UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

MONALISA CRISTINA MOURA DOS SANTOS

A PROPOSED METHODOLOGY USING ARTIFICIAL INTELLIGENCE TO ESTIMATE THE STATE OF CHARGE FOR BATTERIES OF ELECTRIC VEHICLE

Recife
2020

MONALISA CRISTINA MOURA DOS SANTOS

A PROPOSED METHODOLOGY USING ARTIFICIAL INTELLIGENCE TO ESTIMATE THE STATE OF CHARGE FOR BATTERIES OF ELECTRIC VEHICLE

The master thesis presented to UFPE for the master's degree attainment as part of the requirements of the Programa de Pós-Graduação em Engenharia de Produção.

Research Field: Operational Research.

Advisor: Profa. Dra. Isis Didier Lins.

Recife

2020

Catalogação na fonte
Bibliotecário Gabriel Luz, CRB-4 / 2222

S237p Santos, Monalisa Cristina Moura dos.
A proposed methodology using artificial intelligence to estimate the state of charge for batteries of electric vehicle / Monalisa Cristina Moura dos Santos – Recife, 2020.
81 f.: figs., tabs., abrev. e siglas.

Orientadora: Profa. Dra. Isis Didier Lins.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção, 2020.
Inclui referências.
Textos em inglês.

1. Engenharia de Produção. 2. Veículos elétricos. 3. Estado da carga. 4. Machine learning. 5. Prognóstico. 6. Health management. I. Lins, Isis Didier (Orientadora). II. Título.

UFPE

658.5 CDD (22. ed.) BCTG / 2020-263

MONALISA CRISTINA MOURA DOS SANTOS

A PROPOSED METHODOLOGY USING ARTIFICIAL INTELLIGENCE TO ESTIMATE THE STATE OF CHARGE FOR BATTERIES OF ELECTRIC VEHICLE

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Engenharia de Produção.

Aprovada em: 18/02/2020.

BANCA EXAMINADORA

_____

Profª. Drª. Isis Didier Lins (Orientadora)

Universidade Federal de Pernambuco

_____

Prof. Dr. Márcio das Chagas Moura (Examinador Interno)

Universidade Federal de Pernambuco

_____

Prof. Dr. Flamarion Borges Diniz (Examinador Externo)

Universidade Federal de Pernambuco

I would like to offer this work to my dear grandmother, Isaura, who is a fighter and helped me, enormously, in my growth as person and professional.

ACKNOWLEDGEMENTS

I would like to thank, firstly, God for allowing me to conclude this work besides all difficulties that I had through my journey – this work helped me to know myself more, professionally and personally. I would also like to offer my gratitude to my advisor, Professor Isis, who was always patient through this process and helped me to build and review this work.

Secondly, I would like to thank my parents, especially my mother – who, besides all the bad moments I had this year, was always supporting me. To my father – who not always comprehend my job but gives me support in his away. To my grandparents, Isaura and Marliberto; they always gave the support I needed. To my dear dog Leia, which gave me unconditional love in times I was doubting myself.

To my co-advisor Professor Márcio, who is always pulling up me and my friends at CEERMA. An especial thanks to Leonardo, who helped me through the process, giving opinions and suggestions to build this work. He was a very good friend and supported me as fewer friends ever had. I would like to thank July, for being available to help me. Also, Hiago, Rafael and Thais, dear friends of mine, that supported me when I thought I was going to finish this research. Finally, to all CEERMA team.

I would like to thank Professor Flamarion for his availability to discuss topics involving chemistry and physics that I had difficulties at the early stages of this work.

Finally, I would like to thank PPGEP and CNPq for academical and financial support.

“I am no longer accepting the things I cannot change. I am changing the things I cannot accept”. (Angela Davis, 1981).

ABSTRACT

In order to decrease the emission of greenhouse gases and propose alternatives to the environmental effect of it, the development and improvement of "green technologies" have received special attention due to their utility to prevent the impacts caused by those gases. Thus, electric vehicles (EVs) were, also, an important advancement in this area. To work, the EVs need a reliable battery source and, for most EVs, a lithium-ion battery is used as a power source. Some advantages of lithium-ion batteries are high specific energy density, high cycle life, and low self-discharge. In the context of Prognostic and Health Management (PHM), estimation of the SOC (State of Charge) – which is the remaining charge within the battery and is defined as the ratio of the residual capacity of the battery to its nominal capacity – based on data-driven methods (e.g. Machine Learning – ML, Deep Neural Networks – DNN) and data storage (e.g. Big Data) has come as a suitable alternative to identify patterns in its degradation over time, also being much less time-consuming than physics of failure (e.g. coulomb counting and open circuit approaches) methods, which needs full discharging to estimate SOC. In this work, a methodology using DNN and Machine Learning (ML) algorithms is proposed to predict battery SOC. At first, the input – current and voltage – and the output – SOC – each given in the form of time series, are replicated using Maximum Entropy Bootstrap (MEB), a sampling technique used with non-stationary time series- this technique is used to further compute confidence interval of the remaining time until the next recharge. Afterward, the input dataset is processed using a windowing model as the pre-processing step; this processed dataset is used to train a DNN model. For purposes of comparison, the data is also fed into an ML model, with each replication training the model. Following the training phase, the predicted SOC, for both the DNN and ML model, is filtered by an Unscented Kalman Filter (UKF), which processes the predicted SOC time series in terms of its mean and covariance. Then, the remaining time until the next recharge is computed and compared with the real discharge time. Finally, the confidence interval of the remaining time until the next discharge is calculated for the DNN and ML models. Analyzing the results, the DNN model, which is performed by the Multi-Layer Perceptron, has better results compared with the other applied methods – Support Vector Machines, Random Forest and XGBoost – with lower root mean squared error results and percentage errors for the remaining time until the next discharge – for both non and post-processed results. These results are achieved due to the complexity of the DNN model. However, further analysis in terms of the number of layers for the DNN method needs to be

operated. For the Random Forest and XGBoost methods, which obtain the worst results, they are, generally applied for classification tasks, explaining the observed results.

Keywords: Electric vehicles. State of charge. Machine learning. Prognostic. Health management.

RESUMO

A fim de diminuir a emissão de gases de efeito estufa e propor alternativas para os efeitos no meio ambiente causado pelos menos, o desenvolvimento de "tecnologias verdes" tem recebido uma atenção especial devido a sua importância para prevenir e evitar os impactos causados por esses gases. Então, os veículos elétricos (VEs) são um avanço nessa área. Para funcionar, os VEs precisam de uma bateria que seja confiável, sendo as baterias de lítio as mais utilizadas como fonte de energia. Algumas vantagens que as células de lítio possuem são a alta densidade energética específica, um alto ciclo de vida e baixa auto descarga. No contexto de *Prognostic and Health Management* (PHM), a estimação do estado da carga – que é a carga remanescente da bateria definida pela razão da capacidade residual da bateria e a capacidade nominal – baseada em métodos conduzidos por dados (e.g. *Machine Learning* – ML, *Deep Neural Networks* – DNN) e armazenamento de dados (e.g. *Big Data*) vem como uma alternativa para identificar padrões de degradação através do tempo, sendo um procedimento que usa menos modelos baseados em física de falha (e.g. contagem de coulomb e métodos de circuitos abertos), eles precisam da descarga total para realizar a estimativa do estado da carga. Nesse trabalho, uma metodologia usando DNN e ML foi proposta para fazer a estimativa do estado da carga. Primeiramente, a entrada – corrente e voltagem – e a saída – estado da carga – cada uma dada em formato de série temporal, e, posteriormente, replicada utilizando o *Maximum Entropy Bootstrap* (MEB), sendo utilizada para realizar a estimativa do intervalo de confiança do tempo até a próxima descarga. Depois, os dados de entrada são processados utilizando um modelo de janelamento como etapa de pré-processamento; estes dados pré-processados são utilizados para treinar o modelo DNN. Depois da previsão do estado da carga, os resultados dos modelos de DNN e ML serão filtrados utilizando o *Unscented Kalman Filter* (UKF), que processa a série temporal do estado da carga previsto em termos da sua média e covariância, Então, o tempo restante até a próxima descarga é calculado e comparado com o tempo de descarga real. Finalmente, o intervalo de confiança do tempo restante até a próxima descarga é computado para o DNN e o ML. Analisando os resultados e comparando com outros métodos utilizados, o modelo DNN, representado por um *Multi-Layer Perceptron*, obteve os melhores resultados se comparados com os outros métodos aplicados – *Support Vector Machines*, *Random Forest* e *XGBoost* – com um menor erro médio quadrático e um erro percentual menor para o tempo remanescente até a descarga – para os resultados sem e com pós-processamento. Esses resultados foram alcançados devido a complexidade do modelo DNN. Entretanto, análises posteriores necessitam ser executadas em termos do número de camadas da DNN, entre

outros parâmetros. Para o *Random Forest* e o *XGBoost*, que obtiveram os piores resultados, eles são, geralmente, aplicados para tarefas de classificação, o que explica, em parte, os resultados obtidos.

Palavras-chave: Veículos elétricos. Estado da carga. *Machine learning*. Prognóstico. *Health management*.

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACF	Autocorrelation function
AI	Artificial Intelligence
API	Application Programming Interface
AR	Autoregressive
BMS	Battery Management System
BPTT	Backpropagation Through Time
CALCE	Center for Advanced Life Cycle
CNN	Convolutional Neural Networks
CV	Cross-Validation
DFS	Deep Feedback Theory
DST	Dynamic Stress Test
ECM	Equivalent Circuit Models
EKF	Extended Kalman Filter
EVB	Electrical Vehicle Battery
EVs	Electric Vehicles
FC	Fully Connected
FUDS	Federal Urban Driving Schedule
GFT	Global Feedback Theory
GHG	Greenhouse Gases
GPU	Graphics Processing Units
GRB	Gaussian Radial Basis
GRU	Gated Recurrent Unit
HEV	Hybrid Electric Vehicles
i.i.d.	Independent and Identically Distributed
ICC	Internal Combustion Car
KF	Kalman Filter
Li-ion	Lithium-ion
MEB	Maximum Entropy Bootstrap
ML	Machine Learning
MLP	Multi-layer perceptron
MSE	Mean Squared Error

NN	Neural Networks
PHM	Prognostic and Health Management
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Networks
RUL	Remaining Useful Life
SEI	Solid Electrolyte Interface
SGD	Stochastic Gradient Descent
SOC	State of Charge
SOH	State of Health
SVM	Support Vector Machines
SVR	Support Vector Regression
TS	Time Series
UKF	Unscented Kalman Filter
USABC	US Advanced Battery Consortium
UT	Unscented Transformation
WKK	Wiener Kolmogorov-Khintchine

SUMARY

1 INTRODUCTION

Electric Vehicles (EVs) use rechargeable batteries and electric energy as supply; thus, they are an important development in the automobile industry. If clean energy is used, the EV is an alternative to reduce the consumption of fossil fuels and, therefore, reduce pollution caused by these energy resources. EVs include road and rails vehicles, surface and underwater vessels, electric aircraft and spacecraft (HANNAN *et al*., 2017).

The first electric car, a prototype, was built in 1835 by professor Sibrandus Straight of the University of Groningen, Netherlands (GUARNIERI, 2012). However, the improvement of the road infrastructure and the discovery of large petroleum storage made the industry of internal combustion cars (ICC) cheaper than the electric ones and another point to consider was the low potential of energy storage of the batteries from this epoch. Thus, it was hard for the EVs to attend long distances due to the low capacity of batteries back then in opposite of ICC (LOEB, 1885).

Those vehicles use one or more electric motors or traction motors for propulsion and can be powered either by an electric collector system from off-vehicle source or a self-contained system with a battery, solar panels or an electric generator which converts fuel to electricity. Despite the many power sources for the EVs, the most common ones are batteries. Lithium, nickel-cadmium, lead-acid and alkaline are some of the types of electric batteries which can supply those vehicles, but lithium-ion batteries are the most promising ones because of its features such as high energy density, long life cycle, high efficiency and sustainable performance (HE *et al.*, 2016) (LU *et al.*, 2013) (CHENG *et al.*, 2011).

The battery is a critical component of an EV, so, for the entire electric system to work safely, it must have its reliability guaranteed. It requires continuous monitoring and control systems to prevent abnormal degradation and catastrophic failures. When the operational condition is not supervised properly, the equipment is susceptible to failure which can cause an explosion, fire, release of toxic gases, or other negative impacts (HENDRICKS *et al.*, 2015). As an example, in 2013 the occurrence of two separate lithium-ion battery incidents lead to a grounding of an entire fleet of a Boeing 787 (TOPHAM; SCOTT, 2013) (WILLIARD *et al.*, 2013).

Hence, to indicate the battery state and avoid its failure, a battery management system (BMS) works as a link between the battery and the vehicle. Using the concepts of Prognostics and Health Management (PHM) (ELATTAL *et al.*, 2016) - a technology used to monitor degradation in engineering systems and to provide information such as fault detection and

failure prognostics - , one of the concerns of BMS is to estimate the battery state of charge (SOC) and the state of health (SOH). SOC measures the amount of usable energy at the present cycle and the SOH denotes the remaining performance of the battery over its entire life cycle.

SOC acts as a “fuel gauge” in an EV, it shows how long the battery will be available before it is recharged. Also, it can ensure that batteries operate properly under the desired limits and consequently prolong the battery useful life preventing over-charging or over-discharging. However, battery SOC is inferred using observed variables such as current and voltage. Some SOC estimation techniques are: Coulomb counting via the integration of the loading current (KONGSOON et al., 2009) (ZHANG et al., 2014) (LENG et al., 2014), data-driven methods (e.g. Support Vector Machines (SVM) (VAPNIK; LERNER, 1963) (VAPNIK; CHERVONENKIS, 1964) (HANSEN; WANG, 2005), Neural Networks (NN) (HAYKIN, 1999) (ZHAO et al., 2018) (LI et al., 2014)), and physical model-based methods via equivalent circuit models (ECM) and electrochemical models (GOMEZ et al., 2011) (CHO et al., 2012) (RAHMAN et al., 2016) (STETZEL et al., 2015) (DI DOMENICO et al., 2008).

Data-driven methods, such as SVM and NN, are widely used with the purpose of estimating the SOC or SOH of a battery. These methods consider characteristics given by sensors such as temperature, voltage, current, and capacity. However, these sensors register a considerable amount of data, filtering and selecting the important features is an exhaustive job. To avoid this previous step, Deep Neural Networks (DNN) (LE CUN *et al.*, 1998) (LE CUN *et al.*, 2015) comes as a strategy due to its property to process and extract features from the data in terms of its spatial – Convolutional Neural Networks (CNN) (LE CUN *et al.*, 1998) and its derivations – and time – Recurrent Neural Networks (RNN) (RUMELHART *et al.*, 1986) and its derivations – characteristics (ZHAO *et al*., 2018) (CHEMALI *et al*., 2018).

Chaoui & Ibe-Ekeoch (2017) present an application of dynamically driven recurrent networks (DDRN) in online EV battery analysis. They aim to estimate the SOC and SOH, promising to have both good computational intelligence and robustness while maintaining some simplicity due to the global feedback theorem (GFT). Zhao *et al.* (2018) model a BMS of a lithium-ion battery using the gated recurrent unit (GRU), a type of RNN, and deep feature selection (DFS). These works illustrate how wide are the possibilities of using data-driven methods.

He *et al.* (2014) proposed an approach using NN and Unscented Kalman Filter (UKF) in order to predict the SOC. They mention how easy is to estimate SOC through a data-driven approach using only current and voltage as inputs. Then, to reduce the prediction error they applied UKF, which helps to smooth and predict time-series of nonlinear state-space systems.

UKF is also used to estimate and smooth the SOC in He *et al.* (2016), Partovibakhsh & Liu (2014) and Sun *et al*. (2011). On the other hand, data used to train these models is developed through simulation and, thus, data augmentation techniques can be explored as an approach to approximate better real-life scenarios.

In order to estimate the confidence interval, the maximum entropy bootstrap (MEB) is applied. This method replicates nonstationary time-series, which are, basically, the inputs (e.g. current, voltage, impedance) of the data-driven algorithm applied in this work (VINOD; LOPEZ-DE-LACALLE, 2009).

Thus, to avoid accidents and losing profits, BMS is essential to know the behavior of the studied system and to building appropriated maintenance policies. Hence, this work proposes an applied methodology using ML and DNN - which are data-driven and uses data from monitoring systems (e.g. current, voltage, temperature) to estimate SOC and other metrics related to BMS; the data-driven approach is generally used to continuous monitoring systems due to its practicality and it does not require full battery discharge as Coulomb Counting and models based of thermochemistry and thermodynamics data. Therefore, the proposed methodology uses data from a Li-ion battery monitoring system to estimate SOC and the remaining time until the next recharge. As a further step, the estimated SOC results are going to be post-processed using an Unscented Kalman Filter (UKF) to minimize the effect prediction error and the confidence interval of the remaining time until the next discharge is going to be calculated using Maximum Entropy Bootstrap (MEB).

1.1 Justification

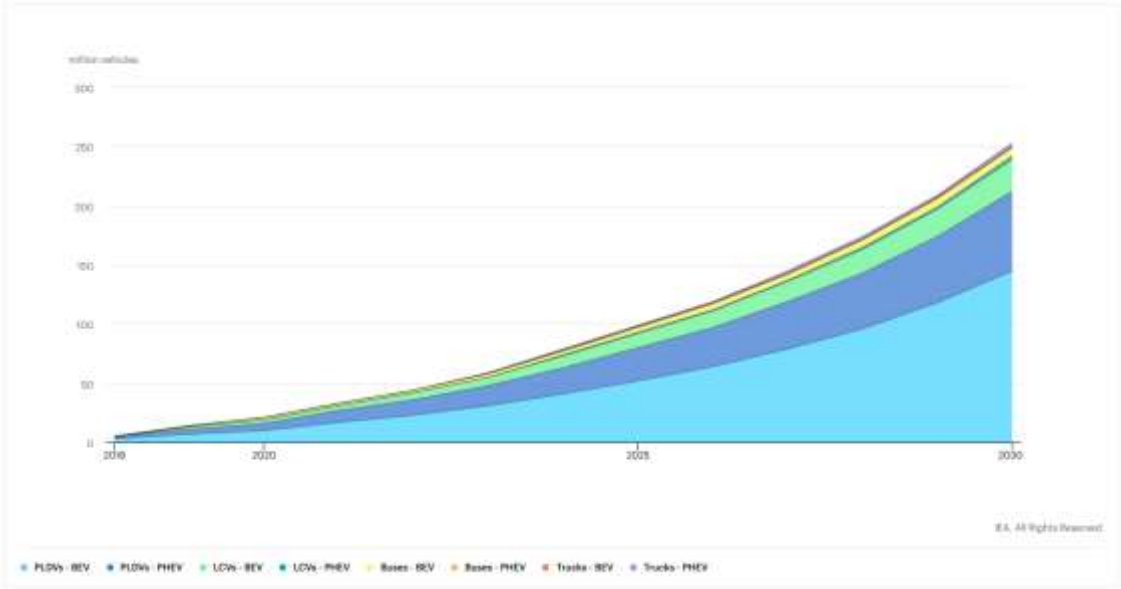
With the growing awareness of global warming and climate changes, the concern of the consumption of fossil fuels and vehicles based on this power source has grown due to GHG emissions (HANNAN *et al.*, 2014) (BUDZIANOWSKI *et al.*, 2012) (SULAIMAN, 2015). Furthermore, with the effects caused by these emissions (e.g. overheating, flooding and melting glaciers), the petroleum-based industry, including the automotive, is trying to develop alternative solutions that use other energy sources (POULLIKKAS, 2015) (HOFMANN *et al.*, 2016) (ABDUL-MANN, 2015) (CANALS CASALS *et al.*, 2016). Thus, EVs come as an alternative to reduce the impact of these pollutants.

Nowadays, lithium-ion batteries are the most common energy storage because of their advantages, such as high energy densities, cycling durability, no memory effect, and low self-discharge. However, some physical-chemical events such as thermal and electrochemical instability of the electrode and the flammability of the electrolyte can cause catastrophic

failures. When a battery fails, fires and explosions are some of the effects of these events causing accidents for the consumers.

Research on EV is mainly focused on increase vehicle efficiency, reduce price of components and develop methods for an effective and efficient charging system. A recent survey conducted by IEA (International Energy Agency) – Global EV Outlook 2019 – discusses recent developments in electric mobility around the world. This survey is supported by the Electric Vehicles Initiative and combines projections, policy recommendations and other economics perspectives. According to this survey, electric mobility is expanding rapidly; in 2018, for example, the global EVs fleet exceeded 5.1 million – almost doubling the number of new sales in this field, being China the largest market, having 45% of the fleet, followed by Europe, with 24%, and United States, with 22% (Figure 1).

Figure 1 - Number of EVs with a projection until 2030.

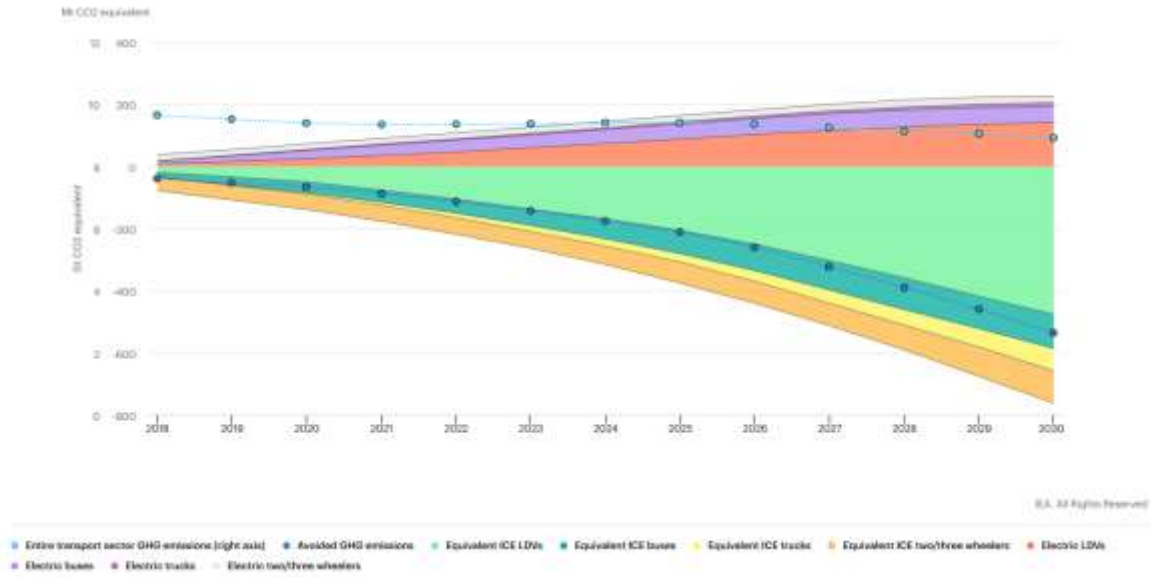


Source: IEA, 2017.

This research also does an outlook of growth of EV – passenger light duty vehicles, light commercial vehicles, buses and trucks – with and without hybrid systems; thus, around 300 million vehicles will compose the fleet of EVs. However, even if all these advantages are considered, only a few countries are fully included in the EV market, being the limiting the high cost and the production numbers. Considering the producers perspective to reduce the impacts limitations, they set a production target (IEA, 2017; DAS *et al.*, 2019), and according to the same survey – from IEA – the production is set to 548 million by 2040 (DAS *et al.*, 2019).

The evolution of GHG emissions from EV fleet is given by the combined evolution of the energy used by an EV and the carbon intensity of electricity generation; thus, if the grid becomes less carbon intensive, also the EVs are. Even though is well-known that EVs are a progress in decreasing GHG emissions, it requires a decarbonization of power systems (Figure 2) (IEA (2017)).

Figure 2 - A well-to-wheel net and avoided GHG emissions form EV fleets by transport sectors based on the IEA scenario – until 2030.



Source: IEA, 2017.

To build a PHM strategy, some steps are required including the choice of the method that will process the data. Due to the increasing power of data storage and computer processing, ML methods (including DNN and SVM) - with the right data preprocessing - are arising due to their capability of processing a large amount of data while also being precise during machine prognosis and diagnosis. These methods are widely used due to the possibility of updating continuously the model and the predictions using new data and training. Also, compared to other models such as coulomb counting, adaptative filters (e.g. UKF) and thermochemical models – which requires the full discharge of the system or accumulates errors through the predictions – a data-driven approach is constantly trained and updated, which allows precision and effectiveness during the equipment's operation.

1.2 Objectives

In the following topics are presented the General Objective and the Specific Objectives.

1.2.1 General Objectives

The main purpose of this study is to develop a data-driven prognostic strategy comparing different DNN and ML methods to estimate the SOC of lithium-ion batteries and compare its results and performance.

1.2.2 Specific Objectives

a) Literature review: a study about battery management systems (BMS), types of batteries, how to proceed with BMS studies, DNN and ML.
b) Data collection: search for EVs batteries data sets, with the focus being on prognostics data sets;
c) Feature selection: analyze what type of metrics (e.g. current, voltage, temperature) are important to develop the DNN and ML models from the selected battery dataset;
d) Data Augmentation: perform the data augmentation using MEB;
e) Algorithm implementation: implement the DNN and ML algorithms to estimate the SOC;
f) Test the model efficiency: compare different models, their metrics and prediction results, considering their processing (train and test) time and choose the more efficient one;
g) Estimate SOC and its prediction error;
h) Calculate the confidence interval of the remaining time until the next discharge using Maximum Entropy Bootstrap (MEB);
i) Improve the output (SOC) and predicted time through a filtering strategy such as Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF).

1.3 Methodology

Conforming to Fontelles *et al.* (2009), this research can be classified as applied. The approach is considered quantitative research owing to work with data that can be translated in numbers to be classified and evaluated. According to the objectives, the study can be classified as explanatory. As technical procedures, this work is documentary research.

This study aims to develop a BMS focused on prognostics of EVs batteries based on DNN and ML to estimate SOC. Firstly, the dataset from train and test is going to be replicated to calculate the confidence interval and perform data augmentation for the DNN network. Then, an autoregressive model is going to work as a preprocessing step, and, for the DNN network is

going to, also, work as data augmentation. After the preprocessing step, the ML and DNN algorithms are going to be trained and this training process is going to be evaluated. As a result, the SOC and the remaining time until the next discharge is going to be estimate and evaluated. To achieve this purpose the study will involve the following steps:

a) Literature review;
b) Data collection;
c) Feature selection;
d) Feature extraction;
e) Algorithm implementation;
f) Test model efficiency;
g) Estimate SOC;
h) Compare models;
i) Process the output;
j) Evaluate the methodology;
k) Write the master's thesis.

1.4 Thesis' Structure

The content of the following chapters of this dissertation are briefly described below:

Chapter 2: the theoretical background and literature review of essential concepts related to BMS, ML, DNN, UKF, and MEB.

Chapter 3: gives a detailed description of the proposed methodology for predicting SOC;

Chapter 4: presents the results of the implementation of the proposed method on a battery dataset;

Chapter 5 provides some concluding remarks and comments about future works.

2 THEORETICAL BACKGROUND AND LITERATURE REVIEW

In this section, are presented the following topics: Battery Electric Vehicles – Lithium-ion Batteries and Battery Management System -, Machine Learning – Neural Networks and Multi-layer Perceptron, Support Vector Machine, Random Forest Regression and Extreme Gradient Boosting Regression -, Unscented Kalman Filter, Maximum Entropy Bootstrap and a brief Literature Review.

2.1 Battery Electric Vehicle

According to 2016, Special Report Energy and Air Pollution made by World Energy Outlook, in 2015 half of all pollution caused by nitrogen oxide were attributed to the transportation sector. Also, 50% of the overall health-related economic cost (about $865 billion) is due to air pollution. Norway and some other countries are acting in order to prevent the negative effects by prohibiting new fossil-based vehicles by 2030. Thus, electric vehicles (EVs) and hybrid electric vehicles (HEVs) (Figure 3) are one of the promises of green travel where fuel-based transports will be gradually substituted by electric-based ones reducing the dependence and consumption of fossil energy which reduces the GHG emission. These vehicles can be easily charged with a structure that can be plugged in the streets with power source stations at parking lots during the day or overnight at home with energy generated by a power station or renewable energy. However, the key component is the battery which can store the energy that will help using EVs (REN *et al.*, 2017).

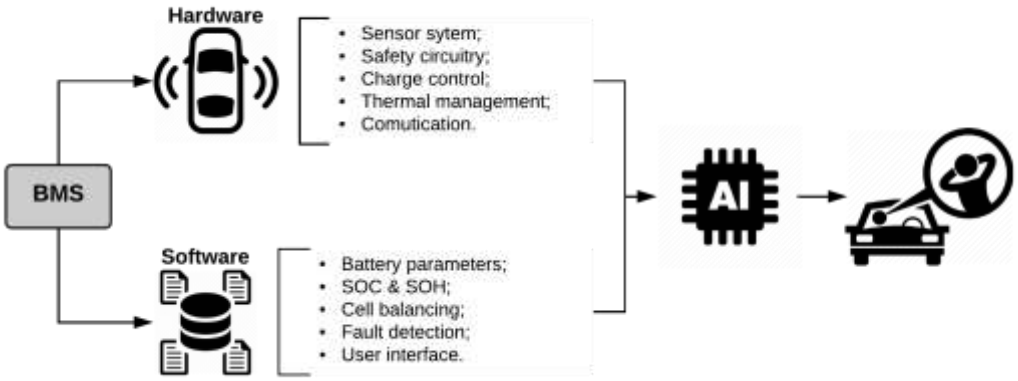
Figure 3 - The charging station of EVs

Source: LOFGREN, 2013.

Li-ion batteries are the common choice for portable electronic devices such as cellphones, personal computers, tablets as well as EVs because of its high energy-to-weight and power-to-weight ratios and low self-discharge rates (LINDEN; REDDY, 2002) (CAPASSO; VENERI, 2014). Also, considering all rechargeable electrochemical systems, Li-ion batteries are the ones chosen as power source of EVs and HEVs. Even though these batteries have a certain number of advantages, issues such as safety, cost, recycling, and infrastructure are still a concern nowadays.

In order to ensure the safety of the battery, a battery management system (BMS) is necessary for an EV, and it is composed of sensors, controllers, actuators, which are controlled by algorithms, and signal. According to Xiang *et al.* (2011), BMS has two structures (Figure 4): hardware and software. The hardware part is basically the sensors and controllers which guarantee to measure, control and communicate the state of the battery. The software part is responsible for controlling the hardware operation, making decisions and estimating the states of all sensors; furthermore, it performs the data analysis determining fault identification and state estimation – including estimating SOC – helping the user to extract information from the battery health condition through an interface.

Figure 4 - BMS elements to identify system failure.



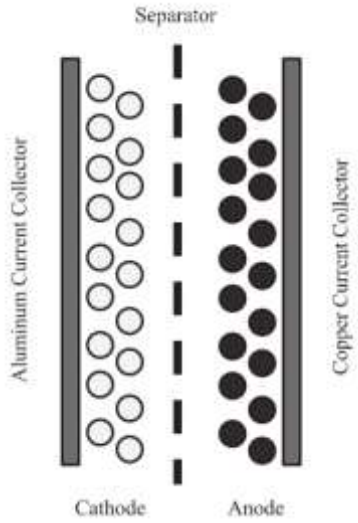
Source: adapted from Hannan *et al.*, 2017.

2.1.1 Lithium-Ion Batteries

The first lithium-ion (li-ion) battery was made in 1962 and it was composed of a negative electrode and a positive electrode made of manganese dioxide; however, the first rechargeable battery was released in 1985 by the Moli Energy. This kind of battery was successfully used in 1991 by SONY in mobile devices due to its high energy density, high

operating voltage, and low self-discharging rate. Besides, it is found in different types of models such as cylindrical, coin, pouch and prismatic (Figures 7 to 10) (Figure 5). In the automotive industry, li-ion batteries are applied in different kinds of hybrid vehicles, plug-in hybrid vehicles, buses and trucks and electric vehicles (DU, CAO, ZHANG, 2018).

Figure 5 - Battery structure.



Source: Cui *et al*., 2017.

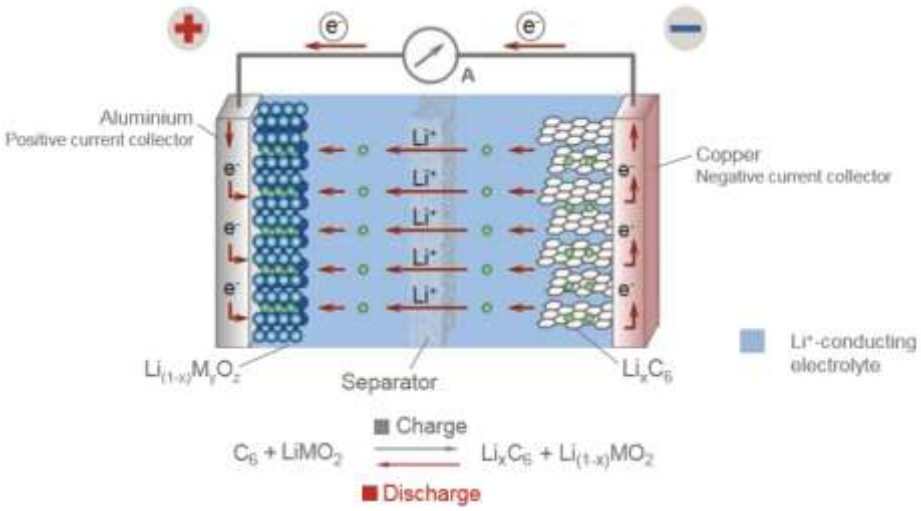
A lithium-ion battery can be charged and discharged repeated times; though, the amount of electrical charge that it can store decreases over time. It happens due to different degradation phenomena such as chemical side reactions or loss of conductivity (WRIGHT *et al.,* 2003) (WANG *et al.,* 2015). Also, catastrophic failures can occur unexpectedly as a result of fire or explosion and this often happens owing to overstress conditions (the battery operates outside the recommended conditions of current, voltage or temperature). So, this type of equipment requires continuous monitoring and control, i.e., a good BMS (KORTHAUER, 2018).

Figure 6 shows how a rechargeable li-ion battery works. At first, an ion electrolyte with dissociated lithium conducting salt, which is a conductor, is between the two electrodes. A separator - a porous membrane that electrically isolates that allows ion transportation (HENDRICKS *et al.,* 2015). Consequently, lithium ions flow between the electrodes of the battery during the charging and the discharging process.

The discharging process is when the lithium from the negative electrode (anode) is deintercalated and electrons are released. Thus, these electrons carrying electricity migrates via an external connection (a cable) to the positive electrode (cathode). The anode can be made of graphite or carbon compounds and the cathode can be made of mixed oxides. When the

charging occurs, this process is reversed. Depending on the application, how batteries are designed is important to guarantee efficiency, reliability, and safety on the vehicle. A single battery can be used (e.g., cellphones) or multiple batteries are organized in series or parallel.

Figure 6 - How a rechargeable li-ion battery works.



Source: Korthauer, 2018.

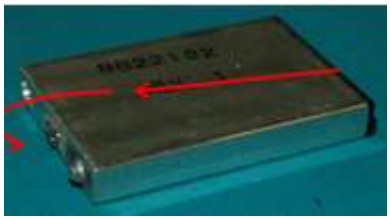
Figure 7 - Cylindrical cell.

Source: CALCE, 2014.

Figure 8 - Coin cell.

Source: BBC- UK.

Figure 9 - Prismatic cell.



Source: CALCE.

Figure 10 - Pouch cell.

Source: CALCE.

The batteries can also be modeled using two types of design: block and modular design (Figures 11 and 12). On one hand, the block design and the storage components, altogether, build a single block with structures such as collectors, sensors and other components. On the other hand, in the modular design, some individual cells are combined to build a sub-unit, which are used in the battery units, that constitute a module.

Figure 11 - Li-ion battery system with block design

Source: KORTHAUER, 2018.

Figure 12 - Li-ion battery system with a modular design.

Source: KORTHAUER, 2018.

The useful life of a battery relies on how the system operates, how they are made and how good the production process is. Hence, the BMS works as a condition controller that monitors the cell system (e.g. voltage, current, and temperature), allowing safety by using on and off systems and controlling the temperature with cooling and heating systems.

2.1.2 Battery Management System

One of the functions of a BMS is to estimate the SOC which is the remaining charge capacity of the battery. The SOC is basically the ratio between the residual capacity and the nominal capacity and when estimated, it is dependent on the temperature and on the current (CHEMALI *et al*., 2016) (AHMED *et al.*, 2014) (LI *et al*., 2017). The capacity is a measurement of how much electric charge a power source can give under certain discharge conditions; also, it depends on the discharge current, the cut-off voltage, the temperature and the type of material which it is made of.

As already defined before, SOC is the relation between the remaining capacity and the nominal or the maximum capacity of the battery. It is represented by Equation 2.1, for continuous measurements, being: SOC_t the present SOC, SOC_0 the initial SOC, $I_{L,\tau}$ the instantaneous load current, η_i the Coulomb efficiency and C_a the maximum available capacity (KORTHAUER, 2018).

$$SOC_t = SOC_0 - \int_0^t \frac{\eta_i I_{L,\tau} d\tau}{C_a} \tag{2.1}$$

For discrete measurements, a sample period $\triangle t$ is considered and Equation 2.2 represents the SOC in each time interval k.

$$SOC_k = SOC_{k-1} - \eta_i I_{L,k} \frac{\Delta t}{C_a} \quad (2.2)$$

The SOC can be estimated through a diverse number of methods from chemical methods (e.g. Open Circuit Voltage – OCV) to data-driven methods (e.g. SVM, RF); thus, some of them can be emphasized. However, due to, firstly, try to estimate this metric without experiments, ML models are used – because of its data-driven characteristics; also, this approach can be rearranged to fit new monitoring data, which allows parameters recalculating and model updates. To compute the SOC, the following equation was used: $SOC_k = 1 - \frac{Capacity_k\,(Ah)}{Nominal\ Capacity}$, being the capacity in $t = k$ and the nominal capacity given by the chosen dataset.

- Open Circuit Voltage (OCV) method: a method to estimate SOC using an open circuit. In order to do the estimative, a linear approximation is used between SOC and OCV, being this relationship depend on the type of battery (SNIHIR et al., 2006) (DONG et al., 2011) (TANG et al., 2015) (ZHENG et al., 2016). It depends on capacity and material of the battery. Even though it is a simple method with high precision, the main disadvantage of this method is that is takes a long time to reach for the reaction to reach the equilibrium. Hence, for purposes of testing driving profiles, this method is only applied when the vehicle is placed in a parking, which do not allow the evaluation during operational time.
- Coulomb Counting method: being the easiest way to estimate SOC, the Coulomb counting method is implemented with low computational power and is based on the integration of battery current with respect to time while the battery is charging or discharging (ZHANG et al., 2014) (LENG et al., 2014) (KONGSOON et al., 2009). However, this method can incur in inaccuracies caused by uncertain disturbances and variables such as noise, temperature and current. Other difficulty that was found is the determination of the initial SOC causing a cumulative effect. Also, this method needs the complete discharge of the cell and periodic capacity calibration to achieve the maximum capacity, which decrease the lifespan of the battery.
- Model-based SOC estimation: since the OCV cannot do online estimation and needs a long pause to monitor the SOC, this method does not have a good performance while the vehicle is operating. Therefore, a model is built to estimate SOC. These models are, mostly, electrochemical or uses equivalent circuit model (GOMEZ et al., 2011) (CHO et al., 2012) (RAHMAN et al., 2016) (STETZEL et al., 2015) (DI DOMENICO et al., 2008); they are, also, used to analyze the

performance related to the internal material and consider the effects of electrodynamics and thermodynamics. One of the disadvantages of this model is the lack of detailed explanation on the electrochemical reactions of the battery; also, is a complex and not easy to generalize.

- Adaptative filter algorithm (e.g. Kalman Filter, Extended Kalman Filter and Unscented Kalman Filter): The Kalman Filter (KF) is widely used to estimate the dynamic state of charge of the battery (XU et al., 2012) (TING et al., 2014) (URBAIN et al., 2007) (YATSUI et al., 2011) (HU et al., 2013) (PLETT et al., 2004, Parts 1, 2 and 3) (LEE et al., 2007) (CHEN et al., 2013) (MASTALI et al., 2013) (ZHU et al., 2012) (HE et al., 2011) (XIONG et al., 2013) (HE et al., 2012) (HE et al., 2013). This method filters the parameters from observation with noise having a self-correcting nature, which helps to tolerate a high variation of current. However, KF is not able to perform accurate estimative in non-linear application; thus, the Extended Kalman Filter (EKF) is applied. The main difference between these two methods is that the EKF uses partial derivatives and first order Taylor series expansion to linearize the battery model. Even though, EKF has a better performance compared to KF, the linearization error can occur if the system in highly non-linear condition. Considering that the EKF only operates in the first and second order of a non-linear model, which results in a significant error in non-linear state-space models, the Unscented Kalman Filter (UKF) come as an alternative to these problems. The UKF is a version of KF which applies discrete-time filtering algorithm and unscented transform (UT) to solve the filtering problem. Thus, this method is better than the EKF because it can predict accurately the system state until the third order of any non-linear system. For this reason, the UKF can perform the filtering without needing the computation of the Jacobian matrix and noise of the process does not need to be Gaussian; hence, performing in highly non-linear system. On the other hand, this method suffers from poor robustness because of the modeling uncertainty and system disturbances. These methods are, generally, applied as a recursive TS filter – predicting the next SOC – or as a preprocessing step, in hybrid methods, to minimize the effects of prediction error.
- Data-driven methods (e.g. Neural Networks, Support Vector Machines): these methods are based on condition monitoring data (SALKIND et al., 1999) (CHARKHGARD et al., 2011) (CHEN et al., 2011) (XUAN et al., 2011)

(ANTÓN et al., 2013). They work well to solve the nonlinearity problem; however, a large amount of data is needed to train and test these models. Machine Learning (ML) methods such as Neural Networks (NN) and Support Vector Machines (SVM) were applied (HE *et al.*, 2014), (LI *et al*., 2018).

- Hybrid methods: these methods are, basically, the combination of two or three algorithms in order to increase the efficiency and accuracy of the battery model (CHENG et al., 2011) (LI et al., 2013) (XU et al., 2009). This results in effective and reliable results, also decreasing the cost of the BMS. However, these methods have high computational cost and require memory from the device. This approach was the one chosen in this work.

2.2 Machine Learning

As can be seen in Kumar (2017), artificial intelligence (AI) techniques were developed in order to reproduce human characteristics such as perception, analysis, reasoning, and others (LI; ZHANG, 2017). For Jordan and Mitchell (2015), it aims to improve the performance measurements using the training step as a basis to model the technique. Thus, ML learn knowledge from the real world and represents it through its learning ability (PORTUGAL; ALENCAR; COWAN, 2018).

ML methods consider inductive inference to learn a set of examples from an environment, which can be used in supervised and unsupervised learning methods. The first step, the training, used a set of n examples (e.g. $D = \{(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \dots, (\boldsymbol{x_n}, y_n)\}$) to infer a function f. In supervised problems, the mapping function f learns the patterns of inputs and outputs and, generally, returns a function that represents the behavior of such data (CHAPELLE *et al.*, 2002) (LINS *et al.*, 2013). So, the learning problem can be defined based on the output:

Regression problems: assumes that the output are real values;

Classification problems: assumes that the output is a set of categories. It can characterize a binary classification when there are only two categories or a multi-class problem.

In this work, regression models of ML were used in order to predict the SOC of li-ion batteries in different temperature, current and voltage conditions.

2.2.1 Neural Networks and Multi-Layer Perceptron

A Neural Network (NN) has as main motivation the functioning of the human brain which has some typical characteristic such as high complexity, nonlinearity and the capacity of parallel processing; it also has the capacity to organize its basic structure, the neuron, in order

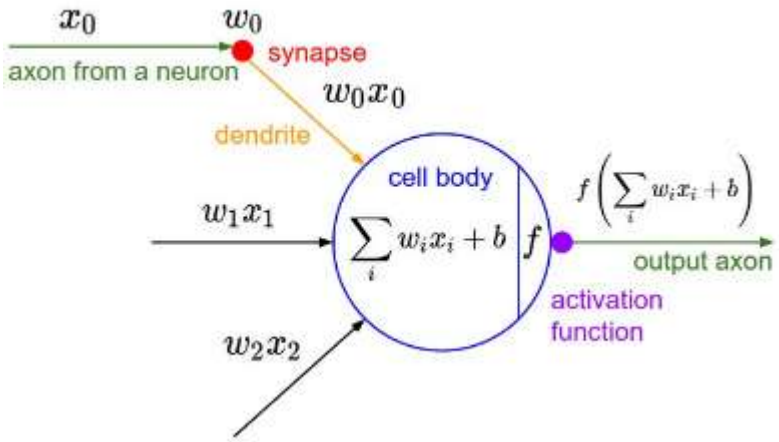
to perform activities like pattern recognition, perception, and motor control. Thus, an NN can be described as a heavily parallel distributed processor made of simple processing units that have the power to store knowledge through a learning process with weights that save this acquired knowledge (HAYKIN, 2008).

The NN has some properties and capabilities as nonlinearity – the artificial neuron can be either linear or nonlinear -, input-output mapping – which is the capacity to learn a unique input and link it with the desired output -, adaptivity – the capacity to change the synaptic weights and adapt it to the new environment, and other properties (HAYKIN, 1998) Hence, as an information-processing unit, the neuron, is basic for the fundamental operation of a NN and has a structure composed of:

- A set of synapses or connecting links which are characterized by a weight;
- A sum of the input signal and the weights by a linear combiner;
- And an activation function in order to restrict the output of the neuron.

NN whose simplest component is a neuron. Figure 13 shows how a neuron operates; it receives information as an input x_n, being $n = 1, 2, 3, ...$, and this input is multiplied by a weight w_n. Then, all these inputs are summed up and added to a bias b. This constitutes the cell body of a neuron; the activation function f will process this input and give the output that will enter the next neuron.

Figure 13 - Basic neuron mode.



Source: LI *et al.*, 2018.

Thus, the multilayer perceptrons (MLPs), is a variant of the perceptron model proposed by Rosenblatt (1950), come as an advancement of the NN theory and is consisted of a set of sensory units, known as nodes, which is the input layer, a set of hidden layers, which can be one or more, and an output layer to compute the results from the nodes. So, the input propagates through the network in a forward direction within each layer. The MLPs have several applications and are well-known for their backpropagation algorithm, which is based on the error-correction learning rule (HINTON, 2012).

The error back-propagation learning has two basic steps: the forward step and the backward step. The forward step is when the input data is applied to the nodes of the network and the effects of it are propagated layer by layer and a response is the output of the network. In this step, the weights are fixed. Then, in the backward step, the weights are adjusted obeying the error-correction rule and the output is subtracted from the desired value producing an error. Thus, this error is propagated back through the network and the synaptic weights are adjusted to predict better results. This algorithm is a so-called backpropagation (RUMELHART et al., 2016).

Given an input layer with n_0 neurons and input vector $X = (x_0, x_1, \dots, x_{n_0})$ and an activation function called sigmoid $f(x) = 1/1 + e^{-x}$, to compute the network output of each unit in each layer, it is necessary to consider a set of hidden layers $(h_1, h_2, \dots, h_N)$ and to presume that n_i is the number of neurons by each layer h_i. So, the output of the hidden layer is given by Equation (2.3):

$$h_i^j = f\left(\sum_{k=1}^{n_{i-1}} w_{k,j}^0 x_k\right) \qquad j = 1, \dots, \ n_i \tag{2.3}$$

And it can also be represented as (Equation (2.4)):

$$h_i^j = f\left(\sum_{k=1}^{n_{i-1}} w_{k,j}^{i-1} h_{i-1}^k\right) i = 2, \dots, N \; and \;\; j = 1, \dots, \;\; n_i \tag{2.4}$$

The $w_{k,j}^i$ is the weight between the neuron k in the hidden layer i and the neuron j in the hidden layer $i+1$, and n_i is the number of neurons in the ith hidden layer. So, the output can be represented as follows:

$$h_i = \left(h_i^1, h_i^2, \dots, h_i^{n_i}\right) \tag{2.5}$$

So, the network output is

$$y_i = f\left(\sum_{k=1}^{n_N} w_{k,j}^N h_N^k\right) \tag{2.6}$$

$$Y = \left(y_1, \dots, y_j, \dots, y_{N+1}\right) = F(W, X) \tag{2.7}$$

The $w_{k,j}^N$ is the weight between the neuron k in the Nth hidden layer and the neuron j in the output layer, n_N is the number of neurons of the Nth hidden layer, Y is the vector of the output layer, F is the transfer function and W is the matrix of weights, which is defined by the following formulation:

$$W = \left[W^0, \dots, W^j, \dots, W^N\right] \tag{2.8}$$

$$W^i = \left(w_{j,k}^i\right) \tag{2.9}$$

$$w_{j,k}^i \in \mathbb{R} \tag{2.10}$$

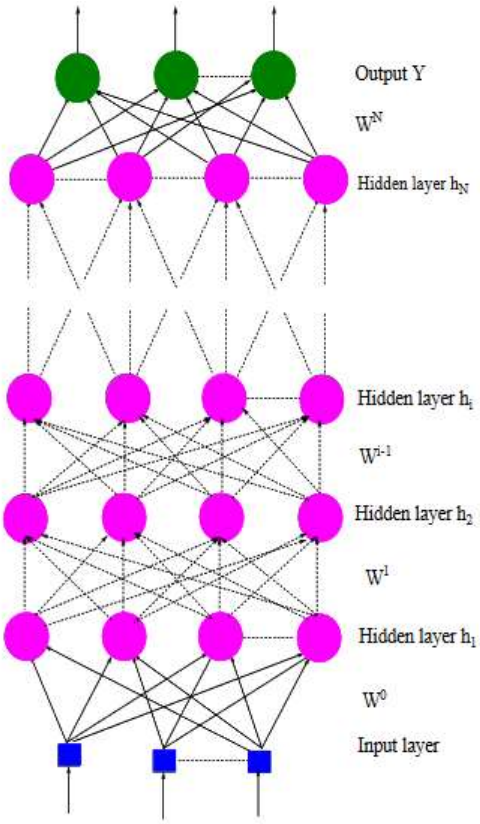
$$0 \leq i \leq N \tag{2.11}$$

$$1 \leq j \leq n_{i+1} \tag{2.12}$$

$$1 \leq k \leq n_i \tag{2.13}$$

For simplicity, $n = n_i, \; i = 1, \dots, N$ can be considered, the X as the input of the NN, f as the activation function, W^i as the matrix of weights between consecutive hidden layers, W^0 as the matrix of weights between the input layer and the first hidden layer and W^N as the matrix of weights between the last layer and the output layer. Figure 14 shows the scheme of an NN.

Figure 14 - MLP structure.



Source: RAMCHOU *et al.*, 2016.

2.2.2 Support Vector Machine

The SVM (Support Vector Machines) is a supervised and unsupervised learning method widely used to recognize patterns and regression. This algorithm was, firstly, developed in Russia on the 60's by Vapnik and Lerner (1963) and Vapnik and Chervonenkis (1964) and is based on statistical learning theory (SLT) developed by Vapnik and Chervonenkis (1974) and Vapnik (1995). In this work, a regression problem is considered, thus, specifically SVR (Support Vector Regression) is described below.

SVR is a version of SVM for regression problems. Thus, after the learning phase, the input data from the SVM are used in a quadratic mathematical programming problem which is convex (SCHÖLKOPF & SMOLA, 2002). The KKT (Karush Kuhn Tucker) conditions guarantee the global optimum, the first-order conditions are needed to solve the linear programming problem satisfying the regularization conditions (ZHAO & DIMIROVKI, 2004).

The principal aim is to find the hyper plan that best represents the regressors given a dataset D. The mapping representation of the data must be the most fitting in order to optimize the SVR objective function. Thus, the regression equation of the hyperplan is given by (Equation 2.14):

$$f(x) = w^T x + b \tag{2.14}$$

Being x the input data and $\mathrm{w^T}$ and b the coefficients that will be estimated by the following regularized risk function (Equation 2.15):

$$R(C) = C\frac{1}{m}\sum_{i=0}^{m}\psi_{\varepsilon}(y_i, f_i) + \frac{1}{2}w^T w \tag{2.15}$$

Where

$$\psi_{\varepsilon}(\mathrm{y_i}, \mathrm{f_i}) = \begin{cases} |\mathrm{y_i} - \mathrm{f_i}| - \varepsilon \mathrm{if} |\mathrm{y_i} - \mathrm{f_i}| \geq \varepsilon \\ 0 \quad \mathrm{otherwise} \end{cases} \tag{2.16}$$

Being $\mathrm{y_i} - \mathrm{w^T x_i} - \mathrm{b} \leq \varepsilon + \xi_\mathrm{i}$ and $\mathrm{f_i}$ the estimated values for the same variable at the same time. Equation (2.16) is called the ε- insensitive loss function and it indicates the estimated points that are inside the boundaries of the SVM with a ray of ε that do not suffer penalization. To facilitate the calculations, ξ_i is defined when they are above the boundaries and ξ_i^* is when it is under the boundaries. Thus, ε is a predefined tolerance.

The second part of Equation (2.17) is used as a smooth function since one of the aims of the SVM is to obtain a function $f(x)$ as flatter as possible to minimize $w^T w$ which is related to the capacity of the ML model. Therefore, C measures the trade-off between the empiric risk and the model smoothness. The primal formulation can be given by:

$$\min_{\mathrm{w,b,\xi,\xi^*}} \frac{1}{2}\mathrm{w^T w} + \mathrm{C} * \sum_{\mathrm{i=0}}^{\mathrm{l}}(\xi_\mathrm{i} + \xi_\mathrm{i}^*) \tag{2.17}$$

where i express the number of penalized points and is subject to (Equations 2.18, 2.19, 2.20, 2.21):

$$y_i - w^T x_i - b \leq \varepsilon + \xi_i \tag{2.18}$$

$$w^T x_i + b - y_i \leq \varepsilon + \xi_i^* \tag{2.19}$$

$$\xi_i \geq 0 \tag{2.20}$$

$$\xi_i^* \geq 0 \tag{2.21}$$

Being the primal Lagrangean function (Equation 2.22):

$$L(w, b, \xi, \xi^*, \alpha, \alpha^*, \beta, \beta^*) = \frac{1}{2}w^T w + C\sum_{i=0}^{l}(\xi_i + \xi_i^*) - \sum_{i=0}^{l}(\beta_i \xi_i + \xi_i^* \beta_i^*) - \sum_{i=0}^{l}\alpha_i(w^T x_i + b - y_i + \varepsilon + \xi_i) - \sum_{i=0}^{l}\alpha_i^*(y_i, w^T x_i - b + \varepsilon + \xi_i^*) \tag{2.22}$$

Where $\alpha, \alpha^*, \beta,\ \beta^*$are the vectors of dimension l of the Lagrangean multipliers associated with the constraints (Equations 2.18, 2.19, 2.20, 2.21). Equation (2.22) needs to be

reduced with respect to the primal variables w, b, ξ_i, ξ_i^* and maximized with respect to the dual variables $\alpha, \alpha^*, \beta, \beta^*$. Satisfying the remain KKT conditions, the initial problem becomes in a dual problem with α_i and α_i^* variables. The dual problem is modeled as following (Equation 2.23):

$$maxL_D(\alpha, \alpha^*) = \frac{-1}{2}\sum_{i,j}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T x_j - \sum_i(\varepsilon - y_i)\alpha_i - \sum_i(\varepsilon + y_i)\alpha_j^* \quad (2.23)$$

Subject to (Equations 2.24, 2.25, 2.26):

$$\sum_i(\alpha_i - \alpha_i^*) = 0 \quad (2.24)$$

$$0 \leq \alpha_i \leq C \quad (2.25)$$

$$0 \leq \alpha_i^* \leq C \quad (2.26)$$

The solution of the dual problem is the following regression (Equation 2.27):

$$f(x) = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)x_i^T x + b \quad (2.27)$$

To solve the linear regression of the SVR is necessary to calculate $x_i^T x_j$ and $x_i^T x$ (equations (2.23) and (2.27)), called the mapping functions. To decrease the computational effort caused by the searching of the best mapping function, $x_i^T x_j$ and $x_i^T x$ are replaced by a kernel function $K(x_i, x_j)$. The regression problem can be solved by this kernel function and is represented by (Equation 2.28):

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x_o, x_i) + b \quad (2.28)$$

The kernel function adopted is "Gaussian Radial Basis" (GRB), given by $K(x_i, x_j) = exp\left(-\gamma\left\|x_i - x_j\right\|^2\right)$, where $\gamma = 1/2\sigma^2$ is also a model parameter. After the parameter is chosen, is necessary to evaluate the estimated function. Consequently, the prediction errors are compared with the original values y_i and the predicted values $\hat{y}_i$. One of the measurements of error more common is given by the Mean Squared Error (MSE) (Equation 2.29), which is described by a dataset with size m as:

$$\text{MSE} = \left(\frac{1}{m}\right)\sum_{i=0}^{m}(y_i - \hat{y}_i)^2 \quad (2.29)$$

The performance of the SVR depends on the parameters *C*, *ε*, and *γ*, which are defined a priori.

2.2.3 Random Forest Regression

Random Forest (RF), firstly proposed by Breiman (2001), is a ML algorithm that is a combination of tree predictors in which each of them depends on the value of a random vector that is sampled independently and has the same distribution of all the trees in the forest. Its generalization error depends on how the trees converge and on the strength of the individual

trees in the forest and their correlation. Thus, the RF is an ensemble technique that generates multiple independent decision trees from a given random vector of parameters.

Each input vector gives to its correspondent tree an output related to classification or regression. For regression problems, as the one related to this work, the random forest are composed by trees depending on a random vector Θ, in which the tree predictor $h(x, \Theta)$ is a numerical value opposed to the class labels. The outputs are numerical, and it is assumed that the training set is independent from the distribution of the random vector (X, Y). It can be characterized as an ensemble of B decisions trees $\{T_1(X), \dots, T_B(X)\}$, where $X = \{x_1, \dots, x_B\}$ is a p-dimensional vector. Its predictor is formed by calculating the average over B of the trees $\{h(x, \Theta_B)\}$

2.2.4 Extreme Gradient Boosting Regression

Extreme Gradient Boosting (XGBoost) (CHEN & GUESTRIN, 2016) is an ML algorithm based on decision-trees ensemble that uses Gradient Boosting (GB) (FRIEDMAN *et al.*, 2000; FRIEDMAN, 2001) as its framework. XGBoost is considered an optimized and scalable ML system for tree boosting, which can perform parallel processing, handle missing values and avoid overfitting and bias. This method is widely applied for problems involving unstructured data (e.g. images, text), yet it is also used to solve basic problems such as classification and regression – which is applied in this work.

XGBoost is widely applied to solve a different range of problems – such as store sales prediction, high energy physics event classification, web text classification, customer behavior prediction, motion detection, hazard risk prediction and other fields (LE *et al.*, 2019) (WANG *et al.*, 2017) (TORLAY *et al.*, 2017). One of the most important factors that helps the success of XGBoost is the scalability due to several important systems and algorithmic optimizations. It applies the principle of boosting weak learners using gradient descent architecture.

2.3 Unscented Kalman Filter

The Kalman filter (KF) is well-known for formulating state-space linear dynamical systems and can be sum up as a recursive solution to linear optimal filtering problems. It can be applied for stationary and nonstationary systems and gives a recursive solution in which each updated estimate of the state is calculated from the last estimate and the new input data; so, it is only required to store the previous data. The KF is more efficient compared to the direct estimation through the whole entire dataset (KALMAN *et al.*, 1960) (LEWIS *et al.*, 1986) (GREWAL & ANDREWS, 1993).

A linear, discrete-time dynamical system is exposed in the following block diagram (Figure 15). As a first step to describe the KF, there are a few concepts that need to be clarified. The state vector or state, given by x_k, is the minimal set of that which is enough to describe the unforced dynamical behavior of the system and k is the discrete time. Thus, the state is the minimum amount of data that is needed to predict its future behavior of the input data. Generally, the state x_k is undefined; so, to estimate the observed data y_k is used (HAYKIN, 2001).

Thus, Figure 2.13 can be explained as follows. Firstly, there is the process equation (Equation 2.30) in which $F_{k+1,k}$ is the transition matrix that takes the state x_k from time k to time $k+1$. The process has a noise which is assumed as w_k and is additive, white and Gaussian, with zero mean and covariance given by Equation 2.31.

$$x_{k+1} = F_{k+1,k}x_k + w_k \tag{2.30}$$

$$E[w_n w_k^T] = \begin{cases} Q_k, & n = k \\ 0, & n \neq k \end{cases} \tag{2.31}$$

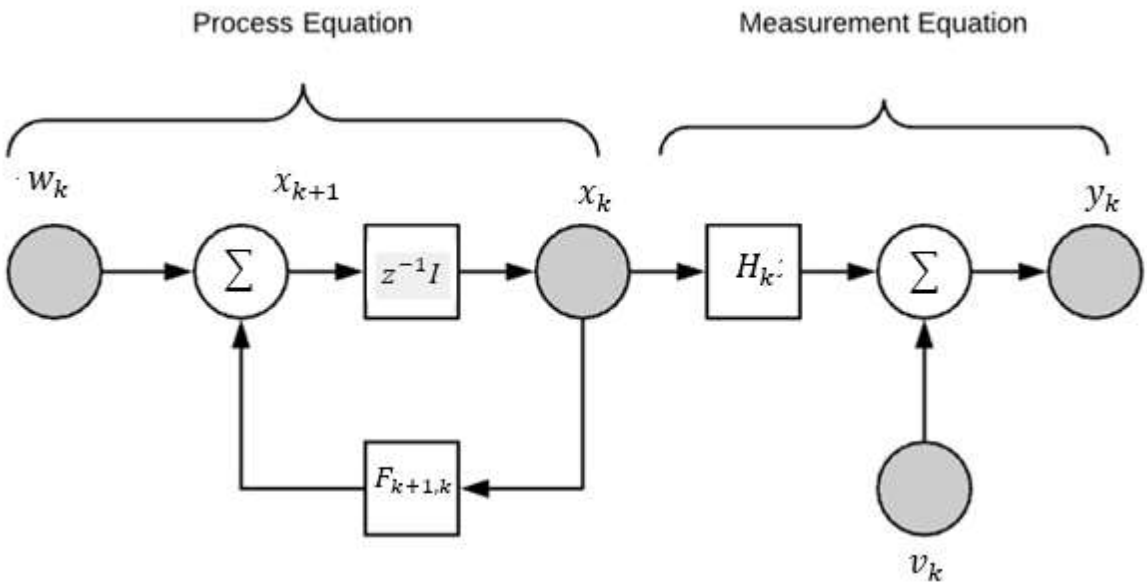
T is the notation for matrix transposition and the dimension of the state space being M. Then, there is the measurement equation (Equation 2.32), given by y_k, and H_k is the measurement matrix. The measurement noise is v_k and has the same properties as w_k. Its covariance is formulated as following (Equation 2.33):

$$y_k = H_k x_k + v_k \tag{2.32}$$

$$E[v_n v_k^T] = \begin{cases} R_k, & n = k \\ 0, & n \neq k \end{cases} \tag{2.33}$$

The measurement noise, v_k, is not correlated with the process noise w_k, and the dimension of the measurement space is N. This problem can be classified as filtering if $i = k$, prediction if $i > k$ and smoothing if $1 \leq i \leq k$.

Figure 15 - Block diagram of a dynamical process.



Adapted from (HAYKIN, 2001).

In this work, was applied the unscented Kalman filter (UKF). UKF was proposed by Julier et al. (JULIER *et al.*, 1995) (JULIER & UHLMANN, 1996) and developed by Wan and van der Merwe JULIER & UHLMANN, 1997) (WAN et al., 2000) (WAN & VAN DER MERWE, 2000) (VAN DER MERWE et al., 2000) to correct the faults that has been noticed in the extended Kalman filter (EKF) -which the difference between them and the UKF is that the last one uses a sample of specific points to calculate the mean and the covariance of the Gaussian random variables (GRV). Then, this is propagated through the system, which captures the following mean and covariance precisely to the second order of a Taylor series expansion in which learns the nonlinearity of the system. Thus, the basis of the UKF is the unscented transformation (UT) (HAYKIN, 2001).

The UT is used to calculate statistics of random variables that pass through nonlinear transformation. So, propagating a random variable x, with dimension L, in a nonlinear function $y = f(x)$ and if x has a mean $\bar{x}$ and a covariance P_x. To compute the statistics of y, there is the matrix $\mathcal{X}$ of $2L + 1$ sigma vectors $\mathcal{X}_i$ are represented in (Equations 2.34, 2.35., 2.36):

$$\mathcal{X}_0 = \bar{x} \tag{2.34}$$

$$\mathcal{X}_i = \bar{x} + (\sqrt{(L+\lambda)P_x})_i \qquad i = 1, \dots, L \tag{2.35}$$

$$\mathcal{X}_i = \bar{x} - (\sqrt{(L+\lambda)P_x})_{i-L} \qquad i = L+1, \dots, 2L \tag{2.36}$$

Being $\lambda = \alpha^2(L + \kappa) - L$ a scaling parameter. The constant α sets the spread of sigma points around $\bar{x}$ and is, generally, a small positive value between 10^{-4} 1. Then, the constant κ

is a secondary scaling parameter, which is, in general, set to $3 - L$, and β incorporates the prior knowledge of the x, and $(\sqrt{(L+\lambda)P_x})_i$ is the ith column of the matrix square root. Thus, the sigma vectors are propagated through the nonlinear function (Equation 2.37)

$$\mathcal{Y}_i = f(\mathcal{X}_i), \qquad i = 0, \dots, 2L \tag{2.37}$$

The mean and the covariance of y are approximated through a weighted sample mean and covariance of the subsequent sigma points (Equations 2.38 and 2.39),

$$\bar{y} \approx \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_i \tag{2.38}$$

$$P_y \approx \sum_{i=0}^{2L} W_i^{(c)} (\mathcal{Y}_i - \bar{y})(\mathcal{Y}_i - \bar{y})^T \tag{2.39}$$

And the weights are given by (Equations 2.40, 2.41, 2.42),

$$W_o^{(m)} = \frac{\lambda}{L+\lambda}, \tag{2.40}$$

$$W_o^{(c)} = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta, \tag{2.41}$$

$$W_i^{(m)} = W_i^{(c)} = \frac{\lambda}{L+\lambda}, \qquad i = 1, \dots, 2L. \tag{2.42}$$

Hence, the UKF is an extension of the UT for the KF, and, for a nonlinear discrete-time system state equation and measurement equation (Equations 2.43 and 2.44):

$$x_{k+1} = F(x_k, u_k, v_k) \tag{2.43}$$

$$y_k = H(x_k, n_k) \tag{2.44}.$$

F and H are nonlinear system function, and x_k is the unobserved state vector, $u_k$is the external input, y_k is the observed measurement signal, v_k is the process noise and n_k is the observed noise – the noises are, in general, assumed as additive Gaussian white noise. So, $x^a = [x^T\ v^T\ n^T]^T$ and $x^a = [(x^x)^T\ (x^v)^T\ (x^n)^T]^T$; so, the UKF equation is as follows:

The initialization is (Equations 2.45, 2.46, 2.47, 2.48)

$$\hat{x}_0 = E[x_0] \tag{2.45}$$

$$P_0 = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \tag{2.46}$$

$$\hat{x}_0^\alpha = [\hat{x}_0^T \quad 0 \quad 0\ ]^T \tag{2.47}$$

$$P_0^\alpha = E[(x_0^\alpha - \hat{x}_0^T)(x_0^\alpha - \hat{x}_0^T)^T] = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & R_v & 0 \\ 0 & 0 & R_n \end{bmatrix} \tag{2.48}$$

Computing the matrix $\mathcal{X}_k^\alpha$, $k = 1, 2, \dots \infty$, with sigma points (Equation 2.49)

$$\mathcal{X}_{k-1}^\alpha = [\hat{x}_{k-1}^\alpha \qquad \hat{x}_{k-1}^\alpha + \sqrt{(L+\lambda)P_{k-1}^\alpha} \qquad \hat{x}_{k-1}^\alpha - \sqrt{(L+\lambda)P_{k-1}^\alpha}\ ] \tag{2.49}$$

And the time-update equations are Equations 2.50, 2.51, 2.52, 2.53, 2.54,

$$\mathcal{X}_{k/k-1}^x = F(\mathcal{X}_{k-1}^x, u_{k-1}, \mathcal{X}_{k-1}^v) \tag{2.50}$$

$$\hat{x}_{k/k-1} = \sum_{i=0}^{2L} W_i^m \mathcal{X}_{i,k/k-1}^x \tag{2.51}$$

$$P_{k/k-1} = \sum_{i=0}^{2L} W_i^c (\mathcal{X}_{i,k/k-1}^x - \hat{x}_{k/k-1})(\mathcal{X}_{i,k/k-1}^x - \hat{x}_{k/k-1})^T \quad (2.52)$$

$$Y_{k/k-1} = H(\mathcal{X}_{k/k-1}^x, \mathcal{X}_{k-1}^n) \quad (2.53)$$

$$\hat{y}_k = \sum_{i=0}^{2L} W_i^m Y_{i,k/k-1}^x \quad (2.54)$$

And the measurement equations are given by Equations 2.55, 2.56, 2.57, 2.58, 2.59.

$$P_{\tilde{y}} = \sum_{i=0}^{2L} W_i^c (Y_{i,k/k-1} - \hat{y}_{k/k-1})(Y_{i,k/k-1} - \hat{y}_{k/k-1})^T \quad (2.55)$$

$$P_{x_k,y_k} = \sum_{i=0}^{2L} W_i^c (\mathcal{X}_{i,k/k-1}^x - \hat{x}_{k/k-1})(Y_{i,k/k-1} - \hat{y}_{k/k-1})^T \quad (2.56)$$

$$K_k = P_{x_k,y_k} {P_{\tilde{y}}}^{-1} \quad (2.57)$$

$$\hat{x}_k = \hat{x}_{k/k-1} + K_k[y_k - \hat{y}_{k/k-1}] \quad (2.58)$$

$$P_k = P_{k/k-1} - K_k P_{\tilde{y}_k} K_k^T \quad (2.59)$$

2.4 Maximum Entropy Bootstrap

Bootstrap is a well-known statistic technique that has diverse applications such as confidence interval construction, bias, and standard deviation estimation; also, one of the advantages of this method is that it does not need an assumption over the dataset distribution. Firstly, i.i.d. (independent and identically distributed) bootstrap was proposed by Efron in 1979 and it is a powerful tool for statistical inference and is appropriated for complex problems where traditional confidence intervals are difficult to construct and are unreliable when it is not known the distribution of the dataset.

However, if the studied data is not i.i.d., a Maximum Entropy Bootstrap (MEB) is more indicated. Vinod (2006) says that, in the 1930s, time-series inference was based on the Wiener-Kolmogorov-Khintchine (WKK) theory where, based on this theory, was constructed a population called ensemble Ω, which was heavily relied on stationary assumptions. Then, the MEB comes as new computer-intensive construction of Ω. MEB can be applied when the time series are short, non-stationary, with some regime changes and jump discontinuities. It is a tool for highly dependent non-stationary time series (THEIL, 1980) (THEIL & LAITINEN, 1980).

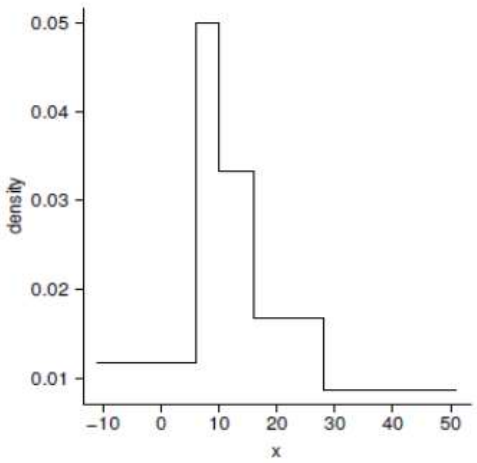
MEB, developed by Vinod & Lopez-de-Lacalle (2009), is an adequate method to replicate non-stationary time-series (TS) and it is applied to study real-world situations. In addition to these characteristics, MEB also does a more precise analysis due to its tests based on asymptotic theory and just uses transformations that do not destroy the TS format.

The initial population which will be analyzed is the original TS, MEB will generate many replicas of the dataset using the algorithm given below in which the original format will

be maintained at the replica and, also, the time dependence of autocorrelation function (ACF). The step-by-step is given below (VINOD & LOPEZ-DE-LACALLE, 2009):

a) Sort the original data, obtaining a new TS, x_t, and save the position if these values before the transformation;
b) Calculates the medians between the sorted TS value using $z_t = \frac{(x_t - x_{t-1})}{2}$, $t = 2, 3, \ldots, T-1$;
c) Compute the trimmed mean of the errors $x_t - x_{t-1}$ and compute the value of z_1 and z_t using: $z_1 = x_1 - m_{trm}$ e $z_t = x_T - m_{trm}$;
d) Build a Maximum Entropy density function using the values of z as limits of the intervals. The distribution is constructed under a uniform distribution of the intervals. The distribution with equal probabilities (Figure 16) and it grants that the mean of any interval satisfies the mean preservation constraints which obey the ergodic theorem;
e) Simulate, using inverse transform method, T values of constructed TS and soft them in ascending order;
f) Sort the original values again using the saved position in step 1. Maintaining the original format of the TS according to its dimension, i.e. the biggest point at the original and replicated TS will be at the same location;
g) Repeat steps 2 to 6 for analysis until it is necessary.

Figure 16 - Distribution from MEB.



Source: VINOD & LOPEZ-DE-LACALLE, 2009.

2.5 Literature Review

As mentioned before in the Theoretical Background section, the estimation of the SOC can occur through a diversity of methods such as data-driven methods (e.g. NN, SVM, RF),

adaptative filters (e.g. EKF and UKF) traditional methods and hybrid methods. Thus, this step of the BMS is important in order to prevent further accidents and unwanted costs. Since the vehicles are continuously monitored, data-driven approaches are preferable than other methods due to its rapiDNNy adaptability and generalization power. Then, in this section, some previous works in the field of data-driven methods are discussed.

He *et al.* (2014) proposed a SOC estimation for li-ion batteries using NN and UKF to filter the prediction error. Firstly, they used as training set a dataset from a Dynamic Stress Profile (DST) with temperature of 0ºC, 10ºC, 20ºC, 30ºC, 40ºC and 50ºC and, as test set , they used two different profiles with the same temperature as the training set: the US06 Driving Schedule and the Federal Urban Driving Schedule. This train and test set were processed through a time series autoregressive model with step 4, and this processing was optimized with the NN model. After the predictions were done, the SOC time was processed by a UKF in order to reduce the prediction error. This work is the main basis for this dissertation, thus contributing with the windowing and filtering ideas. However, due the simulating characteristics of the dataset, real-life patterns might not have been identified by the trained model. Thus, the confidence interval of the remaining time until the next discharge is estimated using Maximum Entropy Bootstrap (MEB), which allows a better range of prediction and confidence in the estimative.

Li *et al*. (2018) proposed a battery capacity estimation using RF regression, in which this technique was able to learn the dependencies of the features, such as charging voltage and capacity measurements, based on a signal that was available during typical battery operation. This method is applied and validated using lithium nickel manganese cobalt oxide batteries with diverse aging patterns. The results showed that the proposed methodology was able to evaluate the health states of different batteries under diverse cycling conditions with promising applications for online battery capacity estimation. This work shows that traditional ML models can lead to great performances even on diverse cycling conditions, contributing to this dissertation as an encouragement to try different models, such as SVM.

Tao *et al.* (2017) proposed a five-state nonhomogeneous Markov chain model to design Li-ion batteries and investigate the capacity fading dynamic of different formulations. Thus, the state and the behavior of the active materials in the Li-ion battery were modeled and to verify its efficiency a dataset of almost 3 years of cycling capacity fading experiments was used.

Zhao *et al*. (2018) investigated the modeling of Li-ion batteries using RNN which were trained with dynamic battery data (e.g. vehicle cycle test results). This was used to simplify the process of battery modeling because there was no need for battery parameterization and

specialized characterization of tests. Thus, a model consisting of an RNN with a gated recurrent unit (GRU) and a deep feature selection (DFS) were implemented. Therefore, two RNNs were performed, one with current as input and the other with power as input and, then, were compared with equivalent circuit models (ECMs).

Tao and Chen (2017) proposed a methodology that considers the capacity recovery effect in dynamic fading processes. Some features that effects this process were analyzed and extracted and, then, a Random Forest (RF) was applied to model and predict the capacity degradation recovery effect.

Chemali *et al*. (2018) used DNN to estimate the SOC and the training data was from a drive cycle loads of Li-ion at various temperatures. The DNN can encode the dependencies in time into the network weights and provide an accurate estimation. Batteries exposed between -20ºC to 25ºC were used as input. In Chen *et al.* (2011) proposed an EKF based battery model, which considered the effects of hysteresis in an open circuit voltage, that was integrated with an NN in order to estimate the SOC.

Li *et al.* (2007) used a merged Fuzzy Neural Network, that has a superior performance compared with traditional neural networks, to estimate SOC of a li-ion battery combined with a reduced form of a genetic algorithm. The proposed methodology used twelve inputs and one output to approximate a continuous non-linear function. Thus, the validation results showed that the method is effective and accurate. Then, a more advanced algorithm was implied, which was named adaptative neuro-fuzzy inference system, being more efficient than the previous one described (AWADALLAH & VENKATESH, 2016).

Hansen & Wang (2005) used SVM to estimate the SOC of a large-scale li-ion-polymer battery pack, and it proposed to remove the drawbacks of traditional methods to estimate the SOC (e.g. coulomb counting). They used measurements such as voltage and current as input of SVM. Antón *et al.* (2013) also estimates the SOC of a $LiFePO_4$ battery cell using SVM and obtained effective and accurate results.

Zheng *et al.* (2016) proposed an estimation of SOC using online inference of battery OCV. Two types of OCV were tested – low-current and incremental OCV – and this method observes the relationship between OCV and SOC; also, in these two experiments, three temperatures were tested and two SOC estimators were compared in terms of accuracy, convergence, time, and robustness for online estimation. Even though their results were good, OCV-SOC estimation is done when the battery is fully discharged, which does not allow the prediction during operational time of the EV. The data-driven approach, however, allows to

perform continuous training with new data and predict more precisely the present conditions of the battery.

Sbarufatti *et al.* (2017) proposed a method for prognosis of Li-ion batteries using particle filter combined with a radial basis NN. This proposed methodology can be trained online which allows flexibility and adaptability to new information. As well as this methodology, this present work also allows an online estimation due to the fact of applying data-driven techniques.

Zou *et al.* (2015) suggested a methodology to estimate SOC and SOH (State of Health). As a first step, the dependency of the nominal parameters of a first-order resistor-capacitor model is determined, and the degradation is quantified. Secondly, two EKF with different time scales are used to combine SOC/SOH monitoring – with SOC being estimated in real-time, and SOH updated online. Even though this present work does not propose the SOH prediction, the combined approach proposed is easily reproduced due to the characteristics of data-driven models, which does not require a wide knowledge about the physics of failures and the degradation process of the battery.

As a contradiction to most methodologies presented above, due to simulating characteristics of some datasets used, using methods such as Maximum Entropy Bootstrap (MEB), the confidence interval of the predicted time until the next recharge can be performed in order to approximate even better real-life conditions and its variance. Also, filtering techniques have proven to be effective in reducing prediction's errors and increasing model's performance (He *et al.,* 2014) and, thus, could be explored by most methodologies presented

3 APPLIED METHODOLOGY

The following sections is going to present the dataset description and the proposed methodology applied in this work.

3.1 Dataset Description

The dataset used to validate the proposed methodology was from the CALCE (Center for Advanced Life Cycle Engineering) Battery Research Group – University of Maryland. This research group is focused on developing state of art BMS to single and multi-cell systems, estimating the SOC and the state of health (SOH) of batteries and studying the degradation process.

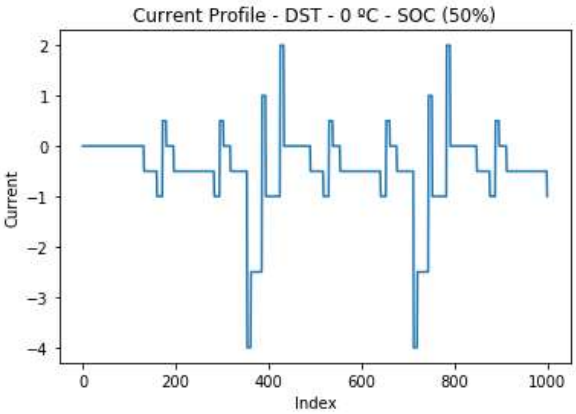
Thus, to model data-driven methodologies, the conditions of the measurements are important. It is advisable to simulate the dataset with real-life conditions such as road conditions, speeds, and driving styles in order to cover as many conditions as possible in terms of current, voltage, loading charging rates, and SOC. Hence, the dataset provided by the CALCE group use battery testing simulating driving cycles. The training data was collected using the dynamical stress testing (DST) profile (Figure 18), which is specified by the US Advanced Battery Consortium (USABC). This test consists of various current steps with different amplitudes and lengths being a simplification of real-life conditions. The DST was conducted under 0 °C, 10 °C, 20 °C, 30 °C, 40 °C, and 50 °C. The test dataset was collected under the Federal Driving Urban Driving Schedule (FUDS) - FUDS an urban driving profile - which uses the same profile as DST, but simulates a driving condition schedule based on an operation of a vehicle in a urban road - (Figures 19) and these tests were conducted under 0 °C, 10 °C, 20 °C, 30 °C, 40 °C. FUDS is more complex than the DST in order to provide robustness and generalization to the proposed methodology. The index in the Figures 3.3 and 3.4 indicates the number of registered points by the test platform.

The batteries cells are composed of $LiFePO_4$ (Lithium Ion Phosphate) with 76g and a rating capacity of 2230 mAh (Figure 17). It was placed in a temperature chamber and an Arbin BT2000, a battery test equipment, was controlling the charging and discharging (Figure 20).

Figure 17 - Studied battery.

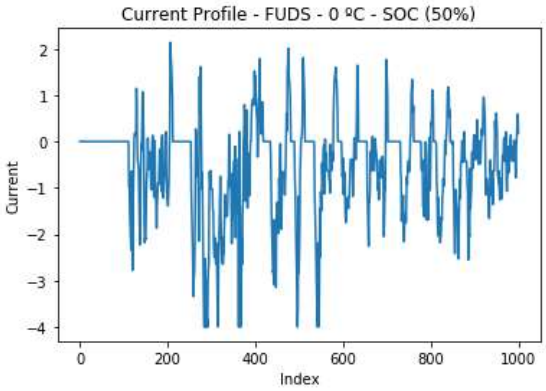
Source: CALCE.

Figure 18 - Current profile of DST at 0ºC and SOC = 50%.



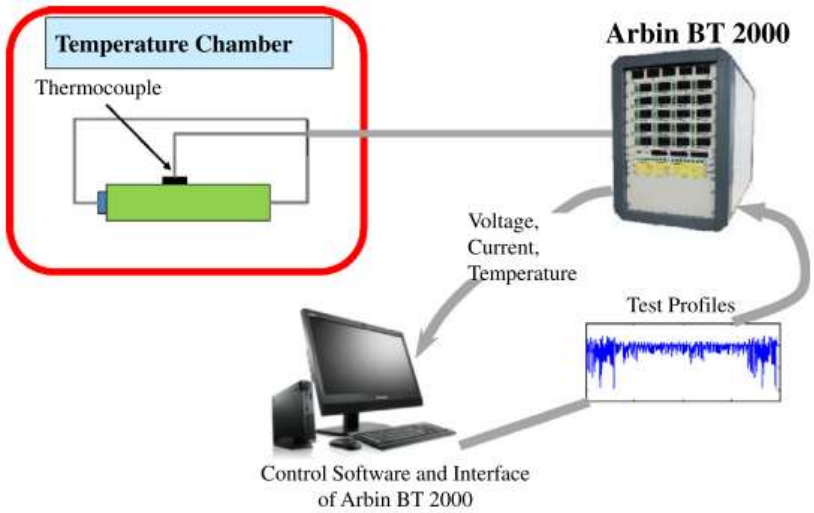
Source: The Author, 2019.

Figure 19 - Current profile of FUDS at 0ºC and SOC = 50%.



Source: The Author, 2019.

Figure 20 - Test procedure.



Source: HE *et al.*, 2014.

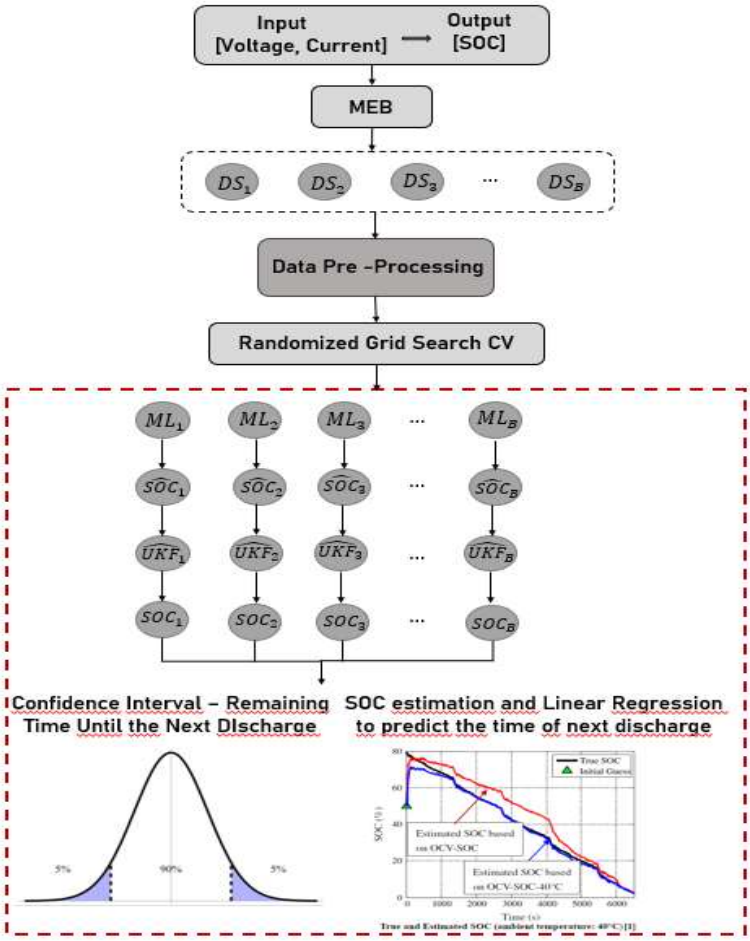
This dataset has, originally, 8084 registered points for the DST – which was used for training – and 9057 points for the FUDS – which was used for testing the proposed

methodology. In the training phase, 20% of the training dataset (DST) was used to validate the algorithm (HE *et al.*, 2014) (HU *et al.*, 2012) (HU *et al.*, 2012) (XIONG *et al.*, 2014)

3.2 Methodology

In order to predict the SOC, the methodology in Figure 21 is proposed. Firstly, the input is composed of current and voltage and the output is the SOC. Each input and output are replicated 5000 times using MEB. It maintains the main characteristics of a non-stationary time-series and works as a data augmentation strategy to reproduce the train and test data in order to calculate multiple predictions using an ML algorithm or accurately estimate the SOC using DNN. The main objective of the bootstrap technique applied in this work is to estimate the confidence interval of the predicted remaining time until the next discharge. The replication is not only used to predict the confidence interval of the recharge time, but also to guarantee a better generalization of the tested algorithms.

Figure 21 - Proposed methodology.



Source: The Author, 2019.

Then, due to the capacitive resistance in the battery, both current and voltage in previous stages will affect the present state of the battery. Thus, a windowing time-series model is used as preprocessing. It uses steps of size 5 – the step size was chosen due to empirical tests evaluating the prediction performance, which was tested a size of 30, proposed by He *et al.* (2014), of 7, 10 and 6, obtaining better results to size length 4 - is applied to the current and voltage replicated data to assess the time dependencies between each measurement. So, the inputs at a time i are $[\,I(i), I(i-4), \ldots, I(i-4k), V(i), V(i-4), \ldots, V(i-4k)]$ (He *et al.*, 2014) – k is a parameter to determine the size of the input vector -and the output is $SOC(i)$.

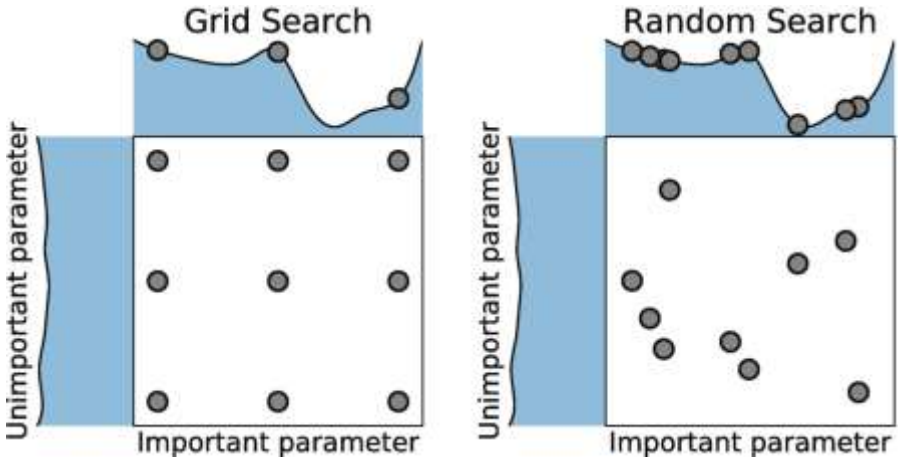
This approach is applied because of the capacitive resistance in the battery, which the current and voltage of previous samples has effect on the present battery state. Hence, due to the windowing processing, each of the 5000 samples of the input and output loses the first 20 – the multiplication of the size length versus the windowing steps.

This approach is also applied in He et al. (2014), however, in this work, only the windowing model is applied without the ML parameters optimization. Also, for propose of a better performance, testing different lengths of windowing through each proposed data-driven algorithm allows to evaluate individually each performance during training. As a further step, the implementation of this optimization algorithm evaluating the performance of each algorithm simultaneously is required.

After pre-processing the data through the windowing model, these replicated datasets are going to enter two models: the ML and the DNN. The parameters of the DNN and the ML methods are selected by means of randomized grid search cross-validation (CV). This search method uses random hyper parameters in a chosen set to test the ML algorithm; thus, not all parameters are tried out, but a fixed number of parameters settings is sampled from the specified distributions. The randomized search CV was chosen over the grid search cross validation due to its computational cost and better performance. Figure 22 shows the grid search CV versus the randomized search CV, which demonstrate how both of them performs the parameters search – the first one only combines each of the suggested parameters and test their performance during the validation phase; however, the randomized search CV combine random values of a given interval and test these parameters during the validation phase.

The results are given in Tables 1, 2, 3 and 4 for the SVM, MLP, RF and XGBoost, respectively. Then, the SOC is predicted for each replication and a histogram with the results is built to construct the confidence interval.

Figure 22 - Randomized Search Cross Validation



Source: Bergstra; Bengio, 2012.

Table 1 - SVM hyperparameters

	SVM
Kernel	RBF
C	6.5643
epsilon	0.0192
gamma	0.0010

Source: The Author, 2019.

Table 2 - MLP hyperparameters

	MLP
Activation	ReLU
$\boldsymbol{\beta_1}$	0.9000
$\boldsymbol{\beta_2}$	0.9990
E	1×10^{-8}
Hidden Layers	15
Learning Rate	Constant, 0.001
Validation Split	0.1

Source: The Author, 2019.

Table 3 - RF Hyperparameters.

	RF
Number of Estimators	1400
Minimum Sample Split	2
Minimum Sample Leaf	1
Maximum Features	'Auto'
Maximum Depth	100
Bootstrap	'True'

Source: The Author, 2019.

Table 4 - XGBoost Hyperparameters.

	XGBoost
Random State	42
Column Sample by Tree	0.9041
Gamma	0.2252
Learning Rate	0.03398
Maximum Depth	3
Number of Estimators	113
Subsample	0.9233

Source: The Author, 2019.

To implement these algorithms, Python language, using Anaconda as API (Application Program Interface), and the following packages were used: sklearn (PEDREGOSA *et al.*, 2006), numpy (OLIPHANT, 2006), matplotlib (HUNTER, 2007), seaborn (WASKON *et al.*, 2018) and keras (CHOLLET, 2016). The replications were made in R (RStudio) using the package meboot (VINOD & LÓPEZ-DE-LACALLE, 2009). All experiments were performed in a GPU GEFORCE GTX 2080 Ti.

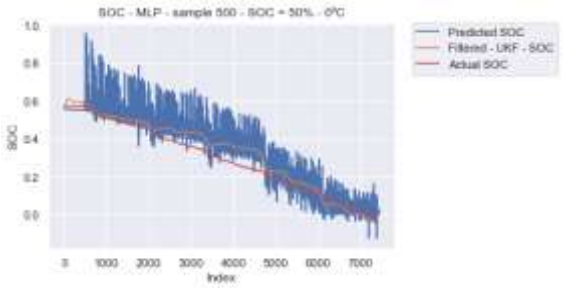
4 RESULTS

As a first step, the inputs, which are current and voltage for a given temperature and SOC, are replicated using MEB. The chosen conditions, of the $LiFePO_4$ battery, were a temperature of 0 ºC and a SOC of 50 %, in other words, the battery is half discharge; however there are other conditions offered by the data basis, that condition was chosen due to the fact of being the first one of the dataset and to validate, in a simpler way, the DNN and ML methods. Furthermore, in order to test the generalization power of the proposed methodology, other temperature and SOC conditions will be tested including other datasets of accelerated tests of EVs batteries.

Thus, the replication step is to calculate the confidence intervals. Then, these replicated time series (TS) are pre-processed through a windowing model based on He *et al.* (2014). Also, according to this author, this type of model, to pre-process battery data, can capture the thermodynamic and physicochemical behavior of the device. Thus, with the proposed windowing (HE *et al.,* 2014) of size 4, this process was performed with a step size of 5. This step size was chosen under several tests with the dataset and the artificial intelligence algorithms.

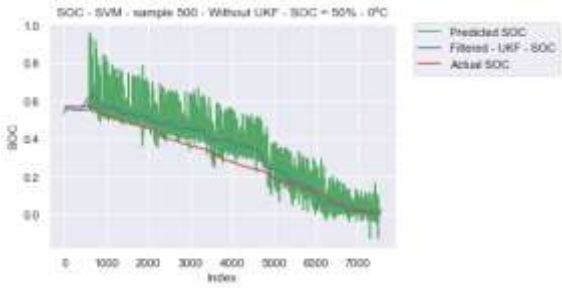
Then, with the original training dataset, the Randomized Grid Search CV was performed and the hyperparameters were calculated. After that, each replicated sample from the training set enters the DNN – MLP model – and the ML models – SVM, RF and XGBoost. So, the SOC is estimated and the time until the next recharge for each sample. And, finally, the confidence interval is computed. In order to post-process the SOC results, the UKF is applied in each of the samples. Then, the filtered SOC, the remaining time until the next discharge and the confidence interval are calculated after the post-processing. Figures 23, 24, 25 and 26 show the comparison of the predicted SOC, filtered SOC, and actual SOC for MLP, SVM, RF and XGBoost for the sample number 500.

Figure 23 - Predicted and actual SOC – Without and With UKF - MLP.



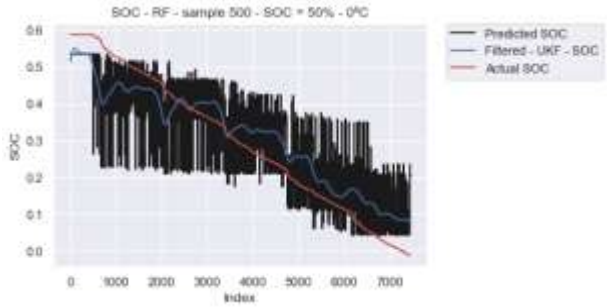
Source: The Author, 2019.

Figure 24 - Predicted and actual SOC – Without and With UKF - SVM.



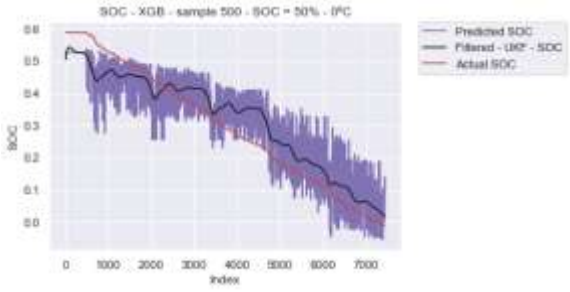
Source: The Author, 2019.

Figure 25 - Predicted and actual SOC – Without and With UKF - RF.



Source: The Author, 2019.

Figure 26 - Predicted and actual SOC – Without and With UKF - XGB



Source: The Author, 2019.

Observing Figures 27 to 34, for the SVM, two outliers were found in the model without UKF, and when the predicted SOC were post-processed, the number of outliers increased to seven. However, as well as the MLP the RMSE values decreased and the range of this values also reduced. In the RF was observed a considerable decrease in the numbers of outliers when the UKF is applied, and as the two previous models, the RMSE also reduced with the post-processing. For the XGBoost, the same behavior was observed – the RMSE values reduced – which can partially validate the application of UKF as a post-processing method.

Figures 35 to 42 show the distribution of the RMSE for the estimated SOC without and with UKF. For all the tested data-driven methods, the RMSE value range reduced. If all the tested method is visually compared, considering the predictions without and with UKF, the MLP obtained better results in terms of RMSE than the other methods. If the graphics of each method is analyzed, some outliers can be found, which is good considering that the RMSE is a metric that analyses the performance of the training phase. The RMSE is, basically, the difference between the predicted value and the real observed value and evaluate the trained model and its accuracy. For the MLP, Figures 27 and 28, due to the application of the filter, is

observed a decay of the number of outliers samples – the outliers samples are those ones which had the worst results if compared with the average samples -; also, the values of the RMSE and the range of these values reduced if compared with the ones without filter.

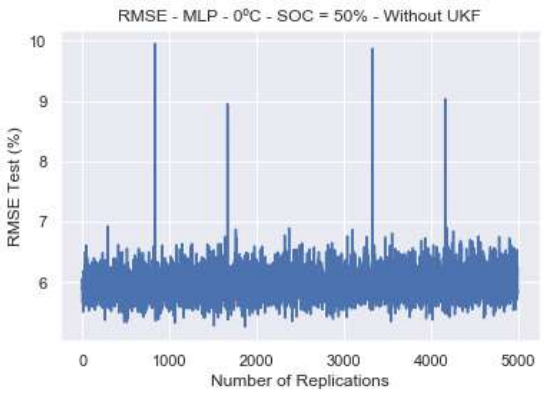
Thus, analyzing the range between the maximum and the minimum RMSE for the proposed algorithms (Table 5), the MLP was 0.0467, SVM was 0.1034, the RF was 5718.7119 and for the XGBoost was 0.021. As for the filtered output, the MLP still performs better than the SVM, with the range between the maximum and the minimum 0.028291 and 0.036202, respectively. The results for the RF and XGBoost had worst results due to two reasons, the first one was that the discrepant value is an outlier – which was not taken off to respect the original number of the proposed samples – and this can occur as a result of algorithm initialization. Other reason is the fact that RF and XGBoost are generally applied for classification models rather that regressions – they are models based on decisions trees, which are commonly used for classification. The MLP and SVM, however, are more versatile in terms of which task is chosen to perform – classification or regression – which leads to better results. This result shows that the post-processing technique had a great influence on the improvement of the predicted SOC.

Table 5 - Maximum and minimum values between each sample - RMSE - SOC = 50% - 0ºC – MLP, SVM, RF and XGBoost.

	MLP	**SVM**	**RF**	**XGBoost**
Maximum – Without UKF	0.0994	0.1803	5718.7951	0.0881
Minimum – Without UKF	0.0526	0.0769	0.7285	0.0671
Maximum – With UKF	0.0611	0.0809	5718.7967	5718.7921
Minimum – With UKF	0.0328	0.0447	0.05887	0.05637

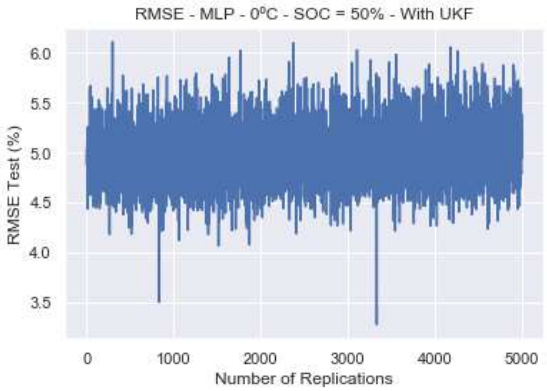
Source: The Author, 2019.

Figure 27 - RMSE Test – considering each sample - Without UKF- MLP.



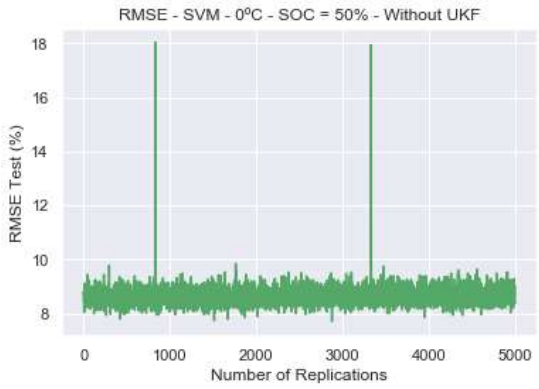
Source: The Author, 2019.

Figure 28 - RMSE Test – considering each sample - With UKF- MLP.



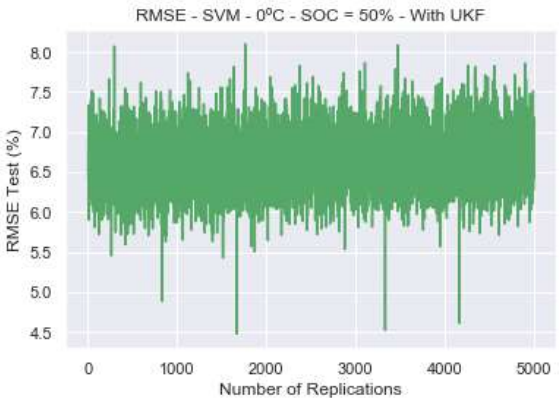
Source: The Author, 2019.

Figure 29 - RMSE Test – considering each sample - Without UKF- SVM.



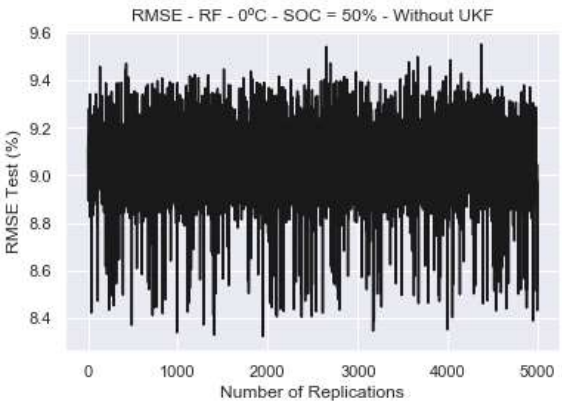
Source: The Author, 2019.

Figure 30 - RMSE Test – considering each sample - With UKF- SVM.



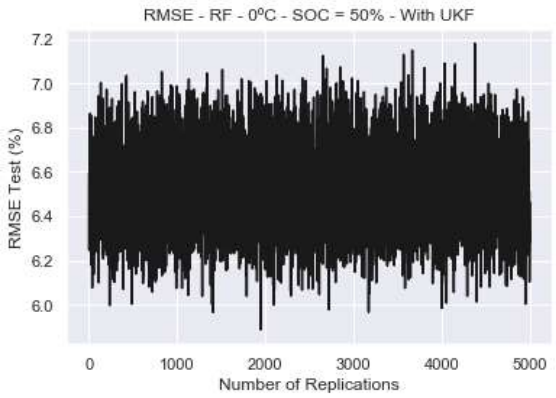
Source: The Author, 2019.

Figure 31 - RMSE Test – considering each sample – Without UKF- RF.



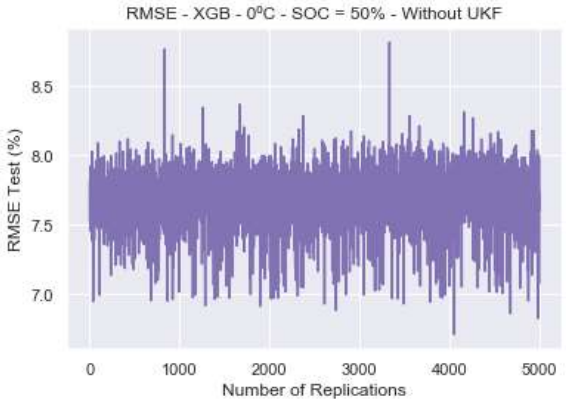
Source: The Author, 2019.

Figure 32 - RMSE Test – considering each sample – With UKF- RF.



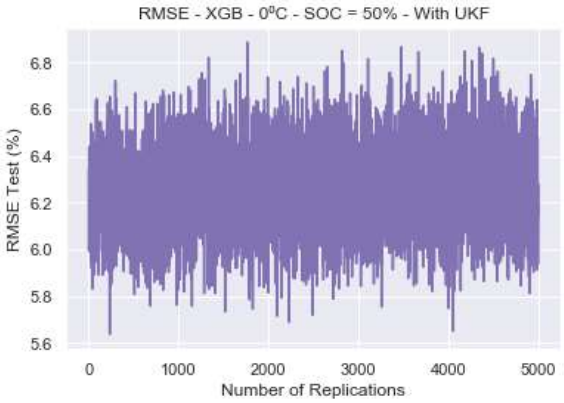
Source: The Author, 2019.

Figure 33 - RMSE Test – considering each sample – Without UKF- XGBoost



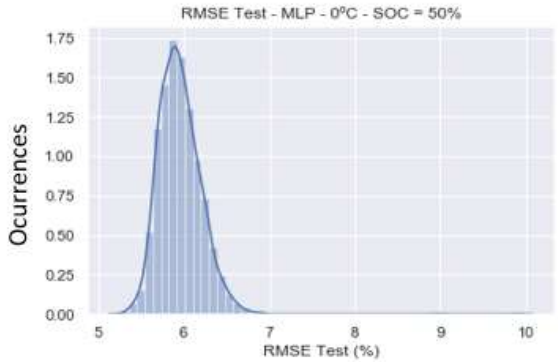
Source: The Author, 2019.

Figure 34 - RMSE Test – considering each sample – With UKF- XGBoost.



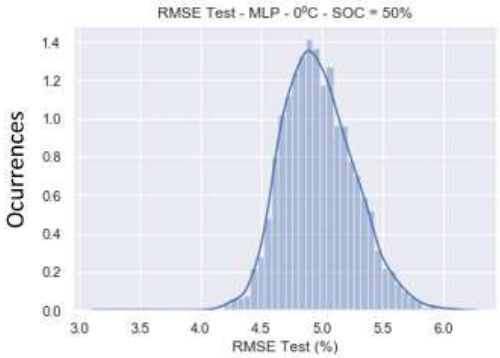
Source: The Author, 2019.

Figure 35 - Test RMSE distribution – considering each sample - Without UKF - MLP.



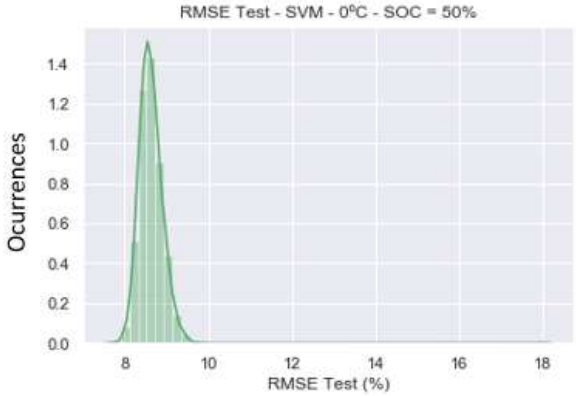
Source: The Author, 2019.

Figure 36 - Test RMSE distribution – considering each sample - With UKF - MLP.



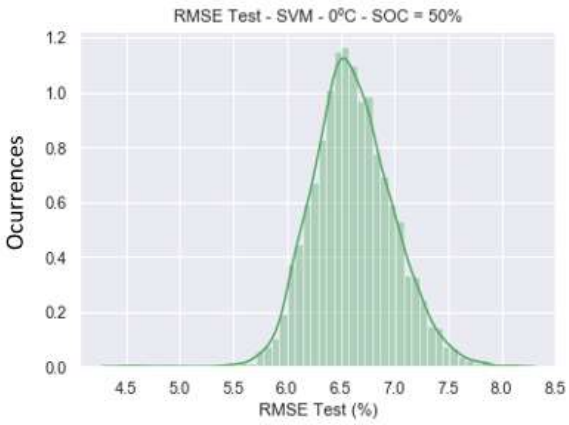
Source: The Author, 2019.

Figure 37 - Test RMSE distribution – considering each sample – Without UKF- SVM.



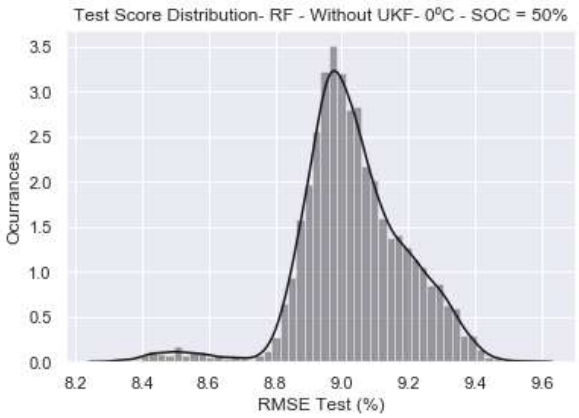
Source: The Author, 2019.

Figure 38 - Test RMSE distribution – considering each sample - With UKF - SVM.



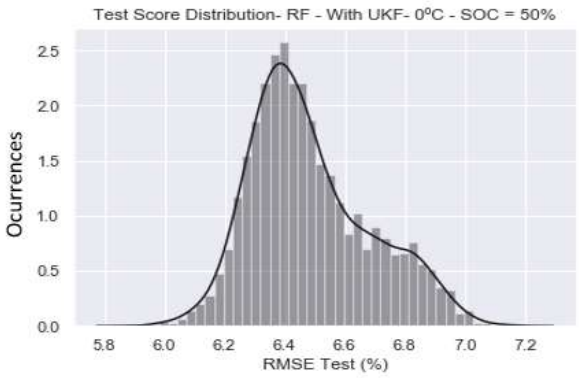
Source: The Author, 2019.

Figure 39 - Test RMSE distribution – considering each sample – Without UKF - RF.



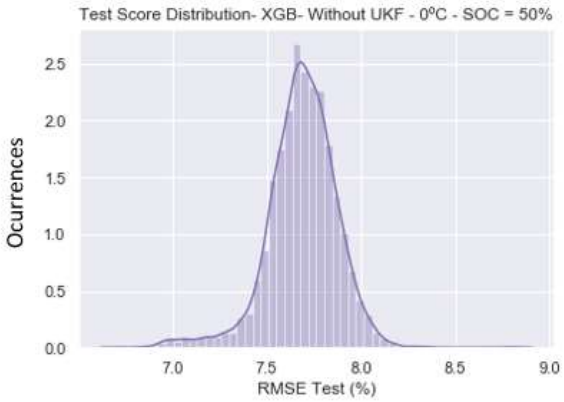
Source: The Author, 2019.

Figure 40 - Test RMSE distribution – considering each sample – With UKF - RF.



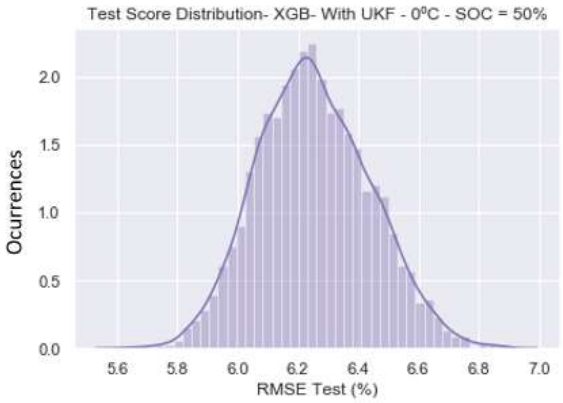
Source: The Author, 2019.

Figure 41 - Test RMSE distribution – considering each sample – Without UKF - XGBoost.



Source: The Author, 2019.

Figure 42 - Test RMSE distribution – considering each sample– With UKF - SVM.



Source: The Author, 2019.

Then, the remaining time until the battery discharge, which is considered as failure, is predicted. In this case, the SOC reaches 0% and the actual value of time for when this SOC is reached is 28020.6105 s. Table 6 shows the maximum, minimum and mean of the percentage error of this predicted time at SOC = 0%, for the non-filtered and filtered output.

Comparing these results, it can be assured that the MLP has a better performance than the SVM, the RF, and the XGBoost, with a mean percentage error of 3.2604% for MLP for the non-filtered output. For the filtered output, there is a mean percentage error of 2.0788 %, also a better result than the others ML approaches, reassuring the better performance of the MLP and ensuring the improvement of the results by using UKF. The SVM had the second better performance with a percentage error of 4.5681% for the non-filtered prediction and 3.1341%

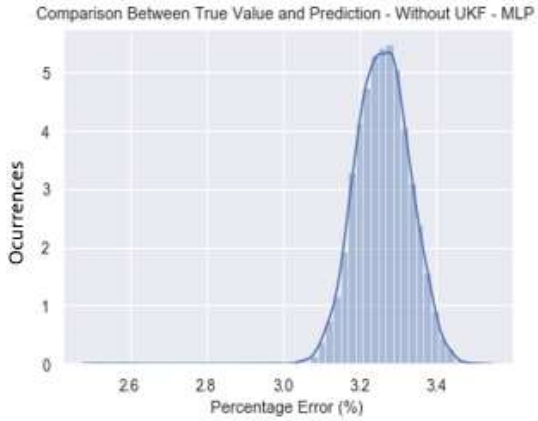
for the filtered prediction. These results occur since the MLP model is more complex, being a DNN model, and the RF and XGBoost are more adapted to classification tasks even though they also perform regression. However, as a further step, a sensitive analysis considering the number of layers of the MLP need to be evaluated. Figures 43 to 50 show the distribution of the percentage error for the non-filtered and the filtered results, which also reaffirm the effectiveness of the UKF in reducing the errors and improving the prediction performance.

Table 6 - Percentage error – maximum, minimum and mean between each sample – SOC = 50% - 0ºC – MLP, SVM, RF and XGBoost.

	MLP	**SVM**	**RF**	**XGBoost**
Maximum – Without UKF	3.5039 %	5.6875 %	10.0395 %	10.0395 %
Mean _ Without UKF	3.2604 %	4.5681 %	9.3776 %	9.3776 %
Minimum – Without UKF	2.5193 %	4.1258 %	8.8476 %	8.8476 %
Maximum – With UKF	2.2872 %	4.3350 %	8.6549 %	7.6150 %
Mean – With UKF	2.0788 %	3.1341 %	7.9633 %	6.7158 %
Minimum – With UKF	1.4996 %	2.7368 %	7.4920 %	6.2594 %

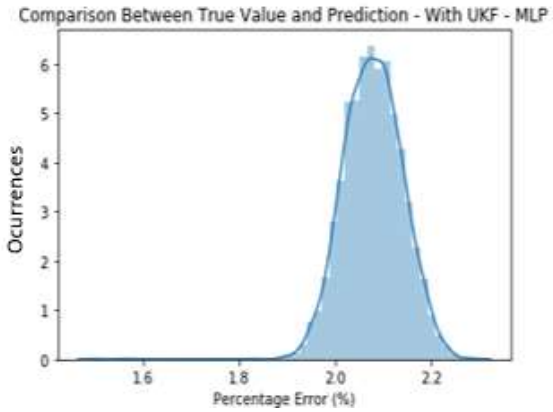
Source: The Author, 2019.

Figure 43 - Distribution of Percentage error – considering each sample - Without UKF- MLP.



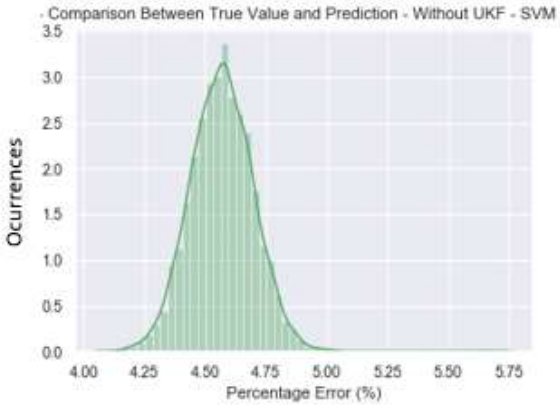
Source: The Author, 2019.

Figure 44 - Distribution of Percentage error – considering each sample - With UKF - MLP.



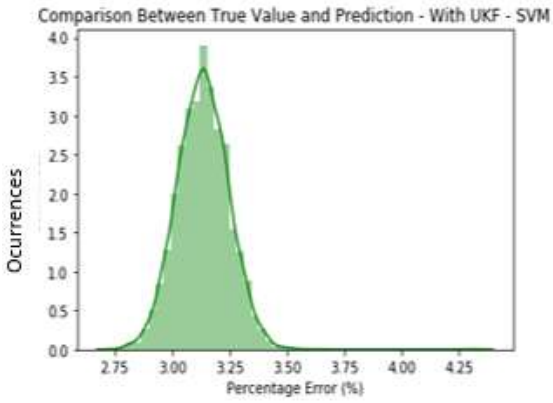
Source: The Author, 2019.

Figure 45 - Distribution of Percentage error – considering each sample - Without UKF - SVM.



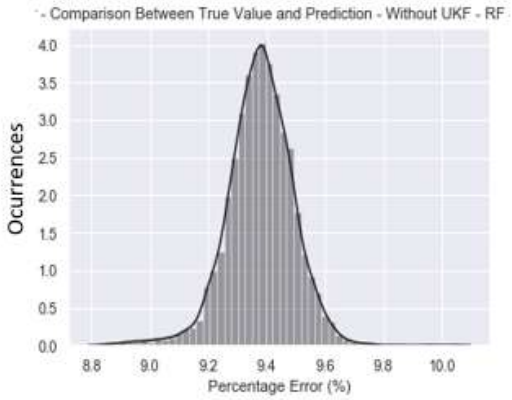
Source: The Author, 2019.

Figure 46 - Distribution of Percentage error – considering each sample - With UKF – SVM.



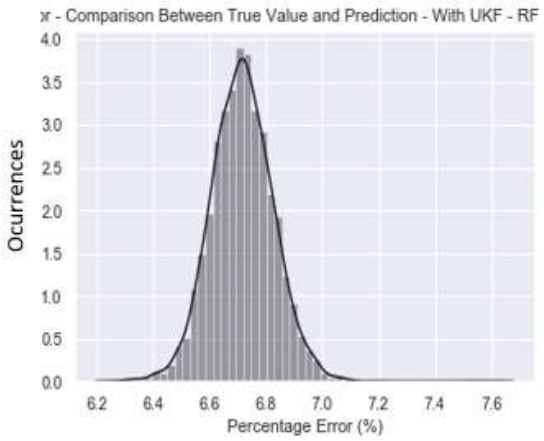
Source: The Author, 2019.

Figure 47 - Distribution of Percentage error – considering each sample - Without UKF – RF.



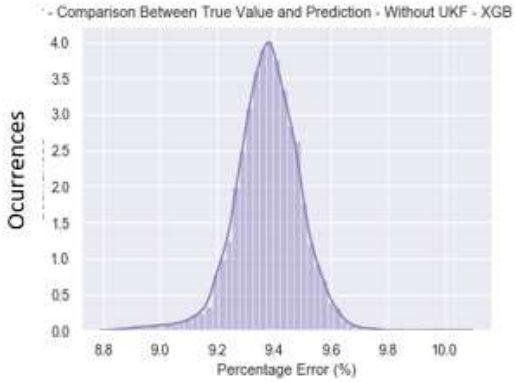
Source: The Author, 2019.

Figure 48 - Distribution of Percentage error – considering each sample - With UKF – RF.



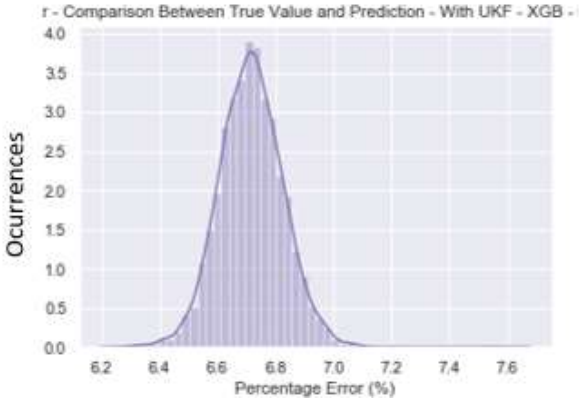
Source: The Author, 2019.

Figure 49 - Distribution of Percentage error – considering each sample - Without UKF – XGBoost.



Source: The Author, 2019.

Figure 50 - Distribution of Percentage error – considering each sample - With UKF – XGBoost.



Source: The Author, 2019.

Table 7 shows the maximum, mean, and minimum values of a confidence interval of 95% for all the tested methods, and Table 8 shows the range of the confidence interval for each method. If each method is compared in terms of the mean value of the predicted remaining time until the next discharge and the absolute range of each confidence interval, there is a proximity in terms of the original value – 28020.6105 s -, and the range is the smallest if compared with the other methods, which assigned a lower variability during the SOC and the remaining time until the next recharge prediction. However, there are still some modifications that need to be done in order to improve the proposed predictions, such as test other types of hyperparameters optimization methods and link the pre and post processing with the algorithm performance in order to run each sample and achieve better results.

Table 7 - Predicted SOC- Confidence Interval of 95% - maximum, minimum and mean between each sample – SOC = 50% - 0ºC – MLP, SVM, RF and XGBoost.

	MLP	**SVM**	**RF**	**XGBoost**
Upper Limit – Without UKF	29002.4432s	29614.2873s	30154.3876s	30833.7596s
Mean _ Without UKF	28934.1956s	29300.6436s	30648.2715s	30648.2715s
Lower Limit – Without UKF	28726.5416s	29176.6894s	29774.5443s	30499.7750s
Upper Limit – With UKF	28661.5054s	29235.2931s	30445.7743s	30154.3876s
Mean – With UKF	28603.1301s	28898.8085s	30251.9770s	29902.4412s
Lower Limit – With UKF	28440.8245s	28787.4842s	30119.9256s	29774.5443s

Source: The Author, 2019.

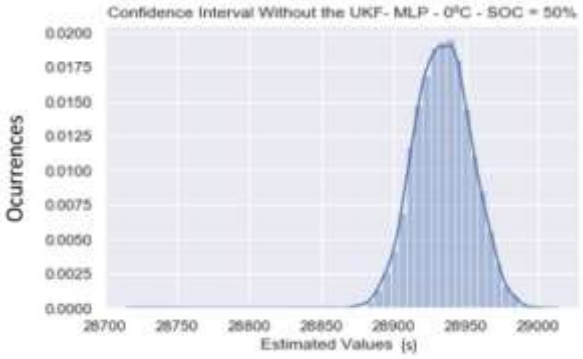
Table 8 - Absolute range of the predicted confidence interval for each method - Confidence Interval of 95%.

	MLP	**SVM**	**RF**	**XGBoost**
Without UKF	275.9016s	437.5979s	379.8433s	333.9846s
With UKF	220.6809s	447.8089s	325.8487s	379.8433s

Source: The Author, 2019.

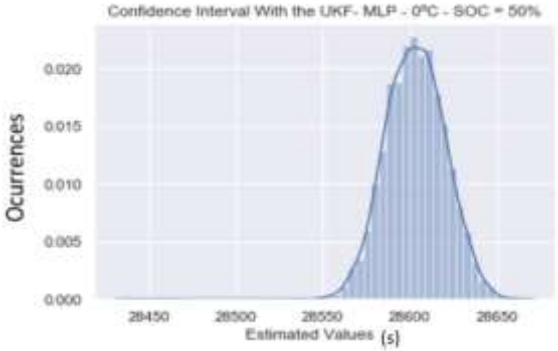
Figures 51 to 58 show distribution of the remaining time until the next discharge for each of the proposed methods.

Figure 51 - Distribution of the predicted remaining time until the next recharge – considering each sample - Without UKF - MLP.



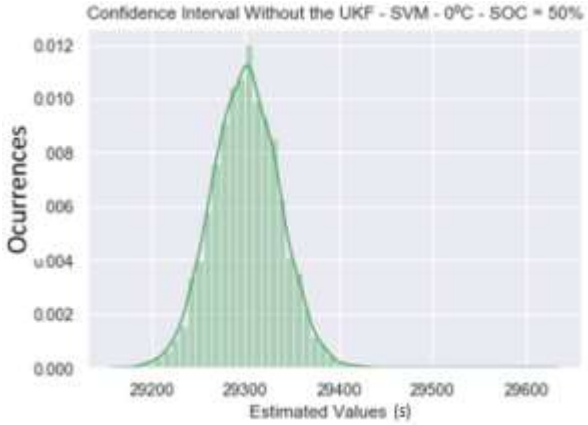
Source: The Author, 2019.

Figure 52 - Distribution of the predicted remaining time until the next recharge – considering each sample - With UKF - MLP.



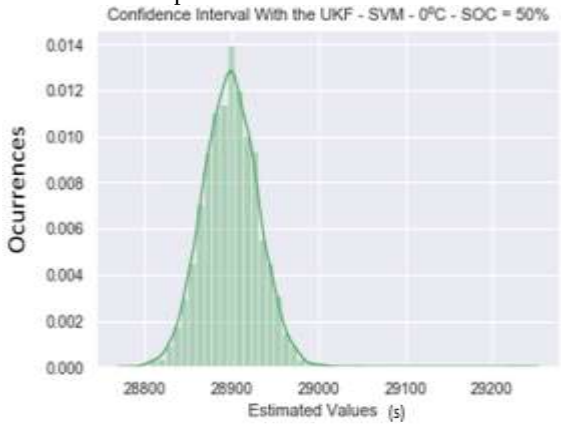
Source: The Author, 2019.

Figure 53 - Distribution of the predicted remaining time until the next recharge – considering each sample - Without UKF - SVM.



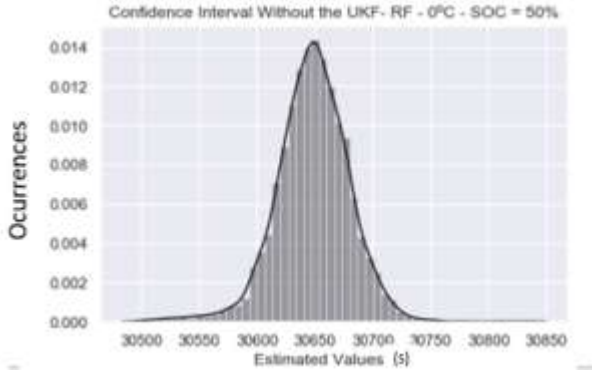
Source: The Author, 2019.

Figure 54 - Distribution of the predicted remaining time until the next recharge – considering each sample - With UKF - SVM.



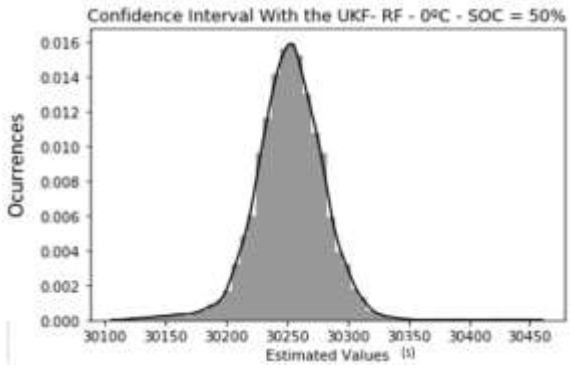
Source: The Author, 2019.

Figure 55 - Distribution of the predicted remaining time until the next recharge – considering each sample - Without UKF - RF.



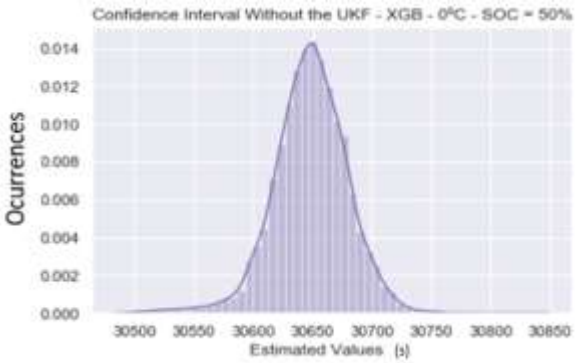
Source: The Author, 2019.

Figure 56 - Distribution of the predicted remaining time until the next recharge – considering each sample - With UKF - RF.



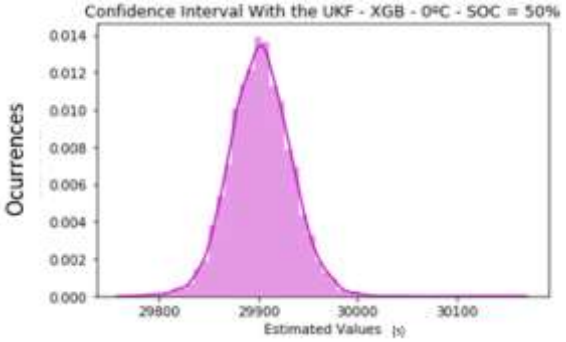
Source: The Author, 2019.

Figure 57 - Distribution of the predicted remaining time until the next recharge – considering each sample - Without UKF – XGBoost.



Source: The Author, 2019.

Figure 58 - Distribution of the predicted remaining time until the next recharge – considering each sample - With UKF - XGBoost.



Source: The Author, 2019.

To test performance of the proposed methodology, the Kruskal- Wallis test was done (KRUSKAL, WILLIAM, WALLIS, 1952) (MCKIGHT & NAJAB, 2010) (Table 4.5 and 4.6). This test is an extension of the Wilcoxon-Mann-Whitney test and is a non-parametric test used to compare three or more populations, it is used to test if all populations that have the same distributions against the null hypothesis, which indicates that the population have different distributions. This test is the nonparametric counterpart of the ANOVA with factor 1.

Test results are given by Table 9 and 10. The test was performed for the RMSE results, without and with UKF, and for the percentage error of the predicted remaining time until the next recharge. For all of them the null hypothesis was rejected, which means that each of these results are from different distributions, ensuring the different performance and results between models.

Table 9 - Results of the Kruskal-Wallis test for the RMSE.

	Statistics	$\boldsymbol{P-value}$
Without UKF	18027.2021	0.0
With UKF	13424.5837	0.0

Source: The Author, 2019.

Table 10 - Results of the Kruskal-Wallis test for the percentage error.

	Statistics	$\boldsymbol{P-value}$
Without UKF	16872.6573	0.0
With UKF	18749.0529	0.0

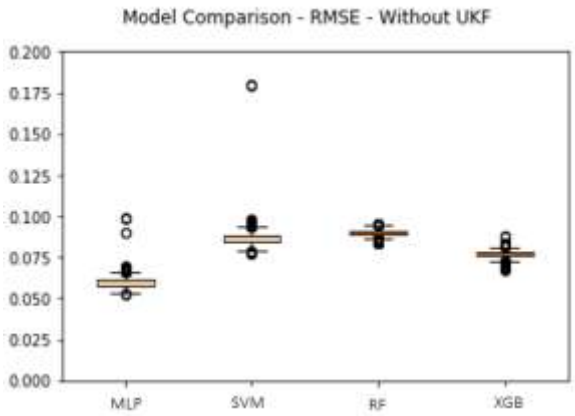
Source: The Author, 2019.

Thus, to compare the tested models, a box plot is outlined using RMSE, to refer the SOC estimation, and PE, to refer the remaining time until the nest discharge. Therefore, to compare

each of the performed models - at the training and test step, without and with UKF – a box plot, from each of these conditions, it is indicated in Figures 59 to 62. Through each box plot, it is observed that MLP had better results if compared with the other three models.

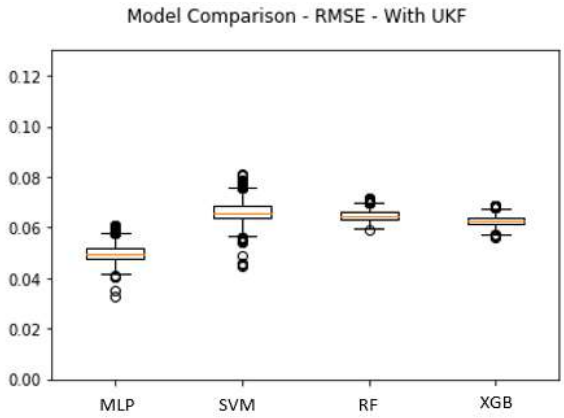
Observing only RMSE results, it can be concluded that XGBoost had better results if compared with SVM and RF, which shows that XGBoost is more effective in training process than SVM and RF; however, MLP stills has better results than the other three, which reinforce choosing MLP as the best performed model. If PE is compared between models, MLP and SVM are the ones with better results, which shows that training performance not always indicates good performance at testing.

Figure 59 - Model Comparison - RMSE - Without UKF.



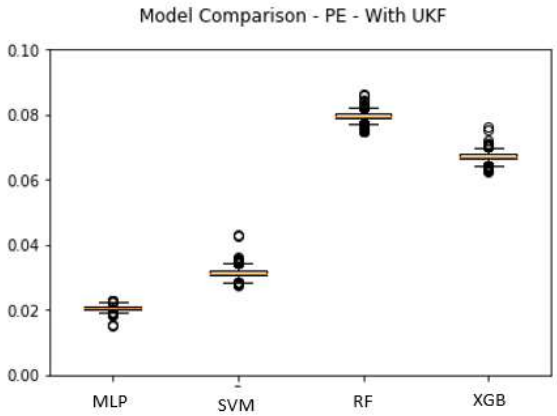
Source: The Author, 2019.

Figure 60 - Model Comparison - RMSE - With UKF.



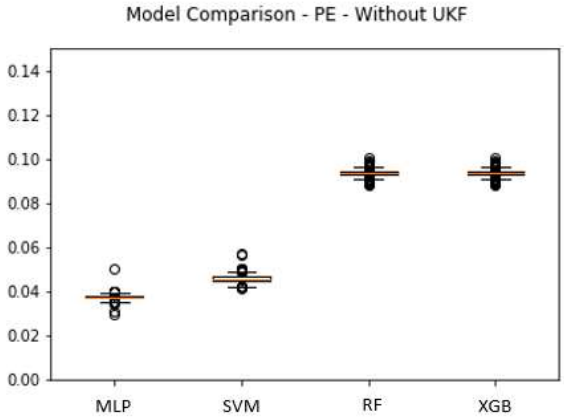
Source: The Author, 2019.

Figure 61 - Model Comparison - PE - Without UKF.



Source: The Author, 2019.

Figure 62 - Model Comparison - PE - With UKF.



Source: The Author, 2019.

5 CONCLUSIONS AND FURTHER WORKS

EVs are one of the possibilities in the automobile industry to reduce the use of fossil fuel and the pollution caused by GHG. Thus, they can be an alternative, also, because of the rechargeable batteries, which can use clean power sources as hydroelectric or eolic and can be applied not only in EVs but surface and underwater vessels, electric aircraft, and spacecraft.

Hence, as a critical component of an EV, the battery needs extra care and monitoring in order to guarantee the safety and reliability of the system. It must be continuously monitored and controlled to avoid abnormal degradation and catastrophic failures. Thus, if the conditions are not properly supervised, the car is susceptible to explosion, fire, release of toxic gases.

Thus, monitoring and knowing when the battery will be discharged is a function of the BMS. Its main function is to show the current conditions of the EV and forecast conditions such as SOC and SOH. So, this work aims to predict the SOC and the remaining time until the next recharging using ML – SVM, RF and XGBoost - and DNN – MLP - techniques. In order to predict the remaining time until the next discharge was predicted using a linear regression of SOC – as ordinate – and remaining time as abscissa. To estimate confidence interval, was used MEB with 95% of confidence.

For the performed techniques, as can be seen, by the results, the MLP had a better performance with higher scores on the test and an interval with an RMSE of 0.046798. Then, the confidence interval of the SOC prediction had a better performance for the MLP with a mean percentage error of 3.260404%.

After these predictions, the UKF was applied to post-process the output in order to decrease the prediction errors. The MLP still had better performance than the other approaches, as expected, with an RMSE of 0.028291 for the MLP. Also, the mean percentage error for the prediction of the failure time decreased being 2.0788 % for the MLP, which validates the efficiency of the UKF.

Thus, as expected, the applied DNN method, MLP, had better results if compared with the ML methods. However, for generalization proposes, another statistical metrics need to be evaluated to guarantee the efficiency of the proposed methodology. Also, other conditions of temperature, SOC, and dataset need to be tested.

As further steps, another DNN architectures – such as CNN and RNN – will be tested. Other ML methods – for example, Adaboost, Gradient Boosting and K-nearest neighbors – should be applied to test the efficiency of the data processing. For the hyperparameters search

of the DNN and ML methods, the windowing pre-processing and the UKF parameters, heuristics techniques (e.g. Genetic Algorithms, Particle Swarm Optimization) can be applied to find the optimum parameters for these each methodology phases.

REFERENCES

ABDUL-MANAN AFN. Uncertainty and differences in GHG emissions between electric and conventional gasoline vehicles with implications for transport policymaking. **Energy Policy** 87:1–7, 2015.

AHMED, R.; EL SAYED, M.; ARASARATNAM, I.; TJONG, J.; HABIBI, S. Reduced-order electrochemical model parameters identification and SOC estimation for healthy and aged Li-ion batteries. Part I: parametrization model development for healthy batteries. **IEEE Journal of Emerging and Selected Topics in Power Electronics,** 2, 3: 659-677, 2014.

ANTÓN, J.C.A.; GARCÍA NIETO, P.J.; JUEZ, F.J.C.; LASHERAS, S.F.; VEGA, M.G.; GUTIÉRREZ, R.M.N. Battery state-of-charge estimator using the SVM technique. **Applied Mathematical Modelling**, v. 37, 9, pp. 6244-6253, 2013. https://doi.org/10.1016/j.apm.2013.01.024.

ANTÓN, J.C.A.; GARCÍA NIETO, P.J.; VIEJO, C.B.; VILÁN, J.A.V. Support Vector Machines Used to Estimate the Battery State of Charge. **IEEE Transaction on Power Electronics**, v. 28, 12, pp. 5919-5926, 2013.

AWADALLAH MA.; VENKATESH B. Accuracy improvement of SOC estimation in lithium-ion batteries. **J Energy Storage**; 6:95–104, 2016.

BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. **Journal of Microbiological Methods**, v. 43, 1, pp. 3-31, 2000.

BREIMAN, L. Random forests. **Machine Learning**, 45, 5-32, 2001.

BRIDNNE, J. S. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. **Neurocomputing**, F. F. Soulié and J. Hérault, Eds. Springer Berlin Heidelberg, pp. 227–236, 1990.

BUDZIANOWSKI WM. Negative carbon intensity of renewable energy technologies involving biomass or carbon dioxide as inputs. **Renew Sustain Energy**;16(9):6507–21, 2012.

'Button battery' warning over child deaths in Manchester. BBC – UK. Web Page: https://www.bbc.com/news/uk-england-manchester-29610570. Accessed: 11/10/2019.

CALCE – Center for Advanced Life Cycle Engineering. 2014. Web Page: https://calce.umd.edu/ - Accessed in: 15/02/2019.

CANALS CASALS L.; MARTINEZ-LASERNA E.; AMANTE GARCÍA B.; NIETO N. Sustainability analysis of the electric vehicle use in Europe for CO2 emissions reduction. **J Clean Prod**; 127:425–37, 2016.

CAPASSO, C.; VENERI, O. Experimental analysis of a zebra battery-based propulsion system for urban bus under dynamical conditions. International Conference on Applied Energy Icae 2014. **Energy Procedia**, v. 61, pp. 1138-1141, 2014.

CHAOUI, H.; IBE-EKEOCHA, C.C. State of Charge and State of Health Estimation for Lithium Batteries using Recurrent Neural Networks. **IEEE Transactions on Vehicular Technology**, 2017.

CHARKHGARD, M.; FARROKHI, M. State-of-Charge Estimation for Lithium-Ion Batteries Using Neural Networks and EKF. **IEEE Transactions on Industrial Electronics**, v.57, pp. 4178 – 4187, 2010. 10.1109/TIE.2010.2043035.

CHEMALI, E.; KOLLMEYER, P.; PREINDNN, M.; EMADI, A. State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. **Journal of Power Sources**, v. 400, pp. 242-255, 2018.

CHEMALI, E.; PREINDNN, M.; MALYSZ, P.; EMADI, A. Electrochemical and electrostatic energy storage and management systems for electric drive vehicles: state-of- art review and future trends. **IEEE Journal of Emerging and Selected Topics in Power Electronics 4**, v. 4, pp.1117-1134, 2016.

CHEN Z.; QIU S.; MASRUR MA.; MURPHEY YL. Battery state of charge estimation based on a combined model of extended Kalman filter and neural networks. **2011 Int Jt Conf Neural Netw**, 2156–63, 2011.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD **International Conference on Knowledge Discovery and Data Mining, San Francisco**, California, pp. 785-794, 2016.

CHEN, Z.; DENG, S.; CHEN, X.; SANCHEZ, R.V.; QIN, H. Deep neural networks-based rolling bearing fault diagnosis. **Microelectronics Reliability**, 75, pp – 327-333, 2017.

CHEN, Z.; FU, Y.; MI, C.C. State of charge estimation of lithium-ion batteries in electric drive vehicles using extended Kalman filtering. **IEEE Trans Veh Technol 2013;**62:1020–30, 2013.

CHEN, Z.; QIU, S.; MASRUR, M. A.; MURPHEY, Y. L. Battery state of charge estimation based on a combined model of extended Kalman filter and neural networks. **2011 Int Jt Conf Neural Netw** 2011:2156–63, 2011.

CHENG KWE.; DIVAKAR BP.; WU HJ.; DING K.; HO FH. Battery-management system (BMS) and SOC development for electrical vehicles. **IEEE Trans Veh Technol** ;60:76–88, 2011.

CHO, S.; JEONG, H.; HAN, C.; JIN, S.; LIM, J.H.; OH, J. State-of-charge estimation for lithium-ion batteries under various operating conditions using an equivalent circuit model. **Comput Chem Eng**, 41:1–9, 2012.

CHOLLET, F. Xception: Deep Learning with Depthwise Separable Convolutions. **CoRR**, 2016.

CIODARO, T.; DEVA, D.; DE SEIXAS, J.; DAMAZIO, D. Online particle detection with neural networks based on topological calorimetry information. **J. Phys. Conf. Series 368**, 012030, 2012.

CUI. Q.; LI, Z.; YANG, J.; LIANG, B. Rolling Bearing Fault Prognosis Using Recurrent Neural Network. **29th Chinese Control And Decision Conference (CCDC)**, 1196-1201, 2017.

DAS, H.S., RAHMAN, M.M., LI, S. TAN, C.W. Electric Vehicles standards, charging infrastructure, and impact on grip integration: A technological review. **Renewable and Sustainable Energy Reviews**, 1-27. 2019.

DI DOMENICO, D.; FIENGO, G.; STEFANOPOULOU, A. Lithium-ion battery state of charge estimation with a Kalman Filter based on a electrochemical model**. Control Appl 2008 CCA 2008 IEEE International Conference**:702–707; 2008.

DONG, T.; LI, J.; ZHAO, F.; YI, Y.; JIN, Q. Analysis on the influence of measurement error on state of charge estimation of LiFePO4 power Battery. **ICMREE2011 - Proc 2011 Int Conf Mater Renew Energy Environ**, 1:644–9, 2011.

DU, H.; CAO, D.; ZHANG, H. Modeling, Dynamics, and Control of Electrified Vehicles. **Woodhead Publishing**. 1st Edition, 2018.

DU, J.; LIU, Z.; WANG, Y. State of charge estimation for Li-ion battery based on model from extreme learning machine. **Control Eng Pract**, 26:11–9, 2014.

E.A. WAN AND R. VAN DER MERWE. The unscented Kalman filter for nonlinear estimation. **Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC)**, IEEE, Lake Louise, Alberta, Canada, October 2000.

E.A. WAN, R. VAN DER MERWE, AND A.T. NELSON. Dual estimation and the unscented transformation. S.A. Solla, T.K. Leen, and K.-R. Müller**, Eds. Advances in Neural Information Processing Systems 12**, Cambridge, MA: MIT Press, pp. 666–672, 2000.

EPA US06 or Supplemental Federal Test Procedure (SFTP)
https://www.epa.gov/emission-standards-reference-guide/epa-us06-or-supplemental-federal-test-procedures-sftp

F.H. LEWIS, Optical Estimation with an Introduction to Stochastic Control Theory. **New York: Wiley**, 1986.

FONTELLES, M.J.; SIMÕES, M.G.; FARIAS, S.H.; & FONTELLES, R.G.S. Metodologia da pesquisa científica: diretrizes para a elaboração de um protocolo de pesquisa. **Revista Paraense de Medicina**, 23(3): 1-8, 2009.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine**. Annals of Statistics**, pp. 1189–1232, 2001.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R.; et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). **The annals of statistics**, 28 (2), 337–407, 2000.

Global EV. Outlook. **International Energy Agency (IEA)**; 2017.

GOMEZ, J.; NELSON, R.; KALU, E.E.; WEATHERSPOON, M.H.; ZHENG, J.P. Equivalent circuit model parameters of a high-power Li-ion battery: Thermal and state of charge effects. **Journal of Power Sources**, Vol. 196, Issue 10, pp. 4826-4831, 2011.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning (Adaptive Computation and Machine Learning series). **The MIT Press**. 2016.

GRAVES, A. Supervised Sequence Labelling with Recurrent Neural Networks. Studies in Computational Intelligence, **Springer**, 2012.

GU, J.; WANG, Z.; KUEN, J.; MA, L.; SHAHROUDY, A.; SHUAI, B.; LIU, T.; WANG, X.; WANG, G.; CAI, J.; CHEN, T. Recent advances in convolutional neural networks. **Pattern Recognition**, v. 77, pp. 354-377, 2018.

GUARNIERI, M. Looking back to electric cars. **IEEE History of Electro - Technology Conference: The Origins of Electrotechnologies.** pp. 1–6, 2012. 10.1109/HISTELCON.2012.6487583.

H. M. ELATTAR; H. K. ELMINIR; AND A. M. RIAD, "Prognostics: a literature review," **Complex Intell. Syst.**, vol. 2, no. 2, pp. 125–154, 2016.

HANNAN MA; AZIDIN FA; MOHAMED A. Hybrid electric vehicles and their challenges: a review. **Renew Sustain Energy**; 29:135–50, 2014.

HANNAN, M.A.; LIPU, M.S.H.; HUSSIAN, A.; MOHAMED, A. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations. **Renewable and Sustainable Energy Reviews.** 78 834–854, 2017. http://dx.doi.org/10.1016/j.rser.2017.05.001.

HANSEN, T.; WANG, C. Support vector-based battery state of charge estimator. **Journal of Power Sources.** V. 141, 2, pp. 351-358, 2005.

HASSAN R.; MOHAMMED A.; JANATI I.; YOUSSEF G.; MOHAMED E. Multilayer Perceptron: Architecture Optimization and Training. **International Journal of Interactive Multimedia and Artificial Intelligence**, Vol. 4, Nº1, pp. 26-30, 2016. DOI: 10.9781/ijimai.2016.415.

HAYKIN, S. Kalman Filtering and Neural Networks. John Wiley & Sons, Inc. 2001. DOI:10.1002/0471221546.

HAYKIN, S. Neural Networks and Learning Machines. Prentice Hall. 3rd edition, 2008.

HAYKIN, S. Neural networks. 2ed. Upper SadDNNe River: Prentice Hall, 1999.

HAYKIN, S. Neural Networks: A Comprehensive Foundation. Prentice Hall. 1998.

HE, H.; XIONG, R.; GUO, H. Online estimation of model parameters and state-of-charge of LiFePO4 batteries in electric vehicles. **Appl Energy**, 89:413–20, 2012.

HE, H.; XIONG, R.; ZHANG, X.; SUN, F.; FAN, J.; MEMBER, S. State-of-Charge Estimation of the Lithium-Ion Battery Using an Adaptive Extended Kalman Filter Based on

an Improved Thevenin Model. **IEEE Transactions on Vehicular Technology.** 60:1461–9, 2011.

HE, HW; XIONG, R; PENG, JK. Real-time estimation of battery state-of-charge with unscented Kalman filter and RTOS mu COS-II platform. **Appl Energy**; 162:1410–8, 2016.

HE, W.; WILLIARD, N.; CHEN, C.; PECHT, M. State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation. **International Journal of Electrical Power and Energy Systems**. v. 62. pp. 783-791. 2014.

HE, W.; WILLIARD, N.; CHEN, C.; PECHT, M. State of charge estimation for electric vehicle batteries under an adaptive filtering framework. In: **IEEE Conference on Prognostics and System Health Management (PHM)**. China: Beijing, 2012.

HE, W.; XILLIARD, N.; CHEN, C.; PECHT, M. State of charge estimation for electric vehicle batteries using unscented kalman filtering. **Microelectron Reliab**, 53:840–7, 2013.

HE, Z.; CHEN, D.; PAN, C.; CHEN, L.; WANG, S. State of charge estimation of power Li-ion batteries using a hybrid estimation algorithm based on UKF. **Electrochimica Acta**. V. 211, 1, pp. 101-109, 2016.

HELMSTAEDTER, M. et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. **Nature**. 500, 168–174 (2013).

LEUNG, M. K., XIONG, H. Y., LEE, L. J. & FREY, B. J. Deep learning of the tissue-regulated splicing code. **Bioinformatics.** 30, i121–i129 (2014).

HENDRICKS, C., WILLIARD, N., MATHEW, S., AND PECHT, M. A failure modes, mechanisms, and effects analysis (FMMEA) of lithium-ion batteries. **Journal of Power Sources** 297: 113–120, 2015.

HINTON, G. ET AL. Deep neural networks for acoustic modeling in speech recognition. **IEEE Signal Processing Magazine 29**, 82–97 (2012).

HOFMANN J, GUAN D, CHALVATZIS K, HUO H. Assessment of electrical vehicles as a successful driver for reducing CO2 emissions in China. **Appl Energy**, 2016.

HU, C.; YOUN, B.D.; CHUNG, J. A multiscale framework with extended kalman filter for lithium-ion battery soc and capacity estimation. **Appl Energy**, 92: 694–704, 2012.

HU, X., LI, S., & PENG, H. A comparative study of equivalent circuit models for Li-ion batteries. **Journal of Power Sources**, 198, 359–367. 2012.

HU, X., LI, S., PENG, H., & SUN, F. Robustness analysis of State-of-Charge estimation methods for two types of Li-ion batteries. **Journal of Power Sources**, 217, 209–219. 2012.

HU, X.; SUN, F.; ZOU, Y. Comparison between two model-based algorithms for Li-ion battery SOC estimation in electric vehicles. **Simul Model Pract Theory 2013**; 34:1–11, 2013.

HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, pp. 90-95. 2007. 10.1109/MCSE.2007.55.

J. WANG, J. PUREWAL, P. LIU, J. HICKS-GARNER, S. SOUKAZIAN, E. SHERMAN, A. SORENSON, L. VU, H. TATARIA, M.W. VERBRUGGE, Degradation of lithium ion batteries employing graphite negatives and nickel-cobalt-manganese oxide þ spinel manganese oxide positives: part 1, aging mechanisms and life estimation, **J. Power Sources** 269, 937e948, 2014.

JULIEN, C., MAUGER, A., VIJH, A., ZAGHIB, K. Lithium Batteries – Science and Technology. Springer. 2016.

KANDEPU, R., FOSS, B., IMSLAND, L. Applying the unscented Kalman filter for nonlinear state estimation. **Journal of Process Control**, v. 18, 7-8, pp. 753-768. 2008.

KONGSOON, N., CHIN-SIEN, M., YI-PING, C., YAO-CHING, H. Enhanced coulomb counting method for estimating state-of-charge and state-of -health of lithium-ion batteries. **Applied Energy**, v. 86, pp. 1506-1511, 2009.

KORTHAUER, R. Lithium-ion batteries: basics and applications. Springer Berlin Heidelberg. 2018.

KRIZHEVSKY, A. ImageNet Classification with Deep Convolutional Neural Networks, **Advances in Neural Information Processing Systems 25**. 2013.

KRUSKAL, WILLIAM H.; WALLIS, W. ALLEN. Use of Ranks in One-Criterion Variance Analysis. **Journal of the American Statistical Association**. 47 (260): 583–621. 1952.

LAIFA, T, CHEN, L. Random forest based li-ion battery capacity prediction subject to capacity recovery effect. **2017 IEEE Vehicle Power and Propulsion Conference**, VPPC 2017 – Proceedings, pp. 1-6. 2018.

LAMBERT, F. Tesla battery reignited days after catching on fire in crash as NTSB investigate fire response to another Tesla crash. Eletreck, 2018. Web Page: https://electrek.co/2018/05/10/tesla-battery-reignited-days-after-catching-fire-crash-ntsb-investigate-fire-response/. Accessed: 10/02/2019.

LE, T.L., NGUYEN, H., ZHOU, J., DOU, J., MOAYEDI, H. Estimating the Heating Load of Buildings for Smart City Planning Using a Novel Artificial Intelligence Technique PSO-XGBoost. **Appl. Sci.**, 9(13), 2714, 2019.

LECUN, Y. et al., Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, 86(11), pp.2278–2323, 1998.

LECUN, Y., BENGIO, Y. & HINTON, G. Deep learning. **Nature**, 521(7553), pp.436–444, 2015.

LEE, J.; NAM, O.; CHO, B.H. Li-ion battery SOC estimation method based on the reduced order extended Kalman filtering. **J Power Sources**, 174(1):9–15, 2007.

LEE, S.J.; KIM, J.H. LEE, J.M.; CHO, B.H. The State and Parameter Estimation of an Li- Ion Battery Using a New OCV-SOC Concept. **Power Electron Spec Conference 2007 PESC 2007 IEEE**:2799–2803, 2007.

LENG, F.; TAN, C. M.; YAZAMI, R.; LE, M. D. A practical framework of electrical based online state-of-charge estimation of lithium ion batteries. **Journal of Power Sources**, 255:423–30, 2014.

LI IH, WANG WY, SU SF, LEE YS. (2007). A merged fuzzy neural network and its applications in battery state-of-charge estimation. **IEEE Trans Energy Convers**; 22:697–708.

LI, A., HU, X., LI, T., ZHANG, H., Research on the Prediction Method of Power Battery SOC Based on Deep Learning. **IEEE Third International Conference on Data Science in Cyberspace**, 2018. 10.1109/DSC.2018.00107.

LI, F, JOHNSON, J., YEUNG, S. Neural Networks Part 1: Setting Up Architecture. CS231n Convolutional Neural Networks for Visual Recognition. Web Site: http://cs231n.github.io/neural-networks-1/ Accessed: 12/02/2019.

LI, X., DING, Q., SUN, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. **Reliability Engineering and System Safety**, 172, pp. 1-11, 2018.

LI, Y.; WANG, L.; LIAO, C.; WANG, L.; XU, D. State-of-charge estimation of lithium-ion battery using multi-state estimate technic for electric vehicle applications. In: **Proceedings of the 9th IEEE Veh Power Propuls Conference IEEE VPPC 2013**:316–320, 2013.

LI, Z., HUANG, J., LIAW, B.Y., ZHANG, J. On state-of-charge determination for lithium-ion batteries. **Journal of Power Sources**, v. 348, pp. 281-301. 2017.

LI, Z., WILLIARD, N., CHEN, C., PECHT, M. State of charge estimation for Li-ion batteries using neural networks modeling and unscented Kalman filter-based error cancellation. **International Journal of Electrical Power and Energy Systems**, v. 62, pp-783-791. 2014.

LINDEN, D., REDDY, T. B. Handbook of batteries. McGraw-Hill Professional; 3rd edition. 2001.

LOEB, A.P., "Steam versus Electric versus Internal Combustion: Choosing the Vehicle Technology at the Start of the Automotive Age," Transportation Research Record. **Journal of the Transportation Research Board of the National Academies**, 1885.

LOFGREN, K. EV ARC: World's first self-contained mobile solar charging station for electric vehicles. Inhabitat. 2013. Web Page: https://inhabitat.com/ev-arc-worlds-first-self-contained-mobile-solar-charging-station-for-electric-vehicles/. Accessed: 12/06/2019.

LU LG, HAN XB, LI JQ, HUA JF, OUYANG MG. A review on the key issues for lithium-ion battery management in electric vehicles. **J Power Sources**; 226:272–88, 2013.

M.S. GREWAL AND A.P. ANDREWS, Kalman Filtering: Theory and Practice. Englewood Cliffs, NJ: Prentice-Hall, 1993.

MA, J., SHERIDAN, R. P., LIAW, A., DAHL, G. E. & SVETNIK, V. Deep neural nets as a method for quantitative structure-activity relationships. **J. Chem. Inf. Model.** 55, 263–274 (2015).

MASTALI, M.; VAZQUEZ-ARENAS, J.; FRASER, R.; FOWLER, M.; AFSHAR, S.; STEVENS, M. Battery state of the charge estimation using Kalman filtering. **J Power Sources**, 239:294–307, 2013.

MCKIGHT, P., NAJAB, J. Kruskal-Wallis Test. **The Corsini Encyclopedia of Psychology**. 2010.

MIKOLOV, T., DEORAS, A., POVEY, D., BURGET, L. & CERNOCKY, J. Strategies for training large scale neural network language models. In **Proc. Automatic Speech Recognition and Understanding** 196–201 (2011).

OLIPAHNT, T.E. Guide to Numpy. 2016.

PARTOVIBAKHSH M., GUANGJUN L. (2014). An Adaptive Unscented Kalman Filtering Approach for Online Estimation of Model Parameters and State-of-Charge of Lithium-Ion Batteries for Autonomous Mobile Robots. **IEEE Transactions on Control Systems Technology**. V. 23, 1, pp. 357 – 363.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A.,COURNAPEAU, D., BRUCHER, M., PERROT, M., DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 132, pp. 2825-2830. 2011.

PLETT, G.L. Extended Kalman filtering for battery management systems of LiPB- based HEV battery packs. **J Power Sources**, 134:277–92, 2004.

PLETT, G.L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background. **Journal of Power Sources**, v. 134, 2, pp. 277-292. 2004.

PLETT, G.L. Extended Kalman filtering for battery management systems of LiPB- based HEV battery packs: Part 2. Modeling and identification. **J Power Sources**, 134, 2, pp. 262–76, 2004.

PLETT, G.L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation. **Journal of Power Sources**, v. 134, 2, pp. 277-292. 2004.

POULLIKKAS A. Sustainable options for electric vehicle technologies. **Renew Sustain Energy**, 41, 1277–87, 2015.

R. VAN DER MERWE, J.F.G. DE FREITAS, D. DOUCET, AND E.A. WAN, ‘‘The unscented particle filter,’’ **Technical Report CUED/F-INFENG/TR 380**, Cambridge University Engineering Department, August 2000.

R. VAN DER MERWE AND E.A. WAN, ‘‘Efficient derivative-free Kalman filters for online learning,’’ in **Proceedings of European Symposium on Artificial Neural Networks (ESANN)**, Bruges, Belgium, April 2001

THEIL, H. (1980), “The Symmetric Maximum Entropy Distribution”, **Eco-nomics Letters**, 6, 53–57.

R.B. WRIGHT, J.P. CHRISTOPHERSEN, C.G. MOTLOCH, J.R. BELT, C.D. HO, V.S. BATTAGLIA, J.A. BARNES, T.Q. DUONG, R.A. SUTULA, Power fade and capacity fade resultingfrom cycle-life testing of advanced technology development program lithium- ion batteries, **J. Power Sources**, 119e121, 865e869, 2003.

R.E. KALMAN, ‘‘A new approach to linear filtering and prediction problems,’’ Transactions of the ASME, Ser. D, **Journal of Basic Engineering**, 82, 34 –45 (1960).

RAHMAN, M.A.; ANWAR, S.; IZADIAN, A. Electrochemical model parameter identification of a lithium-ion battery using particle swarm optimization method. **J Power Sources**, 307:86–97, 2016.

REN, L., CUI, J., SUN, Y., & CHENG, X. (2017). Multi-bearing remaining useful life collaborative prediction: A deep learning approach. **Journal of Manufacturing Systems**, 43, 248–256. https://doi.org/10.1016/j.jmsy.2017.02.013.

ROSENBLATT, F. The Perceptron—a perceiving and recognizing automaton. **Report 85**-460-1. Cornell Aeronautical Laboratory. 1957.

RUI-HAO, L.; YU-KUN, S.; XIAO-FU, J. Battery state of charge estimation for electric vehicle based on neural network. **2011 IEEE 3rd Int Conf Commun Softw Netw** 2011:493–6, 2011.

RUMELHART, D. E., HINTON, G.E. WILLIAMS, R.J. Learning representations by backpropagation errors. **Nature**, 323, 533-536, 1986.

S.J. JULIER AND J.K. UHLMANN, ‘‘A general method for approximating nonlinear transformations of probability distributions,’’ **Technical Report**, RRG, Department of Engineering Science, University of Oxford, November 1996.

S.J. JULIER AND J.K. UHLMANN, ‘‘A new extension of the Kalman filter to nonlinear systems,’’ in **Proceedings of AeroSense**: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, 1997.

S.J. JULIER, J.K. UHLMANN, AND H. DURRANT-WHYTE, ‘‘A new approach for filtering nonlinear systems’’. **Proceedings of the American Control Conference**, 1995, pp. 1628–1632.

SAINATH, T., MOHAMED, A.-R., KINGSBURY, B. & RAMABHADRAN, B. Deep convolutional neural networks for LVCSR**. In Proc. Acoustics, Speech and Signal Processing** 8614–8618 (2013).

SALKING, A. J.; FENNIE, C.; SINGH P.; ATWATER, T.; REISNER, D. E. Determination of state-of- charge and state-of-health of batteries by fuzzy logic methodology. **J Power Sources**, 80:293–300, 1999.

SBARUFATTI, C., CORBETTA, M., GIGLIO, M., CADINI, F. Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. **Journal of Power Sources**. 344. 128-140. 2017.

SCHÖLKOPF, B. & SMOLA, A.J., Learning with kernels: support vector machines, regularization, optimization, and beyond. MA: **The MIT Press**, Cambridge, 2002.

SCROSATI B, HASSOUN J, SUN YK. Lithium-ion batteries. A look into the future. **Energy Environ Sci**, 4, 3287–95, 2011.

SNIHIR, I.; REY, W.; VERBITSKIY, E.; BELFADHEL-AYEB, A.; NOTTEN, P.H.L. Battery open-circuit voltage estimation by a method of statistical analysis. **J Power Sources**, 159:1484–7, 2006.

STETZEL, K.D.; ALDRICH, L.L.; TRIBOLI, M.S.; PLETT, G.L. Electrochemical state and internal variables estimation using a reduced-order physics-based model of a lithium-ion cell and an extended Kalman filter. **J Power Sources**, 278:490–505, 2015.

SULAIMAN N, HANNAN MA, MOHAMED A, MAJLAN EH, WAN DAUD WR. A review on energy management system for fuel cell hybrid electric vehicle: issues and challenges. **Renew Sustain Energy**, 52, 802–14, 2015.

SUN, F., HU, X., ZOU, Y., LI, S. Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles. **Energy**. V. 36, 5, pp. 3531-3540, 2011.

TANG, X.; WANG, Y.; CHEN, Z. A method for state-of-charge estimation of LiFePO4 batteries based on a dual-circuit state observer. **J Power Sources**, 296, 23–9, 2015.

TAO, L., CHENG, Y., CHEN, L., SU, Y., CHONG, J., JIN, H., LIN, Y., NOKTEHDAN, A. Lithium-ion battery capacity fading dynamics modelling for formulation optimization: A stochastic approach to accelerate the design process. **Applied Energy**, v. 202, pp. 138-152. 2017.

Tesla model 3. 2018. Available in: https://www.tesla. com/model3; 2017. Accessed in: 21/02/2019.

The EPA Urban Dynamometer Driving Schedule (UDDS). Available: https://www.epa.gov/emission-standards-reference-guide. Accessed in: 21/01/2019.

THEIL, H. AND LAITINEN, K. Singular Moment Matrices in AppliedEconometrics. "Multivariate Analysis – V," Krishnaiah, P., New York, USA: **North USA-Holland Publishing Co.**, pp. 629–649, 1980.

TING, T.O.; MAN, K.L.; LIM, E.G.; LEACH, M. Tuning of Kalman Filter Parameters via Genetic Algorithm for State-of-Charge Estimation in Battery Management System, **Scientific World Journal.** 2014.

TOPHAM, J., AND SCOTT A. Boeing Dreamliners grounded worldwide on battery checks, **Reuters**, 2013. Web page: http://www.reuters.com/article/us-boeing-dreamliner-idUSBRE90F1N820130117. Accessed: 02/02/2019.

TORLAY, L., BERTOLOTTI, M.P., THOMAS, E., BACIU, M. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. **Brain Informatics** volume 4, pages 159–169, 2017.

URBAIN, M.; RAEL, S. State estimation of a lithium-ion battery through kalman filter. … **Conference 2007** PESC:2804–2810, 2007.

USABC Electric Vehicle Battery Test Procedures Manual. **US Department of Energy**. 1996.

VAPNIK, V.; CHERVONENKIS, A., A note on one class of perceptrons. **Automation and Remote Control**, 25, 1964.

VAPNIK, V.; LERNER, A., Parttern recongnition using generalization portrait method. **Automation and Remote Control**, 24: 774-880, 1963.

VINOD, H. D. (2006), "Maximum entropy ensembles for time series inference in economics". **Journal of Asian Economics**, 17(6), 955–978.

VINOD, H. D.; LOPEZ-DE-LACALLE, J. Maximum Entropy Bootstrap for Time Series: The meboot R Package. **Journal of Statistical Software**, v. 29, n. 5, p. 1–19, 2009.

WANG, W., SHI, Y., LYU, G., DENG, W. Electricity Consumption Prediction Using XGBoost Based on Discrete Wavelet Transform. COMPUTER SCIENCE and ENGINEERING. **2nd International Conference on Artificial Intelligence and Engineering Applications (AIEA 2017)** ISBN: 978-1-60595-485-1. 2017.

WASKOM et al. mwaskom/seaborn: v0.9.0. **Zenodo**. 2018, https://doi.org/10.5281/zenodo.1313201.

WILLIARD, N., HE, W., HENDRICKS, C., AND PECHT, M. Lessons learned from the 787 Dreamliner issue on lithium-ion battery reliability. **Energies** 6: 4682–4695, 2013.

WINGFIELD-HAYES, R. Dreamliner: Boeing 787 planes grounded on safety fears. **BBC News**. Web page: https://www.bbc.com/news/business-21054089. Accessed: 12/02/2019.

World Energy Outlook, Special Report Energy and Air Pollution, **International Energy Agency**, 2016.

X. HU, C. ZOU, C. ZHANG, Y. LI, Technological developments in batteries: a survey of principal roles, types, and management needs, **IEEE Power Energy Mag**. 15 (5) (2017) 20–31.

XING Y, MA EWM, TSUI KL, PECHT M. Battery management systems in electric and hybrid vehicles. **Energies**, 4, 1840–57, 2011.

XIONG, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease**. Science 347**, 6218 (2015).

XIONG, R., SUN, F., GONG, X., & GAO, C. A data-driven based adaptive state of charge estimator of lithium-ion polymer battery used in electric vehicles. **Applied Energy**, 113, 1421–1433. 2014.

XIONG, R.; HE, H.; SUN, F.; ZHAO, K. Evaluation on State of Charge estimation of batteries with adaptive extended kalman filter by experiment approach. **IEEE Trans Veh Technol**, 62:108–17, 2013.

XU, J.; GAO, M.; HE, Z.; HAN, Q.; WANG, X. State of charge estimation online based on EKF-Ah method for lithium-ion power battery. In: **Proceedings of the 2009 2nd International Congr Image Signal Process CISP'09**; 2009.

XU, L.; WANG, J.; CHEN, Q. Kalman filtering state of charge estimation for battery management system based on a stochastic fuzzy neural network battery model. **Energy Convers Manag**, 53, 33–9, 2012.

XUAN, W.; LIN, M.; WEI, T.; JIALEI, Q.; MENGNA, Z. State of charge (SOC) estimation of Ni- MH battery based on least square support vector machines. **Adv Mater Res**, 211–212:1204–9, 2011.

YANG, N., ZHANG, X., LI, G. State of charge estimation for pulse discharge of a LiFEPO4 battery by a revised Ah counting. **Electrochim. Acta** 151, pp. 63-71.

YATSUI, M.W.; BAI, H. Kalman filter based state-of-charge estimation for lithium-ion batteries in hybrid electric vehicles using pulse charging. **Veh Power Propuls Conf (VPPC)**, 2011 IEEE 2011:1–5, 2011.

ZHANG, Y., SONG W., Lin S, Feng Z. A novel model of the initial state of charge estimation for LiFePO 4 batteries. **J Power Sources**, 248:1028–33, 2014.

ZHANG, Y., XIONG, R., HE, H., OU, X., PETCH, M. State of charge-dependent aging mechanisms in graphite/Li (NiCoAl)O_2 cells: Capacity loss modeling and remaining useful life prediction. **Applied Energy**, v. 255. 2019. https://doi.org/10.1016/j.apenergy.2019.113818

ZHAO, J., & DIMIROVSKI, G. M., Quadratic stability of a class of switched nonlinear systems. **IEEE Transactions on Automatic Control**, *49*(4), 574–578, 2004.

ZHAO, R. KOLLMAYER, P., LORENZ, R.D., JAHNS, T. A Compact Methodology via a Recurrent Neural Network for Accurate Equivalent Circuit Type Modeling of Lithium-Ion Batteries. **IEEE Transactions on Industry Applications**, v. 9994. 2018.

ZHENG, F., XING, Y., JIANG, J., SUN, B., KIM, J., PECHT, M. Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries. **Applied Energy**, 183, 513-525. 2016.

ZHENG, L.; ZHANG, L.; ZHU, J.; WANG, G.; JIANG, J. Co-estimation of state-of-charge, capacity and resistance for lithium-ion batteries based on a high-fidelity electro- chemical model. **Appl Energy**, 180:424–34, 2016.

ZHU, Z.; SUN, J.; LIU, D. Online state of charge EKF estimation for LiFePO4 battery management systems. **ISPACS 2012 - IEEE Int Symp Intell Signal Process Commun Syst**, 609–14, 2012.

ZOU, Y., HU, X., MA, H., LI, S. E. Combined State of Charge and State of Health estimation over lithium-ion battery cell cycle lifespan for electric vehicles. **Journal of Power Sources**, 273, 793-803. 2015.