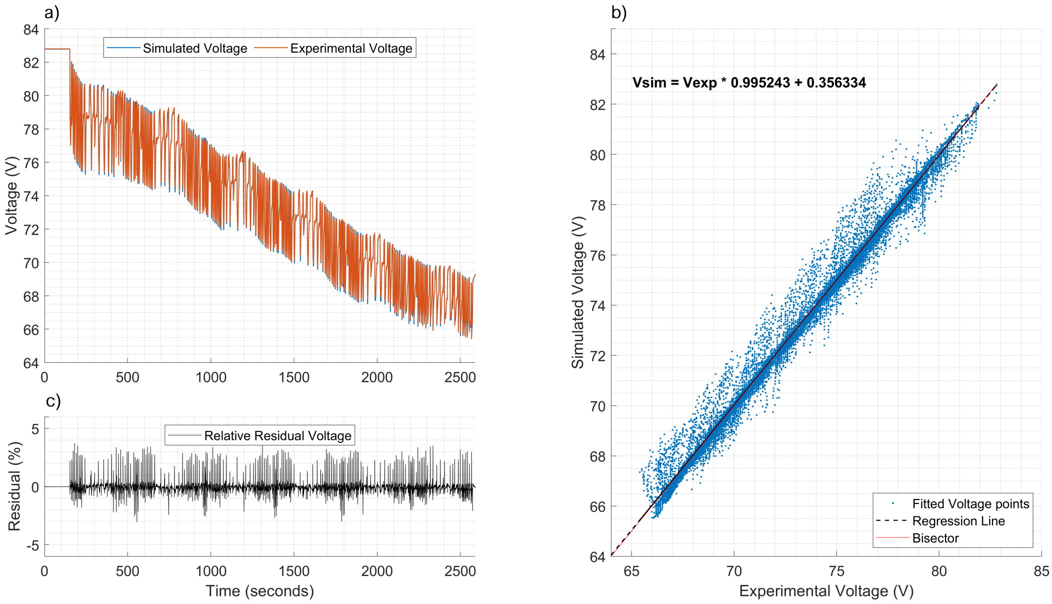


Fig. 3.14 Test #1.1: regression process results for estimation of model voltage. (**a**) The simulated voltage and experimental voltage on real driving mission and (**c**) residual error of estimation. (**b**) Regression process results in terms of fitting equation and bisector comparison.

As per the depicted diagram, the simulated voltages exhibit a relative inaccuracy that remains below 1% in absolute value for the majority of the data points. However, during transient discharging cycles, a few peaks of 4% are achieved. This holds true for the majority of the simulated voltages. Furthermore, the outcomes of the regression analysis are exhibited in Figure 3.14 in the shape of fitted points. The equation displayed for the regression line exhibits a satisfactory level of overlap with the bisector, with a angular coefficient of about 0.995 and a relatively low value of the intercept. The performance results in terms of metrics seen in previous section are reported in Table 3.9.

Table 3.9 Performance metric results for battery model.

Test #	RMSE (V)	R^2
1	0.383	0.993

Table 5 highlights that the optimized battery parameters enable us to confine the RMSE with empirical data to less than 0.38 V. Furthermore, the predictive abilities have been validated by the R-squared metric, which indicates a value of 99.3%. To assess the scope and efficacy of the two-wheeler, the battery's electrical power output is utilized as an input for the vehicle model. Therefore, a similar analysis is conducted to assess the residual error that arises from battery modeling and is subsequently transmitted through the overall model. Figure 3.15 displays the estimation outcomes computed for the battery power.

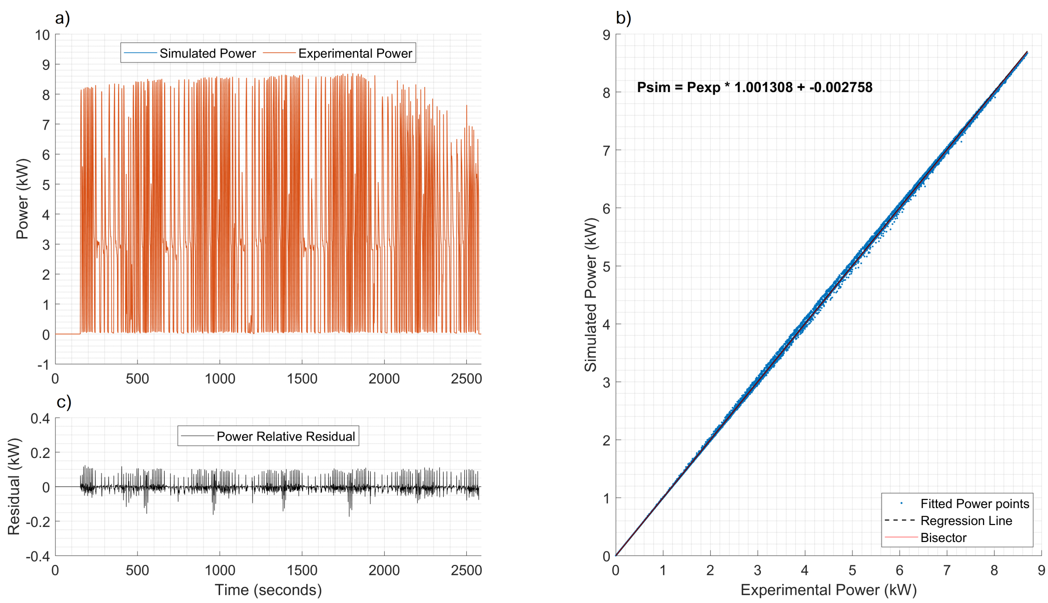


Fig. 3.15 Test #1.1: Regression process results for estimation of model power. (**a**) The simulated and experimental power on real driving mission and (**c**) residual error of estimation. (**b**) Regression process results in terms of fitting equation and bisector comparison.

It is evident that the error in the simulated power residual is confined to a narrow range of values, specifically between $0.1kW$ and $-0.1kW$. Furthermore, it is noteworthy that the residual error for the majority of the time intervals is markedly less than 0.02 kW. The fitting process displayed in the right-hand sub-figure of

Figure 3.15 underscores the consistency of the estimation results. Table 3.10 presents a summary of the statistical metric parameters computed, which serve to further validate the accuracy of the estimation.

Table 3.10 Battery Parameter Identification Power and Energy Metrics.

Test #	RMSE (kW)	R^2	Experimental Energy Delivered (Wh)	Simulation Energy Delivered (Wh)	Relative Error Energy (%)
1	0.022	0.999	2409.3	2407.9	0.058

The battery model's introduced error is deemed to be minimal and satisfactory, thereby resulting in the absence of significant adverse effects of error propagation. Additional information pertaining to the global models is presented in the Results section. The aforementioned findings have been validated through the execution of analogous analyses on disparate datasets in relation to those utilized for parameter identification during the calibration stage. The evaluation metrics for the model testing are presented in Tables 3.11 and 3.12.

Table 3.11 Battery voltage performance on different datasets.

Test #	RMSE (V)	R^2
3.2	0.344	0.984
4.2	0.75	0.983
6.1	0.393	0.960
6.2	0.568	0.989
6.3	0.727	0.984
6.4	0.815	0.987
7	0.181	0.981

Table 3.12 Battery power performance in different datasets.

Test #	RMSE (kW)	R^2	Experimental Energy Delivered (Wh)	Simulation Energy Delivered (Wh)	Error (%)
3.2	0.035	0.999	158.8	157.9	0.527
4.2	0.076	0.999	158.6	161.5	1.852
6.1	0.006	0.999	54.5	54.9	0.573
6.2	0.022	0.999	114.8	115.9	0.980
6.3	0.051	0.999	190.7	193.2	1.300
6.4	0.079	0.999	308.7	312.9	1.382
7	0.007	0.999	86.3	86.5	0.182

The battery model exhibits noteworthy outcomes in relation to battery voltage and electrical power, as evidenced by Tables 3.11 and 3.12. Regarding the electrical battery power, the prediction capabilities have been demonstrated to be of high quality. This is evidenced by R-squared values exceeding 99% for every real-world driving mission in the conducted tests. The findings emphasize the uniformity in estimating overall energy dispensed across on-road empirical routes, with the percentage discrepancy persisting below 2%. The model exhibits certain limitations. Initially, it should be noted that the influence of temperature on battery parameters has not been considered as a result of the unavailability of empirical data. Consequently, the forthcoming activities will entail conducting novel experimental assessments concomitant with the monitoring of the temperature of pertinent powertrain components. It may be of interest to consider additional environmental factors, such as ambient humidity, to assess the potential improvement in precision with respect to an expanding model complexity.

3.3.2 Longitudinal dynamic model of the vehicle

As shown in the Methods section, the electric scooter has been modeled through a longitudinal dynamic approach, which has been represented through Equation 3.18. The test protocol was executed in accordance with the simplified vehicle schematic depicted in Figure 3.16. The input to the model is the battery power obtained from experimental tests and computed by experimental voltage and current, while the output of the model is the predicted speed of the vehicle, Simulated Velocity. The simulation of vehicle speed is evaluated against empirical on-road data.

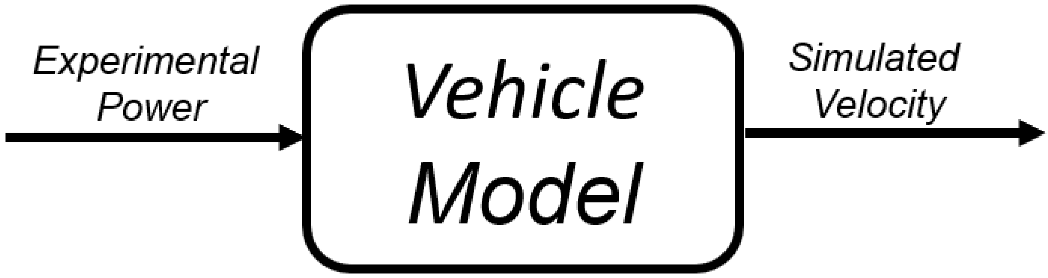


Fig. 3.16 Vehicle model scheme with Input/Output representation.

The process of defining the vehicle model occurred through the identification of parameters across different running scenarios. Ultimately, an iterative methodology facilitated the attainment of parameter value convergence. As previously stated, the load-resistive power of the road can be assessed through the analysis of coastdown test outcomes. Through the utilization of the fitting process, the coefficients of the polynomial Equation 3.16 can be determined. This equation serves as a model for the resistive phenomena that are encompassed within Equation 3.15. The computed values of the coastdown coefficients are hereafter reported in Table 3.13, while the resistive power contribute and the total amount as functions of vehicle speed are shown in Figure 3.17.

Table 3.13 Coastdown coefficients values. A, B and C are the coefficients describing the resistive power contributes.

Parameter	**Value**	**U.d.m.**
A	41.8	N
B	0	$N \cdot \frac{s}{m}$
C	0.3	$N \cdot \frac{s^2}{m^2}$

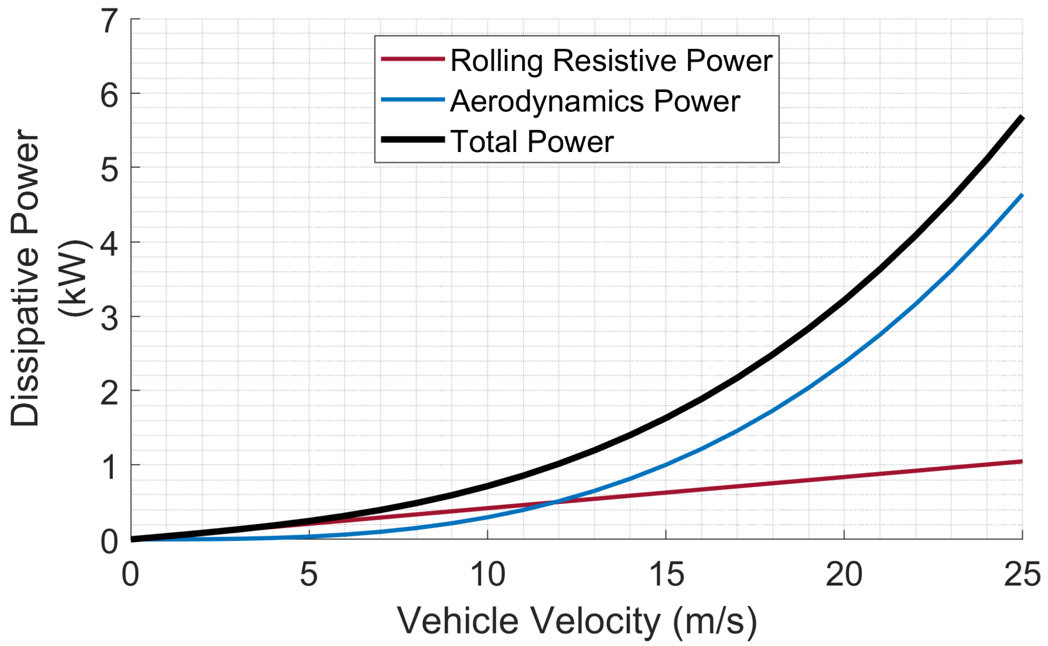


Fig. 3.17 Road load power terms acting on vehicle body as a velocity-dependent function.

The road load power is comprised of two distinct components, namely the rolling and aerodynamic phenomena. Figure 3.17 demonstrates that the rolling term is the dominant factor at lower vehicle speeds, whereas at higher speeds, the aerodynamic forces are the predominant factor, which is in accordance with established literature.

Upon completion of the road load assessment for the vehicle, the equivalent inertia of rotating components m_r and the overall battery-to-road efficiency η_{b2r} were derived through the resolution of the cost-function minimization problem. Table 3.14 contains the numerical values obtained as a result.

Table 3.14 Vehicle dynamic model coefficients: m is the sum of the vehicle and driver masses, m_r is the value of the equivalent mass of the rotating components and η_{b2r} is the overall battery-to-road efficiency, including the electric motors and transmission losses.

Name	**Value**	**U.d.m.**
m	184	kg
m_r	16	kg
η_{b2r}	0.75	-

Therefore, based on the analysis of Equation 3.17 and with the knowledge of the vehicle's mass m and frontal area A_f, the aerodynamic drag coefficient Cx and rolling resistance of the tires f have been precisely determined and presented in Table 3.15.

Table 3.15 Aerodynamic and rolling resistance coefficient values of two-wheeler road load model.

Aerodynamic drag coefficient	C_x	0.81	-
Rolling resistance coefficient	f	11.62	$\frac{kg}{t}$

The authors' literature review section exhibits a dearth of comprehensive empirical findings pertaining to the coefficients of drag and rolling resistance of electric scooters. A study similar in nature has demonstrated relatively similar values for the experimentally calibrated coefficients [140]. To assess the performance of the vehicle model, the same procedure used for the battery was utilized, whereby the model output was compared to the corresponding experimental data. The numerical model's projected velocity during the real-world driving task was evaluated, and the residual error was analyzed. As aforementioned, the modeling of braking was unattainable by the two-wheeler prototype, hence it was not included. Consequently, the determination of the braking torque was obtained through empirical observations, specifically in relation to velocity and acceleration, with the aim of guaranteeing conformity with the comprehensive energy equilibrium throughout the entirety of the driving mission. The Figure 3.18 depicts the input, output, and estimation error of the model. It is observed that the absolute error remains substantially below 2 m/s.

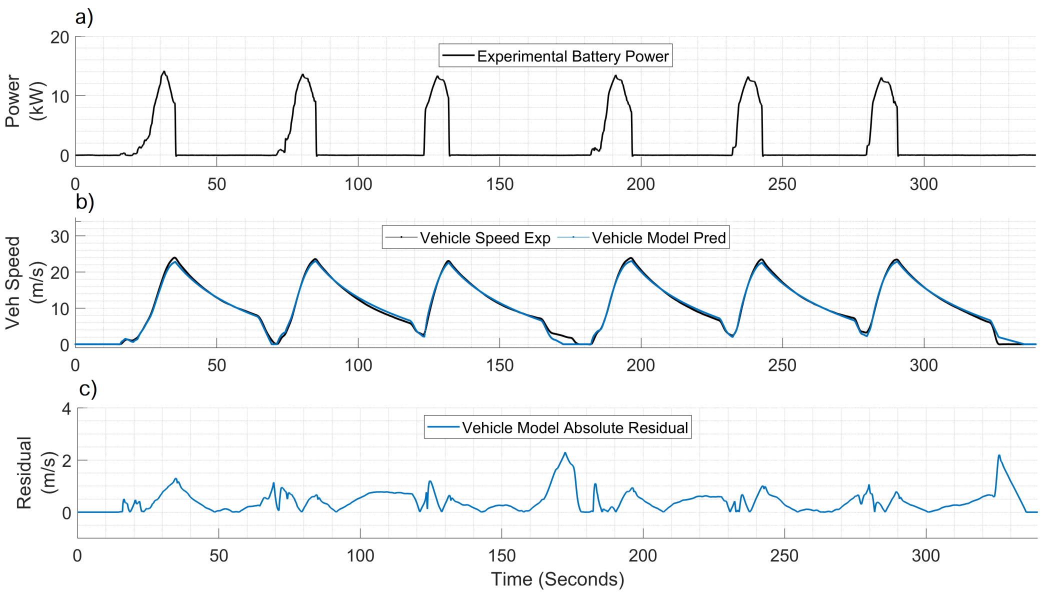


Fig. 3.18 Test #4.1. Standalone vehicle model estimation performance: (**a**) the delivered battery power, (**b**) the predicted and measured speed and (**c**) the residual of the estimation.

Nevertheless, certain inconsistencies between the estimated outcomes and the empirical evidence have been noted. The estimation of braking phases exhibits inadequate conformity with the measured data in certain instances of braking events. Furthermore, the model does not incorporate road slope as a result of the unavailability of appropriate acquisition sensors. Although tests were primarily conducted on roads with a predominantly level surface, it was inevitable that minor variations in road gradient would occur. Ultimately, a more precise evaluation of the efficacy of the battery-to-road conversion process must be undertaken to effectively anticipate the performance of powertrain elements. In summary, notwithstanding the aforementioned reductions in complexity, the model yields favorable outcomes and is deemed appropriate for the intended task. The experimental results obtained from the datasets used for parameter calibration and testing tasks confirm the efficacy of the numerical model's total distance prediction. The vehicle model performance have been assessed computing the electric range prediction in terms of total distance driven estimated along the selected on-road missions. This result is shown in Figure 3.19.

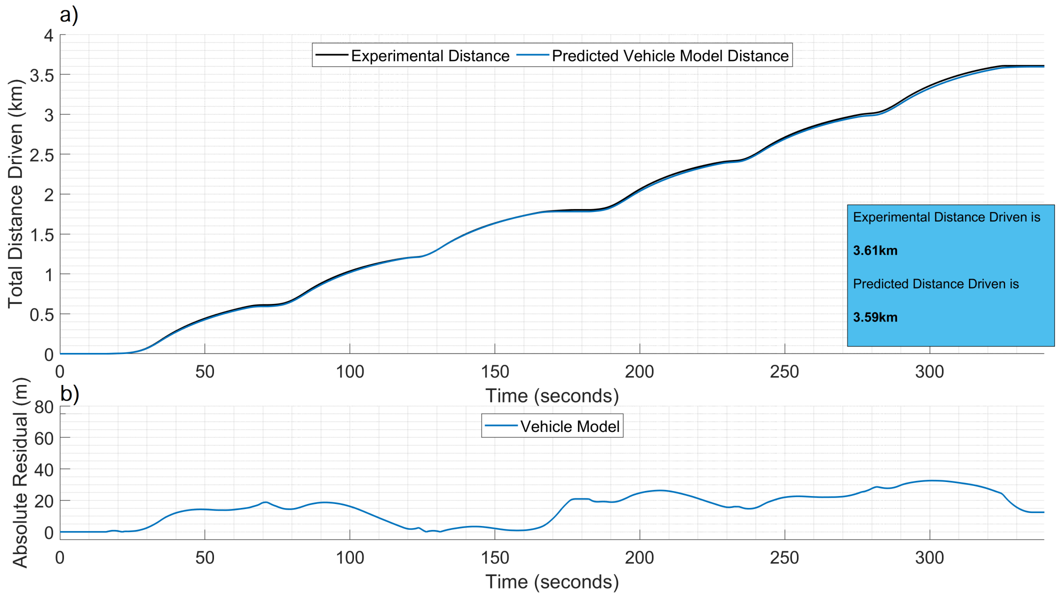


Fig. 3.19 Test #4.1. Electric range prediction of standalone vehicle model: (**a**) the total distance travelled and (**b**) the instantaneous absolute residual.

The numerical model developed demonstrates a high degree of accuracy in predicting spatial distance, as evidenced by the experimental data collected. The instantaneous absolute error appears to remain within acceptable limits, with a maximum relative error of approximately 0.55% observed for the distance traveled. This implies that over a cumulative driving distance of approximately 3.5 kilometers, the actual location of the motorcycle deviates by approximately 20 meters.

The evaluation of the performance of the vehicle model prediction is conducted on various cycles under real-world driving conditions. Therefore, Tables 3.16 and 3.17 present the results of the prediction performance for the datasets used in the identification parameters process and the testing phase, respectively.

Table 3.16 Performance metrics of vehicle model over datasets involved in the identification process.

Test #	RMSE (m/s)	R^2	Total Distance Traveled Experimental (km)	Total Distance Traveled Model (km)	Error Distance (%)
2.1	0.631	0.990	4.100	4.035	1.578
3.1	0.750	0.986	3.006	3.014	0.272
4.1	0.577	0.994	3.607	3.594	0.346
5.1	0.846	0.933	1.200	1.021	15.009
5.2	0.923	0.984	1.334	1.210	9.272
5.3	0.942	0.976	3.730	3.607	3.294
5.4	1.13	0.977	4.236	4.252	0.381

Table 3.17 Performance metrics of vehicle model over test datasets.

Test #	RMSE (m/s)	R^2	Total Distance Traveled Experimental (km)	Total Distance Traveled Model (km)	Error Distance (%)
3.2	0.914	0.978	3.13	3.12	0.496
4.2	0.503	0.996	3.60	3.61	0.334
6.1	0.507	0.974	1.95	1.86	4.90
6.2	0.638	0.987	2.98	2.85	4.41
6.3	0.908	0.982	3.48	3.42	1.89
6.4	0.788	0.991	4.67	4.62	0.93
7	1.171	0.973	1.59	1.53	3.32

Notwithstanding its simplicity, the vehicle model is capable of accurately approximating the distance covered by the electric two-wheeler during on-road experimental trials. The numerical model that was developed enabled an accurate estimation of the electric range.

Monte Carlo-based method for coastdown coefficients validation

The values of the parameters for coast down and their corresponding aerodynamic and rolling coefficients, as presented in Table 3.15, have received limited corroboration in existing literature. It is noteworthy that only a limited number of contemporary investigations have been conducted and have been alluded to in the preceding section. Hence, one of the widely recognized computational techniques in the scientific domain, namely the Monte Carlo method, can be employed to perform a verification analysis of these parameters. The Monte Carlo method [141] is a computational methodology that employs random sampling to address problems that are arduous or unfeasible to solve through analytical means. The application of this concept is pervasive across multiple disciplines such as physics, engineering, finance, and computer science.

The fundamental concept underlying the Monte Carlo technique involves the emulation of a significant quantity of stochastic events or phenomena, and leveraging the outcomes to approximate the performance of a given system or the likelihood of a particular result. The approach relies on the principle of the law of large numbers, which posits that as the quantity of trials or samples escalates, the sample mean gradually approximates the actual mean of the populace. The Monte Carlo method is implemented through a series of sequential steps, which typically include:

1. Define the problem and the system to be modeled.
2. Identify the input parameters and their probability distributions.
3. Generate a large number of random samples for each input parameter using a random number generator.
4. Use the input samples to simulate the behavior of the system and calculate the output or response variable of interest.
5. Analyze the output samples to estimate the behavior of the system or the probability of an outcome.

The utilization of this methodology proves to be advantageous in addressing intricate systems or non-linear associations between input and output variables, and it exhibits the capability to accommodate numerous inputs.

The current study employed the Monte Carlo technique to examine the performance of the resistive power model as characterized by the coastdown parameters. Thus, based on the selected algorithm, the inputs A and C were designated as parameters. The speed of the vehicle was determined by utilizing the model Equation 3.20 and a pre-existing distribution for the inputs. Subsequently, the distribution of the calculated speed was validated. The involved input parameters and their considered normal distribution are reported and show in Table 3.18 and Figure 3.20.

Table 3.18 Input parameters for the Monte Carlo analysis.

Parameters #	**Mean Value**	u.m.	Statistical distribution	Samples size
m	184	$[kg]$	-	1
A	23.23	$[kg/ton]$	Normal Distribution	2500
C	0.3	Ns^2/m^2	Normal Distribution	2500

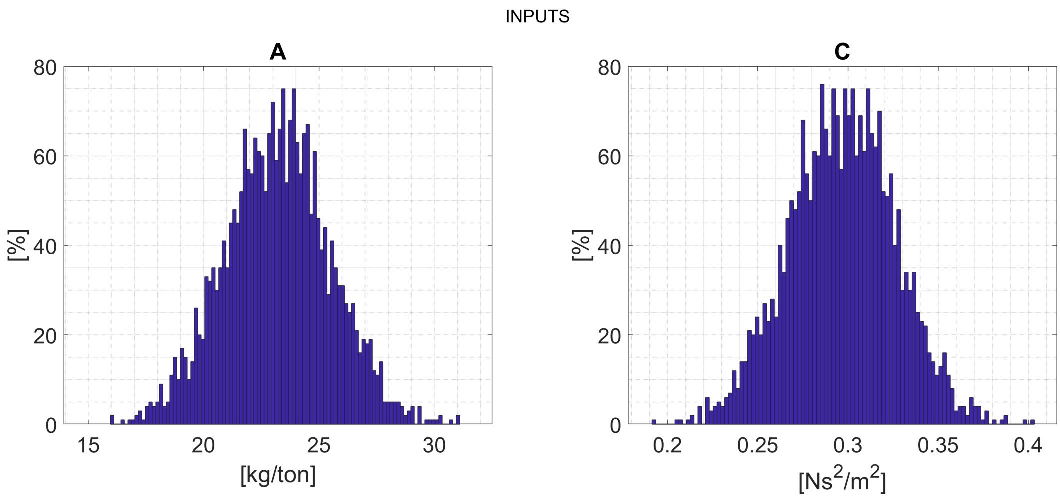


Fig. 3.20 Normal distribution of the input involved in the Monte Carlo analysis.

The study's output is the estimated speed of the vehicle, obtained through the application of the coastdown model with its coefficients featuring a normal distribution. Figure 3.21 displays the distinct curves that were obtained. The aim is to validate the normality of the velocity curve distribution subsequent to the implementation of a model. Hence, Figure 3.22 depicts the projection of instantaneous windows at varying speed levels during the free deceleration test, specifically at 5, 10, 15, and 20 seconds.

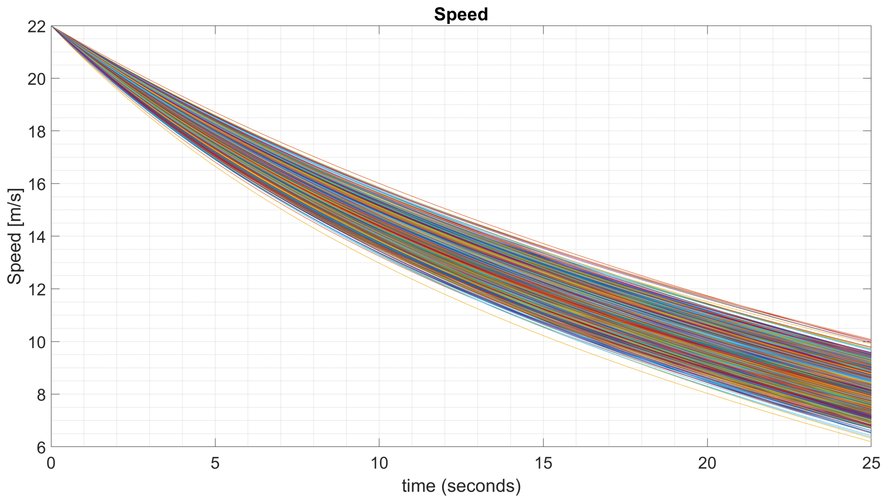


Fig. 3.21 Vehicle speed curves corresponding to the A and C distribution values.

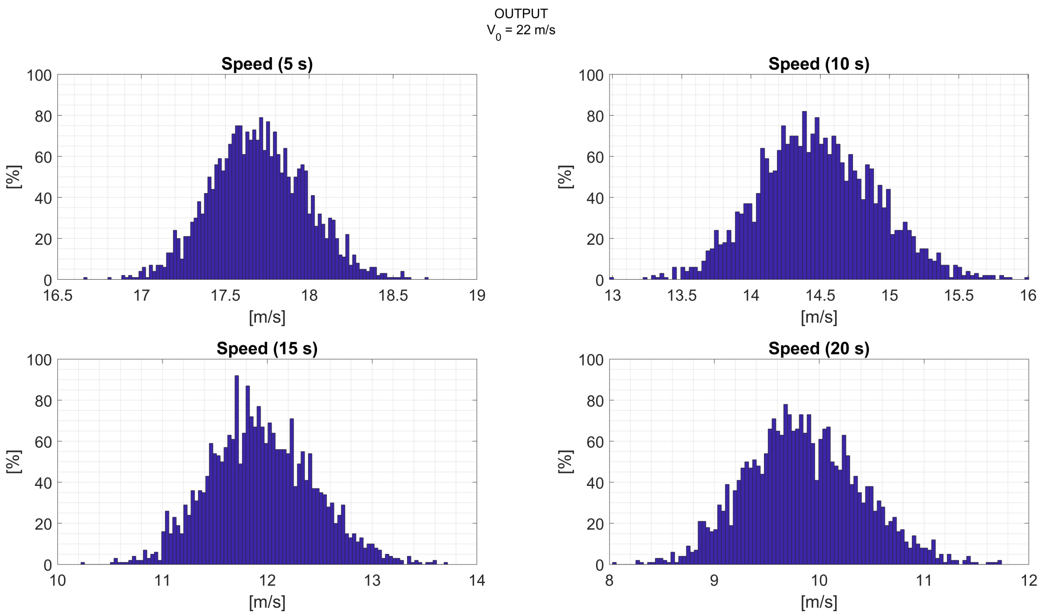


Fig. 3.22 Instantaneous vehicle speed windows at 5, 10, 15 and 20 seconds.

The preservation of the normality of the distribution in the model outputs is evidenced through the verification of the employed methodology. The current methodology was able to exhibit its robustness and reliability. Through the generation of a significant quantity of randomized samples for each input parameter and subsequent simulation of the system's behavior, a probabilistic approximation of the output or response variable was attained.

3.3.3 Global model

Ultimately, the comprehensive evaluation and validation of the global model is conducted by integrating the battery and dynamic vehicle model into a unified sequential model chain. Upon examination of Figure 3.23, the global model was evaluated through the utilization of the experimental current as a universal input and the subsequent comparison of the simulated velocity with the experimental measurements. The analysis conducted on the coastdown coefficients remains applicable to the aerodynamic and rolling coefficients, as these are defined respectively through constants starting from A and C. It is noteworthy that the statistical normal distribution is not affected by a constant.

Fig. 3.23 Global model Input/Output scheme.

As delineated in the preceding sections, the performance capabilities are examined across various on-road driving missions. Test 6.4 has been selected as the designated testing procedure for demonstrating prediction outcomes. The results depicted in Figure 3.24 demonstrate that the prediction of battery performance is adequate, as evidenced by the battery pack voltage and power residual error in relation to the experimental measurements. The analysis of the voltage absolute residual error reveals that only a limited number of peaks, measuring 2 volts, were attained. The aforementioned observation aligns with the findings presented in Figure 3.14 of the section pertaining to the battery model. Regarding the power output of the battery pack, the most significant residual errors are observed during the most extreme transient conditions of the cycle, which are associated with the battery pack's current peaks. Additionally, the largest residual error is observed in the voltage prediction. The voltage prediction discrepancy in transient cycle tracks may be attributed to minor fluctuations in ambient temperature during the day of experimentation. Consequently, the matter of temperature can be resolved through the implementation of supplementary testing procedures aimed at scrutinizing the correlation between battery parameters and temperature. The investigation of the predictive ability of the global model in terms of spatial distance has been investigated. The estimation of vehicle speed during a real-world driving mission is presented in Figure 3.25. The comparison is made between the global model, single vehicle model, and experimental data.

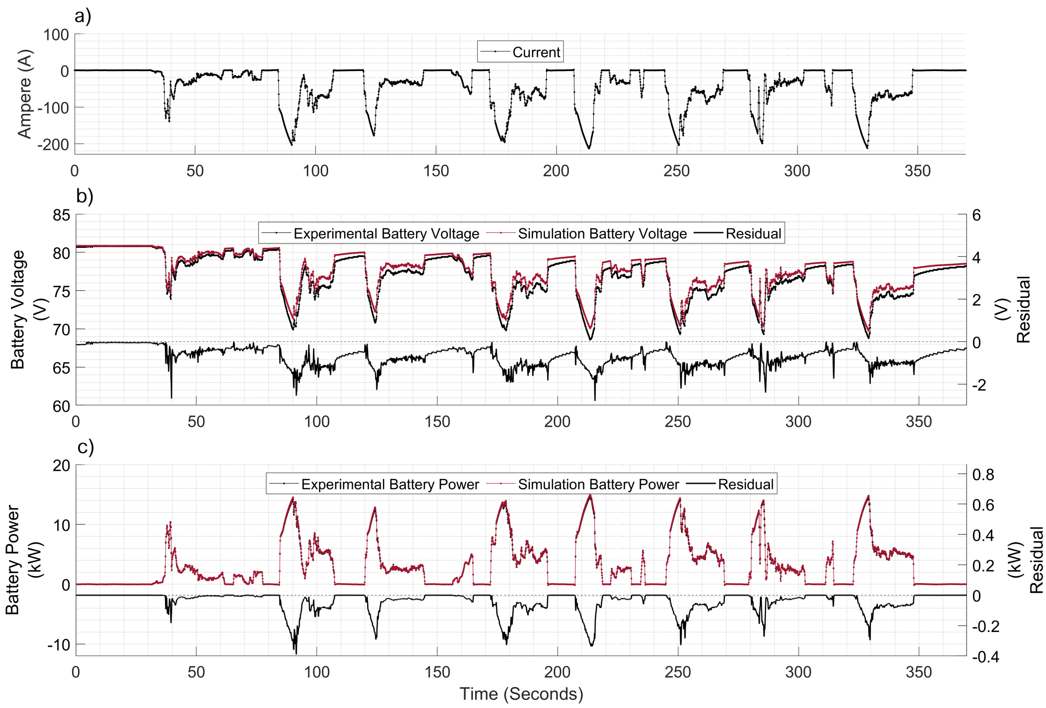


Fig. 3.24 Test #6.4: (**a**) battery pack electrical current (A) as global model input, (**b**) battery pack voltage and (**c**) total battery pack power.

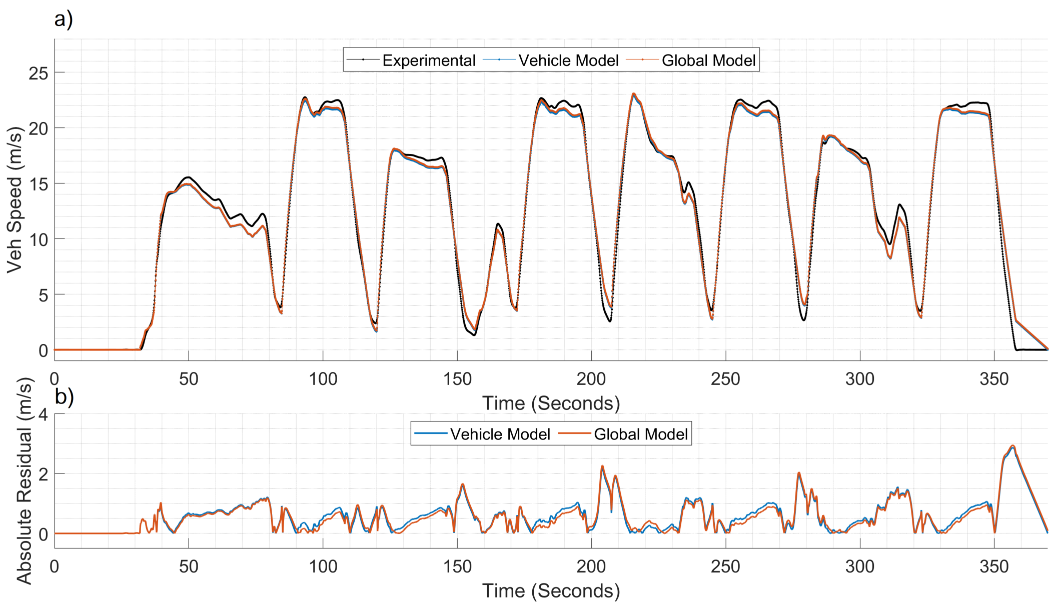


Fig. 3.25 Test #6.4. Two-wheeler performance prediction: (**a**) the vehicle speed (m/s) of the global model compared to that of the single vehicle model and experimental data, (**b**) absolute residual error of global and vehicle models.

Finally, the global model performance results are shown in Figure 3.26 in terms of total distance travel predicted.

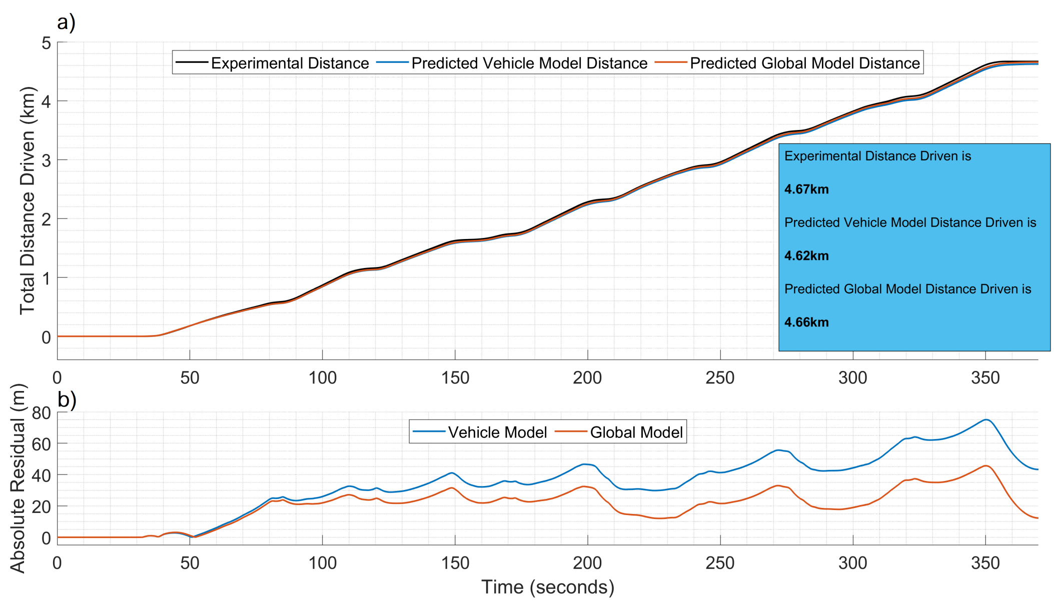


Fig. 3.26 Test #6.4. (**a**) Total distance traveled by the two-wheeler expressed as electric range is estimated by the global model and compared with that of single vehicle model and experimental data. (**b**) Absolute residual error is plotted over the test time.

The computation of the total distance covered by the vehicle during the mission and the prediction of the spatial distance exhibit a comparatively low level of instantaneous error. Table 3.19 presents the comprehensive outcomes obtained from the on-road, mission-based testing dataset.

Table 3.19 Performance metrics of global model over testing datasets.

Test #	RMSE (m/s)	R^2	Total Distance Traveled Experimental (km)	Total Distance Traveled Model (km)	Error Distance (%)
3.2	0.917	0.978	3.13	3.11	0.763
4.2	0.503	0.996	3.60	3.64	1.106
6.1	0.492	0.975	1.95	1.87	4.539
6.2	0.610	0.987	2.98	2.87	3.855
6.3	0.892	0.982	3.48	3.44	1.209
6.4	0.777	0.991	4.67	4.66	0.263
7	1.169	0.973	1.59	1.53	3.225

Upon comparing Tables 3.17 and 3.19 with regards to the testing datasets, it can be observed that the total distance estimation error produced by the global model is comparatively lesser than that of the single vehicle model. The univariate vehicle model exhibits a tendency to underestimate the overall distance covered by the vehicle. The observed phenomenon could potentially be attributed to the consistent magnitude of the global efficiency parameter denoted as η_{b2r}. This suggests that the driveline losses remain constant regardless of the vehicle's operating conditions. Conversely, the battery model exhibits a slight tendency to overestimate the electrical power of the battery pack. Consequently, the two errors in estimation exhibit a tendency to offset one another. Given that the objective of the tool in question is to predict a range, it is preferable to have an underestimation error as opposed to an overestimation error.

Chapter 4

Performance Optimization for Bi-LSTM Neural Network-Based Battery SOH Estimation in Li-ion Batteries

4.1 Introduction

[1] Electric motors have been employed as a means of propelling road vehicles for a considerable duration of time. The vehicle known as ’La Jamais Contente’ achieved a noteworthy milestone in 1899 by surpassing a speed of 100 km/h, thereby becoming the first electric car in history to do so [142]. Nevertheless, the technology was not implemented in everyday use due to the challenges associated with the storage of significant amounts of electrical energy on vehicles. The issue of employing electric motors for automotive traction has been partially resolved by recent technological advancements in Li-ion batteries.
In comparison to various other electrical storage systems, such as lead-acid batteries, it is a well-established fact that Li-ion batteries possess relatively high energy and power density, exhibit low levels of self-discharge, require minimal maintenance,

[1]Part of this chapter has been published in the form of a paper as: Falai, A.; Giuliacci, T.A.; Misul, D.A.; Anselma, P.G. Reducing the Computational Cost for Artificial Intelligence-Based Battery State-of-Health Estimation in Charging Events. *Batteries* 2022, 8, 209.

and demonstrate favorable load characteristics. Furthermore, they are capable of being partially charged and discharged without incurring damage [143–145]. When a battery management system effectively oversees Li-ion batteries, they can guarantee a satisfactory level of safety and a legitimate lifespan, which are essential requirements for automotive purposes. Conversely, the the Li-ion battery pack is unequivocally the most crucial and delicate constituent of the electric vehicle. Monitoring and overseeing the status of battery operation is crucial for the preservation of battery health. The BMS is responsible for ensuring that the battery pack operates within its safe range and under optimal conditions [146]. It is imperative for cells to function within a designated range of temperature and voltage, while refraining from delivering excessively high currents. The variability of these conditions is contingent upon various factors, including the type of chemistry, and may differ across individual cells. In certain circumstances, such as when batteries are subjected to elevated temperatures, they may undergo gas bloating, which can result in leakage or explosion. Additionally, a thermal runaway event may transpire [147, 148]. The management of thermal conditions in cells is a crucial concern in battery technology. Elevated temperatures accelerate battery degradation, resulting in a decline in performance over time, as evidenced by previous studies [149–151]. Conversely, lower temperatures reduce efficiency due to the increased internal resistance of the cell [152]. Typically, batteries tend to experience a faster rate of degradation when operating outside of their optimal temperature range, which is typically considered to be between -20°C and 60°C. As per the existing literature, it has been observed that batteries undergo degradation at different rates based on the stress cycles, which are commonly referred to as cycle aging, even when safe conditions are maintained [153].

The BMS plays a crucial role in preserving the optimal health and performance of the battery. Additionally, it is imperative to have knowledge of the battery's current health status. The degradation of the battery results in a reduction of its capacity, leading to a decline in the range of the vehicle, and an increase in its internal resistance. In typical automotive applications, batteries are deemed to have reached their end of life (EOL) when their capacity has declined to 80% of their initial value or when their internal resistance has increased to 200% of the initial value. These batteries have the potential to serve various stationary purposes, including grid energy distribution, thereby providing them with a secondary utility prior to their recycling [154]. The SOH parameter is employed to characterize the health status of the battery. In certain contexts where power capacity outweighs energy quantity, the internal resistance is commonly considered as SOH metric. As such, SOH is defined as the ratio between EOL and actual internal resistance, and EOL and initial internal resistance as expressed in the Equation 4.1,

$$SOH = \frac{R_{EOL} - R_{actual}}{R_{EOL} - R_{new}}, \tag{4.1}$$

where R_{EOL} is the end of life cell resistance, R_{actual} is the actual cell resistance and R_{new} is the fresh cell resistance. From the other SOH definition, the SOH is characterized as the ratio of the current battery capacity to the capacity at the initial stage of its lifespan, as described in the Equation 4.2,

$$SOH = \frac{Q_{current}}{Q_{new}} x100\%, \tag{4.2}$$

where $Q_{current}$ is the actual battery capacity and Q_{new} is the fresh battery capacity. This metric is particularly relevant in scenarios where the energy availability is of considerable importance [155].
Thus, the measurement of capacity or internal resistance may be necessary depending on the specific application. Various methodologies have been suggested in literature with regards to the assessment of SOH of batteries in electric vehicles [156–159]. The identification of SOH has been addressed in previous studies [160, 161] through the utilization of a semi-empirical formula, which draws upon the principles of the Arrenius equation. Considering the lithium-ion loss as main aging mechanism, the Arrenius equation with a power low relation with the cycles times are reported in equation 4.3 [159].

$$\zeta = Ae^{-\frac{Ea}{RT}} n^{z}, \tag{4.3}$$

where ζ is the relative capacity loss of batteries with unit of %, A is a constant, Ea represents the activation energy [J mol-1]; R is the gas constant [J/(mol-1 K)]; T represents temperature [K]; n is the cycle numbers and z is the power law factor.
Battery aging studies encompass not only the examination of capacity and resistance changes over time but also the analysis of equivalent circuit parameters in both cycling and calendar part. A method for estimating State of Health (SOH) has been developed in [162]. This method exploits a simplified equivalent circuit model to parameterize a single-variable and time-based SOH inference model. Furthermore, additional model-based techniques consider the rise in internal resistance when

analyzing the status of battery aging. The precise quantification of this can be achieved through the utilization of electrochemical impedance spectroscopy (EIS) technology, as evidenced by previous studies [163–165]. From analytical methods view point, one commonly used approach for estimating battery ageing level is known as "coulomb counting." This method involves estimating the state of health (SOH) of the battery by integrating the current over time [166]. Regular calibration is necessary for this process, and it is not feasible to perform it in real-time [167]. A state observations-based model faced the aging estimation considering an input u state vector and an output voltage y and reported in Equation 4.4

$$\begin{cases} \hat{x} = Ax + Bu \\ y = Cx \end{cases} \tag{4.4}$$

The objective is thus to calibrate a model based on empirical data in order to minimize the error between the estimated variable $\hat{x}$ and y using a parameter known as the gain K. Presently, the utilization of this application for online purposes is rare and it is also a high-cost application. Several empirical data-driven models have been developed for the purpose of estimating SOH under complex aging conditions. These models, including NNs, have been identified as research hotspots due to their potential in accurately estimating SOH when sufficient data is available. The advantages of approximation and learning speed make NNs particularly effective in this regard [168, 169]. According to existing literature [170], the feed-forward neural networks (FFNN) [171–173], convolutional neural networks (CNNs) [174, 175], and recurrent long short-term memory (LSTM) [176, 177] are the most effective neural networks. Sungwoo Jo et al. (2021) [178] conducted a comparative analysis, revealing that LSTM outperformed the other two types in terms of performance. LSTM models may incur significant computational expenses and memory utilization owing to the memory cell's size and intricate architecture.

The BMS is responsible for identifying the SOH of the battery, in addition to performing various other tasks such as safety control, failure prevention, and optimization of energy consumption. In general, ARM processors with a 32-bit architecture (multicore) are commonly utilized in automotive boards and are capable of delivering sufficient processing capabilities [179]. As technological advancements continue to progress, it is anticipated that an increasing amount of data and tasks will require more storage and fulfillment [180]. The utilization of external cloud devices has been suggested as a potential resolution for this matter. However, it is contingent upon the presence of dependable and efficient internet connectivity [181]. The authors of reference [182] have presented an LSTM model for estimating the remaining useful life of a system. The model utilizes multichannel full charge profiles and has demonstrated notable enhancements over the baseline LSTM. Additionally, the model has significantly reduced the number of parameters required for its implementation. Nevertheless, the utilization of complete charge cycle information leads to a substantial requirement of memory and processing capacity. Consequently, the author in reference [183] proposes the utilization of a RNN-LSTM model for the prediction of Remaining Useful Life (RUL) through the analysis of partial charge information within the voltage domain range. The model is designed to establish boundary constraints. Nonetheless, a comprehensive investigation of the entire SOC domain has not been conducted, and the potential reduction in memory usage and computational expenses resulting from the modification of SOC window lengths for SOH assessment during charging remains uncertain. Hence, the primary focus of this research pertains to the evaluation of the SOH of a battery through the utilization of incomplete charging information and the modification of the SOC window length via a Bi-LSTM approach. The aim is to decrease the computational burden and memory consumption while preserving a high level of accuracy.

4.2 Material & Methods

The present investigation relies on the necessity to devise an algorithm for estimating the remaining battery life, which is associated with SOH, to be integrated into a forthcoming tool that simulates an entire BMS. The present study centers on the creation of an artificial intelligence (AI) algorithm that is computationally efficient and minimizes hardware memory usage. This is due to the limited computational capacity and memory of contemporary battery management systems.

Numerous aging experiments have been conducted in laboratory settings. Consequently, a substantial volume of literature exists that delineates the performance features of various Li-ion cell chemistries [184]. The data chosen for the purpose of cyclic aging analysis in the current study and in accordance with the stated objectives were sourced from Sandia National Laboratories [185]. The study of long-term degradation in cell operating conditions has prompted an investigation into the potential use of lithium 18650 nickel-manganese-cobalt (NMC) cells. The estimation of SOH in partial recharge conditions has been facilitated by the development of models that utilize bidirectional LSTM networks. These networks are a type of recurrent artificial neural network that can process both individual data points and entire data sequences. In this instance, various time sequences associated with charge-discharge experiments were inputted into these models. The accurate estimation of the state of health (SOH) is an essential task of the BMS utilized in the control and management of battery packs in electric vehicles. It is imperative that the SOH estimation be efficient, reliable, and resilient. Presently, the quantity of in-vehicle functionalities is consistently expanding, thereby augmenting the aggregate of necessary hardware memory. Hence, various models have been devised to estimate SOH in partial recharge scenarios, encompassing diverse SOC intervals, with the aim of achieving a lightweight yet effective algorithm. The optimal range of SOC for accurately estimating SOH has been determined. In this section, the proposed method composed by sequential steps is discussed and shown in Figure 4.1.

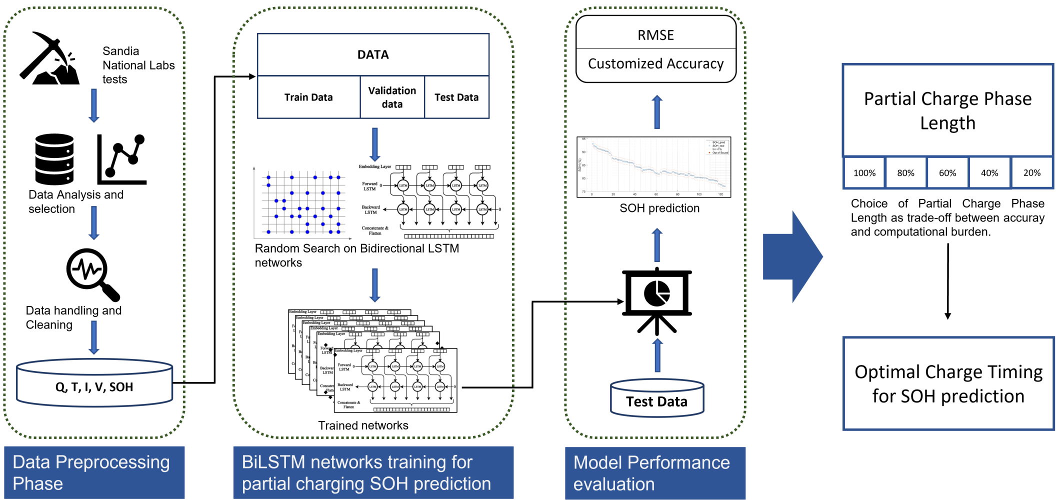


Fig. 4.1 Main steps of the applied methodology workflow in the SOH estimation definition.

During the phase of data processing, the acquired cell signals from cycle aging tests were subjected to analysis, handling, and cleaning. During the second step of training Bi-LSTM networks, the data was partitioned into separate training and validation datasets. These datasets were subsequently utilized to facilitate the learning process of multiple Bi-LSTM architectures. The utilization of the random search algorithm as a powerful hyperparameter tuning methodology facilitated the identification of the optimal network architecture in terms of accuracy. Grid search and random search are commonly employed hyperparameter optimization techniques for this objective. From a computational perspective, the latter approach facilitates the examination of a greater number of neural networks in order to select the optimal hyperparameters, thereby reducing the time required to identify them [186]. In the course of the learning process, the dataset is partitioned into training, validation, and test sets through a random process. This is done to facilitate the training and validation of the chosen artificial intelligence algorithms. The utilization of Bi-LSTM neural networks in this study was motivated by their remarkable proficiency and effectiveness in predicting time-series data and acquiring knowledge about the crucial pathways involved in the aging events of the cell cycle, as previously reported [187]. The Bi-directional Long Short-Term Memory (Bi-LSTM) neural network represents an expanded version of the conventional LSTM neural network. It comprises two LSTM networks that operate in both forward and backward directions to process data. The LSTM model incorporates a gate mechanism that facilitates the retention of extended temporal sequences of information within the memory. The utilization of Bi-LSTM models, which allow for bidirectional data processing during training, has been shown to yield superior performance and predictive capabilities compared to conventional LSTM-based models, as evidenced by prior research [188]. Similar to other AI models, the Bi-LSTM architecture is characterized by a collection of hyperparameters that necessitate specification to customize the model for the particular task at hand. In this study, the hyperparameters were tuned using the random search optimization technique [189]. The ultimate stage involved the assessment of the optimal Bi-LSTM network's performance by analyzing a test dataset using two distinct metrics, namely, RMSE and a customized regression accuracy (CRA) which will be defined in the next sections. Subsequently, the aforementioned approach was utilized to determine the optimal duration for partial charging, which enables precise and computationally efficient evaluation of the state of health (SOH) of the battery. The present study evaluated the optimal SOC window for accurately estimating the state of health (SOH) of a battery during a vehicle charging event.

4.2.1 Cycle aging data preprocessing phase

The current study focused on analyzing a cell of the 18,650 variety, featuring NMC chemistry on the cathode and graphite on the anode. The multi-channel battery testing system was utilized to conduct cycle aging tests. Additionally, the protocol for cycle aging has been documented by Sandia National Laboratories [185]. Tables 4.1 and 4.2 provide a summary of the test equipment and test operating conditions, respectively.

Table 4.1 Sandia National Laboratories equipment for cycle aging experiments.

Cell Type	**Cathode**	**Anode**	**Capacity (Ah)**	**Test Equipment**
18,650 NMC	NMC	graphite	3.00	High-precision Arbin

Table 4.2 Test operating conditions of cycle aging experiments.

Charge C-Rate	**Discharge C Rate**	**SOC Range**	**Environmental Temperature (°C)**	**N° Cycles**
0.50	2.00	0–100	25	661

In the Table 4.2 the *NCycles* is the number of charge and discharge cycles that a cell can process before it reaches its end-of-life condition corresponding to 20% capacity loss of a fresh cell capacity. The cell has been charged through a constant current constant voltage (CCCV) protocol, with 0.5 C current during CC phase and current taper to 0.05 A on CV. The NMC cell has been cycled from 2 to 4.2 V during all cycling tests for the whole SOC domain. A portion of the experimental test acquisition and the exploited CCCV protocol are reported in Figure 4.2.

The data acquisition system collected the following signals over time:

- Cycle index, number of charge–discharge cycle;
- Cell current [A];
- Cell voltage [V];
- Charge and discharge capacity [Ah];

- Charge and discharge energy [Wh];
- Cell temperature [°C];
- Environmental temperature [°C].

The SOH parameter was computed after the cell residual capacity has been determined at the end of each ith cycle by using Equation 4.5 [190],

$$SOH_i = \frac{Q_{actual,i}}{Q_{rated}}, \tag{4.5}$$

where $Q_{actual,i}$ is the capacity computed at the i_{th} charge–discharge cycle and Q_{rated} is the cell nominal capacity. Figure 4.3 depicts the trend of State of Health (SOH) as defined by residual capacity in accordance with the aforementioned equation, and pertains to the entirety of the aging test, encompassing all cycles.

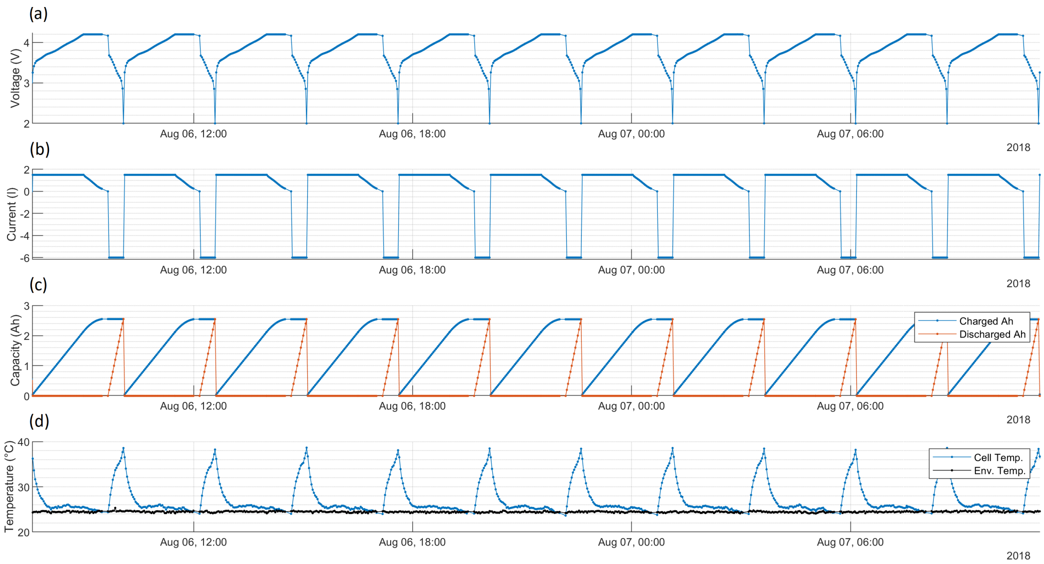


Fig. 4.2 Example of constant current constant voltage charge profiles measured by Sandia National Laboratories during the performed experimental cell aging campaign. The cell operating parameters measured during the tests are (**a**) cell voltge (V), (**b**) current (A), (**c**) charged and discharged capacity (Ah), and (**d**) cell and environmental temperature (°C).

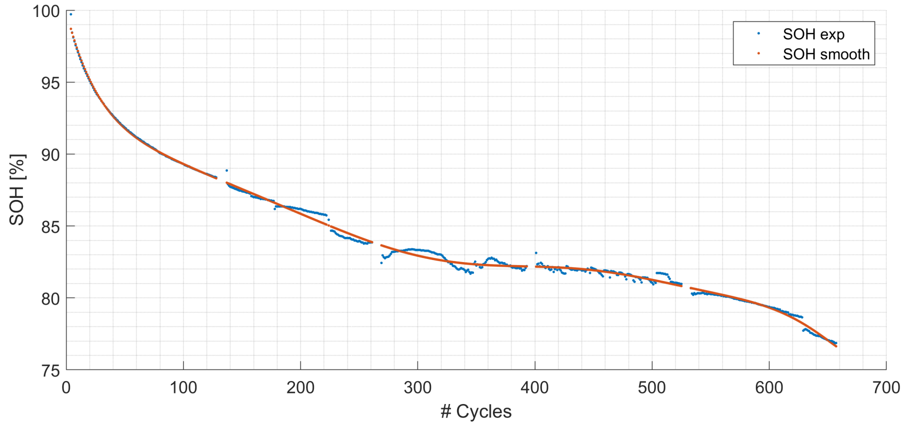


Fig. 4.3 SOH trend along entire aging cycle experimental campaign conducted by Sandia National Lab. The theoretical and smoothed SOH curves are reported.

The graphical representation illustrates that the SOH exhibits a declining pattern until it attains the end-of-life magnitude. In the automotive domain, this magnitude corresponds to approximately 75-80% of the initial cell state. Additionally, it is evident that a smoothed SOH curve was computed, as there were some rises in the capacity over the duration of the aging campaign. In the context of an automotive application, it is possible that the aforementioned statement may not hold true, as a battery pack is likely to undergo discharge-charge cycles within a comparatively brief period. Prior to utilization in AI algorithm training, the obtained data underwent preprocessing procedures to assess their robustness and quality. Illustrative instances encompass the detection and elimination of atypical traces, detected null values, and outliers values of output signals. Anomalous behaviors were observed in the current signal during the experimental tests when applying the CCCV charge-discharge protocol, as reported in the Figure A.2 and A.3 of A. The figure illustrates that sudden peaks were observed, resulting in the exclusion of reference cycles from the algorithm development analysis. Furthermore, the signals were intentionally truncated to enable the selective analysis of relevant data over time, thereby expediting the training process of the neural network for the respective case studies. Subsequently, the collected data underwent a process of resampling, whereby the frequency was standardized from a variable to a constant rate throughout the duration of the study. The data under consideration is interpolated using a sample time of 5 seconds. Ultimately, the acquired dataset encompassed a range of charging cycles spanning from initial cell conditions to their end-of-life state. Every iteration consisted of temporal signals for cell temperature, voltage, current, charged capacity, and the corresponding SOH value.

4.2.2 BiLSTM networks training for partial charging SOH Prediction

At the very beginning, the dataset was partitioned into training, validation, and test sets to facilitate the development of the AI algorithm. As previously expounded in Chapter 2, this partitioning is imperative during the data preprocessing stage for the purposes of learning and validation.

Figure 4.4 depicts the Bi-LSTM architectures utilized in the developed AI model, along with the corresponding hyperparameters investigated by the random-search optimization technique.

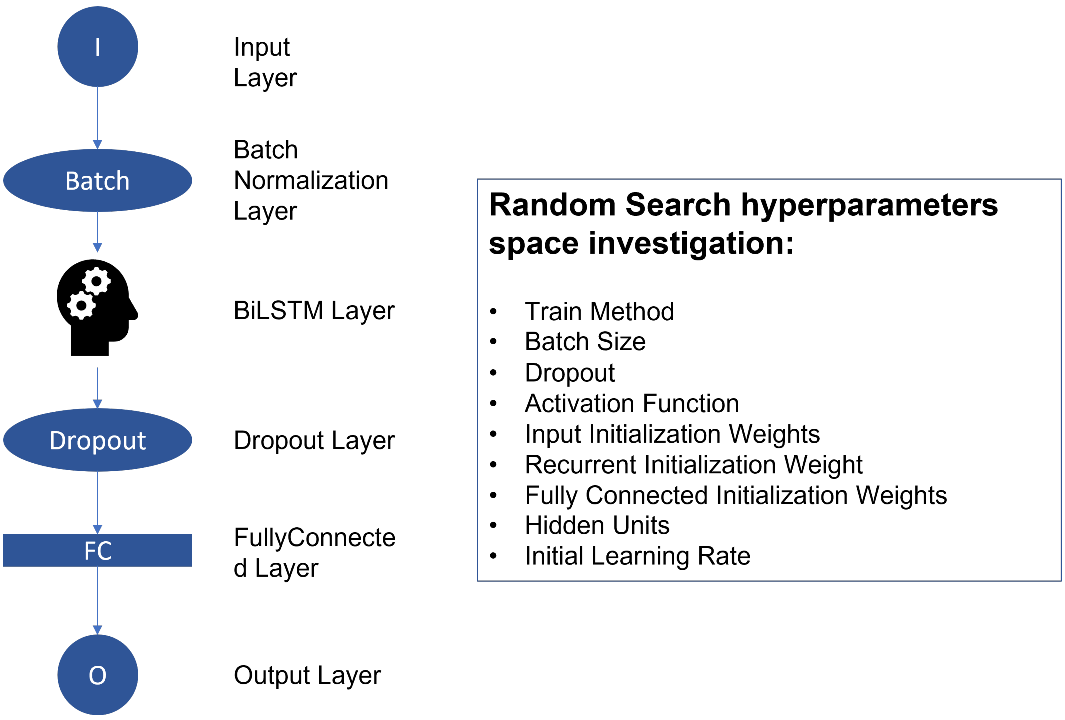


Fig. 4.4 AI model graph based on Bi-LSTM network and hyperparameters investigated by the random-search optimization technique.

The Bi-LSTM model was composed of an input layer which normalizes data with the z-score method and with the dimension of the input data, a batch normalization layer which normalizes a batch of data across all observations, and a Bi-LSTM layer for capturing long-term dependencies between cell parameters, and the predicted SOH value. The Bi-LSTM layer has been defined by the number of hidden units (hidden state, correspondent to the number of information remembered between time steps), the activation function to update the cell and hidden state, and the weights initialization. Moreover, dropout layer randomly drops out input elements and was included to mitigate overfitting [191]. The model also featured a fully connected layer for forecasting the SOH output of single dimension and an output layer, which was a regression layer responsible for computing the half-mean-squared-error loss function for the regression task [192]. Regarding the training process, it is noteworthy that the algorithm employed for optimization, as depicted in the box of Figure 4.4, is widely recognized as the predominant approach for optimizing neural networks.The publication [193] provides a comprehensive survey of various optimization techniques. The duration of the learning phase is determined by a finite number of epochs, indicating the number of complete iterations through the entire dataset.

The present investigation centered on the evaluation of SOH predicted on charge cycles, which represent instances of temporal sequences. The issue at hand pertains to sequence-to-one regression networks, with the corresponding loss function being the half-mean-squared-error as depicted in Equation 4.6,

$$Loss = \frac{1}{2}\sum_{i=1}^{N}\frac{(\hat{y}_i - y_i)^2}{N}, \tag{4.6}$$

where N is the number of responses, y_i is the target output, and $\hat{y}_i$ is the network's prediction for response i. Finally, an early stopping technique was applied when the performance of the validation phase started to degrade in order to avoid overfitting on training dataset [194].

Finally, for the sake of comprehensiveness, Table 4.3 displays the search space encompassing the optimal combination of hyperparameters. A total of 3000 combinations were examined, from which the optimal one was selected.

Table 4.3 The investigated domain values of Bi-LSTM hyperparameters.

Hyperparameters	Domain values
Train Method	"*adam*", "*rmsprop*", "*sgdm*"
Batch Size	$[4, 16, 32, 64, 128, 256]$
Dropout	from 0.1 to 0.5 with step 0.1
Activation Function	"*tanh*", "*softsign*"
Input Initialization Weights	"*glorot*", "*he*", "*orthogonal*", "*narrow − normal*", "*zeros*", "*ones*"
Recurrent Initialization Weight	"*glorot*", "*he*", "*orthogonal*", "*narrow − normal*", "*zeros*", "*ones*"
Fully Connected Initialization Weight	"*glorot*", "*he*", "*orthogonal*", "*narrow − normal*", "*zeros*", "*ones*"
Hidden Units	from 5 to 80
Initial Learning Rate	from 0.0001 to 0.01 with step 0.0001

A complete explanation of involved hyperparameters' values is provided in the [195], where some added papers as reference are listed and the tool has been exploited for the current work. For batch size, dropout and learning rate typical values, a deeper insight can be given to [196–198].

4.2.3 Model Performance Evaluation

The assessment of a model is a crucial stage in its development process. Certain techniques, such as the ANN, LSTM and Bi-LSTM models, conduct assessments during the execution of back-propagation. Nevertheless, the assessment of a model is still carried out manually using diverse techniques. It is noteworthy that the successful evaluation of models is a crucial aspect. Particularly in a supervised learning environment where the true values are accessible. These values facilitate the functioning of evaluation methods. Supervised learning models can be broadly categorized into two types: regression problems and classification problems. Furthermore, the techniques employed for assessing these models can be classified into the aforementioned two categories exclusively. The process of evaluating a model is a crucial component in the development of said model. Identifying the optimal model that accurately represents the data is a crucial step in data analysis. The analysis also centers on the future performance of the selected model. Assessing the efficacy of a model using training data is deemed inappropriate in the field of data science. The generation of overoptimistic and overfit models can occur with ease.

Hence, in the present work the selection of each model was based on the consideration of various performance metrics pertaining to the prediction outcomes of SOH in a regression environment. The performance of all Bi-LSTM architectures that were trained was analyzed based on test data and the following metrics:

- the RMSE considering the test dataset,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} x_i - \hat{x}_i^2}{n}} \tag{4.7}$$

- the coefficient of determination R^2,

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2} \tag{4.8}$$

- the customized regression accuracy (CRA) coefficient, which compared the predicted SOH, $\hat{SOH}$, with the corresponding measured value, SOH through an identified threshold thr. The CRA details are provided in the next section.

The Results and Discussion section extensively examines and analyzes the performance of each neural networks model and the relative optimal selection of hyperparameter values.

4.2.4 Different SOC windows identification during partial charging events

The duration necessary to completely or partially recharge the battery unit of an electric vehicle is a pivotal concern for the majority of drivers. The duration of the battery pack charging process may vary depending on the charging power provided by the grid, potentially lasting for several hours. Consequently, the present endeavor is centered on ascertaining the optimal length of partial charging that strikes a balance between precision and computational expenditure for the purpose of on-board state of health estimation. Limitations in memory usage and data storage capacity are a commonly acknowledged challenge in contemporary on-board control units utilized in passenger vehicles. Thus, a potential solution to this issue could be to decrease the amount of data recorded on the BMS. The memory and computational cost for on-board data processing can be reduced by minimizing the length of partial charging data logged over time. Moreover, the sampling rate of the dataset is a significant factor in reducing memory usage. The sample rate for this activity has been established at 0.2 Hz owing to the signal of interest's limited dynamic range.

The methodology expounded in the preceding section was devised to compute the state of health (SOH) of the cell by exclusively utilizing a subset of the information pertaining to the complete battery charging procedure. Prior to executing the train-test split procedure on the data, the preprocessing stage involved the consideration of different durations of partial charging segments over time in the initial test scenario. To account for fixed segments within the 0-100% state of charge (SOC) range, the duration of time under investigation was established as a percentage. It is important to acknowledge that as a cell ages, the duration of data logging for a specific percentage of the state of charge (SOC) window decreases due to capacity degradation, as illustrated in the accompanying figure 4.5.

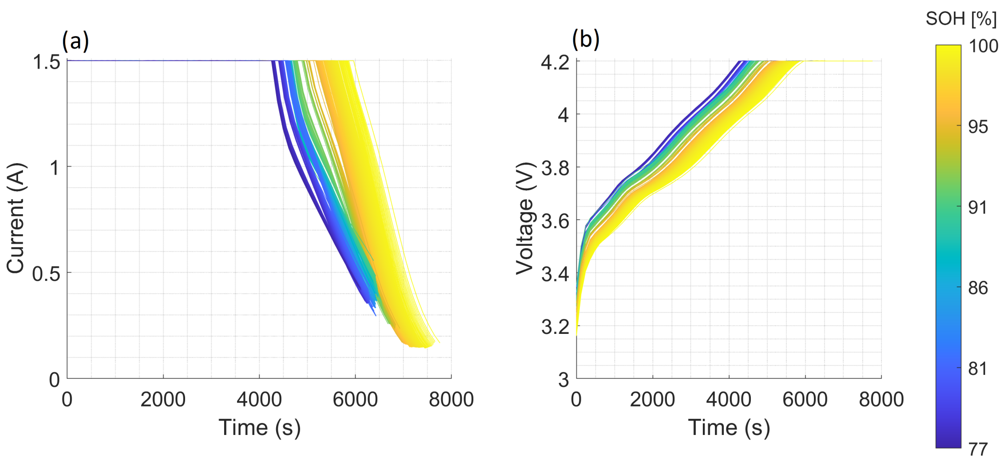


Fig. 4.5 The acquired (**a**) current (A) and (**b**) voltage (V) are plotted as time series for each independent charging cycle with SOH reference values. The lower the SOH value, the faster the maximum voltage value is reached. This translates into a smaller amount of energy injected into the cell, as the charging current is constant.

It is evident that a decrease in the state of health (SOH) of a single cell results in shorter recharge curves. This is attributed to the reduced capacity of the cell to undergo recharging.

In addition to complete recharge data, multiple time series of partial recharge data were acquired as proportions of the entire SOC domain. The SOC intervals that were taken into account had the dimensions of:

- 80%
- 60%
- 40%
- 20%.

Partial charge segments were gathered for each individual cycle over a period of time, with a starting point chosen at random. To ensure that the segments aligned with the entire length of the data, a specific area of points was randomly selected as the cut point. If we denote by n_k the length of the k-th cycle in terms of sample data over the entire state of charge (SOC) domain, and by L the selected length as the partial charging size related to the specific SOC window, then the cut space S from which the segment starting point was randomly selected can be defined by the Equation 4.9.

$$0 \leq S \leq n_k - L. \tag{4.9}$$

Figure 4.6 showcases instances of partial charging segments that have been retained, corresponding to SOC ranges of 80%, 60%, 40%, and 20% in relation to the complete SOC window.

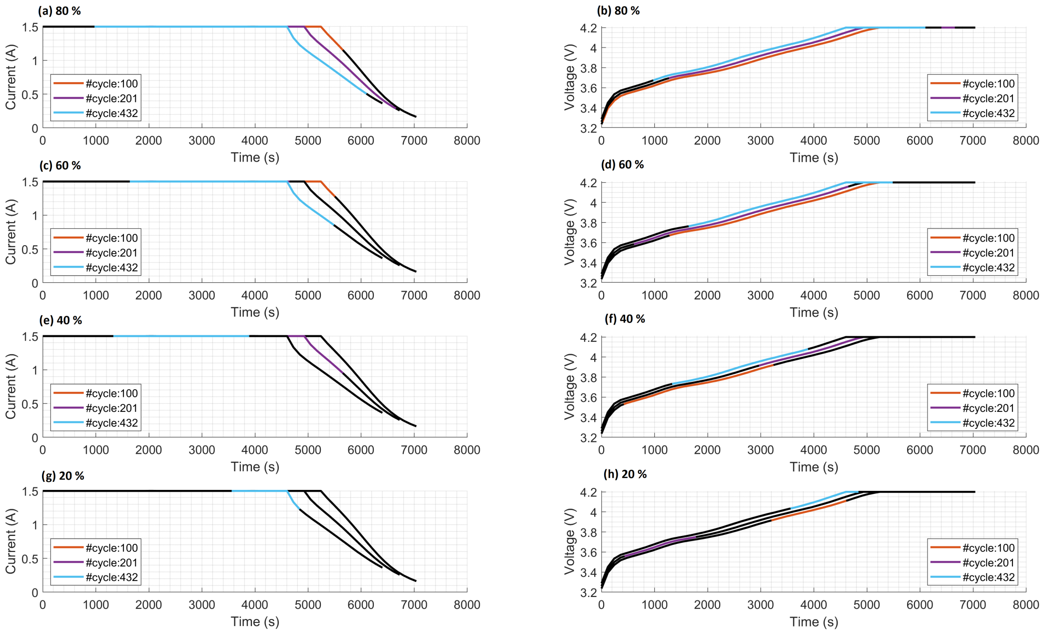


Fig. 4.6 Operating current and voltage of few partial charging segments. (**a**,**b**) are operating current and voltage of the partial charging length equal to 80%. (**c**,**d**) are operating current and voltage of the partial charging length equal to 60%. (**e**,**f**) are operating current and voltage of the partial charging length equal to 40%. (**g**,**h**) are operating current and voltage of the partial charging length equal to 20%. In the graphs, the black lines is the full size of data (equal to 100% of length).

To facilitate the development of knowledge pertaining to Bi-LSTM neural networks, the dataset corresponding to each partial charge segment was partitioned into training, validation, and test data in a randomized manner, as documented in reference [199]. The study utilized an 80-20% partitioning scheme, whereby the training and validation dataset were allocated 80% of the data, while the remaining 20% was assigned to the test dataset. The features encompassed the operational parameters of the cell, including voltage (V), current (A), charged capacity over time (Ah), and temperature (°C). The objective of the study was to predict the SOH values using the developed model.

4.2.5 Optimal partial charging SOC length and optimal SOC Window identification for SOH estimation

This section trained Bi-LSTM networks and determined the best network architectures for each charge length under consideration. The investigation's goal was to identify the ideal SOC range for calculating a cell's remaining life while it was being charged. In order to analyze the ideal SOC window for the SOH estimation, the ideal partial charging length L_{opt} has to be determined. For the on-board SOH estimation by control units of Li-ion battery packs, a trade-off between prediction accuracy (RMSE, CRA), computational cost, and memory consumption was actually examined.

The top 1, 5, and 10 Bi-LSTM networks have generated a sensitivity analysis that takes into account the RMSE and CRA for each charge length. While keeping in mind the time needed to execute the numerical models for the cell SOH estimate, the computational expenses for each charging length taken into consideration were examined. Additionally, the memory storage capacity was examined in light of the data logged and the memory consumed by the models. Finally, the ideal input length based on trade-offs was identified and used for the best SOC window analysis. In the Results and Discussion section, sensitivity and trade-off analyses will be fully explained.

After determining the ideal partial length L_{opt}, data was cut from the full-size data of cycles at various starting points to create a new dataset. To assure the contiguous size of the data, the cut points were specified at each 10% step in the SOC window until reaching the last point, while adhering to the 0% and 100% SOC limits. Hence, the SOC windows SOC_{win} among the entire domain are shown in Equation 4.10) and expressed as a percentage of size data of a single charging cycle:

$$SOC_{L_{opt}} = [0; L_{opt}], [10; L_{opt} + 10], ..., [100 - L_{opt}; 100]. \tag{4.10}$$

For instance, if the optimal length L_{opt} was observed to be 40%, then the $SOC_{L_{opt}}$ is reported in Equation 4.11:

$$SOC_{40} = [0; 40], [10; 50], [20; 60], [30; 70], [40; 80], [50; 90], [60; 100]. \tag{4.11}$$

Considering the example of dataset shown in Figure 4.6 for L_{opt} equal to 40%, the related charging data are illustrated in Figure 4.7.

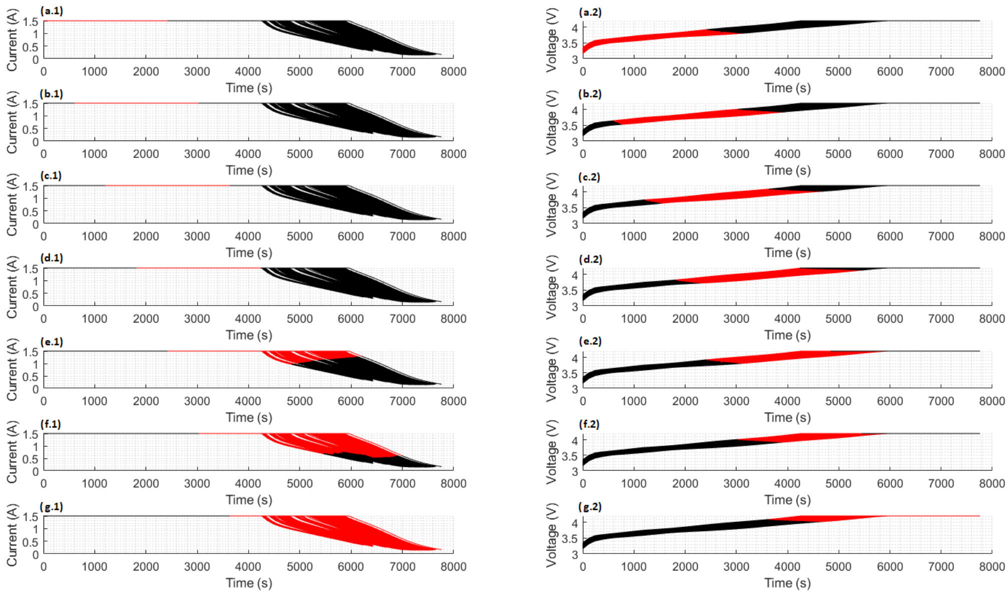


Fig. 4.7 Fixed SOC window equal to 40% moving over the entire domain, for the generation of datasets. (**a.1**,**a.2**) are, respectively, current and voltage of SOC window [0,40]. (**b.1**,**b.2**) are, respectively, current and voltage of SOC window [10,50]. (**c.1**,**c.2**) are, respectively, current and voltage of SOC window [20,60]. (**d.1**,**d.2**) are, respectively, current and voltage of SOC window [30,70]. (**e.1**,**e.2**) are, respectively, current and voltage of SOC window [40,80]. (**f.1**,**f.2**) are, respectively, current and voltage of the SOC window [50,90]. (**g.1**,**g.2**) are, respectively, current and voltage of SOC window [60,100]. In the graph, curves for the entire cycles are plotted in black.

Once L_{opt} was identified in this study, the same data split between training, validation, and test was maintained. As indicated in the Results and Discussion section, this made it possible to carefully analyze the conclusions of the choice of the threshold value *thr* used in the definition of the CRA. The top 30 Bi-LSTM-trained networks from the previous section were employed for a new learning process in terms of neural network learning. The target variable, the feature definitions, and the regression task, however, were all the same. The best SOC range for estimating capacity degradation during charging events was then established.

4.3 Results & Discussions

Prior to delving deeper into the numerical results, it is necessary to present a comprehensive overview of the performance metrics that were utilized to analyze the outcomes. The evaluation of the trained models was conducted by considering their ability to accurately estimate values on the test dataset. The evaluation of the model outputs was conducted through a comparison with actual targets, utilizing the absolute error metric as defined in Equation 4.12.

$$E_i = x_i - \hat{x}_i \tag{4.12}$$

The primary parameters taken into account for assessing the caliber of the model forecasts were the *RMSE* as stated in Equation 4.7 and the coefficient of determination R^2, shown in Equation 4.8. Additionally, the efficacy of the forecasts was assessed using a precision measure specifically developed for evaluating outcomes in the domain of regression and named in the present work as Customized Regression Accuracy (*CRA*). The final parameter was designed to assess accuracy by treating the problem as a classification task, wherein the algorithm's output is deemed correct if the absolute error value is less than a predetermined threshold. The aforementioned expression is presented as Equation 4.13

$$CRA = \frac{\sum_{i=1}^{n} T_i}{n} \times 100 \ with: \begin{cases} T_i = 1 & if \, |E_i| < threshold \\ T_i = 0 & if \, |E_i| > threshold \end{cases} \tag{4.13}$$

x_i was the target value, $\hat{x}_i$ was the model output, $\bar{x}_i$ was the mean of the dataset label considering that each experimental test in the dataset had a number of samples equal to n, and E_i is the residual between target and predicted values. The equation demonstrates that the predicted SOH values were accurately classified by establishing a threshold value, thr. Specifically, if the prediction error was below the threshold value, the SOH values were classified as correct ($T_i = 1$); otherwise, they were classified as incorrect ($T_i = 0$). All the results shown in this section are derived from the validation of the model on the testing data.

A sensitivity analysis was performed on the parameter "thr" to assess the precision of the prediction model, as illustrated in Figure 4.8.

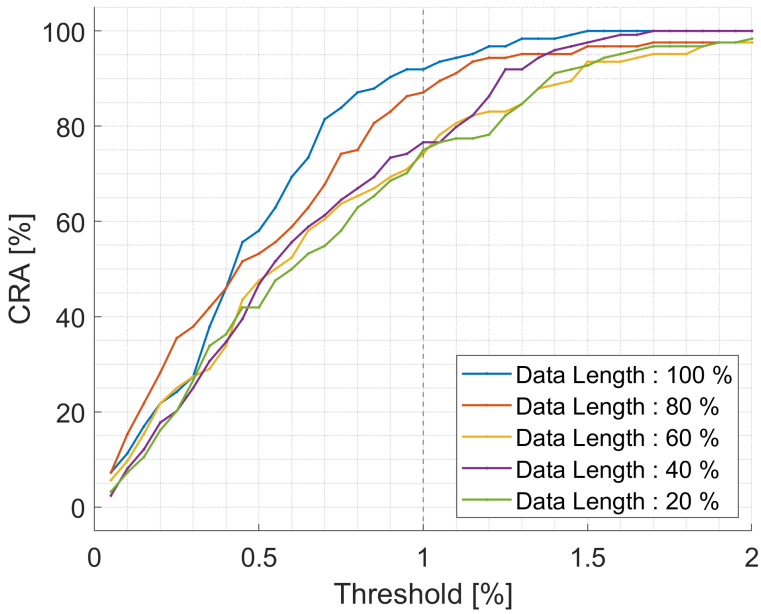


Fig. 4.8 Sensitivity analysis of the neural network's accuracy depending on the threshold value and for each SOC window data length.

In order to perform sensitivity analysis, the optimal neural network was identified for each charging data length through the minimization of the RMSE metric. Upon examining Figure 4.8, it is evident that the charging length equivalent to 40% exhibits the most rapid growth and is the sole length that attains 100% accuracy among the partial charging data lengths, within the threshold variability range of 0.1% - 2%. The threshold in this context denotes the degree of tolerance pertaining to the estimated state of health of the cell in comparison to the corresponding measurements. Regarding the sensitivity analysis, a threshold of approximately 1% was selected. The consistency of the 1% tolerance value with the literature [200] is evidenced by its ability to restrict the error in cell SOH estimation to a maximum of 2.2%. At the chosen threshold level and with a partial length of 40%, the CRA achieves a nearly 80% success rate. The findings depicted in Figure 4.8 were obtained through analysis #1 of 4.3.1, wherein the Bi-LSTM training dataset was generated by randomly selecting multiple segments across the SOC spectrum. The aim was to explore the minimum charging duration that ensures satisfactory precision in estimating the cell SOH. The comparative study reveals that analysis #1 of 4.3.1 exhibits a significantly lower level of accuracy in contrast to analysis #3 of 4.3.3. The latter approach involves training each network with data from a specific SOC window, as a means of determining the optimal SOC charging window for SOH estimation.

Therefore, in the subsequent analysis, the predictive outcomes exhibit a significant improvement, and the threshold value can be substantially reduced. The accuracy values presented in this section are reported solely for the purpose of analyzing the sensitivity of the *thr* value and justifying the choices made. Subsequent sections contain additional information and evaluations regarding the obtained predictive performances, along with corresponding analyses.

4.3.1 Analysis #1 : SOC windows length identification for partial charging events

Following the explication of the metrics implicated in the evaluation of the accuracy of SOH estimation, this section centers on examining the impact of individual partial charging lengths on the prediction of cell SOH. Figure 4.9 displays the prediction outcomes of the trained model as measured by *CRA* and *RMSE*. To determine the most effective combination of hyperparameters for the various Bi-LSTM models, the random-search algorithm was utilized. This involved exploring a broad range of parameters, resulting in the examination of 3000 models with varying hyperparameter combinations.

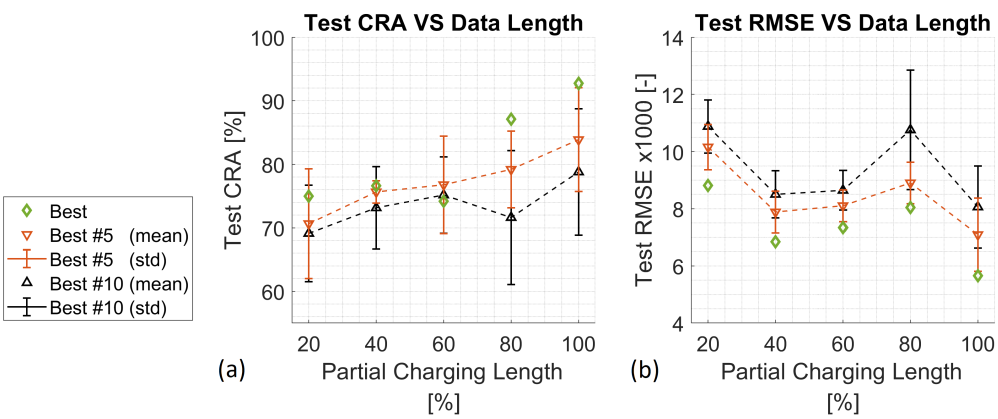


Fig. 4.9 (**a**) CRA and (**b**) RMSE results of the best one, five and 10 trained neural networks according to the RMSE on testing dataset. For partial charging lengths, the minimum RMSE is equal to 0.0068 corresponding to 40% data length.

The study presents the Bi-LSTM networks that were trained at the top one, five, and 10 positions for each analyzed percentage of charging length, based on their performance standard deviations. The general tendency CRA exhibits a positive correlation with the extent of the partial charge taken into account. By extending the charging duration from 20% to 40%, there is an observed increase of approximately 7% in the test CRA of the top five networks. In terms of $RMSE$, it can be observed that while the boundary cases of 20% and 100% represent the worst and best options, respectively, the trend of the top five networks undergoes a change. As an AI model that relies on data, a substantial quantity of observations is necessary to identify and discern particular patterns, particularly when utilizing random processes for the purpose of generalization. The analysis of Figure 4.9 reveals that a network configuration utilizing 40% of the SOC window as input during the charge phase can achieve an accuracy level of approximately 77%. The term "accuracy" is defined in relation to the CRA, where a threshold parameter value of 1% is utilized. Considering the customized characteristic of this metric and the stochasticity linked to the choice of hyperparameters and training data for Analysis #1, a CRA score of 77% is sub-optimal (superior values are attained for Analysis #3). Nevertheless, the utilization of a 40% data length yields highly favorable RMSE and R2 scores in comparison to existing literature. In addition, it is noteworthy that the 40% data length scenario exhibits the lowest RMSE among all cases, except for the 100% length scenario which pertains to the entire SOC domain. It can be observed that the random process approach did not yield an optimal Bi-LSTM configuration for each charging length case. However, it is plausible that such an optimal configuration could be achieved through multiple additional iterations. The analysis yielded a noteworthy finding indicating that the SOH of a cell can be accurately detected by monitoring only 40% of the entire 0-100% SOC charging process.

Figure 4.10 displays the performance outcomes of the regression analysis conducted on the estimation of state of health for each case study, which corresponds to the percentage length of the SOC window. The results are associated with the optimal Bi-LSTM network found by the random search approach.

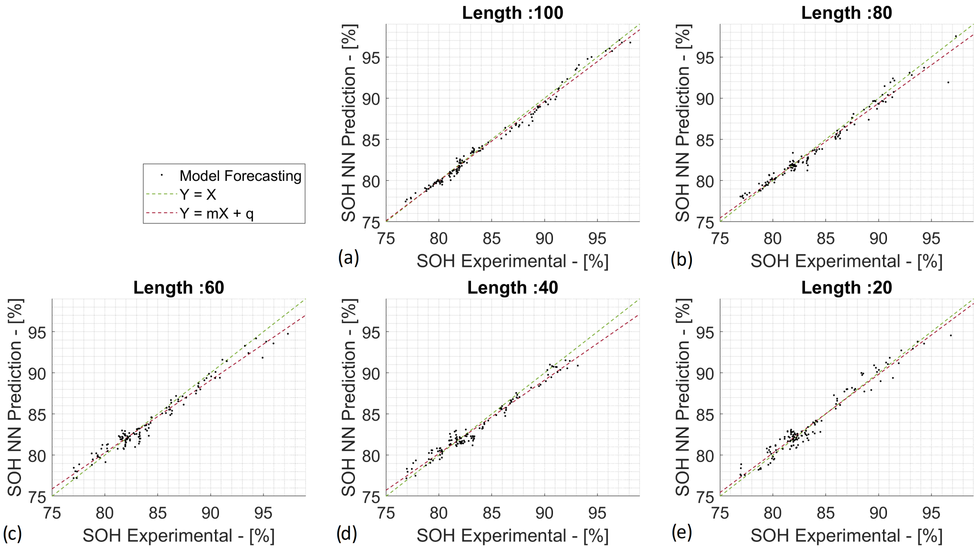


Fig. 4.10 Best neural network regression performance. (**a**) refers to the full charging length equal to 100%, (**b**) refers to the partial charging length equal to 80%, (**c**) refers to the partial charging length equal to 60%, (**d**) refers to the partial charging length equal to 40%, (**e**) refers to the partial charging length equal to 20%.

The correlation points between predicted and target values are represented by the black points in the figure. The bisector, denoted by a green dashed line, and the fitting regression line, represented by a red dashed line, are both present in the given context. Table 4.4 presents the findings pertaining to the RMSE and R2, as well as the regression line parameters, for the optimal neural networks, in conjunction with the two preceding figures. The predicted SOH data points throughout the complete cycle aging examination are representative of the test set utilized for validating performance.

Table 4.4 Analysis #1: Best neural network regression statistics.

Data Length [% of SOC]	100	80	60	40	20
m	2.67	5.82	9.92	8.81	3.68
q	0.97	0.93	0.88	0.89	0.96
Test RMSE $\times$ 1000	5.65	8.04	7.33	6.80	8.93
Test R^2	0.99	0.97	0.96	0.96	0.96

The R2 metric results are similar across all case studies. However, the SOC percentages pertaining to partial recharges indicate that the 40% scenario yields the lowest RMSE value. This suggests that it is feasible to achieve exceptional performance even with limited data availability. Table 4.5 presents the optimal neural networks identified by the optimization technique in the context of the learning process of Bi-LSTM models. The table presents a comprehensive summary of the Bi-LSTM architecture configurations achieved in different scenarios of partial and full recharge. This enables the replication of the achieved performances with clarity.

Table 4.5 Analysis #1: Best trained neural network and relative architecture hyperparameters.

Data Length [% of SOC]	**100**	**80**	**60**	**40**	**20**
Hidden Layers	1	1	1	1	1
Hidden Neurons	59	30	29	47	52
State Activation Function	tanh	tanh	tanh	tanh	softsign
DropOut	0.2	0.3	0.2	0.1	0.1
Batch Size	128	32	64	32	64
Learning Rate	0.0090	0.0089	0.0060	0.0069	0.0044
Optimization Algorithm	sgdm	sgdm	adam	sgdm	adam
Training Epochs	190	84	108	264	186

As a result of optimization approaches that aim to minimize a loss function, the process of determining an optimal (or optimal-like) set of hyperparameters that defines the network architecture is derived from these techniques. As a consequence, we have obtained values for each architecture of the Bi-LSTM that are sometimes even significantly different because of the broad defining space domain for each hyperparameter. In practical terms, as far as the training epochs are concerned, the optimization method has stopped after just 84 or even more than 250 epochs. This is because the process was terminated during the learning phase when the loss function did not continue to show significant improvements, allowing the computing time to be significantly decreased. Figure A.4 in the Appendix A summarizes the learning phases of each example studied in terms of loss function trend.

4.3.2 Analysis #2: Computational cost and memory occupancy for the best SOC window length identification

The on-board implementation of a SOH estimator for lithium-ion cells in automotive applications poses significant challenges with respect to computational power, cost, and memory occupancy. These challenges are attributed to various factors. The automotive industry necessitates swift and effective real-time processing capabilities in its control units to manage time-sensitive duties, including the supervision and regulation of diverse vehicular systems. The accurate and timely estimation of battery health is a crucial aspect of the SOH estimator, which necessitates the use of intricate calculations and analysis of battery data. Hence, it is imperative to possess adequate computational capacity to execute these computations within the stipulated time limitations. Automotive control units are subject to resource constraints, which result in limited computational capabilities when compared to those of desktop computers or servers. The constraint in question arises from considerations pertaining to expenses, dimensions, energy usage, and thermal dissipation. The aforementioned limitations necessitate the optimization of computational algorithms and models employed in the SOH estimator to reduce computational load while upholding satisfactory precision. Additionally, the SOH estimator mandates data storage for past battery measurements and memory allocation for trained models or algorithms. The memory capacity of automotive control units is frequently constrained, and the acquisition of supplementary memory may prove to be costly or subject to limitations. Hence, it is imperative to devise effective methodologies for storing data and creating concise models that can be contained within the confines of restricted memory capacity.

This study aimed to evaluate the effectiveness of profiling analysis as a performance metric in conjunction with *CRA* and *RMSE* for determining the optimal SOC window length for monitoring capacity fade. Therefore, a profiling methodology was devised to measure the advantages of the suggested technique in relation to computational expenses and memory consumption. Figure 4.11 illustrates the computational performance necessary for the processing phase of the electronic control unit. The computation of the elapsed time depicted in Figure 4.11 involved the consideration of the mean duration for 10 iterations across the top 30 neural networks for every input length of the charging dataset. The observed trend in the duration appears to follow a nearly linear pattern. The processing time was computed using a laptop equipped with an Intel(R) Core (TM) i7-10510U CPU @ 1.80 GHz and 16 GB of RAM. A linear relationship between input charging length and memory occupancy of stored data is evident from the plotted figure. As the duration of the time series analyzed during the processing phase increases, there is a corresponding increase in the amount of memory space required. The memory allocation for the Bi-LSTM network size was determined by calculating the necessary storage capacity for the top 10 neural network architectures for each data length.

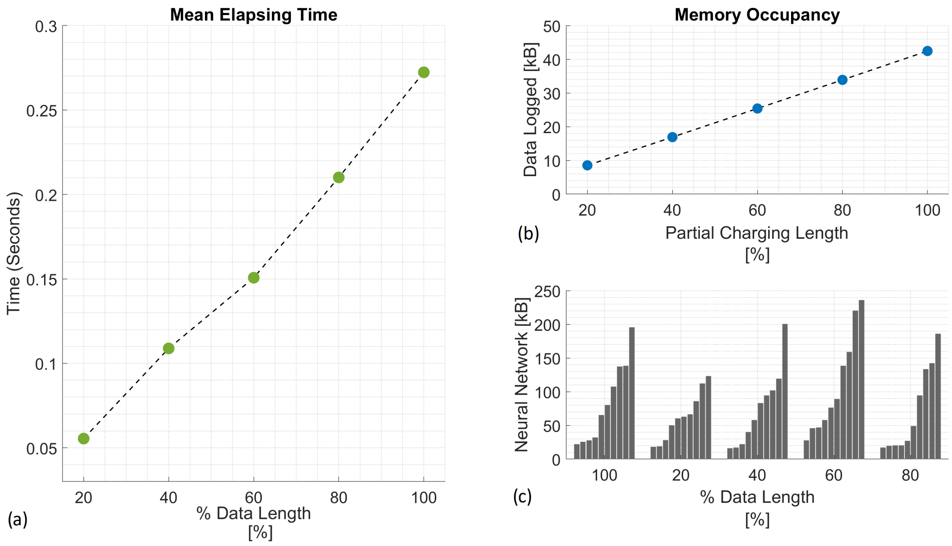


Fig. 4.11 The plotted values are referred to the results shown in Analysis #1. (**a**) Computational cost of Bi-LSTM prediction on different dataset length. (**b**) Memory occupancy by logging different SOC window length data. (**c**) Memory occupancy by Bi-LSTM network architectures.

The primary contributor to the decrease in memory usage is attributed to the reduction in dataset size, which exhibits a clear linear relationship. Conversely, the dimensions of the Bi-LSTM model exhibit fluctuations within the identical spectrum (ranging from 20 kB to 230 kB) across all input data lengths. The memory utilization of Bi-LSTM networks in Matlab is subject to variability based on multiple factors such as network dimensions, parameter count, and implementation particulars. In a broad sense, it is expected that the memory utilization would increase proportionally with the magnitude of the networks. The predominant factor that influences the memory consumption of a neural network is its parameter count, encompassing the network's weights and biases. The memory usage of a network is contingent upon the specific parameters associated with that network, thereby rendering it variable depending on the size of the network in question. If the architectures and dimensions of the networks exhibit similarity, the memory utilization may not demonstrate substantial variation among them. By conducting a cursory examination of the architectures of the top-performing networks presented in Table 4.4, it is evident that each network comprises a single hidden layer and a similar number of neurons. Consequently, the selection of an appropriate SOC window length is independent of the model dimension in this scenario.

4.3.3 Analysis #3 : Best SOC window identification for optimal SOH estimation

Based on the findings of Analysis #1 and Analysis #2, it can be inferred that the SOC window length L_{opt} that offers the optimal balance between computational efficiency and accurate estimation of cell SOH is the configuration that utilizes 40% of the input SOC data length. Thus, the current section is centered on the aforementioned charging length parameter. This section of the study explores the aspect of the charging process that provides more comprehensive insights into the SOH of the battery, thereby enabling a more accurate prediction of the battery's residual lifespan. Figure 4.12 displays the analysis outcomes for $L_{opt} = 40\%$. As shown in figure, higher performance in terms of CRA and RMSE are obtained due to an accurate reporting of the partial charging start points. The analyses presented indicate that the final charging state of charge (SOC) interval, spanning from 60% to 100%, yields the maximum *CRA* and the minimum *RMSE* in estimating the state of health of the cell. Therefore, the aforementioned range is deemed as the optimal SOC window for estimating the state of health of the cell. With reference to the data representation depicted in Figure 4.7, it can be inferred that the charging process's constant voltage (CV) phase aligns with the optimal state of charge (SOC) charge range between 60% and 100%. The CCCV tests are commonly utilized in the estimation of cell state of health and evaluation of battery performance during the process of aging [201]. Empirical evidence suggests that the partial CV charging phase is the optimal method for estimating SOH due to its ability to provide a greater amount of information and robustness in relation to capacity degradation [202]. In contrast to examining a single CV trace, the current study undertook a comparison and analysis of various partial charging lengths across the complete charging process spectrum. Regarding the optimal Bi-LSTM neural network, it is noteworthy to emphasize that the remaining sectors of the domain exhibit noteworthy outcomes, as evidenced by a CRA precision that consistently approximates 90%. Additionally, it is notable that the RMSE of the 0-40% trace exhibits a modest rise of 0.0054 in comparison to the optimal value of 0.0012 observed in the 60-100% trace.

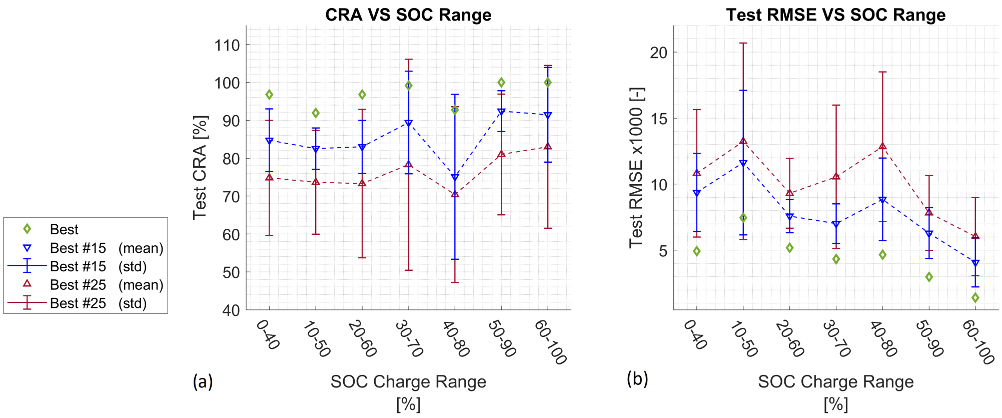


Fig. 4.12 Sensitivity analysis of the best 15 and 25 neural networks according to the RMSE on the testing dataset for $L_{opt} = 40\%$. (**a**) CRA trend depending on the input SOC window selected. (**b**) Test RMSE $\times 1000$ depending on the input SOC window selected.

The regression outcomes for the optimal Bi-LSTM network for each SOC charge range were depicted in Figure 4.13. It is evident that the range of 60–100% exhibits the least deviation in the regression line by bisector, and the predicted points are closely clustered on the bisector.

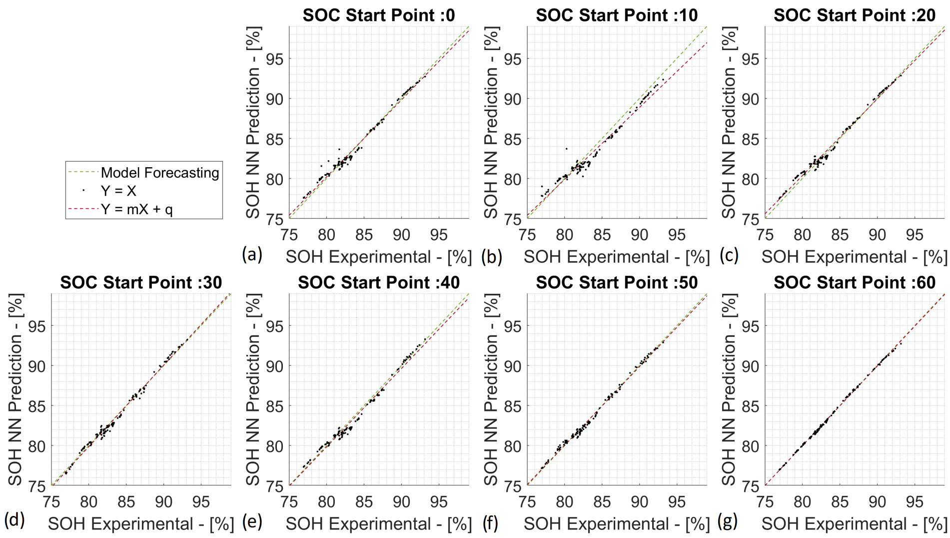


Fig. 4.13 Best neural network regression performance. (**a**) refers to the SOC window [0,40], (**b**) refers to the SOC window [10,50], (**c**) refers to the SOC window [20,60], (**d**) refers to the SOC window [30,70], (**e**) refers to the SOC window [40,80], (**f**) refers to the SOC window [50,90], (**g**) refers to the SOC window [60,100].

The diagram illustrates the correlation points between predicted and target values, with the black points serving as the visual representation of this relationship. The bisector, denoted by a green dashed line, and the fitting regression line, indicated by a red dashed line, are present in the given context. In a nutshell Tables 4.6 and 4.7 displays the effectiveness of the proposed models' validation through the achieved fitting regression parameters. The estimated SOH points during the course of the cycle aging test correspond to the tests used to validate performance.

Table 4.6 Analysis #3: Best neural networks regression parameters statistic of first four fixed SOC length windows over the entire SOC domain.

Charge Segment SOC [%]	**0–40**	**10–50**	**20–60**	**30–70**
m	0.96	0.90	0.96	1.02
q	3.29	7.99	3.82	−1.59
Test RMSE ×1000	5.45	7.52	5.23	4.09
Test R^2	0.98	0.95	0.98	0.99

Table 4.7 Analysis #3: Best neural networks regression parameters statistic of last three fixed SOC length windows over the entire SOC domain.

Charge Segment SOC [%]	**40–80**	**50–90**	**60–100**
m	0.98	0.98	0.99
q	1.34	1.35	0.30
Test RMSE ×1000	4.81	3.12	1.27
Test R^2	0.98	0.99	0.99

Similar to the preceding section, the Table 4.8 presents the specifics of the optimal BiLSTM structures for each individual segment of the SOC field. It is noteworthy to state that the lengths under consideration are equivalent to 40% of the SOC and span the entire domain with increments of 10 from the starting point.

Table 4.8 Analysis #3: Best neural network architectures for the 40% SOC length across the domain span.

Charge Segment SOC [%]	0–40	10–50	20–60	30–70	40–80	50–90	60–100
Hidden Layers	1	1	1	1	1	1	1
Hidden Neurons	15	15	65	54	42	51	51
State Activation Function	tanh	tanh	softsign	softsign	tanh	softsign	softsign
DropOut	0.3	0.3	0.3	0.2	0.5	0.5	0.5
Batch Size	16	16	32	64	64	32	32
Learning Rate	0.0098	0.0098	0.0055	0.0099	0.0086	0.0083	0.0083
Optimization Algorithm	sgdm	sgdm	rmsprop	sgdm	sgdm	sgdm	sgdm
Training Epochs	92	30	43	155	103	91	62

4.3.4 Optimal SOH estimation: performance analysis on the best charge segment of SOC

This concluding section presents a detailed analysis of the training and validation performance of the optimal Bi-LSTM network, with a focus on the most effective SOC window, ranging from 60% to 100%. As already explained, the performance in terms of *CRA* and *RMSE* of this trained network with the homogeneous dataset are much more analytically compared with Figures 4.9 and 4.10 from Analysis #1. The statistical metrics and Bi-LSTM architecture details are presented in Tables 4.6, 4.7 and 4.8. Figure 4.14 presents the training history across epochs, depicting a comparison of the loss function between the training dataset and validation dataset. The observed trend in the learning process appears to align with favorable fit outcomes, thereby precluding the possibilities of overfitting or underfitting during the training phase. Figure 4.15 depicts the predicted points of the SOH for the test set across all aging cycles. The portrayed outcomes in the figure demonstrate a favorable level of effectiveness in the estimation of SOH, with a maximum CRA of 100% and a minimal residual error for each cycle projection, indicating a prediction uncertainty of no more than 1%. Due to the robust predictive capabilities, it is possible to attain a comparable level of precision, with an accuracy rating of 100%, in a region of uncertainty (threshold) that has been reduced to 0.4% through an examination of the residual depicted in the figure. The definition of the percentage by which a model estimation may fall within a given uncertainty range of the target is provided by the CRA for the purpose of enhancing clarity. The threshold parameter value is indicative of the amplitude of the range.

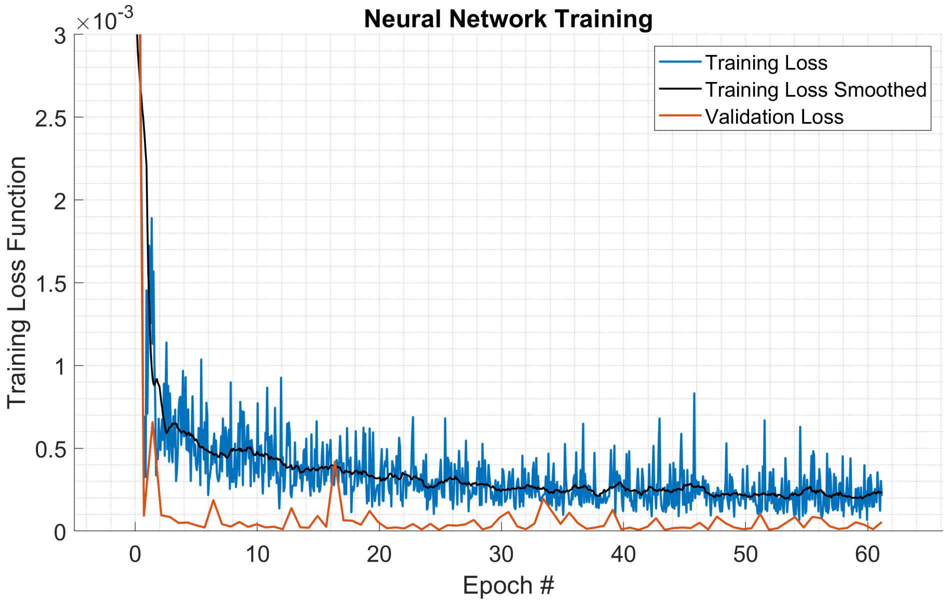


Fig. 4.14 Loss function trend of the best SOC window-related Bi-LSTM architecture during training process.

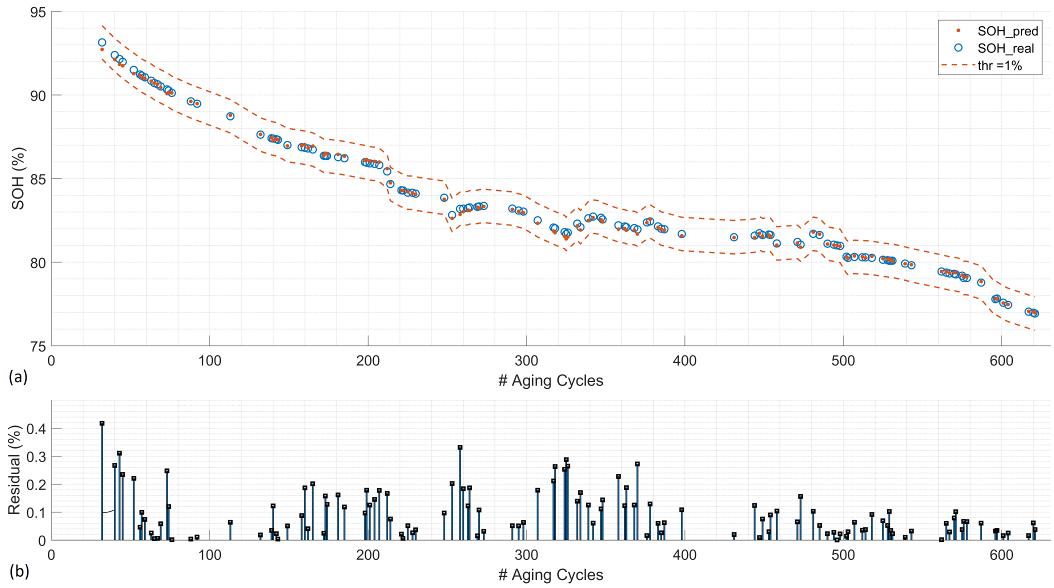


Fig. 4.15 Performance of Bi-LSTM network-based SOH estimation on a testing dataset. # Aging Cycles is the number of cycles that one cell has cycled.

In the Appendix A, Figure A.6 displays the learning performances in relation to the loss function. Additionally, Figure A.7 exhibits the prediction performance of SOH on aging cycles across all segments of the SOC domain, excluding the optimal segment.

Chapter 5

Conclusions & future works

In this final Chapter, the main conclusions following from the previous chapters are summarized and the possible indications for future works are given. The significance of data-driven models in the automotive industry is noteworthy from a scientific and academic research perspective due to various reasons. The utilization of data-driven models facilitates the enhancement of vehicle performance by optimizing various parameters, including fuel efficiency, reduction of polluting emissions, and road safety. Through the analysis of extensive data gathered by vehicle sensors, intricate patterns, correlations, and relationships can be discerned, which can be leveraged to enhance vehicle design and efficacy. From the forecasting and diagnostic capabilities point of view, utilizing data-driven models facilitates the ability to anticipate forthcoming vehicle performance and states. As an illustration, these can be employed to anticipate the decline of crucial components, such as the battery. The utilization of this data can be employed for the purpose of preemptive maintenance, mitigating the possibility of unforeseen malfunctions and enhancing overall safety measures.

In the realm of virtual sensing, data-driven models are extensively employed for the purpose of estimating or monitoring physical quantities. In practical applications, the utilized models have the ability to acquire knowledge from the interrelationships among the data obtained from the pre-existing sensors and the variables that require estimation or measurement. Consequently, advantages can be derived in relation to cost reduction by substituting costly physical sensors. Additionally, non-invasive sensors can be utilized in situations where it may be arduous or unfeasible to install physical sensors directly on specific vehicle components or under certain conditions.

The implementation of virtual sensing techniques can facilitate the acquisition of precise measurements or estimations of physical parameters, thereby enhancing the management and optimization of vehicular performance.

In summary, the dissertation contributes by proposing:

- The development of a virtual sensor that can be utilized for real-time prediction and monitoring of NOx emissions in diesel engine applications, particularly in the context of dynamic road driving conditions, and implementable on-board on ECU;
- A virtual environment for the simulation of an electric vehicle with the purpose of performance estimation of an electrified two-wheeler, integrating a battery model and a dynamic vehicle model into a global performance analysis;
- An optimal SOH estimation tool during charging partial phases that can be integrated on-board on BMS. The optimal SOC window has been identified as the most effective means of predicting battery cell lifetime.

This thesis work presents methodologies aimed at aiding OEMs in addressing the challenges posed by stringent regulations on polluting emissions and the complexities of electrification. Specifically, the proposed approach involves the development of advanced performance estimation tools that model the behaviour of vehicle subsystems in steady-state and dynamic operating conditions with the aim of achieving computational cost savings, performance improvements, and enhanced monitoring capabilities.

5.1 Conclusions

The present sub-section entails the derivation of conclusions for each topic expounded in this dissertation, in accordance with the sequence of the chapters.

5.1.1 Real-time pollutant emissions estimation in heavy-duty CI engines using data-driven approach

This chapter presents the implementation of a virtual sensor for the purpose of estimating and monitoring NOx emissions in diesel vehicular applications. The utilization of AI algorithms, specifically the XGBoost machine learning model, has demonstrated exceptional suitability and reliability in the execution of this task. The model has exhibited remarkable flexibility, robustness, and outstanding performance in predicting NOx engine-out. The results obtained exhibited a favorable predictive performance in both steady-state and dynamic conditions. A campaign of experimentation was conducted to gather data from the engine test bench and the ECU, with the aim of developing and calibrating the virtual sensor. The virtual sensor was subsequently subjected to testing under transient on-road driving conditions to assess its predictive capabilities in dynamic scenarios, via real-world driving missions. The creation of the virtual sensor was preceded by a data preprocessing procedure that was deemed essential for the proper acquisition of knowledge and assessment of the tool. The XGBoost architecture definition hyperparameters were optimized using the GridSearchCV tool to identify the optimal combination from the analyzed options. During the experimental tests, engine parameters were recorded and a feature extraction process was conducted in order to identify the most influential and weighted variables during the development of the predictors. The optimization of model accuracy and reduction of computational load are both crucial factors in enhancing the applicability of the model in practical scenarios.

Under steady-state conditions, the predictive accuracy was approximately 98%, whereas it was 85% during transient conditions. The findings suggest that the virtual sensor exhibits significant promise as a valuable instrument for monitoring and regulating emissions in diesel engine applications in real-time. The integration of the virtual sensor into the ECU presents a viable option for reducing NOx emissions. This approach involves utilizing the ECU control system to make real-time adjustments to the engine parameters, in collaboration with the virtual sensor.

5.1.2 Modeling and on-Road Testing of an Electric Two-Wheeler towards Range Prediction and BMS Integration

The research pertaining to Chapter 3 expounds upon a comprehensive, data-oriented modeling methodology for prognosticating the performance during operation of a two-wheeled electric vehicle. The study employed a comprehensive approach by integrating a dynamic vehicle model with a second order Thevenin ECM model to assess the predictive capabilities over real-world driving missions under dynamic conditions. The present study has developed a global model that utilizes the current discharged by the battery pack as an input parameter, and subsequently forecasts the instantaneous velocity of the vehicle. The battery model characteristic parameters, including the OCV, the cell static resistance R_0, the resistance R_1, R_2, and the capacitance C_1 and C_2 of each pair of RC circuit, as well as the parameters of the vehicle model, such as coastdown coefficients A, B and C, the aerodynamic drag coefficient Cx, the rolling resistance coefficient f, the equivalent mass associated to the rotating components m_r, and the overall battery-to-road efficiency η_{b2r}, were determined via a calibration procedure that relied on data obtained from both acquisition and on-road experimental measurements. The calibration procedure entailed an optimization stage, whereby the model parameters were varied to minimize an objective function.

The model's predictive capabilities have been validated via performance analysis metrics applied to the estimation of vehicle speed. The R2 values consistently exceeded 97% and the RMSE errors remained consistently below 1 m/s (with the exception of a single dataset) for all road driving tracks. The outcomes of the evaluated numerical model indicate that the forecaster tool has the potential to be embedded in BMS-integrated systems for the purpose of devising diverse management strategies for battery packs. The utilization of a sophisticated vehicle model facilitates the subsequent implementation of control logic for the efficient management of Li-ion batteries in both Software-in-the-Loop and Hardware-In-The-Loop environments.

5.1.3 Computational Cost Reduction for Artificial Intelligence-based Battery SOH estimation in Li-ion Battery Cells

The final chapter presents the findings related to the creation of an AI-driven model that is computationally efficient for predicting the remaining useful life of lithium-ion cells utilized in electric vehicle batteries. The proposed methodology offers an evaluation of state of health during incomplete recharging cycles, by selecting the state of charge range that yields precise and efficient prognostic estimates. The development of the estimator involved the utilization of multiple Bi-LSTM neural networks. These networks were employed to leverage various datasets that contained time series of charge data with varying lengths across the entire SOC domain. The methodology that was proposed has determined the most suitable length of the SOC window, which strikes a balance between the accuracy of predictions and the computational cost for the purpose of estimating the SOH on-board. As a result, we were able to identify the optimal SOC range that enables enhanced performance in SOH estimation. The input model utilized during the training process comprised of current, voltage, and charged capacity, whereas the output estimation was based on the battery cell's state-of-health. The aforementioned study pertains to an 18,650 cell that possesses a 3 Ah capacity and employs nickel-manganese-cobalt chemistry. The dataset in question is a constituent of a series of cycle aging experiments conducted by the Sandia National Laboratories. The findings demonstrate a high level of consistency and indicate that it is possible to forecast the SOH of a battery with a maximum margin of error of ±0.4%. This can be accomplished by monitoring solely the final 40% of the SOC window, specifically during the constant voltage (CV) phase of the entire constant current constant voltage (CCCV) charging process. This approach involves reducing the memory usage in the battery management system for charge data logging and decreasing computational time by approximately 2.3 times.

The computational lightness exhibited by the model and the findings of the current study render it highly appropriate for on-board implementation. This would enable the BMS to ensure optimal performance and enhance the battery's longevity.

5.2 Future works

This subsection presents recommendations for future research pertaining to each topic covered in the dissertation, as summarized in Figure 5.1.

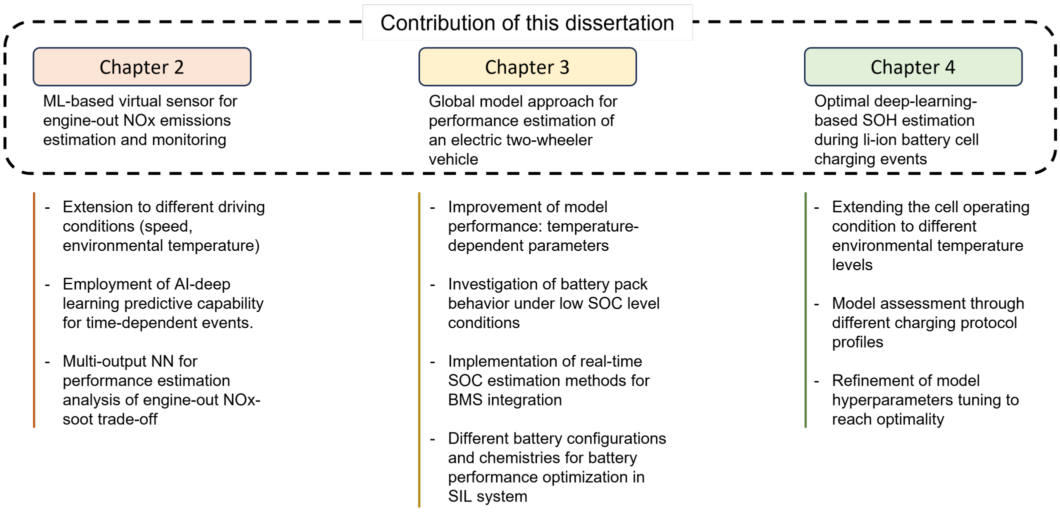


Fig. 5.1 Recommendations for future work divided by each research topic.

The central focus that unites all research subjects in this dissertation pertains to the creation of on-board implementable models utilizing data-driven techniques. The ensuing recommendations will highlight enhancements to modeling features, thereby augmenting the predictive capacity of the model through improved generalization across specific applications.

As far as the development of a virtual sensor for estimating and monitoring NOx emissions is concerned, a potential avenue for future research could involve expanding the analysis to encompass various driving scenarios. In this case, conducting a comprehensive experimental study involving varied controlled ambient temperatures might result in a significant enhancement of the estimator. During its operational lifespan, a land vehicle may experience fluctuating temperatures that can impact the engine's efficiency. This can result in the engine operating at significantly different points, with consequent effects on polluting emissions, particularly during cold starts. Regardless of the AI model employed, an increased diversity of operating conditions utilized for training may potentially result in an enhancement in the precision of predicted NOx level concentrations.
The increasing amount of data generated by on-board vehicle sensors due to tech-

nological advancements has necessitated the development of algorithms that can effectively process voluminous and time-sensitive data. The implementation of deep learning algorithms, namely artificial neural networks (ANN), recurrent neural networks (RNN), convolutional neural networks (CNN), and generative neural networks (GEN), has the potential to significantly enhance NOx estimation by acquiring complex representations and detecting concealed patterns within the data. It is noteworthy that the efficacy of said algorithms is contingent upon the sufficiency of the training data and the accurate configuration of the models.
Our current research focuses on a significant advancement in NOx prediction through the development of a deep learning model that utilizes an artificial neural network. This model is designed to accurately predict NOx and soot emissions in the 11.0 L Diesel engine seen in chapter 2. The model that has been developed comprises a neural network for multi-output regression and can briefly observed in [203]. This enables the estimation of both NOx and soot quantities simultaneously. The present study utilized experimental data consisting of combustion parameters within the cylinder, which were obtained under stationary conditions on the test bench. The present application incorporates a defined parameter, denoted as sample weight α, that specifies the proportion of impact exerted by the output on the computation of the loss function. The loss function is expressed as $Loss = \alpha \cdot loss_{NOx} + (1-\alpha) \cdot loss_{Soot}$. The study determined that the optimal point for the simultaneously estimation of the two pollutants is $\alpha = 0.5$. This value was assigned to ensure that both outputs have equal impact on the overall loss. MSE is employed as the loss function in this particular instance. For the sake of clarity and completeness, Figure 5.2 provides a comprehensive overview of the initial outcomes obtained with respect to loss and R2, as the parameter α undergoes variation.

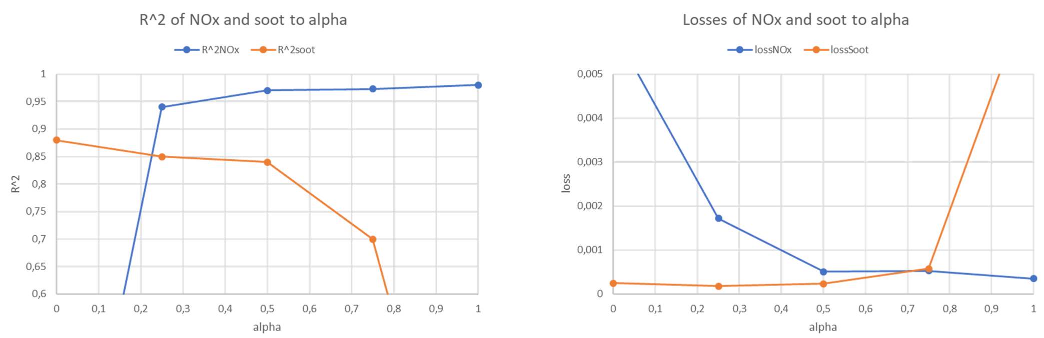


Fig. 5.2 Loss and R^2 as functions of sample weight α parameter.

Considering that the outcomes of the study were obtained from an experimental campaign conducted in stationary engine bench conditions, a potential avenue for further research might involve the validation of this model in transient conditions. This would entail taking into account real-world driving missions and incorporating the findings from Chapter 2. The aforementioned data-driven model exhibits potential for comprehensive analysis and may be deemed suitable for prospective integration into on-board ECUs to facilitate the real-time assessment of said pollutants.

Likewise, in terms of enhancing prediction accuracy, the modeling methodology employed in the research discussed in Chapter 3 could be refined to account for the influence of temperature and the complete SOC range on battery model parameters. In fact, the behavior of batteries is significantly influenced by their SOC and temperature. By taking into account the complete SOC domain, we can encompass the entire spectrum of battery performance, thereby enabling the development of more precise models. The properties of batteries, such as voltage, capacity, internal resistance, and self-discharge rate, may exhibit notable fluctuations across distinct states of charge. Through the process of modeling these variations, it is possible to precisely capture the actual behavior of the battery in real-world scenarios. During operation, batteries undergo dynamic changes in their SOC. The inclusion of parameters that exhibit variability across the entire SOC domain facilitates the acquisition of the battery's behavior as it changes over time. This holds significant importance for applications that entail recurrent cycles of charging and discharging or situations where the SOC may exhibit considerable variation over a period of time. In terms of on-board estimation, accurate SOC estimation is imperative for effective battery management systems. Given that the model parameters are depending on distinct SOC levels, it is feasible to derive a more precise estimation of current SOC by utilizing the measured voltage, current, and temperature of the battery. Incorporating temperature as a variable within the model parameters enables the consideration of the temperature-sensitive characteristics of the battery. Temperature can cause significant changes in battery capacity, internal resistance, and efficiency. In a nutshell, a battery model that incorporates parameters as a function of the entire SOC range and temperature offers a more robust and flexible representation. It allows for better adaptation to different operating conditions and enhances the model's generalization. The model can be made applicable to diverse scenarios, including but not limited to varying battery chemistries, cell designs, and operating environments, thereby enabling the capture of a broader spectrum of battery behaviors.

With regards to the estimation of SOC, the current investigation employs the Coulomb Counting (CC) technique, which is extensively utilized and yields a highly precise estimation, subject to certain boundary conditions that may frequently fluctuate or remain unknown. The primary constraints of CC pertain to the accumulation of errors resulting from imprecise measurements, calibration issues, or fluctuations in battery efficiency. Additionally, CC exhibits limited capacity for real-time adaptation and fails to capture non-linear behavior, as it is unable to account for variations in the behavioral dynamics of the battery caused by temperature, aging, and different operating conditions at extreme SOC levels. Lastly, CC is susceptible to measurement noise, which can introduce errors due to noise in the current and voltage readings. Hence, in order to address the constraints associated with CC in the estimation of SOC, multiple techniques have been developed to effectively manage the aforementioned complexities. The Kalman filter and Extended Kalman filter are recursive estimation methods that consider measurement noise, model uncertainties, and variations in battery characteristics. In addition, the utilization of data-driven methodologies, such as artificial neural networks, can effectively utilize extensive datasets to construct precise SOC estimation models by means of past data, which can effectively manage diverse battery chemistries and the impacts of aging. These approaches are particularly effective and suitable for on-board implementation on BMS.

The utilisation of recursive estimation techniques, such as the Kalman filter, and data-driven approaches, such as artificial neural networks, for the estimation of State-of-Charge (SOC) in on-board Battery Management Systems (BMS) does present enhanced precision. However, it is important to acknowledge the existence of different challenges that must be overcome in order to guarantee reliability and the ability to adapt in real-time within practical vehicular contexts. Several prominent challenges that have been identified, with the relative proposed solutions are here briefly mentioned.

- Measurement accuracy and noise: the precision of SOC estimation is contingent upon the accuracy of current and voltage measurements, which may be subject to noise and sensor inaccuracies. In order to mitigate measurement mistakes, it is possible to utilize sophisticated sensor calibration and filtering methods.

- Model complexity: the complexity of mathematical models tends to increase when employing more advanced estimation approaches. It is of utmost importance to ensure that these models exhibit both accuracy and computational efficiency in order to facilitate their deployment in real-time applications. It may be important to employ simplified models that strike a balance between accuracy and computational load.

- Training data and Neural Network adaptability: ANNs necessitate a substantial amount of training data and may exhibit limited adaptability in response to dynamic settings. The utilization of real-world driving data for continuous online training or retraining of ANNs has the potential to enhance their adaptability.

- Fault tolerance: battery systems may encounter sensor failures or exhibit drift phenomena as time progresses. The ability to effectively detect and handle malfunctioning sensors, as well as the capacity to withstand sensor failures, is of utmost importance in ensuring correct SOC assessment.

- Environmental variability, including temperature and other external conditions, has the potential to exert a substantial influence on battery performance. It is imperative for models and algorithms to consider these variances in order to uphold accuracy throughout diverse operating situations.

- Integration with BMS: implementing these methods within existing BMS hardware and software architectures can be a challenge. Careful system integration and optimization are necessary to ensure compatibility and real-time performance.

- Safety considerations are of utmost importance when it comes to estimating SOC of batteries. This is due to the potential risks associated with overcharging or deep discharging, which can lead to battery damage. It is crucial to priorities safety and incorporate fail-safe features in these estimation approaches.

- Regulatory compliance: automotive systems are required to comply with safety and regulatory standards in order to ensure regulatory compliance. SOC estimation methods need to comply with relevant standards and certification processes.

The present discourse outlines the key considerations that must be taken into account with regards to the design and construction of the energy storage system, specifically in relation to its developmental and production phases. These factors have been identified based on the requirements of the OEM. The positioning of lithium cells can be varied to form modules, taking into account practical constraints such as weight, size, costs, and performance for the intended use. These modules can be arranged in series and parallel connections to compose the entire battery pack, while considering both dimensional and operational constraints. For instance, a cooling system can be provided to monitor and control the system temperature and keep it within the optimal operating range. Hence, the provision of a Software-In-the-Loop (SIL) virtual setting for examining and assessing diverse tracks and chemistry is highly practical in enhancing drum efficiency. From this viewpoint, it is plausible to integrate the battery model as a software constituent in a SIL for the purpose of investigating vehicle performance. The utilization of virtual testing methods for vehicle behavior, configurations, and chemistry might potentially yield several advantages, including enhanced efficiency, reduced costs, and increased flexibility. This approach eliminates the necessity for physical testing and enables the execution of diverse driving simulations and scenarios, thereby facilitating the evaluation of performance across various road and environmental conditions. An additional benefit, which is not of lesser importance, pertains to the aspect of safety. The integration of a battery model within a Software-in-the-Loop environment enables the evaluation of diverse driving scenarios, encompassing hazardous or exceptional scenarios, without exposing the system to potential safety hazards, such as abrupt deceleration or collision simulations. The utilization of the SIL system can prove to be highly advantageous in the examination of diverse battery configurations and chemistry. The parameters of a battery model have the ability to be modified in order to accommodate varying specifications, including but not limited to capacity, internal resistance, voltage curve, and charge loss. The performance of the aforementioned battery configurations can be assessed through the use of simulations within the vehicular context. This approach allows for the observation of the impact on battery longevity, power availability, and energy efficiency.

Regarding the previous research topic pertaining to the formulation of a methodology for assessing the SOH in lithium battery cells, a crucial initial advancement might be to broaden the range of environmental temperatures under which the model operates. The susceptibility of Li-ion cells to temperature makes them vulnerable to thermal variations. Therefore, conducting tests on the model across a broader spectrum of temperatures would enable us to assess its capacity to gauge SOH amidst diverse thermal circumstances. In order to accommodate these enhancements, it is imperative to obtain fresh data or employ alternative experimental procedures. Additionally, it is necessary to modify the model to factor in the impact of thermal effects on battery efficacy. Similarly, the assessment of the AI model's performance can be conducted by employing various charging protocol profiles. Within the battery industry, it is customary to employ high-current charging or pulse charging protocols, which can be analyzed under dynamic conditions. This would allow for evaluating the model's ability to generalize and provide accurate estimates of SOH under a variety of charging conditions. Ultimately, a significant and subsequent advance could pertain to the enhancement of the calibration of the model's hyperparameters in order to attain maximum efficacy. The optimization of hyperparameters for the estimation model of SOH can be enhanced through various techniques, such as refining the grid utilized in RandomSearchCV, combining it with GridSearchCV, or utilizing Bayesian optimization. These methods can aid in the search for the optimal hyperparameter combination that maximizes the precision of the estimation model. This will aid in guaranteeing that the model is configured optimally and capable of delivering dependable estimations of SOH.

References

[1] Samsung. Specification of product. Technical report, 2019.

[2] Peter Weill and Stephanie L. Woerner. Optimizing your digital business model. *IEEE Engineering Management Review*, 43(1):123–131, 2015.

[3] Gruendinger Wolfgang Seiberth Gabriel. Data-driven business models in connected cars, mobility services and beyond. *BVDW Research*, 01(18):52, 2018.

[4] Milan Koch., Hao Wang., Robert Bürgel., and Thomas Bäck. Towards data-driven services in vehicles. In *In Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2020)*, pages 45–52. SciTePress, 2020.

[5] Kun Liao, Xiaodong Deng, Ying Liao, and Qingyu Zhang. Supplier empowerment: Mediating situational factors and perceived performance. *Journal of Purchasing and Supply Management*, 26(3):100611, 2020.

[6] Ishak Bin Aris, Ratna Kalos Zakiah Sahbusdin, and Ahmad Fairuz Muhammad Amin. Impacts of iot and big data to automotive industry. In *2015 10th Asian Control Conference (ASCC)*, pages 1–5. IEEE, 2015.

[7] Ziran Wang, Rohit Gupta, Kyungtae Han, Haoxin Wang, Akila Ganlath, Nejib Ammar, and Prashant Tiwari. Mobility digital twin: Concept, architecture, case study, and future challenges. *IEEE Internet of Things Journal*, 9(18):17452–17467, 2022.

[8] Toshika Srivastava. Data driven development for automotive domain. Virtual conference, 2021. QCon Plus - AI, ML & Data Engineering.

[9] Library of Congress Science Reference Section. Who invented the automobile?

[10] Matt Wolfe. The first sports car?

[11] History.com Editorsr. Automobile history. 2010.

[12] J. Choi D. An, N.H. Kim. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Elsevier*, 133:223–236, 2015.

[13] M.J. Roelle M.G. Shaver, J.C. Gerdes. Physics-based modeling and control of residual-affected hcci engines. *Dynamic Systems, Measurement and Control*, 131:12, 2009.

[14] L. Guzzella J. Asprion, O. Chinellato. A fast and accurate physics-based model for the nox emissions of diesel engines. *Applied Energy*, 103:221–233, 2013.

[15] P. Tunestal R. Johansson A. Widd, K. Ekholm. Physics-based model predictive control of hcci combustion phasing using fast thermal management and vva. *IEEE Transactions on Control Systems Technology*, 20:688–699, 2012.

[16] D. Bestle T. Funke. Physics-based model of a stroke-dependent shock absorber. *Multibody Syst. Dyn.*, 30:221–232, 2013.

[17] S.D. Fassois J.S. Sakellariou, K.A. Petsounis. Vibration based fault diagnosis for railway vehicle suspensions via a functional model based method: A feasibility study. *J Mech Sci Technol*, 29:471–484, 2015.

[18] P.S. Els G.S. Heymans, J.F. Grobler. Physics based modelling of a magneto-rheological equipped hydro-pneumatic semi-active suspension system. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Charlotte, North Carolina, USA, August 21–24, 2016. ASME.

[19] R. Alaggio G. Quaranta A. Aloisio, A. Contento. Physics-based models, surrogate models and experimental assessment of the vehicle–bridge interaction in braking conditions. *Mechanical Systems and Signal Processing*, 194:110276, 2023.

[20] Hyung Chul Kim and Timothy J. Wallington. Life cycle assessment of vehicle lightweighting: A physics-based model to estimate use-phase fuel consumption of electrified vehicles. *Environmental Science & Technology*, 50:11226–11233, 2016.

[21] H. DaCosta J. Chi. Modeling and control of a urea-scr aftertreatment system. *SAE Technical Paper*, 2005-01-0966:18, 2005.

[22] G. Stewart T. Samad. Systems engineering and innovation in control—an industry perspective and an application to automotive powertrains. *University of Maryland Model-Based Systems Engineering Colloquia Series*, Washington, DC:USA, 2013.

[23] X. Xu X. Liu J. Liu Y. Xiao, X. Zhang. Deep neural networks with koopman operators for modeling and control of autonomous vehicles. *IEEE Transactions on Intelligent Vehicles*, 8:135–146, 2023.

[24] Zhixiong Li Shuitao Gu Y. Pan, X. Nie. Data-driven vehicle modeling of longitudinal dynamics based on a multibody model and deep neural networks. *Measurement*, 180:109541, 2021.

[25] M. Battistoni F. Mariani, C.N. Grimaldi. Diesel engine nox emissions control: An advanced method for the o2 evaluation in the intake flow. *Applied Energy*, 113:576–588, 2014.

[26] H. Kim J. Keel T. Yoon J. Lee J. Lee, S. Kwon. Machine learning applied to the nox prediction of diesel vehicle under real driving cycle. *Applied sciences*, 11:3758, 2021.

[27] M.S.H. Lipu M.A. Hannan, D.N.T. How. Deep learning approach towards accurate state of charge estimation for lithium-ion batteries using self-supervised transformer model. *Sci Rep*, 11:19541, 2021.

[28] Nahiduzzaman M Ahsan M Haider J. Shahriar SM, Bhuiyan EA. State of charge estimation for electric vehicle battery management systems using the hybrid recurrent learning approach with explainable artificial intelligence. *Energies*, 15:8003, 2022.

[29] S.P. Nangrani S. Dewalkar. Artificial intelligence-based state of charge estimation of electric vehicle battery. *Smart Technologies for Energy, Environment and Sustainable Development*, 2:699–705, 2020.

[30] V.G. Priya K. Sekar K. Srinivasan R.M.S. Moorthy M. Suresh, B.K. Selvi. Ai based battery life estimation of electric vehicle. In *Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Dharan, Nepal, 2022, pp. 1009-1014, 2022. IEEE.

[31] Tejaswini P. and Sivraj P. Artificial intelligence based state of charge estimation of li-ion battery for ev applications. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 1356–1361. IEEE, 2020.

[32] S. Khaleghi Rahimian and Y. Tang. A practical data-driven battery state-of-health estimation for electric vehicles. *IEEE Transactions on Industrial Electronics*, 70:1973–1982, 2023.

[33] C.D. Kim S.M. Hell. Development of a data-driven method for online battery remaining-useful-life prediction. *Batteries*, 8:192, 2022.

[34] Yang Y. Shi Y. Zeng J. Tian Y., Wen J. State-of-health prediction of lithium-ion batteries based on cnn-bilstm-am. *Batteries*, 8:155, 2022.

[35] P. Bracinik D. Motyka M. Tomasov, M. Kajanova. Overview of battery models for sustainable power and transport applications. In *13th International scientific conference on sustainable, modern and safe transport (TRANSCOM 2019)*, pages 548–555. Elsevier, 2019.

[36] Aziz J. Sutikno T. Yao L.W., Wirun A. Battery state of charge estimation with extended kalman filter using third order thevenin model. *Telkomnika*, 13:401–412, 2015.

[37] Zhang R. Lao Z. Xia B., Sun Z. A cubature particle filter algorithm to estimate the state of the charge of lithium-ion batteries based on a second-order equivalent circuit model. *Energies*, 10:457, 2017.

[38] Lund P.D. Lindgren J., Asghar I. A hybrid lithium-ion battery model for system-level analyses. *Int. J. Energy Res.*, 40:1576–1592, 2016.

[39] Y. Xiao J. Hou Z. Fu L. Yao, Z. Fang. An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy*, 214:118866, 2021.

[40] Yu X. Dong X. Ren X., Liu S. A method for state-of-charge estimation of lithium-ion batteries based on pso-lstm. *Energy*, 234:121236, 2021.

[41] S. Yang X. Liu B. Wu, W.D. Widanage. Battery digital twins: Perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. *Energy and AI*, 1:100016, 2020.

[42] Soumya Singh, Max Weeber, and Kai Peter Birke. Implementation of battery digital twin: Approach, functionalities and benefits. *Batteries*, 7(4), 2021.

[43] European Parliament. Regulation (ec) no 715/2007 of the european parliament and of the council of 20 june 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (euro 5 and euro 6) and on access to vehicle repair and maintenance information (text with eea relevance). Technical report, European Union, 2007.

[44] C. Kannan; T. Vijayakumar. Application of exhaust gas recirculation for nox reduction in ci engines. *Elsevier*, pages 189–222, 2022.

[45] M. Baratta; R. Finesso; D. Misul; E. Spessa. Comparison between internal and external egr performance on a heavy duty diesel engine by means of a refined 1d fluid-dynamic engine model. *SAE International Journal of Engines*, 8:1977–92, 2015.

[46] A. Jain; A.P. Singh; A.K. Agarwal. Effect of split fuel injection and egr on nox and pm emission reduction in a low temperature combustion (ltc) mode diesel engine. *Energy*, 122:249–264, 2017.

[47] A. Maiboom; X. Tauzia; J.-F. Hètet. Influence of high rates of supplemental cooled egr on nox and pm emissions of an automotive hsdi diesel engine using an lp egr loop. *Int. J. Energy Res.*, 32:1383–1398, 2008.

[48] A. Maiboom; X. Tauzia. Nox and pm emissions reduction on an automotive hsdi diesel engine with water-in-diesel emulsion and egr: An experimental study. *Fuel*, 90:3179–3192, 2011.

[49] H.R. Stanglmaier; C.E. Roberts. Homogeneous charge compression ignition (hcci): Benefits, compromises, and future engine applications. *SAE Int.*, 108:2138–2145, 1999.

[50] M. Puškár; M. Kopas. System based on thermal control of the hcci technology developed for reduction of the vehicle nox emissions in order to fulfil the future standard euro 7. *Sci. Total Environ.*, 643:674–680, 2018.

[51] A. J. Torregrosa; A. Broatch; A. Garcìa; L.F. Mònico. Sensitivity of combustion noise and nox and soot emissions to pilot injection in pcci diesel engines. *Appl. Energy*, 104:149–157, 2013.

[52] T. Aoyama; Y. Hattori; J. Mizuta; Y. Sato. An experimental study on premixed-charge compression ignition gasoline engine. *SAE Technical Paper*, 960081:9, 1996.

[53] D.N. Cao; A.T. Hoang; H.Q. Luu; V.G. Bui; T.T.H. Tran. Effects of injection pressure on the nox and pm emission control of diesel engine: A review under the aspect of pcci combustion condition. *Energy Sources*, 0:1–18, 2020.

[54] E. Buyukkaya; C. Muhammet. Experimental study of nox emissions and injection timing of a low heat rejection diesel engine. *Int. J. Therm. Sci.*, 47:1096–1106, 2008.

[55] H.L. Fang; H.F.M. DaCosta. Urea thermolysis and nox reduction with and without scr catalysts. *Appl. Catal. B Environ.*, 46:17–34, 2003.

[56] I. Nova; L. Lietti; P. Forzatti. Mechanistic aspects of the reduction of stored nox over pt–ba/al$_2$o$_3$ lean nox trap systems. *Catal. Today*, 136:128–135, 2008.

[57] P. Jiao; Z. Li; B. Shen; W. Zhang; X. Kong; R. Jiang. Research of dpf regeneration with nox-pm coupled chemical reaction. *Appl. Therm. Eng.*, 110:737–745, 2017.

[58] L. Hofmann; K. Rusch; S. Fischer; B. Lemire. Onboard emissions monitoring on a hd truck with an scr system using nox sensors. *J. Fuels Lubr.*, 113:559–572, 2004.

[59] R. Finesso; G. Hardy; C. Maino; O. Marello; E. Spessa. A new control-oriented semi-empirical approach to predict engine-out nox emissions in a euro vi 3.0 l diesel engine. *Energies*, 10:1978, 2017.

[60] M. Baratta; H. Kheshtinejad; D. Laurenzano; D. Misul; S. Brunetti. Modelling aspects of a cng injection system to predict its behavior under steady state conditions and throughout driving cycle simulations. *J. Nat. Gas Sci. Eng.*, 24:52–6, 2015.

[61] C. Guardiola; B. Pla; D. Blanco-Rodriguez; P.O. Calendini. Ecu-oriented models for nox prediction. part 1: A mean value engine model for nox prediction. *J. Automob. Eng.*, 229:992–1015, 2015.

[62] R. Finesso; D. Misul; E. Spessa. Estimation of the Engine-Out NO2/NOx ratio in a EURO VI diesel engine. *SAE Int.*, 0317:15, 2013.

[63] C. Atkinson; G. Mott. Dynamic model-based calibration optimization: An introduction and application to diesel engines. Detroit, MI, USA, 11–14 April 2005, 2005. SAE World Congress.

[64] N. Donmez; O. Ozener. Modeling of nox emissions in internal combustion engine. *Int. J. Eng. Res. Adv. Technol.*, 5:36–43, 2019.

[65] S. D'Ambrosio; R. Finesso; L. Fu; A. Mittica; E. Spessa. A control-oriented real-time semi-empirical model for the prediction of nox emissions in diesel engines. *Appl. Energy*, 130:265–2793, 2014.

[66] B. Winkler-Ebner; M. Hirsch; L. del Re; H. Klinger; W. Mistelberger. Comparison of virtual and physical nox-sensors for heavy duty diesel engine application. *SAE Int. J. Engines*, 3:1124–1139, 2010.

[67] S. Stadlbauer; D. Alberer; M. Hirsch; S. Formentin; C. Benatzky; L. del Re. Evaluation of virtual nox sensor models for off road heavy duty diesel engines. *SAE Int. J. Commer. Veh.*, 5:128–140, 2012.

[68] R. Fechert; B. Bäker; S. Gereke; F. Atzler. Using machine learning methods to develop virtual nox sensors for vehicle applications. Stuttgart, Germany, 18 August 2020, 2020. Internationales Stuttgarter Symposium.

[69] N. Kempema; C. Sharpe; X. Wu; M. Shahabi; D. Kubinski. Machine-learning-based emission models in gasoline powertrains part 2: Virtual carbon monoxide. *SAE Int. J. Engines*, 16, 2022.

[70] X. Yuan; L. Li; Y. Wang. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Trans. Ind. Inform.*, 16:3168–3176, 2019.

[71] W. Shao; X. Tian. Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development. *Neurocomputing*, 222:91–10, 2017.

[72] C. Liu; M. Guo; Y. Yang; S. Zhang. Nox prediction for diesel engine using improved ga-svr. *Measurement*, 173:108187, 2020.

[73] H. Zhang; Z. Zhang; Y. Liu. Nox emission prediction of diesel engines based on support vector machine. *Measurement*, 125:338–346, 2018.

[74] X. Li; Y. Liu; H. Zhang; Z. Zhang. Nox prediction for diesel engine using multi-input deep learning network based on lstm. *Measurement*, 145:36–43, 2019.

[75] T. Wakjira; A. Rahmzadeh; M. S. Alam; R. Tremblay. Explainable machine learning based efficient prediction tool for lateral cyclic response of post-tensioned base rocking steel bridge piers. *Structures*, 44:947–964, 2022.

[76] T. Chen; C. Guestrin. Xgboost: A scalable tree boosting system. San Francisco, CA, USA, 13–17 August 2016, 2016. 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.

[77] H.J. Friedman. Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 29:1189–232, 2001.

[78] L. Breiman. Random forests. *Mach. Learn.*, 45:5–32, 2001.

[79] C. Bentejac; A. Csorgo; M.G. Munoz. A comparative analysis of xgboost. *Artif. Intell. Rev.*, 54:1937–1967, 2021.

[80] K. B. Altug; S.E. Kucuk. Predicting tailpipe nox emission using supervised learning algorithms. Ankara, Turkey, 11–13 October 2019, 2019. 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies.

[81] L. Hu; C. Wang; Z. Ye; S. Wang. Estimating gaseous pollutants from bus emissions: A hybrid model based on gru and xgboost. *Sci. Total Environ.*, 783:146870, 2021.

[82] R. Finesso; O. Marello. Calculation of Intake Oxygen Concentration through Intake CO2 Measurement and Evaluation of its Effect on Nitrogen Oxide Prediction Accuracy in a Heavy-Duty Diesel Engine. *Energies*, 15:342, 2022.

[83] R. Finesso; G. Hardy; C. Maino; O. Marello; E. Spessa. A New Control-Oriented Semi-Empirical Approach to Predict Engine-Out NOx Emissions in a Euro VI 3.0 l Diesel Engine. *Energies*, 10:1978, 2017.

[84] Shubham Goyal. Boosting performance with xgboost.

[85] Lundberg; Scott M; Lee Su-In. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[86] S.M. Lundberg; G. Erion; H. Chen. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.*, 2:56–67, 2020.

[87] R. Can; S. Kocaman; C. Gokceoglu. A comprehensive assessment of xgboost algorithm for landslide susceptibility mapping in the upper basin of ataturk dam, turkey. *Applied Sciences*, 11:11, 2021.

[88] Ceshine Lee. Feature importance measures for tree models. Technical report, Veritable, Medium, 2017.

[89] Terence Shin. Understanding feature importance and how to implement it in python. Technical report, towardsdatascience, 2021.

[90] S. Ronaghan. The mathematics of decision trees, random forest and feature importance in scikit-learn and spark. Technical report, Medium, 2018.

[91] T. Chen; C. Guestrin. Xg boost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD*, page 785–794, San Francisco, CA, USA, 13–17 August 2016, 2016. ACM Press.

[92] T. Hastie; R. Tibshirani; J.H Friedman. Boosting and additive trees. *The Elements of Statistical Learning, Springer: New York, NY, USA*, 2nd ed.:337–384, 2009.

[93] J. Friedman. Greedy function approximation: A gradient boosting machine. *Comput. Stat. Data Anal.*, 38:367–378, 2002.

[94] J. Friedman. Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38:367–378, 2002.

[95] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. New York, NY, USA, 2016. Association for Computing Machinery.

[96] D. Martins. Xgboost: A complete guide to fine-tune and optimize your model. Technical report, towardsdatascience, 2021.

[97] Jason Brownlee. A gentle introduction to k-fold cross-validation.

[98] M. Peckham; J. Parnell; M. Hammond; B. Mason. The measurement of fast transient emissions during real world driving. *Frontiers*, 6:19, 2020.

[99] T. Sellerei; C. Ferrarese; J. Franzetti; R. Suarez-Bertoa; D. Manara. Real-time measurement of nox emissions from modern diesel vehicles using on-board sensors. *Energies*, 15:8766, 2022.

[100] EV-Volumes Roland Irle. Ev-volumes.

[101] Electric Vehicle Database. Range of full electric vehicles.

[102] M.H. Almannaa S.Q. Liu-S. Glaser A. Rakotonirainy M. Masoud, M. Elhenawy. Optimal assignment of e-scooter to chargers. In *In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference*, page 4204–4209, Auckland, New Zealand, 27–30 October, 2019. IEEE.

[103] N. Kimura T. Morizane-M. Nakaoka K. Kaneko, H. Omori. A novel type of edlc electric motor driven scooter with pulse super-rapied charger. In *In Proceedings of the 2015 International Conference on Electrical Drives and Power Electronics*, pages 476–481, Tatranska Lomnica, Slovakia, 21–23 September, 2015. IEEE.

[104] J.K. Cheng Y.L. Hwang. The dynamic behavior and modal analysis of electric scooter. *J. Vibroeng*, 16:2297–2304, 2014.

[105] L. Berzi M. Pierini-G. Lutzemberger E. Locorotondo, L. Pugi. Online identification of thevenin equivalent circuit model parameters and estimation state of charge of lithium-ion batteries. In *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe*, pages 1–6, Palermo, Italy, 2018. IEEE.

[106] G.A. Rincon-Mora M. Chen. Accurate electrical battery model capable of predicting runtime and i-v performance. *IEEE Transactions on Energy Conversion*, 21(2):504–511, 2006.

[107] K. Steiner H. Walz-T. Soczka-Guth D.U. Sauer D. Andre, M. Meiler.

[108] R. Dougal L. Gao, S. Liu.

[109] S.P. Santiago Pindado J.C. Javier Cubas A.P.-H. Á Porras-Hermoso, B.C.-L. Borja Cobo-Lopez. Analytical models for li-ion batteries developed at the idr/upm institute. In *In Proceedings of the 8th European Conference for Aeronautics and Space Sciences*, Pza. del Cardenal Cisneros 3, Madrid, Spain, 1–4 June, 2019. EUCASS.

[110] Reinhardt Klein, Nalin A. Chaturvedi, Jake Christensen, Jasim Ahmed, Rolf Findeisen, and Aleksandar Kojic. Electrochemical model based observer design for a lithium-ion battery. *IEEE Transactions on Control Systems Technology*, 21(2):289–301, 2013.

[111] Kandler A. Smith, Christopher D. Rahn, and Chao-Yang Wang. Model-based electrochemical estimation and constraint management for pulse operation of lithium ion batteries. *IEEE Transactions on Control Systems Technology*, 18(3):654–663, 2010.

[112] S. Tamilselvi, S. Gunasundari, N. Karuppiah, Abdul Razak RK, S. Madhusudan, Vikas Madhav Nagarajan, T. Sathish, Mohammed Zubair M. Shamim, C. Ahamed Saleel, and Asif Afzal. A review on battery modelling techniques. *Sustainability*, 13(18):10042, 2021.

[113] Model based design of balancing systems for electric vehicle battery packs. *IFAC-PapersOnLine*, 48(15):395–402, 2015.

[114] Ahmad Pesaran. Battery thermal management in evs and hevs: Issues and solutions. *Battery Man*, 43, 01 2001.

[115] Maguire P.-Gebbie J. et al. Janarthanam S., Paramasivam S. Hev battery pack thermal management design and packaging solutions. *SAE Int. J. Engines*, 10(3):785–789, 2017.

[116] E. Locorotondo; L. Pugi; M. Pierini; G. Lutzemberger. Online identification of thevenin equivalent circuit model parameters and estimation state of charge of lithium-ion batteries. In *In proceedings of the 2018 IEEE International Conference on Environment and Electrical Engineering and Electrical Engineering*, page 23, Palermo, Italy, 12–15 June 2018, 2018. IEEE.

[117] J. Lv; B. Jiang; X. Wang; Y. Liu; Y. Fu. Estimation of the state of charge of lithium batteries based on adaptive unscented kalman filter algorithm. *Electronics*, 9:1425, 2020.

[118] R. Jackey; M. Saginaw; P. Sanghvi; J. Gazzarri; T. Huria; M. Ceraolo. Battery model parameter estimation using a layered technique: An example using a lithium iron phosphate cell. *SAE Tech. Pap.*, 1:1547, 2013.

[119] L. Zhang; H. Peng; Z. Ning; Z. Mu; C. Sun. Comparative research on rc equivalent circuit models for lithium-ion batteries of electric vehicles. *Applied Science*, 7:1002, 2017.

[120] M. Garcia-Plaza; J. Carrasco; A. Pena-Asensio; J-Alonso-Martinez; SD. Arnaltes Gomes. Hysteresis effect influence on electrochemical battery modeling. *Electric Power Syst. Res.*, 152:27–35, 2017.

[121] Y. Jin; W. Zhao; Z. Li; B. Liu; L. Liu. Modeling and simulation of lithium-ion battery considering the effect of charge-discharge state. In *In Proceedings of the International Conference on Electronic Materials and Information Engineering*, Xi'an, China, 9–11 April 2021, 2021. IOP Science.

[122] Yen Liang Yeh Hua-Sheng Liao Bin-Hao Chen, Po-Tuan Chen. Establishment of second-order equivalent circuit model for bidirectional voltage regulator converter: 48 v-aluminum-ion battery pack. *Energy Reports*, 79(2629-2637), 2023.

[123] S. Atalay; M. Sheikh; A. Mariani; Y. Merla; E. Bower; W.D. Widanage. Theory of battery ageing in a lithium-ion battery: capacity fade, nonlinear ageing and lifetime prediction. *J. Power Sources*, 478:7753, 2020.

[124] W. Yourey. Theoretical impact of manufacturing tolerance on lithium-ion electrode and cell physical properties. *Batteries*, 6:23, 2020.

[125] S. Santhanagopalan; R.E. White. Quantifying cell-to-cell variations in lithium ion batteries. *Int. J. Electrochem.*, 2012:395838, 2012.

[126] F. Van der Sluis; L. Romers; G. Van Spijk; I. Hupkes. Cvt, promising solutions for electrification. *SAE Tech. Pap.*, 1:0359, 2019.

[127] B. Swieczko-Zurek; P. Jaskula; J. Ejsmont; A. Kedzierska; P. Czajkowski. Rolling resistance and tire/road noise on rubberized asphalt pavement in poland. In *In Proceedings of the Rubberized Asphalt—Asphalt Rubber 2015 Conference*, Las Vegas, NV, USA, 4–7 October 2015, 2015. RAR.

[128] Z. El-Sayegh; M. El-Gindy. Rolling resistance prediction of off-road tire using advanced simulation and analytical techniques. *Applied Science*, 2:1620, 2020.

[129] R. Xiong; J. Cao; Q. Yu; H. He; F. Sun. Critical review on the battery state of charge estimation methods for electric vehicles. *IEEE Access*, 6:1832–1843, 2018.

[130] N. Asep.;R. Estiko; W. F. Danang; N. Prapto. Battery state of charge estimation by using a combination of coulomb counting and dynamic model with adjusted gain. In *In Proceedings of the International Conference on Sustainable Energy Engineering and Application*, Bandung, Indonesia, 5–7 October 2015, 2015. IEEE.

[131] P. Ion; C. Dinu; C. Gheorghe. Coast down test—theoretical and experimental approach. In *In Proceedings of the International Automotive Congress*, Brasov, Romania, 1 October 2010, 2010. Siar.

[132] R. Carlson; H. Lohse-Busch; J. Diez; J. Gibbs. The measured impact of vehicle mass on road load forces and energy consumption for a bev, hev, and ice vehicle. *SAE Int. J. Altern. Powertrains*, 2:105–114, 2013.

[133] Mathworks. *Estimate Model Parameter Values (GUI)*.

[134] Mathworks. *Estimate Model Parameter Values (Code)*.

[135] Mathworks. *Equation Solving Algorithms*.

[136] Mathworks. *Least-Squares (Model Fitting) Algorithms*.

[137] C. Zhang; W. Allafi; Q. Dinh; P. Ascencio; K. Marco. Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. *Energy*, 142:678–688, 2018.

[138] W. Yourey. Theoretical impact of manufacturing tolerance on lithium-ion electrode and cell physical properties. *Batteries*, 6:23, 2020.

[139] S. Santhanagopalan; R.E. White. Quantifying cell-to-cell variations in lithium ion batteries. *Int. J. Electrochem.*, 2012:395838, 2012.

[140] T. Paul A. Samet I. Jorge, T. Mesbahi. Study and simulation of an electric scooter based on a dynamic modelling approach. In *In Proceedings of the 2020 Fifteenth International Conference on Ecological Vehicles and Renewable Energies*, Monte-Carlo, Monaco, 10–12 September, 2022, 2022. EVER.

[141] Z. Mooney Christopher. *Monte Carlo Simulation*. Quantitative Applications in the Social Sciences. Sage, 1997.

[142] Alex Van den Bossche. Light and ultralight electric vehicless. *Journal of Scientific Research*, 2(0):10–13, 2010.

[143] G.E. Blomgren. The development and future of lithium ion batteries. *J. Electrochem. Soc.*, 164:A5019, 2016.

[144] T.S. Ustun H. Keshan, J. Thornburg. Comparison of lead-acid and lithium ion batteries for stationary storage in off-grid energy systems. In *In Proceedings of the 2016 4th IET Clean Energy and Technology Conference*, Kuala Lumpur, Malaysia, 14–15 November, 2019. IEEE.

[145] J. Timmons D. Vutetakis. A comparison of lithium-ion and lead-acid aircraft batteries. *SAE Tech. Paper*, 1:2875, 2008.

[146] Y. Zhang et al. K.W. See, G. Wang. Critical review and functional safety of a battery management system for large-scale lithium-ion battery pack technologies. *Int. J. Coal. Sci. Technol.*, 9:36, 2022.

[147] Xuning Feng, Minggao Ouyang, Xiang Liu, Languang Lu, Yong Xia, and Xiangming He. Thermal runaway mechanism of lithium ion battery for electric vehicles: A review. *Energy Storage Materials*, 10:246–267, 2018.

[148] Manh-Kien Tran, Anosh Mevawalla, Attar Aziz, Satyam Panchal, Yi Xie, and Michael Fowler. A review of lithium-ion battery thermal runaway modeling and diagnosis approaches. *Processes*, 10(6):1192, 2022.

[149] Tan C. & Pecht M. Leng, F. Effect of temperature on the aging rate of li ion battery operating above room temperature. *Sci. Rep.*, 5:12967, 2015.

[150] Shuai Ma, Modi Jiang, Peng Tao, Chengyi Song, Jianbo Wu, Jun Wang, Tao Deng, and Wen Shang. Temperature effect and thermal impact in lithium-ion batteries: A review. *Progress in Natural Science: Materials International*, 28(6):653–666, 2018.

[151] B. Jaydeep. Effect of temperature on battery life and performance in electric vehicle. *Int. J. Sci. Res.*, 2:1–3, 2012.

[152] Dongxu Ouyang, Yaping He, Jingwen Weng, Jiahao Liu, Mingyi Chen, and Jian Wang. Influence of low temperature conditions on lithium-ion batteries and the application of an insulation material. *RSC Adv.*, 9:9053–9066, 2019.

[153] C.W. Monroe D.A. Howey T. Raj, A.A. Wang. Investigation of path-dependent degradation in lithium-ion batteries. *Eur. Chem. Soc. Publ.*, 3:12, 2020.

[154] Mohammad Shahjalal, Probir Kumar Roy, Tamanna Shams, Ashley Fly, Jahedul Islam Chowdhury, Md. Rishad Ahmed, and Kailong Liu. A review on second-life of li-ion batteries: prospects, challenges, and issues. *Energy*, 241:122881, 2022.

[155] J. Shen Y. Zhang G. Li Z. Chen Y. Liu X. Shu, S. Shen. State of health prediction of lithium-ion batteries based on machine learning: Advances and perspectives. *iScience*, 24:11, 2021.

[156] Cheng Lin, Aihua Tang, and Wenwei Wang. A review of soh estimation methods in lithium-ion batteries for electric vehicle applications. *Energy Procedia*, 75:1920–1925, 2015.

[157] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche. Critical review of state of health estimation methods of li-ion batteries for real applications. *Renewable and Sustainable Energy Reviews*, 56:572–587, 2016.

[158] Nassim Noura, Loïc Boulon, and Samir Jemeï. A review of battery state of health estimation methods: Hybrid electric vehicle challenges. *World Electric Vehicle Journal*, 11(4), 2020.

[159] Wenwei Wanga Cheng Lina, Aihua Tanga. A review of soh estimation methods in lithium-ion batteries for electric vehicle applications. *ScienceDirect*, 75(1920-1925), 2015.

[160] Z. Yuan Y. Shen H. Han, H. Xu. A new soh prediction model for lithium-ion battery for electric vehicles. In *In Proceedings of the 2014 17th International Conference on Electrical Machines and Systems (ICEMS)*, Hangzhou, China, 22–25 October, 2014. IEEE.

[161] Pier Giuseppe Anselma, Phillip Kollmeyer, Jeremy Lempert, Ziyu Zhao, Giovanni Belingardi, and Ali Emadi. Battery state-of-health sensitive energy management of hybrid electric vehicles: Lifetime prediction and ageing experimental validation. *Applied Energy*, 285:116440, 2021.

[162] Guangdong Hou Bor Yann Liaw Changshui Zhang Zhen Guo, Xinping Qiu. State of health estimation for lithium-ion batteries based on charging curves. *J. Power Sources*, 249(457–62), 2014.

[163] Matteo Galeotti, Lucio Cinà, Corrado Giammanco, Stefano Cordiner, and Aldo Di Carlo. Performance analysis and soh (state of health) evaluation of lithium polymer batteries through electrochemical impedance spectroscopy. *Energy*, 89:678–686, 2015.

[164] M.R. Kevin K. Tadhg M. Kieran, G. Hemtej. Review—use of impedance spectroscopy for the estimation of li-ion battery state of charge, state of health and internal temperature. *J. Electrochem. Soc.*, 168:080517, 2021.

[165] Jussi Sihvo, Tomi Roinila, and Daniel-Ioan Stroe. Soh analysis of li-ion battery based on ecm parameters and broadband impedance measurements. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 1923–1928, Singapore, 2020.

[166] Y.-P. Chen Y.-C. Hsieh K.S. Ng, C.-S. Moo. Enhanced coulomb counting method for estimating state-of-charge and stateof-health of lithium-ion batteries. *Applied Energy*, 86(1506–11), 2009.

[167] C.-J. Wang T. Hansen. Support vector based battery state of charge estimator. *J. Power Sources*, 141(351–8), 2005.

[168] T. Cher-H. Shyh-Chin S. Preetpal, C. Che. Semi-empirical capacity fading model for soh estimation of li-ion batteries. *Appl. Sci.*, 9:3012, 2019.

[169] Meng J. Wang Y. et al. Huang, H. An enhanced data-driven model for lithium-ion battery state-of-health estimation with optimized features and prior knowledge. *Automot. Innov.*, 5:134–145, 2022.

[170] Weibo Liu Nianyin Zeng-Xin Luo Minzhi Chen, Guijun Mab. An overview of data-driven battery health estimation technology for battery management system. *Neurocomputing*, 532(152-169), 2023.

[171] S. Jiang B. Xu-X. Han X. Lai Y. Zheng T. Sun, S. Wang. A cloud-edge collaborative strategy for capacity prognostic of lithium-ion batteries based on dynamic weight allocation and machine learning. *Energy*, 239(122185), 2022.

[172] Z. Chen J. Wu, C. Zhang. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Appl. Energy*, 173(134-140), 2016.

[173] J. Zhang A. Garg-S. Wang X. Liang, N. Bao. valuation of battery modules state for electric vehicle using artificial neural network and experimental validation. *Energy Sci. Eng.*, 6 (5)(397-407), 2018.

[174] W. Shen J. Lu-X.G. Yang J. Tian, R. Xiong. eep neural network battery charging curve prediction using 30 points collected in 10 min. *Joule*, 5(1521–1534), 2021.

[175] M. Preindl Y. Fahmy-A. Emadi E. Chemali, P.J. Kollmeyer. Convolutional neural network approach for estimation of li-ion battery state of health from charge profiles. *Energies*, 15(1185), 2022.

[176] X. Wang Z. Meng-L. Zhao M.J. Deen L. Ren, J. Dong. A data-driven auto-cmm-lstm prediction model for lithium-ion battery remaining useful life. *IEEE Trans. Ind. Inf.*, 17 (5)(3478-3487), 2021.

[177] D. Karimi H. Behi-S.H. Beheshti J. Van Mierlo M. Berecibar S. Khaleghi, M.S. Hosen. Developing an online data-driven approach for prognostics and health management of lithium-ion batteries. *Appl. Energy*, 308(118348), 2022.

[178] Sungwoo Jo, Sunkyu Jung, and Taemoon Roh. Battery state-of-health estimation using machine learning and preprocessing with relative state-of-charge. *Energies*, 14(21), 2021.

[179] Hossam A. Gabbar, Ahmed M. Othman, and Muhammad R. Abdussami. Review of battery management systems (bms) development and industrial standards. *Technologies*, 9(2):28, 2021.

[180] R. Roncella R. Saletti R. Schwarz V.R. Lorentz E.R.G. Hoedemaekers B. Rosca F. Baronti R. Morello, R. Di Rienzo. Advances in li-ion battery management for electric vehicles. In *In Proceedings of the IECON 2018—44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, DC, USA, 21–23 October, 2018.

[181] Shichun Yang, Zhengjie Zhang, Rui Cao, Mingyue Wang, Hanchao Cheng, Lisheng Zhang, Yinan Jiang, Yonglin Li, Binbin Chen, Heping Ling, Yubo Lian, Billy Wu, and Xinhua Liu. Implementation for a cloud battery management system based on the chain framework. *Energy and AI*, 5:100088, 2021.

[182] Kyungnam Park, Yohwan Choi, Won Jae Choi, Hee-Yeon Ryu, and Hongseok Kim. Lstm-based battery remaining useful life prediction with multi-channel charging profiles. *IEEE Access*, 8:20786–20798, 2020.

[183] Benvolence Chinomona, Chunhui Chung, Lien-Kai Chang, Wei-Chih Su, and Mi-Ching Tsai. Long short-term memory approach to estimate battery remaining useful life using partial data. *IEEE Access*, 8:165419–165431, 2020.

[184] M. Yadavc S. Li G. Dos Reis, C. Strange. Lithium-ion battery data and where to find it. *Energy And AI*, 5:100081, 2021.

[185] Sandia National Lab. Data for degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions.

[186] Peter Worcester. A comparison of grid search and randomized search using scikit learn.

[187] A. Ullah M.Y. Lee S.W. Baik N. Khan, F.U.M. Ullah. Batteries state of health estimation via efficient neural networks with multiple channel charging profiles. *IEEE Access*, 9:7797–7813, 2020.

[188] A.S.N Siami Niamini. S.S. Siami-Namini, N.T. Tavakoli. The performance of lstm and bi-lstm in forecasting time series. In *In Proceedings of the IEEE International Conference on Big Data*, Los Angeles, CA, USA, 9–12 December 2019, 2019. IEEE.

[189] Jason Brownlee. Hyperparameter optimization with random search and grid search.

[190] Wu Ming-hu Zhang Fan, Xing Zi-xuan. State of health estimation for li-ion battery using characteristic voltage intervals and genetic algorithm optimized back propagation neural network. *Journal of Energy Storage*, 57(106277), 2023.

[191] Jason Brownlee. A gentle introduction to dropout for regularizing deep neural networks.

[192] Mathworks. Deep learning with time series and sequence data.

[193] Sebastian Ruder. An overview of gradient descent optimization algorithms.

[194] Jason Brownlee. A gentle introduction to early stopping to avoid overtraining neural networks.

[195] Mathworks. Bidirectional lstm layer.

[196] Caner. Selecting optimal lstm batch size.

[197] Jason Brownlee. How to configure the learning rate when training deep learning neural networks.

[198] Jason Brownlee. A gentle introduction to dropout for regularizing deep neural networks.

[199] Jason Brownlee. Train-test split for evaluating machine learning algorithms.

[200] C. Lin X. Mei M. Shi, J. Xu. A fast state-of-health estimation method using single linear feature for lithium-ion batteries. *Energy*, 256:124652, 2022.

[201] L. Luoshi X. Bin, X. Bing. State of health estimation for lithium-ion batteries based on the constant current-constant voltage charging curve. *Electronics*, 9:1279, 2020.

[202] Z. Wei Z. Quan Y. Li H. Ruan, H. He. State of health estimation of lithium-ion battery based on constant-voltage charging reconstruction. *IEEE J. Emerg. Sel. Top. Power Electron*, 2021.

[203] Laura Travaini. Multiple output nn for the prediction of nox and soot in a diesel engine. Master's thesis, Politecnico di Torino, Torino, Italy, 2020.

Appendix A

Figures

Chapter 3

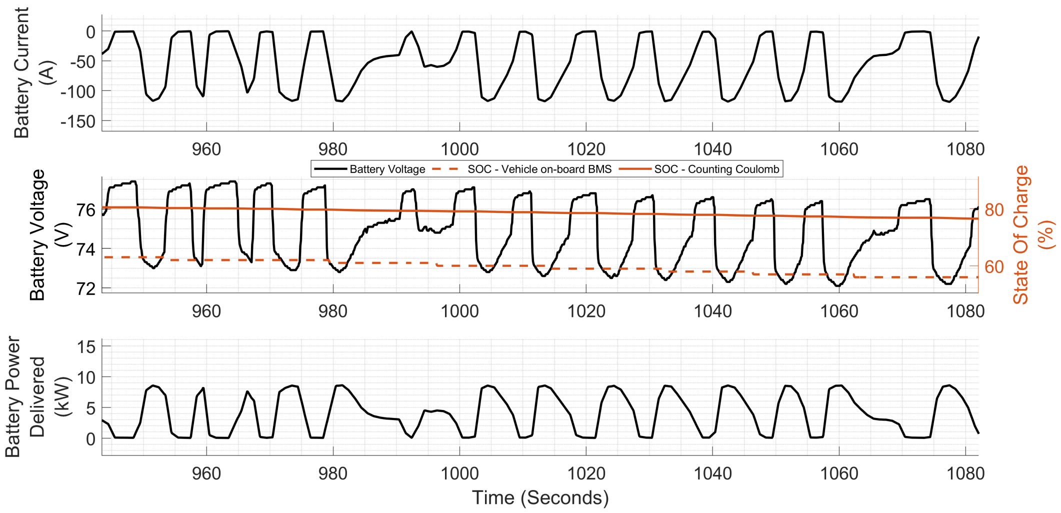


Fig. A.1 Zoom of the operating battery pack voltage, current and power delivered during the tests #1.

Chapter 4

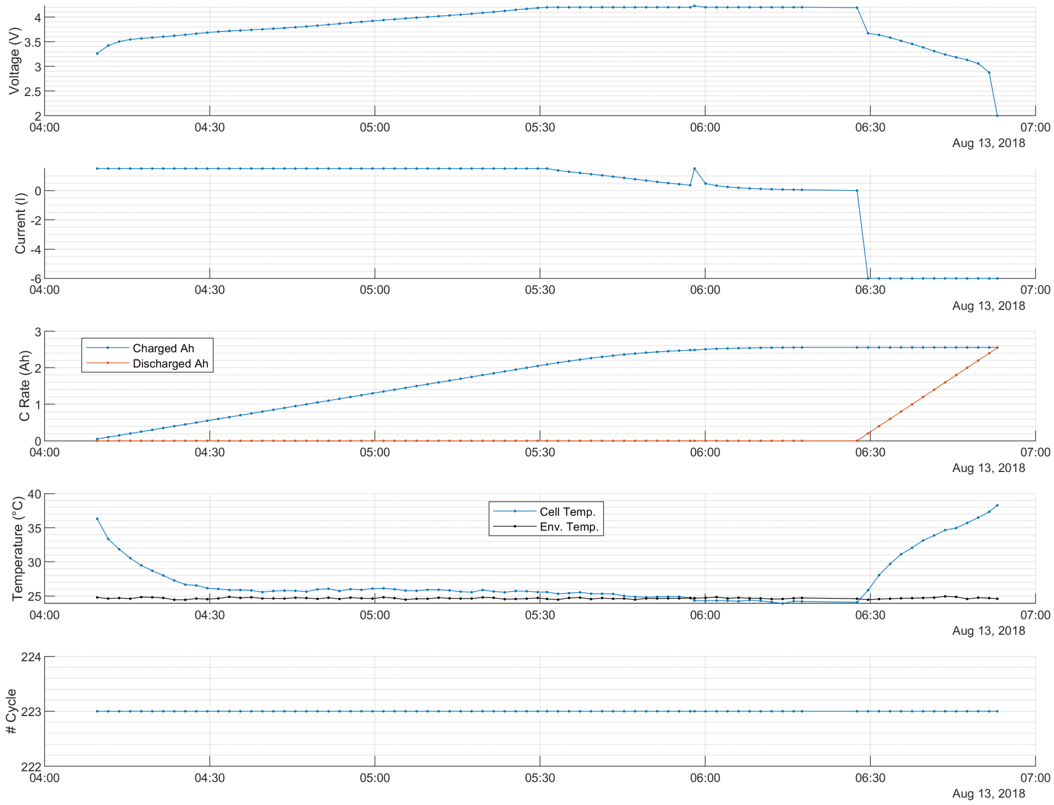


Fig. A.2 Cycle #223 deleted from the whole experimental dataset.

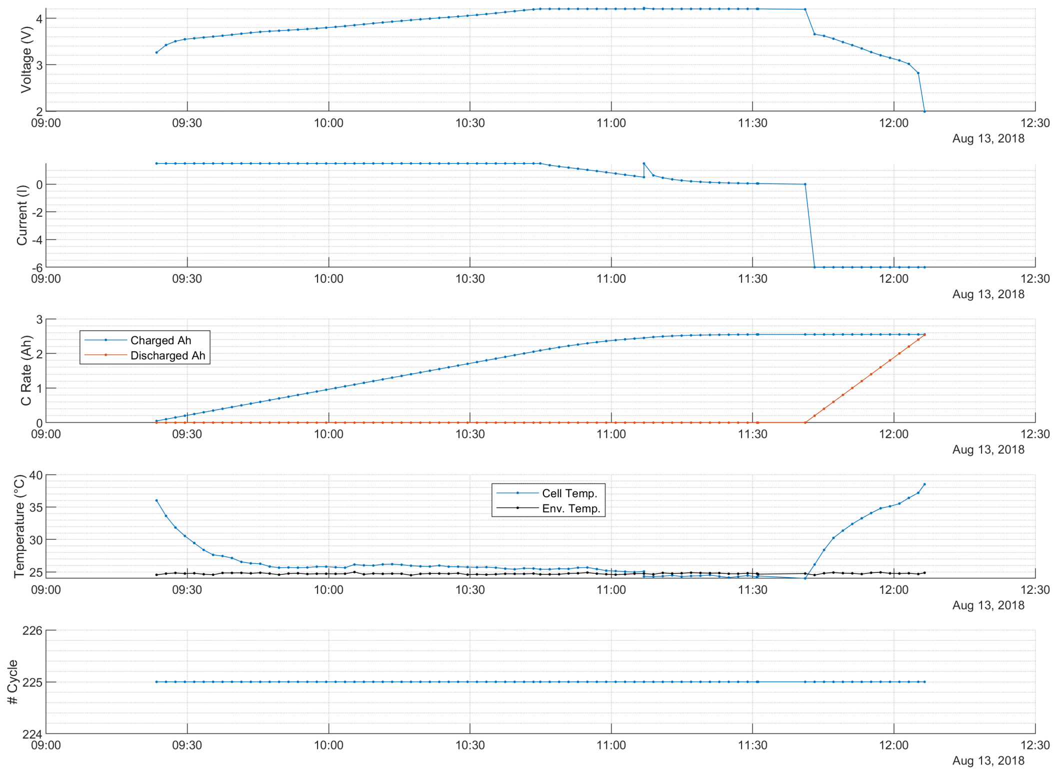


Fig. A.3 Cycle #225 deleted from the whole experimental dataset.

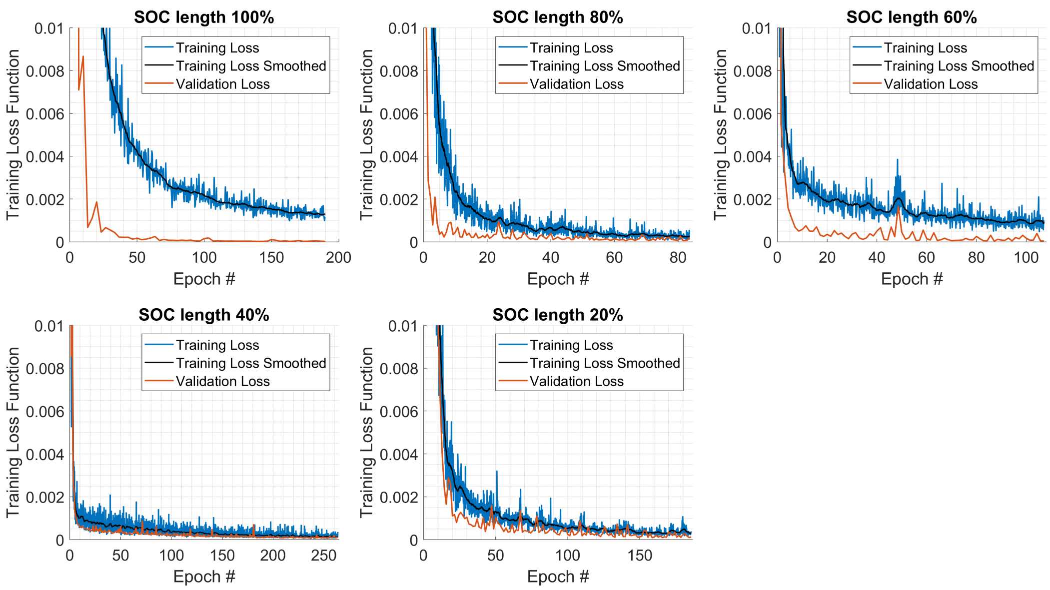


Fig. A.4 Trend in the training loss function for each SOC window length.

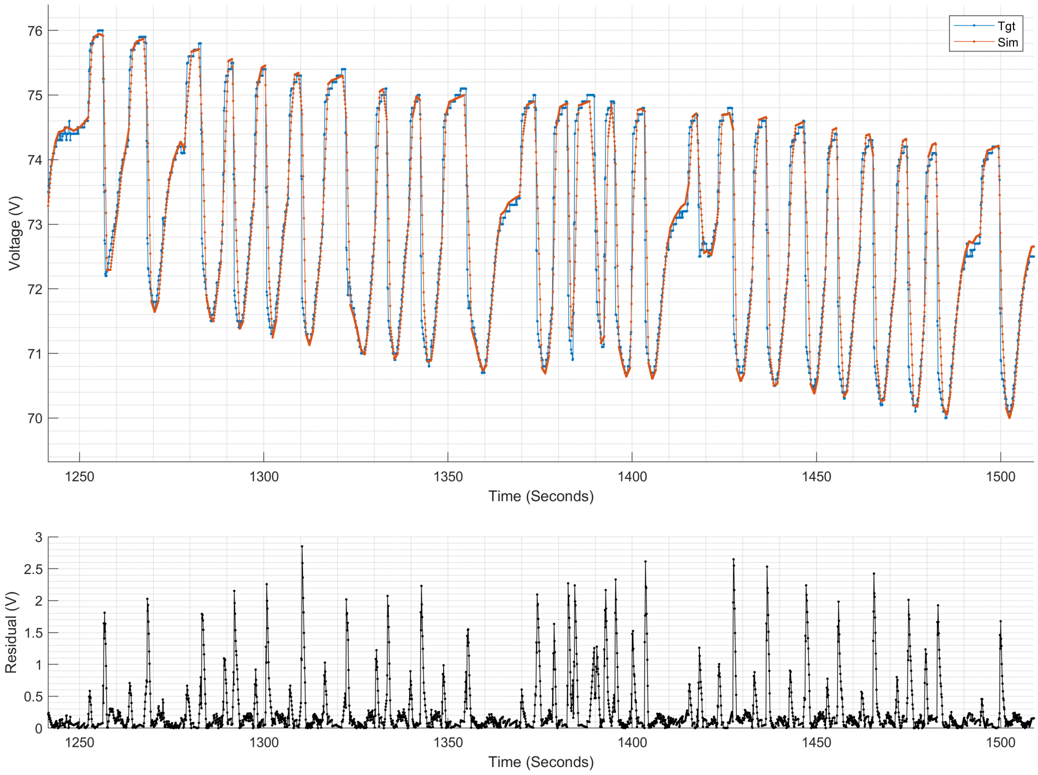


Fig. A.5 Simulated against the measured battery voltage with the puntual error.

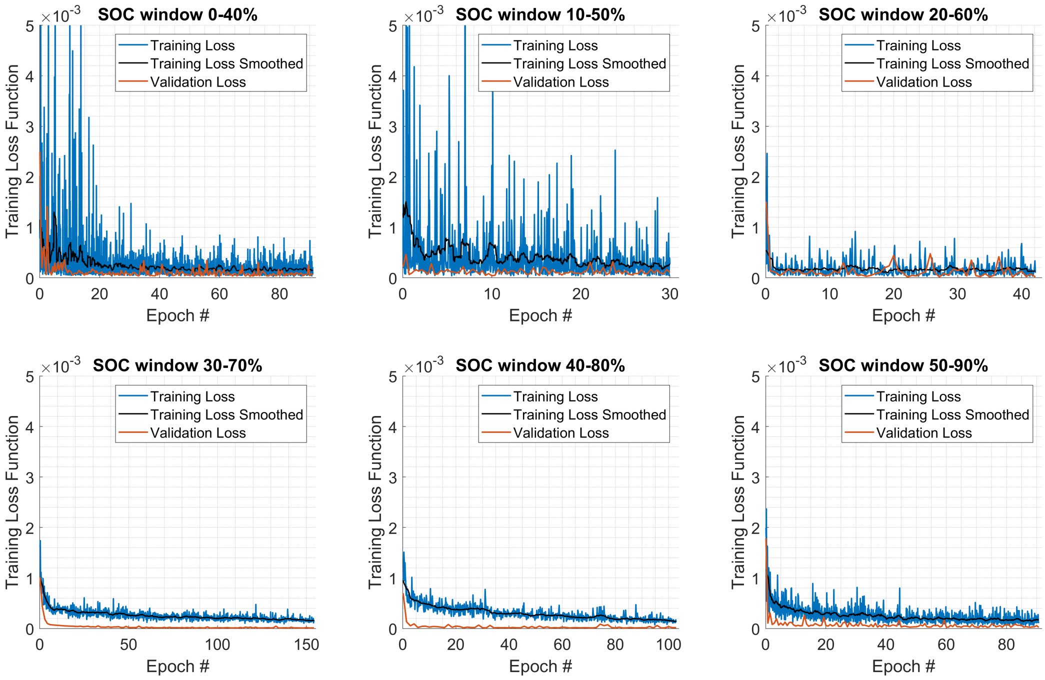
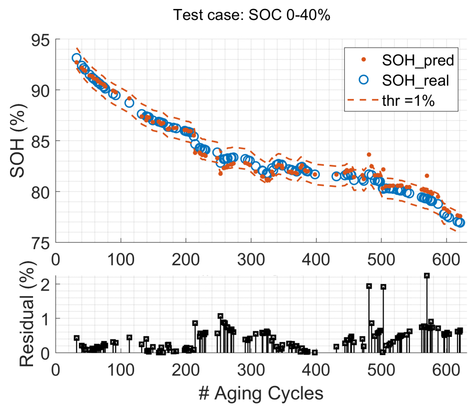
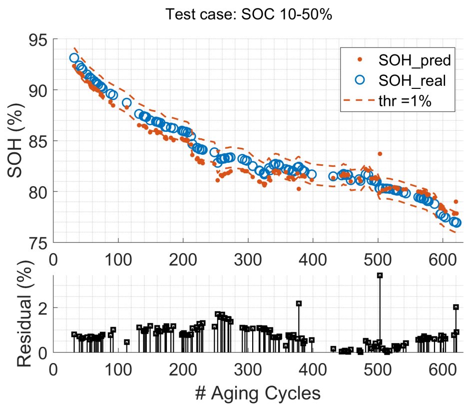


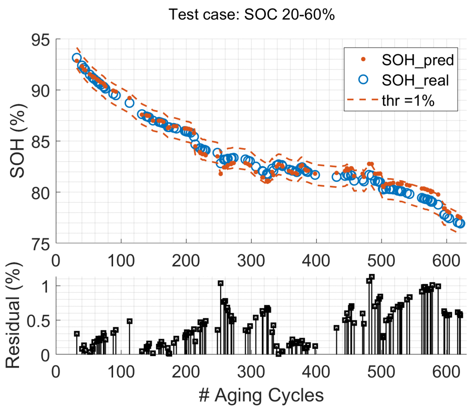
Fig. A.6 Loss function training for SOH prediction across all SOC domain span through optimal SOC window length of 40%.



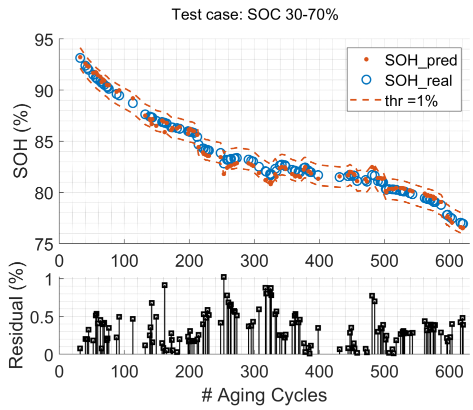
(a) SOC segment [0,40]%



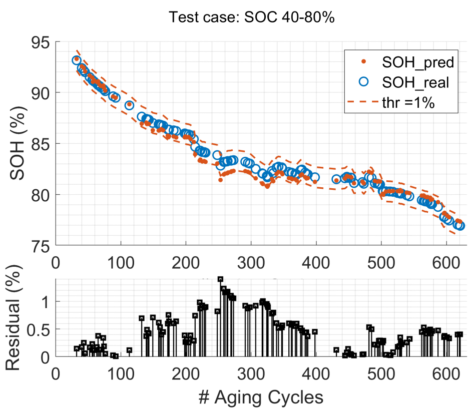
(b) SOC segment [10,50]%



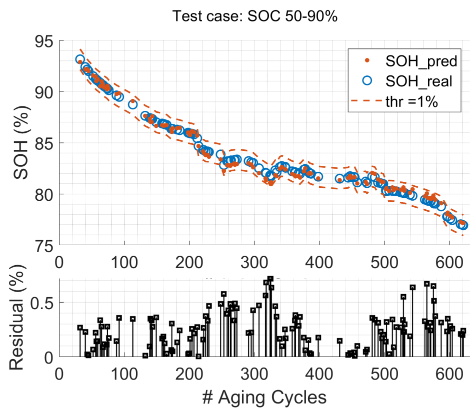
(c) SOC segment [20,60]%



(d) SOC segment [30,70]%



(e) SOC segment [40,80]%



(f) SOC segment [50,90]%

Fig. A.7 Prediction performance of SOH estimation across all SOC domain span.