ONLINE CO-ESTIMATION OF STATE OF CHARGE AND VOLTAGE DYNAMICS OF LI-ION BATTERIES VIA PHYSICS-INSPIRED MODELING

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

---

by
Omidreza Ahmadzadeh
May 2025

Examining Committee Members:

Damoon Soudbakhsh, Advisory Chair, Mechanical Engineering
Yan Wang, Computer and Information Sciences
Philip Dames, Mechanical Engineering
Daniel Jacobs, Mechanical Engineering
Haijun Liu, Mechanical Engineering
Joseph Picone, External Reader, Electrical and Computer Engineering

ABSTRACT

This study presents an interpretable physics-inspired, data-driven approach to discovering governing equations of Li-ion batteries for state-of-charge (SOC) and voltage dynamics. A key parameter for the safe and efficient utilization of these batteries is SOC, which represents the remaining charge in the battery. SOC is not directly measurable and has to be estimated based on other measurements. Despite significant efforts by the industry and academia, the state-of-the-art SOC estimation algorithms have significant errors in low and high SOC regions, leading to original equipment manufacturers (OEMs) limiting the range of charge and discharge of the batteries, hence limiting the endurance and range in high-demand applications such as electric vehicles. Here, we propose a novel approach to SOC estimation.

This research introduces an interpretable physics-inspired, data-driven technique that estimates SOC and discovers the governing equations of battery voltage and SOC. The algorithm employs a sparse identification method designed to uncover governing equations suitable for such nonlinear systems from a library of potential functions. We selected these functions based on the battery's electrochemistry rather than generic terms. This process results in a model with connections to physics. A key to this modeling process is a sparsification step. Here, we proposed a novel formulation for sparsifying the library terms instead of the common approach of using maximum likelihood analysis of training data. We formulated the problem as a regularization with hyperparameters. The new formulation allows for the use of multiple datasets to address the shortcomings of the previous approaches while balancing the model's accuracy and complexity.

The previous parsimonious modeling techniques were sensitive to noisy measurements. To address this issue, we augment the modeling technique with a joint unscented Kalman filter (JUKF), enabling more accurate estimates of SOC and voltage. The JUKF mitigates the effects of noisy voltage measurements. The experimental data demonstrate that the identified model with JUKF achieves a root mean square

error (RMSE) of 1% for SOC prediction, which is a significant improvement over the common approach of equivalent circuit model augmented by extended Kalman filters (RMSE: 6%). Additionally, the model achieves an RMSE of 0.6 mV for voltage correction.

Finally, we address one of the most important and complex issues in the battery management systems of high-demand applications: SOC estimation. We developed a co-estimation framework that utilizes JUKF to update model parameters, account for noise effects, and estimate the SOC. This framework eliminates the need for initial SOC values and ensures convergence by using voltage dynamics as an online SOC-voltage map, making it suitable for real-time applications with uncertain SOC values. In an unseen standard city driving cycle, the model is initialized with a 20% initial SOC error and voltage measurement noise. The SOC values converge to the true values with an RMSE of 1% (Voltage RMSE = 4 mV). The model also performs robustly across temperatures (10°C and 40°C), achieving a SOC RMSE of less than 3%.

In conclusion, this dissertation offers a novel control-oriented data-driven framework for discovering governing equations and state estimation of complex systems. We applied the method to the critical task of accurate and reliable SOC estimation in battery management systems, focusing particularly on low SOC levels where battery behavior is highly nonlinear. By integrating physical insights with data-driven techniques and enhancing robustness with JUKF, this approach advances battery reliability and operational range for electric vehicles and energy storage systems.

KEYWORDS: Nonlinear System Identification, Energy Systems, Estimation and Filtering, Robust Estimation, Autotuning, Machine Learning in Modeling

Dedicated to my family.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my Ph.D. advisor, Dr. Damoon Soudbakhsh, for giving me the opportunity to join his team at Temple University. His unwavering support, insightful discussions, and invaluable guidance have been truly transformative and inspiring throughout this journey. It has been an honor and privilege to work under his mentorship, and I will forever cherish the profound lessons and experiences gained through this research.

I am also sincerely grateful to my dissertation committee members, Dr. Philip Dames, Dr. Daniel A. Jacobs, Dr. Haijun Liu, Dr. Joseph Picone, and Dr. Yan Wang, for their time, helpful feedback, and support in helping me complete this milestone. I was fortunate to take classes from Dr. Dames, Dr. Picone, and Dr. Jacobs during my Ph.D., where their engaging courses laid a strong foundation for my battery modeling research. I also deeply appreciated my discussions with Dr. Wang, whose industrial insights were invaluable.

My heartfelt thanks also go to the current and former members of the DSLab (Mohsen Derakhshan, Renato Rodriguez), the EVSLab, and my friends who supported me throughout this journey. Their camaraderie and collaboration enriched my Ph.D. experience and made it more manageable.

Lastly, I am profoundly grateful to my family, especially my parents, for their unconditional love, understanding, and emotional support. Without them, this achievement would not have been possible.

This accomplishment marks the end of one chapter and the beginning of another. I am eager to apply the knowledge and skills I have gained to new and impactful endeavors. Thank you to everyone who supported me along the way—this success is as much yours as it is mine.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

ADMM	Alternating Direction Method of Multipliers
AIC	Akaike Information Criterion
BMS	Battery Management System
BT	Balanced Truncation
CC	Constant Current
CCCV	Constant Current Constant Voltage
CV	Constant Voltage
DDM	Data-Driven Modeling
DFN	Doyle-Fuller-Newman
DMD	Dynamic Mode Decomposition
DoD	Depth of Discharge
DRT	Distribution of Relaxation Times
ECM	Equivalent Circuit Model
EIS	Electrochemical Impedance Spectroscopy
EKF	Extended Kalman Filter
ERA	Eigensystem Realization Algorithm
FSR	Full Scale Range
JUKF	Joint Unscented Kalman Filter
LASSO	Least Absolute Shrinkage and Selection Operator
LFP	Lithium Iron Phosphate
LiB	Li-ion Battery
MAE	Mean Absolute Error
MPM	Multiple Particle Model
MSE	Mean Square Error
NMC	Nickel Manganese Cobalt
NRMSE	Normalized Root Mean Square Error
OCV	Open-Circuit Voltage

ODE	Ordinary Differential Equation
OKID	Observer Kalman Identification
P2D	Pseudo-Two-Dimensional
PDE	Partial Differential Equation
POD	Proper Orthogonal Decomposition
PyBaMM	Python Battery Mathematical Modelling
RMSE	Root Mean Square Error
SINDy	Sparse Identification of Nonlinear Dynamics
SOC	State-Of-Charge
SOH	State-Of-Health
SOP	State-Of-Power
SPM	Single Particle Model
SPMe	Single Particle Model With Electrolyte
STLS	Sequentially Thresholded Least-Squares
STRidge	Sequentially Thresholded Ridge Regression
SVD	Singular Value Decomposition
UDDS	Urban Dynamometer Driving Schedule
UKF	Unscented Kalman Filter
US06	US-Highway Aggressive Drive Cycle

CHAPTER 1

INTRODUCTION

Batteries are ubiquitous in everyday applications such as electric vehicles, laptops, and cell phones. Lithium-ion batteries are the preferred energy storage solutions due to their high energy and power density [1]. With the global shift toward electrification and sustainable energy solutions, efficiently managing these batteries is becoming increasingly crucial to ensure safe, reliable, and optimized performance [2].

A critical component of battery energy storage systems is the battery management system, which regulates battery performance and ensures that the battery operates within safe limits. This is crucial to avoid conditions that can lead to degradation, failure, and, in extreme cases, thermal runaway, fire, or explosion [3]. A battery management system continuously monitors key battery parameters and states to ensure stable operation.

One of the most important parameters for monitoring battery performance is the state of charge (SOC), which indicates the amount of energy remaining in the battery at any given time. SOC is defined as the ratio of the available charge stored in the battery to its maximum capacity. In Li-ion batteries, SOC is directly related to the amount of lithium ions in the anode. SOC is at 100% when the concentration of lithium ions in the anode is at its maximum while the concentration in the cathode is at its lowest. For example, in an electric vehicle, 100% SOC means that the vehicle is fully charged.

Accurate SOC information is vital for preventing overcharging and over-discharging, both of which can significantly reduce a battery's lifespan and pose safety risks. However, SOC is not measurable directly, and the battery management system relies on sophisticated algorithms and extensive offline experiments for real-time estimation [4]. By accurately estimating SOC, the battery management system can regulate battery performance, maintain safe operation, and keep the system within optimal operating conditions.

While SOC estimation is a critical task in battery management systems, accurately estimating SOC in high-demand applications such as electric vehicles is highly challenging due to the complex electrochemical processes occurring within the battery. This estimation results in relatively large errors at low and high state-of-charge regions due to the highly nonlinear behavior of batteries in these regions. In addition, operating factors such as temperature and battery aging further complicate this estimation. Several considerations, such as avoiding compromising the safety of batteries, lead to limiting the operating state-of-charge region of the batteries, leading to a significant reduction in endurance/range of systems and inefficiency in other tasks such as regenerative braking.

Several approaches have been proposed for SOC estimation. One solution involves modeling the battery dynamics based on its electrochemical processes [5–7]. However, these techniques require detailed knowledge of the battery's internal material properties and are computationally intensive due to the complexity of the battery's dynamics. Another common approach is the equivalent circuit model, which is widely used for online SOC estimation in battery management systems. Equivalent circuit models utilize passive electrical components to develop a simplified, linear model based on measurable data. However, they struggle to accurately estimate SOC across the entire operating range since battery behavior is highly nonlinear, and equivalent circuit models fail to capture this complexity. As a result, these models can limit the battery's usable energy, reducing efficiency in demanding applications, such as electric vehicles.

Therefore, a more reliable and accurate model is needed for SOC estimation that works across the entire operating range using the available data. Data-driven modeling has shown the ability to capture complex battery dynamics for SOC estimation without requiring knowledge of the battery's internal parameters, offering potential improvements for real-time battery safety and performance. However, previous at-

tempts with data-driven models have struggled to generalize well to unseen conditions. This dissertation proposes a novel physics-inspired, data-driven approach to address these limitations, aiming to create an accurate, efficient, and control-oriented SOC model.

The main objective of this study is to develop tractable data-driven models to discover the governing equations of Lithium-ion batteries. Our hypothesis is that The battery's voltage and SOC dynamics can be represented by a few terms from the measured data for input/output, and SOC levels can be accurately estimated via these learned dynamics. To test this hypothesis, the following aims will be pursued:

Aim 1: Discovering a Battery's Voltage and SOC Dynamics

- Create a nonlinear interpretable data-driven model for Lithium-ion battery.
- Enhance the modeling technique by including physics-inspired terms.
- Formulate a multi-objective cost function to capture the dynamics.

Aim 2: Robust Modeling with Noisy Data

- Extend data-driven model using a Joint Unscented Kalman Filter to mitigate effects of noisy measurements.
- Develop a co-estimation framework to update model parameters using measurement data.

Aim 3: Data Generation and Model Development

- Design experiments on a single cell at different temperatures.
- Generate data from a detailed cylindrical cell battery model.
- Conduct experimental studies and collect data.

Figure 1.1 outlines the steps taken to achieve the main objective and fulfill the aims of this study.

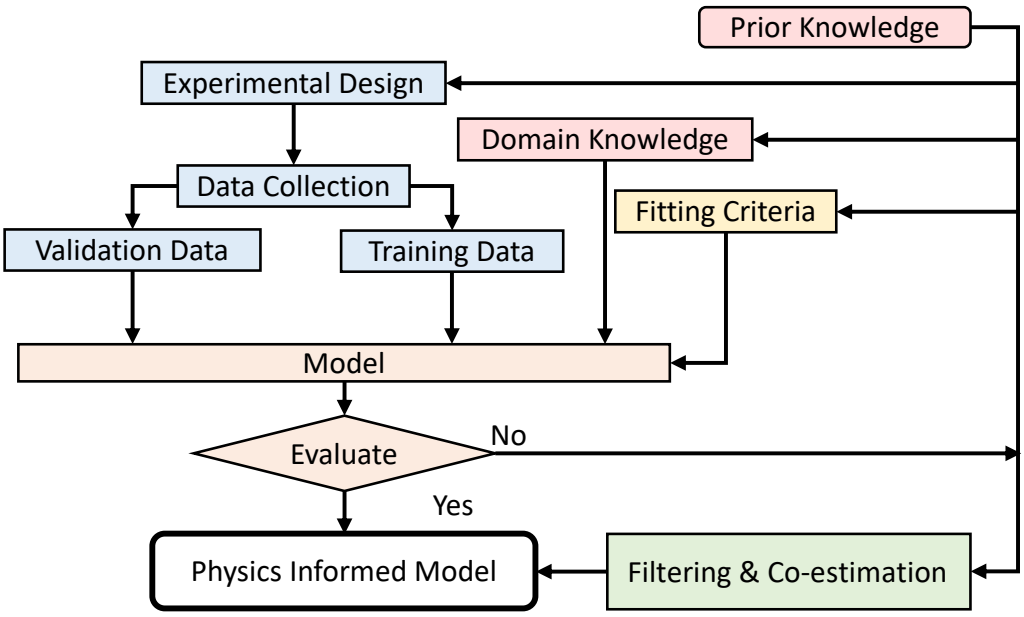


Figure 1.1. Proposed steps for modeling.

1.1 Organization of the Work

This dissertation presents the development of an algorithm for battery modeling in SOC applications, supported by both simulation and experimental analysis. It offers a detailed explanation of interpretable data-driven modeling for accurate SOC estimation in batteries. The remainder of this dissertation is organized as follows:

Chapter 2: Provides a comprehensive literature review on battery modeling for SOC estimation, highlighting the strengths and limitations of existing approaches. We explore the potential of data-driven modeling for battery applications and discuss current gaps in the field, explaining how this research aims to address those gaps using interpretable data-driven methods. This chapter sets the foundation for Aim 1 by identifying key gaps in SOC and voltage dynamics modeling techniques, which the proposed methodology seeks to address.

Chapter 3: Describes a brief overview of interpretable data-driven modeling and explains the proposed method for battery modeling. We present the generic formulation used to build the model and outline the steps involved. The potential of

this methodology to capture battery dynamics is evaluated, along with its limitations. A simplified simulation-based battery model is employed for data generation and model construction. This chapter contributes to Aim 1 by focusing on creating an interpretable data-driven model for capturing battery dynamics.

Chapter 4: Building on the insights from Chapter 3, this chapter addresses the limitations of the initial data-driven model. The generic method lacks generalizability, so we develop a methodology incorporating battery physics knowledge into the data-driven model. We improve the model's foundation and generalizability by analyzing the electrochemical processes within the battery. Furthermore, a novel cost function is proposed to optimize model parameters, enhancing the model's ability to accurately capture battery dynamics. A high-fidelity simulation-based battery model is employed for data generation and model construction. This chapter directly supports Aim 1 by enhancing the modeling technique with physics-inspired terms and formulating a multi-objective cost function to improve model accuracy. It also contributes to Aim 3 by employing a high-fidelity simulation-based battery model for data generation.

Chapter 5: Introduces a joint unscented Kalman filter to achieve robustness and accuracy in the presence of noisy voltage measurements. Additionally, a co-estimation framework is introduced to update model parameters in real-time, addressing uncertainties in the modeled dynamics or the initial SOC value. This chapter addresses Aim 2 by extending the data-driven model to handle noisy data using a JUKF and developing a co-estimation framework.

Chapter 6: Details the experimental setup and procedures used to validate the model. We provide an overview of the equipment, including the battery tester, holder, and thermal chambers. The charge/discharge methods, including current and voltage ranges, rest times, and cutoff parameters, are explained in detail. We present the datasets used to create and evaluate the data-driven model and describe an offline

method for determining SOC as a reference value. This chapter supports Aim 3 by detailing the experimental design and data collection processes.

Chapter 7: The final developed model is applied to experimentally collected data. We present the optimized model parameters and show how the model performs using noisy measured data. We evaluate the model's ability to ensure the convergence of SOC to its true value by filtering out noise from the measured voltage data and updating the model's parameters to address uncertainties. The model is also tested under various operating conditions to verify its generalizability and applicability within battery management systems, ensuring accurate real-time SOC estimation from measurable data. This chapter contributes to Aim 2 and Aim 3 by validating the robustness of the model against noisy data and unknown initial SOC values and verifying its performance across various operating conditions.

Chapter 8: In the final chapter, we summarize the key findings of this dissertation and provide concluding remarks. Recommendations for future work are also discussed.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a comprehensive literature review of battery dynamics modeling approaches used for SOC estimations.

The Battery Management System (BMS) is a critical component of battery energy storage systems, responsible for regulating the charging and discharging processes to ensure safe and reliable battery operation [6,8]. Comprising both hardware and software, the BMS measures key battery outputs, including voltage and temperature, and communicates with power systems to report the remaining energy in the battery, referred to as the state-of-charge (SOC) [9]. Accurate SOC information is essential for the BMS to enhance battery safety by preventing overcharging and discharging, as well as improving its lifespan and performance. However, there is no direct measurement of SOC. Hence, BMS relies on estimating SOC dynamics from an estimated initial SOC value and measured data to regulate the battery's performance and ensure its safe operation. Reliable models and algorithms are needed to achieve accurate SOC estimation. Accurate models of Li-ion batteries (LiBs) for SOC estimation allow for prolonged life and increased performance. Several modeling techniques have been developed to model LiBs behaviors [5, 7, 10–14]. Fig. 2.1 displays different modeling techniques for batteries and Fig. 2.2 presents a comparision of the battery modeling methods. High fidelity indicates the model's capability to capture the dynamics of the battery, while the simple models have less computational cost and are less detailed. In the following sections, we describe each technique along with its merits and demerits.

2.1 Emprical Models

Empirical methods are one of the most widely used techniques to estimate SOC due to their computationally affordable nature. An example of such an approach is coulomb counting, which uses the integral of current as a measure of charge the bat-

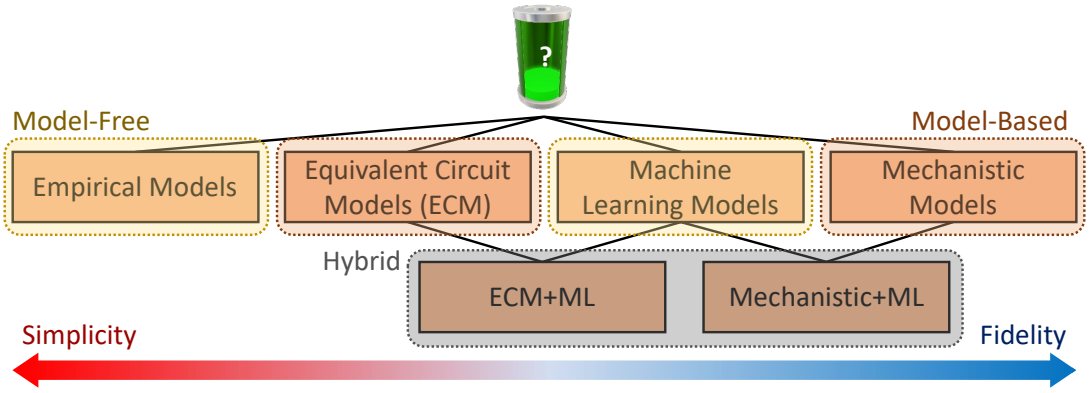


Figure 2.1. Methods for modeling battery systems.

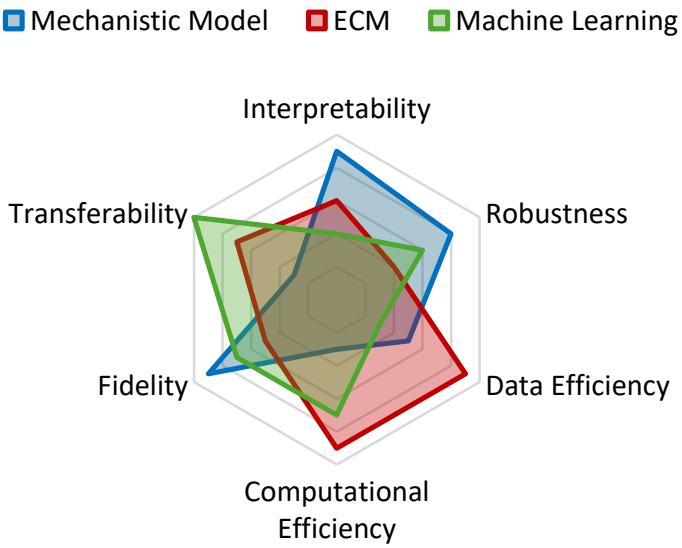


Figure 2.2. Comparison of battery SOC Modeling and Prediction Methods.

tery receives or provides [15,16]. The accuracy of coulomb counting is highly sensitive to the estimated value of the initial SOC values and the precision of measuring the electrical current [17].

SOC can also be estimated using a predetermined (offline) voltage-SOC mapping. For instance, Figure 2.3 illustrates the voltage-SOC mapping for different current rates at 25°C. Although SOC can be estimated based on the voltage at this temperature, it is important to note that voltage behavior is significantly affected by the input current. In high-demand applications where electrical current varies frequently, utilizing voltage-SOC mapping can be challenging.

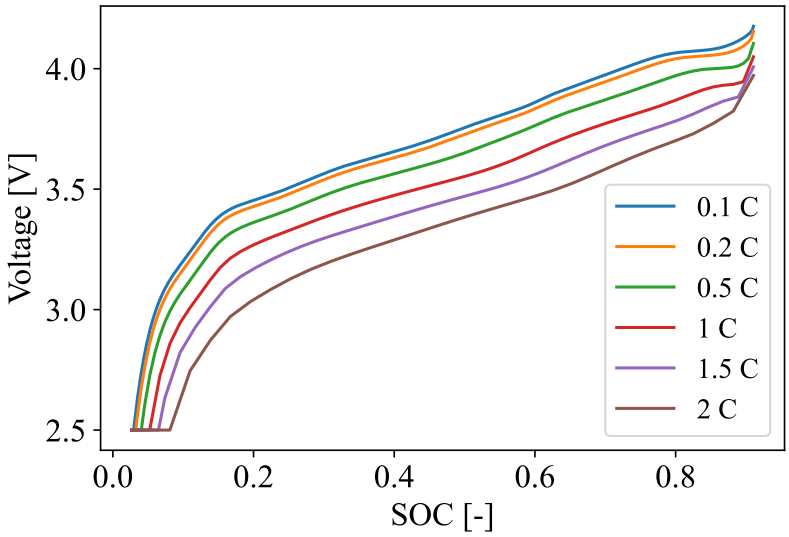


Figure 2.3. Voltage-SOC mapping for different current rates.

Another empirical approach for estimating SOC involves using the open-circuit-voltage (OCV) [18]. This method determines the corresponding SOC values from a predetermined (offline) OCV-SOC mapping. The OCV represents the stable voltage response of the battery when the electrical current is minimal or when the battery is at rest, making it generally independent of the current. However, OCV measurements can only be taken when the electrical current is low, or the battery has rested long enough to allow transient processes to stabilize. Consequently, OCV is not measurable during aggressive current profiles, necessitating the use of alternative models to predict OCV.

Another limitation of this method is the flatness of the OCV-SOC mapping, which makes accurate OCV measurements critical. Even minor fluctuations in OCV can lead to significant changes in the estimated SOC. For example, at an electrical current of 0.1 C (where 1 C represents the current required to fully charge or discharge the battery in one hour), as shown in Figure 2.3, small errors in OCV measurement can cause large errors in SOC estimation.

While empirical methods are popular for their simplicity, they often suffer from low accuracy and robustness, particularly in low-temperature or low SOC regions, as

well as in high-demand applications such as electric vehicles. Below is a summary of the key advantages and disadvantages of empirical models for SOC estimation:

Advantages

- Computationally affordable and efficient.
- They are easy to implement.

Disadvantages

- Highly sensitive to initial SOC values and precise current and voltage measurements.
- Voltage-SOC mapping struggles with varying currents in high-demand applications.
- Errors in OCV measurements can lead to significant SOC estimation errors.
- Low accuracy and robustness in low SOC regions, low temperatures, or aggressive operating conditions.

2.2 Mechanistic Models

Among the most common approaches are mechanistic models such as the pseudo-two-dimensional (P2D) model [1, 7, 10–13, 19, 20]. Traditionally, mechanistic models express the processes inside the battery with analytical equations. These models are commonly referred to as white-box models. The main processes are electrochemical reactions, mass, heat, charge transport, degradation, and deformation.

A limited section of a battery's dynamic is easily observable. However, mechanistic models can effectively capture the dynamics of these processes. Unfortunately, no feasible or usable model describes all the dynamic processes [10, 21]. The mechanistic models need to be designed according to their intended purpose. Models with detailed descriptions of the phenomena and processes are developed for analysis and

diagnosis. In contrast, models for control and optimization purposes need to be computationally fast and accurate. The final goal is to develop or adjust models to be more computationally affordable while maintaining detailed physical phenomena [22].

The P2D model describes the electrochemical behavior based on the porous electrode theory [23] and considers the cathode and anode of the battery composed of a series of spherical particles submerged in a liquid electrolyte. The model captures the spatiotemporal distribution of the states like Lithium concentration in solid and electrolyte phases. The P2D model has been developed by including the thermal dynamics [24] and battery aging [25, 26]. The P2D model provides the high-fidelity battery physical model; however, these models require the tedious task of determining the material and physical parameters of the batteries [14]. Furthermore, it has enormous complex equations and cannot be applied for real-time analysis.

For example, the well-known P2D model with thermal dynamics, also known as the Doyle-Fuller-Newman (DFN) model [21], requires 33 parameters to build the model. Some parameters, such as diffusivity coefficients of solid phases, are hard to measure and need destructive tests. Furthermore, the electrolyte phase's conductivity coefficient and diffusivity coefficient depend on electrolyte concentration [27]; therefore, measuring them is very laborious and cumbersome. More importantly, heat is generated inside the cell during the battery operation, and most of the transport and reaction parameters strongly depend on temperature. Identifying these parameters is particularly challenging.

The single particle model (SPM) [6] is a simplified version of the P2D model that captures the most critical aspects of LiBs while reducing the computational complexity. The SPM assumes no variations in the spatial domain and that the electrolyte concentration remains constant [6]. These assumptions allow for a coarse discretization of the spatial domain, resulting in a single node for each electrode, both positive and negative. Although SPM is computationally fast, it is only accurate for

Table 2.1. Number of required electrochemical parameters for mechanistic models.

Model	Number of parameters	Computational complexity
P2D	33	high
SPMe	31	medium
SPM	20	low

slow dynamic (less than $1C$ rate) [6]. SPM has been extended to work in a broader range of operations by including electrolyte dynamics (SPMe) [7] and multiple particle model (MPM) [28].

The SPM needs 20 parameters to model the battery's dynamics, which is less than the P2D model. However, they still require battery internal immeasurable parameters such as diffusion coefficients; therefore, these models rely on some nominal parameter values [13, 14]. Table 2.1 displays the number of required parameters for different physics-based models. Moreover, in [29], an iterative extended Kalman filter was employed to estimate SOC and model parameters within the SPM framework, while a multiple-model Kalman filter was used to enhance the P2D model for predicting long-term battery performance [30].

Although the P2D model and its variations describe the process precisely, its accuracy is limited to the modeled phenomena. For instance, it only considers the spatial dynamics of the battery in one direction [22]. Additionally, these models are based on first principles that hold true only in idealized conditions, which are difficult to replicate in real systems. Moreover, modeling unobserved or unknown physical processes remains a significant challenge.

In summary, mechanistic models offer a robust framework for understanding and simulating battery processes; however, they come with challenges in terms of complexity and practical applicability. Below is a summary of the key advantages and disadvantages of mechanistic models for SOC estimation:

Advanteges

- Mechanistic models provide detailed insights into the internal processes of the battery, including electrochemical reactions, mass transport, heat transfer, and charge movement, making them ideal for in-depth analysis.
- They offer high interpretability and fidelity, allowing for battery design and performance optimization improvements.
- Mechanistic models are extrapolable and can be applied across a wider range of operating conditions.

Disadvanteges

- Mechanistic models require a large number of parameters, many of which are difficult to measure or estimate accurately.
- They are based on idealized first principles, which may not hold in real-world systems.
- High-accuracy models like DFN involve solving complex nonlinear PDEs, leading to high computational costs that limit their use in real-time applications.
- They cannot account for unknown or unobserved phenomena, restricting their ability to fully capture the system's dynamics.

2.3 Equivalent Circuit Models

Another approach to simulate the battery's dynamic behavior is using equivalent circuit models (ECMs). The ECM represents the dynamics of LiB using linear models constructed by groups of passive electrical components such as resistors and capacitors [31]. The advantage of the ECM is that the measurable data (current and voltage) are enough to describe the behavior of the battery, resulting in a considerable reduction of model order [32]. We can model the battery without considering

electrochemical processes inside the cell using the ECM. This model represents the electrical response of the battery through a series of resistors and capacitors (RC) elements, where each pair of R and C are connected in parallel. In SOC estimation, the goal is to determine the OCV of the battery from the measured terminal voltage and current using the electrical parameters of the ECM. The SOC is then estimated by applying a SOC-OCV mapping. The ECM has received significant attention due to its simplicity and low computational cost [33–40]. However, due to the lack of connection to physics, the validity range of ECMs is limited to narrow operating conditions in which they have been calibrated. To employ ECMs for a broader range, a large amount of experimental data is needed under different operating conditions [41]. Furthermore, ECMs often rely on constant parameters, which limits their ability to capture the battery's nonlinear behavior, especially in the low SOC region or varying temperatures.

An issue with the ECM models is the uncertainty in the SOC values due to the uncertainties in the ECM parameters (R and Cs) and the dependency of these parameters on SOC values. Therefore, Kalman filters have been widely used to adjust these parameters to improve the model performance and adapatability [42]. One of the key advantages of the Kalman filter is its ability to provide real-time updates, allowing for continuous refinement of the estimates based on new measurements. This is particularly important in SOC estimation, where fluctuations in current and voltage can introduce significant uncertainty. For instance, the Kalman filter has been employed for SOC-OCV mapping in studies such as [18, 43], demonstrating its effectiveness in improving estimation accuracy. Additionally, the Kalman filter family has been integrated with ECMs to estimate both SOC and SOH by updating the model parameters dynamically [44–47].

An alternative to simple ECMs, with typically one to five RC elements, is modeling based on electrochemical processes. The electrochemical ECMs let us interpolate

and extrapolate the model's parameters and are typically applied to the components of a LiB, including the anode, cathode, electrolyte, separator, and current collector [48]. The electrochemical ECMs are employed to optimize the battery in terms of energy content, cooling, and overall performance. Identification and parametrization of each process are achieved by analyzing electrochemical impedance spectroscopy (EIS) in the frequency domain or by using the distribution of relaxation times (DRT) method [49, 50]. Such electrochemical-based ECMs are not used for SOC prediction and estimation due to model complexity and EIS measurement requirements.

The accuracy of the model depends on the quality of the measured data. Obtaining precise and reliable measurement curves is challenging due to the numerous disturbances. Working in the frequency domain for EIS needs to adjust the amplitude of the currents to avoid damaging the cell and producing nonlinear responses, which requires specific devices for performing experiments [49]. Furthermore, there is a trade-off between the current amplitude and the resolution of the measured data. In constructing ECM with DRT, a filter is required to separate the processes; however, it can potentially disturb or distort the results and time constants [22]. Below is a brief summary of the key advantages and disadvantages of ECM models for SOC estimation.

Advantages

- ECMs are simple models with low computational cost, using basic electrical components like resistors and capacitors to represent battery dynamics linearly.
- ECMs can describe battery behavior using only measurable data, such as current and voltage, without needing to account for internal electrochemical processes.
- The Kalman filter's ability to continuously refine estimates makes it suitable for dynamic environments where conditions change frequently.

Disadvantages

- ECMs are limited to narrow operating conditions due to their lack of connection to underlying physical processes.
- Extending ECMs to cover a wider range of operating conditions requires extensive experimental data collection, and their accuracy is highly dependent on the quality of the measured data.
- Incorporating electrochemical processes into ECMs involves complex identification and parametrization, often relying on frequency response analysis, which requires carefully controlled experiments and filtering techniques.

To address these limitations, ECMs are typically augmented with additional techniques to improve accuracy, as discussed in Section 2.5.

2.4 Data-driven Models

An alternative approach to mechanistic and ECM approaches is to employ state-of-the-art data-driven techniques for modeling energy storage systems. The main objective of these algorithms is to find the relationship between measured input/output data and essential variables of the battery, like SOC and state of health (SOH).

Data-driven modeling can be classified as black-box modeling. In black-box modeling, the relation between input and output is entirely based on the available data, without any assumptions about the underlying physical processes. Machine learning algorithms, such as support vector regression, have been employed to map the input space with a nonlinear function to a feature space, where regressions are then applied. This approach aims to achieve generalized predictive power based on the data [51]. In [52], support vector-based machine learning techniques were utilized to predict the remaining useful life of LiB. Additionally, support vector machines can be implemented for online estimation of the SOH for LiB [53].

Expanding on machine learning applications, several techniques have been developed to estimate SOC and SOH. Among these, neural networks and their variations

have gained significant attention due to their ability to capture complex battery behavior. Examples include traditional neural networks [54–57], recurrent neural networks [58,59], artificial neural networks [60], long short term memory recurrent neural network [61]. Other approaches, such as clustering combined with genetic algorithms, have also been explored to improve SOC estimation [62].

One key advantage of data-driven models is that they eliminate the need to measure the internal parameters of batteries, which can be difficult and costly to obtain. However, they require extensive data to find the best fit [62]. Furthermore, their performance is usually constrained by data informativeness and reachness, and overfitting is a common concern [63]. Lastly, they can yield physically inconsistent and inexplicable results due to the nature of the machine learning model.

When the battery chemistries change, mechanistic models need to be adapted to capture the new phenomena [64]. Similarly, data-driven models also require adjustments. However, these adjustments are simpler in data-driven models, as their structure is already established, and only the parameters need to be tuned by retraining the model with new data to account for the new phenomena due to their physics-independent nature. In contrast, mechanistic models require estimating or measuring new chemical parameters, which is a more laborious process. This highlights the greater flexibility and transferability offered by data-driven models [65], making them especially valuable for developing new battery chemistries when their properties and physics are not yet fully understood.

For example, Sodium-ion batteries are expected to have a lifespan exceeding 20 years; however, we lack complete physical information about their behavior over such a long period. Data-driven models can predict their performance without requiring a detailed understanding of the underlying dynamics, providing a powerful tool for forecasting the behavior of emerging battery technologies.

One limitation of machine learning models is their reliance on available data, which is often restricted in practice. These models are typically unable to incorporate additional internal information, such as Li-ion concentrations, which could enhance the accuracy of SOC and SOH estimations. However, researchers are exploring new features beyond traditional voltage, current, and temperature measurements to improve predictions. For example, studies have investigated the use of strain sensor data [66] and acoustic signatures of batteries [67] as alternative inputs for SOC estimation. This highlights the advantage of machine learning models in scenarios where the underlying physics is not fully understood; they can still correlate features such as SOC and a battery's acoustic signature without requiring a detailed understanding of the underlying physical processes.

The computational cost of data-driven modeling is higher than ECMs, making real-time model updates challenging as battery behavior evolves. Consequently, the model is updated occasionally when significant changes in battery response are detected. However, once the model is established, the implementation cost is low.

In summary, data-driven modeling has advantages for SOC estimation, but it should be applied cautiously. Below is the summary of the key advantages and disadvantages in the context of machine learning for SOC estimation:

Advantages

- Data-driven models do not require knowledge of internal battery parameters, simplifying the modeling process. They can discern complex battery patterns using measurable data

- They are flexible and can be easily adapted to new battery chemistries or conditions by simply adjusting parameters through retraining rather than relying on extensive measurements or re-estimation.

- They can utilize novel features to improve estimations when traditional features are insufficient.
- Their implementation costs are low, making them suitable for real-time applications.

Disadvantages

- The performance is highly dependent on data quality. A rich and informative dataset is required to capture the battery's operating conditions, as machine learning models typically do not extrapolate well.
- The quantity of data is also crucial; machine learning algorithms often require large datasets for training and validation, which can be difficult to generate in practice.
- Due to the complexity of the algorithms, there is a higher risk of overfitting, which can lead to poor generalization.
- The results can be physically inconsistent and lack interpretability, as the models do not incorporate physical laws.
- While machine learning allows flexibility in feature selection, including too many or unnecessary features can result in ill-conditioned problems due to correlated features, making appropriate feature selection critical.

There have been efforts to address some of the limitations of data-driven models, such as their lack of connection to physics, by combining them with other techniques. The next section focuses on this type of modeling.

2.5 Hybrid Models

There have been efforts to combine machine learning tools with mechanistic models [68] or ECMs [69] to leverage the strengths of both approaches. These models are

commonly referred to as hybrid models. For example, the ECM is incorporated with deep learning for capacity estimation [70]. The ECM and data-driven modeling can be combined for SOC prediction. The data-driven technique is employed to identify the relation between open-circuit voltage and SOC. The OCV is calculated from ECM, and the SOC is determined from OCV [18, 43]. Thermal coupled ECM with neural network is constructed to capture the battery's dynamic and temperature [71]. The ECM with electrochemical processes is augmented with neural networks to provide an accurate model of battery voltage [72].

In [73], a genetic algorithm was used to optimize the Kalman filter parameters for SOC estimation, further enhancing the accuracy of the hybrid model. Similarly, machine learning techniques such as neural networks have been incorporated with ECMs and Kalman filters, especially when the initial SOC is unknown, to improve estimation performance [74].

We can list the merits and demerits of this hybrid ECM modeling technique as follows:

Advantages

- By integrating machine learning techniques, ECMs can better handle variations in operating conditions, extending their usability beyond their intrinsic limitations.

Disadvantages

- Hybrid models using ECM tend to lack physical interpretability, which can make it challenging to understand the underlying processes.
- These models often require large datasets for training and tuning, which may not always be available.

- Due to the linear nature of ECMs, hybrid models struggle to provide accurate SOC estimations, particularly in low SOC regions, where the battery dynamics become highly nonlinear.
- These models rely on an offline SOC-OCV map, which can vary under different operating conditions. In chemistries like Lithium Iron Phosphate (LFP), where the SOC-OCV curve is relatively flat, accurate estimation becomes challenging.

Mechanistic models provide strong extrapolation capabilities and interpretability due to their foundation in physical principles, though they often come with high computational costs. In contrast, data-driven models are computationally efficient but struggle with extrapolation and lack physical interpretability. This complementary nature of mechanistic and data-driven models highlights the potential for hybrid approaches to overcome the limitations of each method and leverage their combined strengths [75].

Recent efforts have focused on integrating these two modeling paradigms. For example, existing mechanistic models like the SPM with thermal dynamics have been enhanced with feed-forward neural networks to improve model accuracy [68]. Another hybrid approach combines a recurrent neural network with the SPM to better capture the battery's voltage dynamics [76]. In [77], a hybrid electrochemical-thermal-neural network integrates a simplified SPM, a lumped thermal model, and a neural network for the co-estimation of SOC and temperature.

Additionally, a physics-informed neural network in [78] was developed to correlate internal battery states using a P2D model and experimental data, showcasing the potential of combining mechanistic insights with machine learning. In another study [79], neural networks were used to solve the PDEs of the SPMe model, reducing computation time and extending the model's operating range.

While these hybrid models offer enhanced performance, they still rely on internal parameters from mechanistic models like the SPM and require substantial amounts of

data to train the machine learning algorithms effectively. We can list their advantages and disadvantages as follows:

Advantages

- By integrating mechanistic models, hybrid approaches can maintain the strong extrapolation and interpretability provided by the underlying physical principles.
- Machine learning algorithms, such as neural networks, improve the accuracy of mechanistic models by capturing complex dynamics that are difficult to model explicitly.
- Hybrid models extend the operational range of mechanistic models such as SPM, particularly by reducing computational costs and improving performance in real-time applications.

Disadvantages

- These hybrid models often require large amounts of data for training the machine learning components, which can be difficult to obtain, especially for high-fidelity mechanistic models.
- Hybrid models still rely heavily on internal parameters from mechanistic models (e.g., SPM or P2D), which are often difficult to measure accurately.
- Although computational efficiency is improved compared to purely mechanistic models, hybrid models still involve significant complexity, particularly in balancing the mechanistic and machine learning components.

Table 2.2 lists the selected papers for battery hybrid modelings.

Table 2.2. Selected studies for battery hybrid models.

Author	Mechanistic	ECM	Data-driven
Tu et al. [68]	*		*
Cui et al. [69]		*	*
Su et al. [70]		*	*
Wang et al. [71]		*	*
Lin et al. [46]		*	*
Borah et al. [72]		*	*
Ting et al. [73]		*	*
Wang et al. [74]		*	*
Park et al. [76]	*		*
Feng et al. [77]	*		*
Xue et al. [79]	*		*
Hofmann et al. [78]	*		*

2.6 Research Objectives

This chapter presented an overview of various modeling approaches for Li-ion batteries, along with their strengths and limitations. Empirical models provide a simple and computationally efficient solution. However, they lack accuracy and robustness, especially in dynamic operating conditions. Mechanistic models offer interpretability and detailed physical insights into battery processes; however, they are computationally expensive and require extensive information on electrochemical parameters, which can be tedious to obtain. ECMs are computationally efficient but struggle to capture the nonlinear dynamics of the battery, depending on predetermined SOC-OCV mappings, and lack a direct connection to the underlying physics. Data-driven models are capable of capturing complex battery behavior. However, they face challenges in terms of generalizability, require large amounts of data, and lack interpretability. Hybrid models that combine ECM with other techniques offer a wider operational range, yet they inherit the limitations of ECMs. Hybrid models that integrate mechanistic and data-driven approaches still rely heavily on the internal parameters of mechanistic models, which can limit their adaptability.

This review highlights several gaps in battery modeling for online SOC estimation. Based on the ability of data-driven models to capture complex battery dynamics from measurable data without requiring information about internal parameters. This dissertation presents a novel approach to battery modeling for SOC estimation. The proposed model is interpretable, physics-inspired, generalizable, and accurate. The key contributions of this dissertation are as follows:

1. **Interpretable Data-Driven Model:** We focus on developing an interpretable data-driven model, moving beyond common machine learning approaches like neural networks, which typically provide only an input-output relationship. The interpretability of the model allows us to uncover governing equations that describe battery behavior, which can be leveraged for other applications, such as designing optimal controllers for fast charging. The interpretable data-driven approach enables us to discover the underlying dynamics of the battery without requiring in-situ measurements of internal parameters.

2. **Physics-Inspired Model:** Incorporating physics into the model prevents physically inconsistent results and enhances accuracy. Unlike approaches that integrate mechanistic models with data-driven methods—requiring difficult-to-access internal parameters and resulting in high computational costs for high-fidelity models—the goal here is to develop a physics-inspired machine learning model that captures battery dynamics meaningfully. The model features are grounded in battery physics to enhance both accuracy and generalizability, ensuring the model reflects the system's true behavior rather than just fitting data patterns.

3. **Improving Generalizability:** One of the primary limitations of data-driven models is their poor generalizability beyond the range of training data. This dissertation addresses this limitation by incorporating insights from battery

physics and proposing a novel cost function to fine-tune the model parameters. The goal is to design a cost function that sacrifices non-essential terms in the model that may cause overfitting to the training data while still capturing the essential, big-picture dynamics of the battery. By focusing on the core features and simplifying the model, we can remove minor details that don't significantly contribute to the understanding of the battery's behavior. This not only enhances the model's interpretability but also ensures that it generalizes well across different operating conditions, especially in regions where the battery exhibits highly nonlinear behavior. Furthermore, reducing the complexity of the model helps to prevent the inclusion of terms that are overly sensitive to noise or anomalies in the training data, improving both its robustness and adaptability to real-world applications.

4. **Mitigating Noise Effects:** As with most data-driven methods, the accuracy of the model is highly dependent on data quality. In practical applications, data is often noisy, so we augment the data-driven model with Kalman filter techniques to mitigate the effects of measurement noise. This not only improves SOC prediction accuracy but also allows the battery management system (BMS) to calculate the required/generated power more effectively by filtering noise from the voltage data.

5. **Co-Estimating SOC and Voltage:** We expand the Kalman filter approach into a co-estimation framework that simultaneously estimates SOC, adapts model parameters, and corrects voltage data. This framework enables more robust estimation, especially when model uncertainties are present. By updating parameters online, the co-estimation framework addresses uncertainties in initial SOC and current measurements, ensuring convergence to the true SOC. This approach enhances the BMS's real-time accuracy and robustness, par-

ticularly in applications like electric vehicles where precise SOC estimation is critical.

2.7 Summary

In conclusion, this chapter explored various modeling approaches for Li-ion battery SOC estimation, highlighting their respective advantages, disadvantages, and gaps in the current literature. This dissertation seeks to address key challenges related to interpretability and generalizability without the need to measure internal battery parameters. We propose an interpretable data-driven model inspired by battery physics that provides an accurate and reliable solution for real-time SOC estimation across the entire operating range. The model is based solely on measurable data and can be transferred to different battery chemistries without the need to reconstruct the model. In addition, the model is augmented with the Kalman filter family to enhance robustness in the presence of system uncertainties. In the next chapter, we will present the general methodology behind our proposed interpretable data-driven model, demonstrating its potential to capture the complex dynamics of the battery.

CHAPTER 3

A DATA-DRIVEN MODEL OF LI-ION BATTERIES

In this chapter, we provide a review of the state-of-the-art interpretable data-driven modeling techniques and utilize them to develop a data-driven battery model. The results serve as motivation for the need to address the drawbacks of these techniques to achieve generalizable data-driven models. We utilized an enhanced single-particle model with electrolyte (SPMe) to represent the actual plant dynamics of the Li-ion battery model.

3.1 Interpretable Data-driven Modeling

Data-driven modeling (DDM) has recently attracted attention for system identification, prediction, and control [4]. With advances in machine learning, interpretable data-driven modeling techniques have emerged to improve our understanding of these predictions. Hence enabling efficient control design and achieving desired system behavior [80].

Modern machine learning techniques, such as neural networks, can provide accurate models for unknown dynamics; however, they often overfit the training data due to excessive degrees of freedom, and their lack of interpretability limits their application [4].

Recently, interpretability and explainability have become central topics in machine learning research [81]. Interpretability refers to making the operations of a system understandable to humans, either through direct introspection or an explanation of the system's output [82]. In this case, interpretability means that we can derive meaningful insights about the system from the relationships between the data and the identified model.

Effective explainability techniques in interpretable models allow users to comprehend, manage, and trust these systems [83]. People are generally reluctant to adopt methods that are not understandable or interpretable [84]. Furthermore, in-

terpretability ensures that only significant variables affect the output, improving robustness by revealing potential model perturbations [85].

One classic example of interpretable modeling is Kalman's minimal realization, which identifies linear empirical models from time series data [86]. Later, statistical techniques such as Proper Orthogonal Decomposition (POD) and Balanced Truncation (BT) emerged, capturing the main dynamics of large-scale, complex systems. More advanced methods like the Eigensystem Realization Algorithm (ERA) [86, 87] and Observer Kalman Identification (OKID) [88, 89] have evolved, utilizing efficient algorithms such as Singular Value Decomposition (SVD) [90] to streamline computations.

Dynamic Mode Decomposition (DMD) [91] is another computationally affordable data-driven technique that leverages SVD. DMD constructs control-oriented linear models from time series data and identifies the primary dynamic modes within high-dimensional nonlinear systems. The identified models exhibit consistent spatio-temporal patterns and approximate the dominant eigendecomposition of the system [91–94].

Sparse Identification of Nonlinear Dynamical Systems (SINDy) [92] is another technique, similar to DMD, that approximates system dynamics using both linear and nonlinear functions. SINDy uses sparse regression to discover a parsimonious model from a library of basis functions.

SINDy offers a control-oriented, interpretable, and generalizable model, allowing for enhanced performance by integrating other techniques. Originally, SINDy was designed to identify the dynamics of autonomous systems [92]. Compared to other machine learning algorithms, such as neural networks, SINDy requires significantly less data due to its reliance on linear regression [95].

SINDy has evolved to handle systems with inputs [96], making it possible to integrate model predictive control [97]. Recent developments have also extended

SINDy to deal with implicit dynamics [98], Monte Carlo library search [99], stability promotion [100], noise robustness [95, 101], and integration with Kalman filters [102, 103].

SINDy's primary assumption is that a small number of functions can describe the system's dynamics. However, selecting these functions often relies on heuristics, and inappropriate choices may prevent SINDy from discovering the correct governing equations. To address this issue, we incorporate system-specific knowledge to guide the selection of basis functions. For example, in the context of Li-ion batteries, we include functions representing electrochemical processes in the base library and supplement this with an extended library of additional terms relevant to battery physics. The Monte Carlo library search is then used to randomly combine these terms, aiming to discover the governing equations while balancing model sparsity and accuracy.

Traditionally, SINDy sparsification is tuned using an Akaike Information Criterion (AIC)-inspired cost function [104]. However, AIC tends to overfit the training data, especially as the data set grows. To counter this, we propose a novel cost function that leverages both training (in-sample) and validation (extra-sample) data and incorporates the number of active terms as a measure of complexity. This approach prevents overfitting and leads to a more generalizable model that performs well across varying operating conditions.

As with many data-driven techniques, SINDy's performance can degrade in the presence of noise. Moreover, SINDy depends heavily on initial conditions and does not update identified dynamics using observation data. To address this, we augment SINDy with a joint unscented Kalman filter (JUKF), creating a co-estimation framework. This framework enables parameter updating, noise mitigation, and improved predictive performance, making it highly suitable for SOC estimation under uncertain and noisy conditions. We chose JUKF because the dynamics discovered

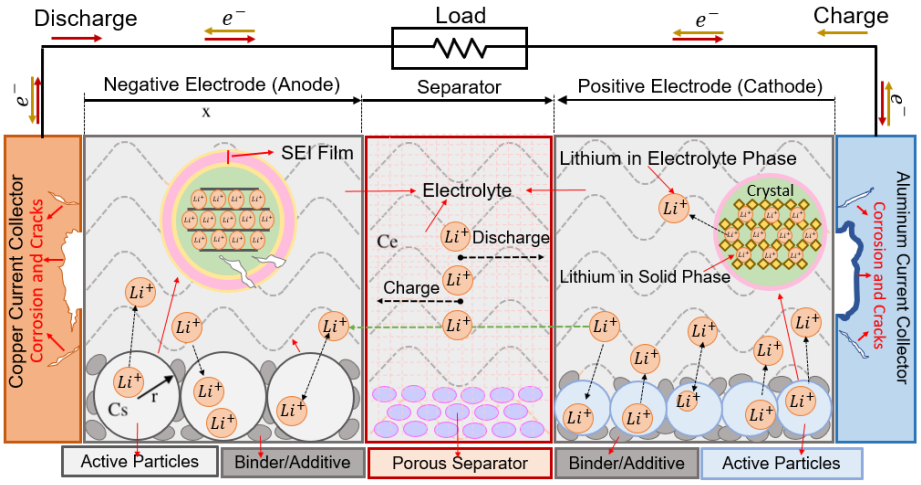


Figure 3.1. Schematic of the internal processes of LiBs.

through SINDy are nonlinear, and JUKF is suitable for nonlinear systems. Unlike the extended Kalman filter (EKF), which linearizes the dynamics and introduces additional errors, JUKF uses a transformation approach that preserves the accuracy of the nonlinear dynamics.

In this chapter, we employed the general SINDy approach to explore its potential to uncover the unknown dynamics of a battery using input-output data generated from simulations. This allowed us to assess the method's capability to model battery behavior and identify key dynamics based solely on input-output data.

3.2 Mathematical Model of Lithium-ion battery SPMe

Here, we provide a summary of the enhanced SPMe [7, 105]. The terminology of the model is presented in Table 3.1. Fig. 3.1 shows a schematic of the main internal processes of LiBs.

Enhanced SPMe reduces the complexity of the full-order representation of LiB models while minimizing the modeling errors [7]. SPMe models the cathode and anode as single spherical particles and incorporates the dynamics of the electrolyte through phase diffusion equations and electrolyte potential. The input and output of enhanced SPMe include the current density $I(t)$ and voltage $V(t)$, respectively.

Table 3.1. SPMe Symbol Definitions

Symbols	Description [units]
$a^{\pm}$	Specific interfacial surface area [m^2/m^3]
brug	Bruggeman porosity [-]
c_e	Li concentration in electrolyte phase [mol/m^3]
$c_s^{\pm}$	Li concentration in solid phase [mol/m^3]
$c_{ss}^{\pm}$	Li concentration at solid particle surface [mol/m^3]
$c_{s,\max}^{\pm}$	Maximum concentration of solid material [mol/m^3]
$D_s^{\pm}$, D_e	Diffusivity of solid, electrolyte phase [m^2/s]
F	Faraday's constant [C/mol]
$i_0^{\pm}$	Exchange current density [A/m^2]
I	Applied current [A/m^2]
$k^{\pm}$	Kinetic reaction rate [$A{\cdot}m^{2.5}/mol^{1.5}$]
L^j	Length of region $j \in \{-, \text{sep}, +\}$ [m]
R	Universal gas constant [J/mol-K]
$R_f^{\pm}$	Solid-electrolyte interphase film resistance [Ω-m^2]
$R_s^{\pm}$	Particle radius in solid phase [m]
t	Time [sec]
T	Temperature [K]
t_c^0	Transference number [-]
$U^{\pm}$	Open circuit potential of solid material [V]
V	Terminal voltage [V]
α_a, α_c	Anodic, cathodic charge transfer coefficient [-]
$\varepsilon_s^{\pm}$, ε_e	Volume fraction of solid, electrolyte phase [-]
ϕ_e	Electrolyte electric potential [V]

The SPMe method is developed using the following underlying assumptions. First, the solid concentration of Lithium in both electrodes is assumed to be constant along the spatial direction x. This allows the lithium-ion concentration to be modeled as a diffusion process in the solid phase (3.1) and in the electrolyte phase (3.2)-(3.3):

$$\frac{\partial c_s^{\pm}}{\partial t}(r,t) = \frac{1}{r^2}\frac{\partial}{\partial r}\left[D_s^{\pm} r^2 \frac{\partial c_s^{\pm}}{\partial r}(r,t)\right], \tag{3.1}$$

$$\frac{\partial c_e^{\pm}}{\partial t}(x,t) = \frac{\partial}{\partial x}\left[\frac{D_e^{\text{eff}}(c_e^{\pm})}{\varepsilon_e^{\pm}}\frac{\partial c_e^{\pm}}{\partial x}(x,t)\right] \mp \frac{1-t_c^0}{\varepsilon_e^{\pm} F L^{\pm}} I(t), \tag{3.2}$$

$$\frac{\partial c_e^{\text{sep}}}{\partial t}(x,t) = \frac{\partial}{\partial x}\left[\frac{D_e^{\text{eff}}(c_e^{\text{sep}})}{\varepsilon_e^{\text{sep}}}\frac{\partial c_e^{\text{sep}}}{\partial x}(x,t)\right], \tag{3.3}$$

where D_e is a function of electrolyte concentration, $D_e^{\text{eff}} = D_e(c_e).(\varepsilon_e^j)^{\text{brug}}$, r represents the radial position within the particle, and the boundary conditions are similar to the full-order model [21]. It is assumed that the true exchange current density along x can be substituted with the average value $i_0^{\pm}(x,t)$. Also, it is assumed that the charge transfer coefficients, α_a and α_c are constant and equal to $\alpha = 0.5$. This yields the following representation of the average exchange current density.

$$i_0^{\pm}(x,t) = k^{\pm}\left[c_{ss}^{\pm}(x,t)\right]^{\alpha_c}\left[c_e(x,t)\left(c_{s,\max}^{\pm} - c_{ss}^{\pm}(x,t)\right)\right]^{\alpha_a}. \tag{3.4}$$

We note that this representation is independent of x. The next step involves calculating the difference in electrolyte potential $\phi_e(x,t)$ at both current collectors [7].

The final step in the development of the SPMe method involves using the implicit finite difference method to solve (3.1) and (3.2), using the Newton method to solve (3.3), and integration by parts to solve the electrolyte potential [7]. Therefore, the relationship between the input $I(t)$ and the output $V(t)$ is given by:

$$\begin{aligned} V(t) = {} & \frac{RT}{\alpha F}\sinh^{-1}\left(\frac{-I(t)}{2a^+L^+i_0^+(t)}\right) + U^+(c_{ss}^+(t)) \\ & -\frac{RT}{\alpha F}\sinh^{-1}\left(\frac{I(t)}{2a^-L^-i_0^-(t)}\right) - U^-(c_{ss}^-(t)) \\ & +\phi_e^+(0^+,t) - \phi_e^-(0^-,t) - \left(\frac{R_f^+}{a^+L^+} + \frac{R_f^-}{a^-L^-}\right)I(t) \end{aligned} \tag{3.5}$$

Additionally, the energy of the battery can be calculated by normalizing the Li-ion concentration in the negative electrode [10]. This calculation yields the SOC as follows:

$$\text{SOC}(t) = \frac{\frac{3}{R_s^-}\int_0^{R_s^-} r^2 c_s^-(r,t)dr - c_{s,\min}^-}{c_{s,\max}^- - c_{s,\min}^-} \tag{3.6}$$

3.3 Sparse Data-Driven Modeling of LiBs

Here, we present a data-driven approach to represent the input-output dynamics of the electrochemical processes of LiBs presented in section 3.2. DDM reduces the costly requirement of identifying complex battery parameters. We investigate data-driven models using a modified version of the SINDyc (SINDy with Control) method [92,96]. SINDyc is a technique that approximates the governing equation through a library of candidate terms (linear and nonlinear) and a set of coefficients. First, we seek governing equations in the form of (3.7) using only electrical current and voltage (Model I). Later, we augment Model I by including current, voltage, and SOC data (Model II). Model I has the following form:

$$\frac{dV}{dt}(t) \stackrel{\text{def}}{=} \dot{V} = f\left(V(t), I(t)\right), \tag{3.7}$$

where $f\left(V(t), I(t)\right)$ is a nonlinear function. We look for a representation of f by considering linear and nonlinear terms of the current and voltage signals and denote it by $\mathbf{\Theta}(\mathbf{V}, \mathbf{I})$. These terms are refined via a sparse regression approach to determine the most dominant terms and develop a reduced-order model of the system. We consider polynomials and trigonometric functions to form the following $\mathbf{\Theta}(\mathbf{V}, \mathbf{I})$ library (potential terms):

$$\mathbf{\Theta}(\mathbf{V}, \mathbf{I}) = \begin{bmatrix} | & | & | & | & | & & | & \\ 1 & \mathbf{V} & \mathbf{V}^2 & \cdots & \sin(\mathbf{V}) & \cdots & \mathbf{I}^2 & \cdots \\ | & | & | & | & | & & | & \end{bmatrix}. \tag{3.8}$$

The columns of $\mathbf{\Theta}(\mathbf{V}, \mathbf{I})$ in (3.8) are time-series data, and they form the potential functions. For example, voltage is a time series vector of m measured samples as follows:

$$\mathbf{V} = \begin{bmatrix} V(t_1) & V(t_2) & \cdots & V(t_m) \end{bmatrix}^T. \tag{3.9}$$

Similarly, the $\dot{V}$ in (3.7) is a time series vector given by $\dot{\mathbf{V}} = \begin{bmatrix} \dot{V}(t_1) & \dot{V}(t_2) & \cdots & \dot{V}(t_m) \end{bmatrix}^T$. A reduced-order model can be obtained by promoting sparsity in $\boldsymbol{\Theta}(\mathbf{V}, \mathbf{I})$ through a sparse vector of coefficients $\boldsymbol{\Xi}$. Therefore, the battery model (3.7) (Model I) can be written as:

$$\dot{\mathbf{V}} = \boldsymbol{\Theta}(\mathbf{V}, \mathbf{I})\boldsymbol{\Xi}. \tag{3.10}$$

The sparse vector of coefficients $\boldsymbol{\Xi}$ can be found using a LASSO (least absolute shrinkage and selection operator) or ridge regression optimization. The LASSO optimization problem has a 1-norm regularization term with $p = 1$ in (3.11), whereas the Ridge Regression has a 2-norm regularization term with $p = 2$. Parameter λ is used to balance the complexity and accuracy of the model.

$$\boldsymbol{\Xi}^* = \arg\min_{\Xi} \left\| \dot{\mathbf{V}} - \boldsymbol{\Theta}(\mathbf{V}, \mathbf{I})\boldsymbol{\Xi} \right\|_2 + \lambda \left\| \boldsymbol{\Xi} \right\|_p \tag{3.11}$$

LASSO regression works well for uncorrelated features [106] but has limitations when the parameters are correlated, which is common in this sparsity-promoting problem. An approach based on Ridge regression, called sequentially thresholded ridge regression (STRidge), is more suitable for this purpose. STRidge nullifies any coefficient that is smaller than a threshold, ξ_{th}, and repeats until the nonzero coefficients of $\boldsymbol{\Xi}$ converge. To implement the regression methods, we approximate $\dot{\mathbf{V}}$ by a first degree finite difference:

$$\dot{V}(t_k) = \frac{V(t_k) - V(t_{k-1})}{T_s}, \tag{3.12}$$

where T_s is the sampling time. Then, we need to find the optimal regularization parameters λ and ξ_{th}. Mean square error (MSE) is a common approach to finding models that best fit the data; however, it does not consider the complexity of the

sparse system and tends to lead to over-fitted solutions. To achieve an accurate ROM, the following AIC-inspired loss function [107, 108] was used.

$$\mathcal{L}\left(\boldsymbol{\Xi}\right) = N \ln\left(\frac{1}{N}\left\|\dot{\mathbf{V}} - \boldsymbol{\Theta}(\mathbf{V},\mathbf{I})\boldsymbol{\Xi}\right\|_2^2 + \varepsilon\right) + 2K. \tag{3.13}$$

In (3.13), K is the number of nonzero coefficients in $\boldsymbol{\Xi}$, and N is the number of measured data in time and the size of $\dot{\mathbf{V}}$. Parameter, ϵ, is a small number ($\ll 1$) that provides the trade-off to avoid overfitting the data and ensuring a parsimonious model [107].

As an alternative approach, we augment model (3.7) with SOC, assuming that besides the voltage and current data, SOC is readily available; therefore, the governing equation has the following form:

$$\dot{V} = f\left(V(t), I(t), \mathrm{SOC}(t)\right). \tag{3.14}$$

Polynomial and trigonometric functions of SOC are added to the library in (3.8), and the sparse model becomes:

$$\dot{\mathbf{V}} = \boldsymbol{\Theta}(\mathbf{V},\mathbf{I},\mathbf{SOC})\boldsymbol{\Xi}. \tag{3.15}$$

The same optimization procedure as STRidge with AIC-inspired loss function (3.13), is implemented for identifying the sparse coefficients.

3.4 Simulation Results

In this section, we show the performance of the data-driven models of section 3.3 to predict the response of the actual plant dynamics using only input-output data. We use the enhanced SPMe of section 3.2 as the plant dynamics to generate the output signal $V(t)$ from a given input signal $I(t)$. The battery parameters used for the data generation process are from [109]. We add a Gaussian noise signal to the simulated

data to account for sensor noise. The applied input $I(t)$ consists of a random signal with an amplitude of 5 A/m^2.

To find the sparsest model that can predict the system dynamics, we explored sparse modeling using both LASSO and STRidge regression. SINDyc is used to identify the noise-free model. AIC-inspired loss function (3.13) was used to find the optimal regularization parameter λ and ξ_{th} in (3.11). Fig. 3.2 shows the AIC-inspired loss function for the STRidge model with $\lambda = 10^{-5}$. It demonstrates that very small ξ_{th} values result in non-sparse models with high AIC-inspired loss, which overfits the data. According to Fig. 3.2, the optimal value for the ξ_{th} parameter is 0.0178 since it results in the smallest loss and promotes the highest level of sparsity in the model by omitting less important terms. The LASSO optimization was solved using the alternating direction method of multipliers algorithm (ADMM) [110]. The same procedure and ADMM were used for the LASSO cost function. Since LASSO did not result in a sparse model, we implemented the sequential threshold for LASSO optimization by defining the tolerance similar to STRidge to promote sparsity in the system (STLASSO). The optimal value of λ and ξ_{th} for STLASSO were found to be 10^{-5} and 1.78×10^{-4} using AIC-inspired loss function, respectively. Table 3.2 presents the sparse model's coefficients derived with STRidge and STLASSO. This Table shows that the STRidge resulted in a more sparse model, which was expected since they had the same λ and the tolerance in STRidge was significantly larger than STLASSO's. In our case, the STRidge resulted in fewer terms involving the electrical current (input).

Fig. 3.3 illustrates that the identified model trained on a dataset consisting of 500 samples can accurately predict the response of the LiB from a random input signal with 1500 samples, yielding a validation error (NRMSE) of 1.21%.

We used a sine-sweep signal input to evaluate Model I's applicability to other types of input. We used the same coefficients found using the stochastic input for

Table 3.2. Comparison of identified model with STRidge and STLASSO for noise-free model

Coeff.	Library terms	STLASSO	STRidge
ξ_1	1	4.1091	144.8080
ξ_2	V	11.2388	264.3248
ξ_3	V^2	26.8896	293.9786
ξ_4	V^3	46.4685	-153.7041
ξ_5	V^4	-13.4855	15.3124
ξ_6	$1/(V^2+1)$	0.4039	22.8214
ξ_7	$\sin(V)$	1.3587	252.5772
ξ_8	$\sin(2V)$	0	0
ξ_9	I	-7.6277	-8.0264
ξ_{10}	I^2	0.0428	0
ξ_{11}	I^3	0.0174	0.0198
ξ_{12}	I^4	-0.0003	0
ξ_{13}	$1/(I^2+1)$	0.0053	0
ξ_{14}	$\sin(I)$	-0.0282	-0.0513
ξ_{15}	$\sin(2I)$	0	0

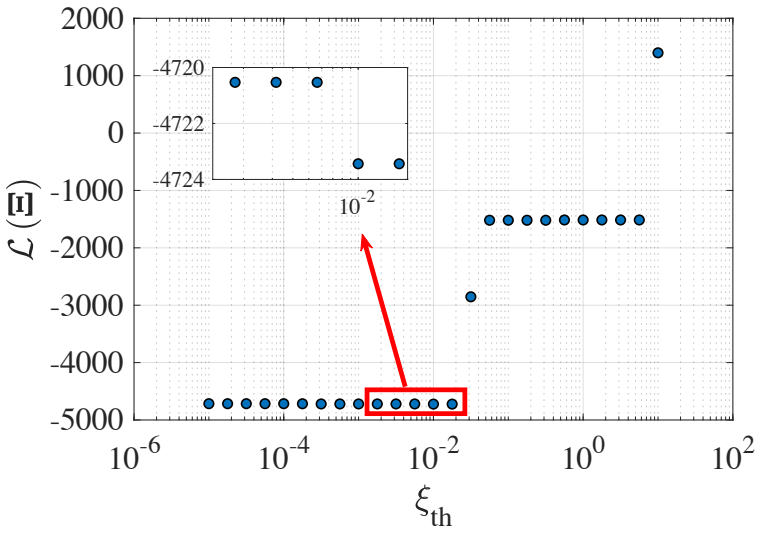


Figure 3.2. Loss function for a number of varying ξ_{th} for the noise-free system.

this purpose. The electrical current $I(t)$ in the form of a sine-sweep signal as shown in Fig. 3.4a was applied to the system. The sinusoidal input was applied for 1 second, and its frequency was changed from 1 Hz to 100 Hz during this time. Fig. 3.4b illustrates

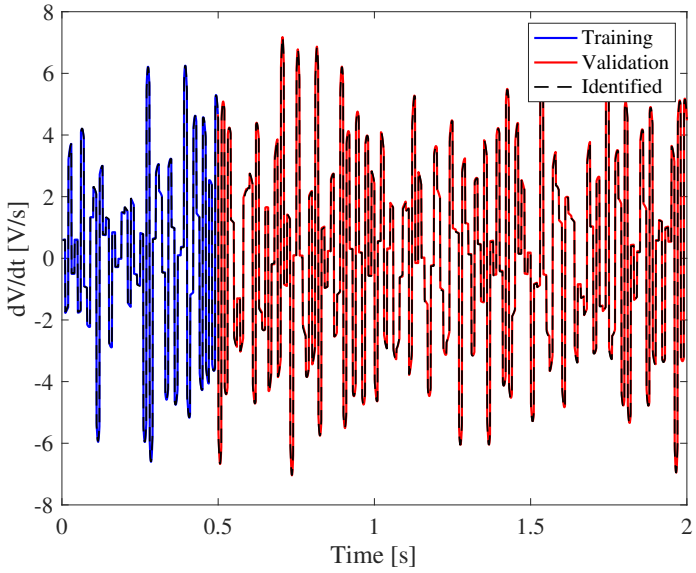


Figure 3.3. Identified system with training data (blue) and validated with validation data (red).

the identified output compared to the system's (SPMe) output. The model predicted the output with NRMSE of 1.59% by only using random input signals for training and without any information about the battery parameters.

We evaluated Model I's robustness to noisy measurements with an added noise with amplitude equal to 15% of the voltage's variance. Fig. 3.5a shows a random current as the model's input and Fig. 3.5b shows that the identified model can track the output in the presence of the noise with NRMSE of 1.97%.

The model parameters are SOC-dependent. This is due to the fact that SOC is a function of concentration (see (3.6)). On the other hand, the concentration changes the output voltage with a given electrical current as described in (3.5). Therefore, Model I parameters are valid only for a given concentration (SOC). We determined Model I parameters with stochastic inputs with 0.75% SOC increments in the range of 0% to 75%. As shown in Fig. 3.6, these parameters follow smooth trends, suggesting the possibility to switch among them as the SOC changes.

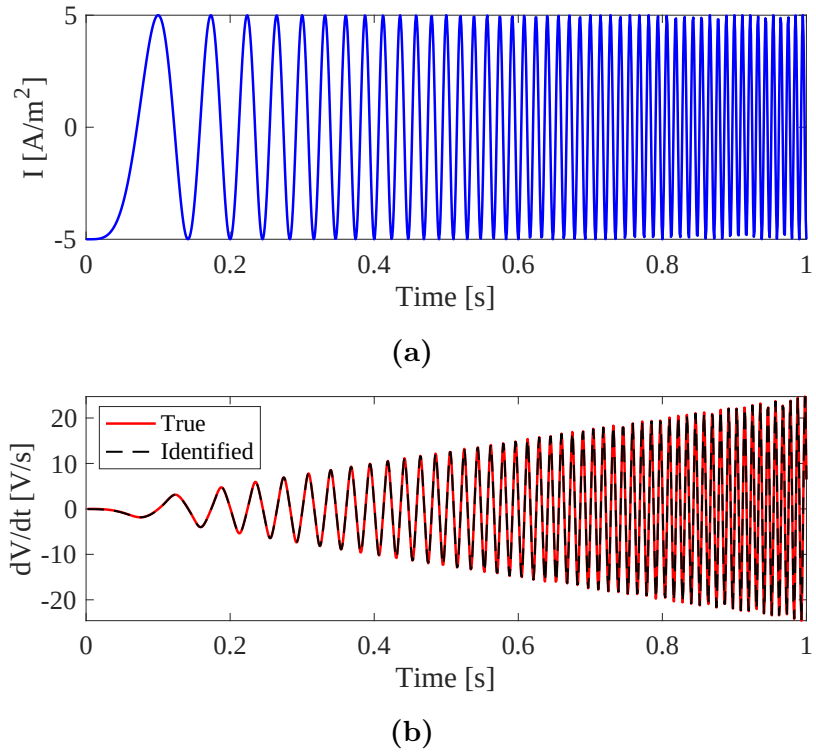


Figure 3.4. Identifying the system with the sine-sweep data. (a) Sine-sweep signal current. (b) Derivative of the voltage.

While Model I worked with varying SOCs, it needed to switch among multiple sets of parameters as the SOC changed. We added SOC-related terms to the model to arrive at a simpler model (Model II) without switching modes. We generate a library $\boldsymbol{\Theta}$ of 22 candidate terms with the same conditions as the previous library. We generated this model and evaluated it with US06 Highway Driving Schedule segments [111]. The SOC changes from 40% charge to 25% charge during this scenario. Fig. 3.7 shows the effect of the US06 input profile on the SOC and the output voltage with a sampling time of 1 second. Fig. 3.8 illustrates that Model II was trained using the first 900 data samples and predicted the following 900 samples for the validation test, where it achieved an NRMSE of 4.78%.

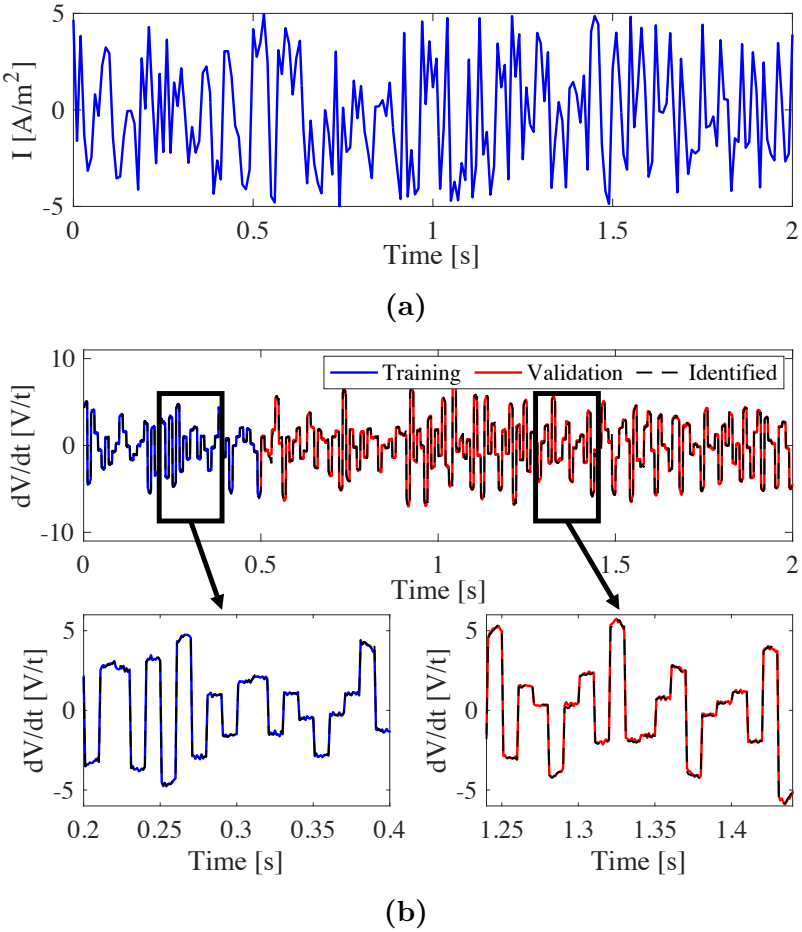


Figure 3.5. Validating the identified system with noisy data. (a) random input for the system with noise, and b) The derivative of the voltage. The blue line in (b) is the training phase, and the red is the validation phase. The bottom graphs are zoomed-in views of the output.

3.5 Summary

In this chapter, we developed data-driven models of voltage prediction using input-output data. We used the enhanced SPMe for the data generation. The identified model was developed using sparse identification. We used sparse regression methods, such as STRidge and LASSO, to balance model performance and complexity by optimizing the AIC-inspired loss function. Model performance and robustness were assessed via validation and generalization tests, where Model I achieved NRMSE values of less than 1.6%. Additionally, Model I was evaluated for robustness to noise, achieving an NRSME of 1.97%. We showed that the Model I parameters are a func-

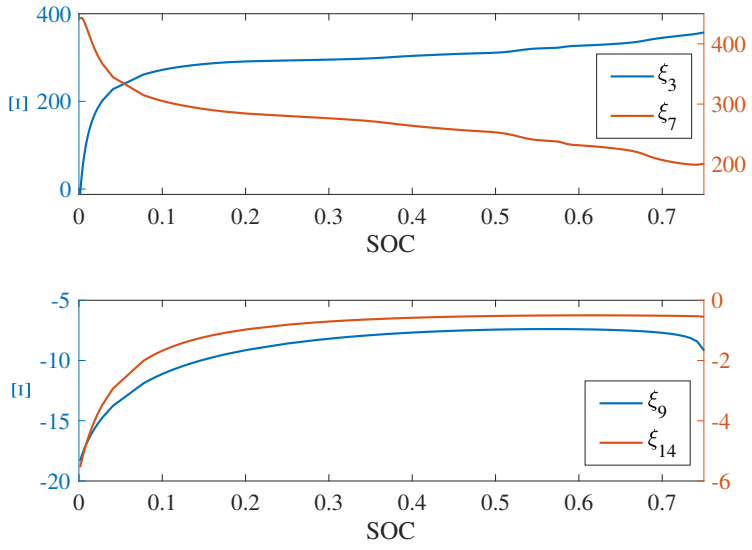


Figure 3.6. Trend of four parameters of Model I versus SOC (ξ_3, ξ_7, ξ_9, and ξ_{14}).

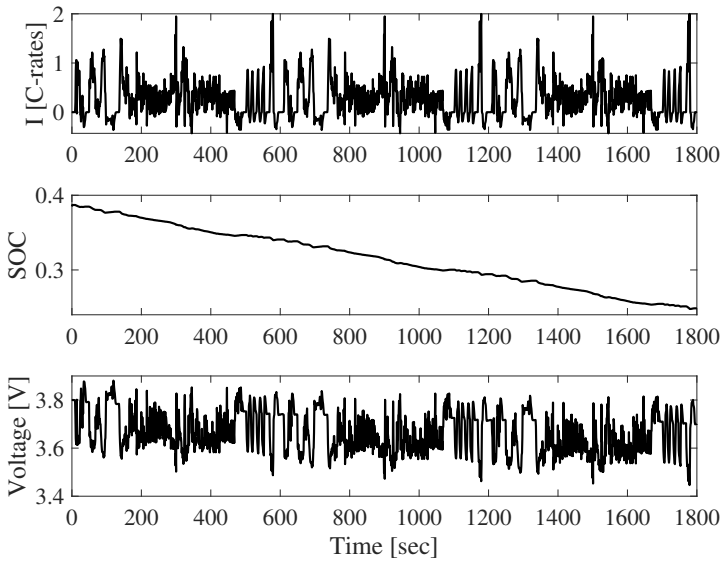


Figure 3.7. US06 Highway Driving profile via enhanced SPMe.

tion of SOC. Lastly, we incorporated SOC in the model and implemented a current signal based on US06. In this simulation, the NRMSE values reach 4.78%.

This chapter demonstrated the potential of the SINDy algorithm for uncovering the battery's underlying dynamics. However, the discovered governing equations have

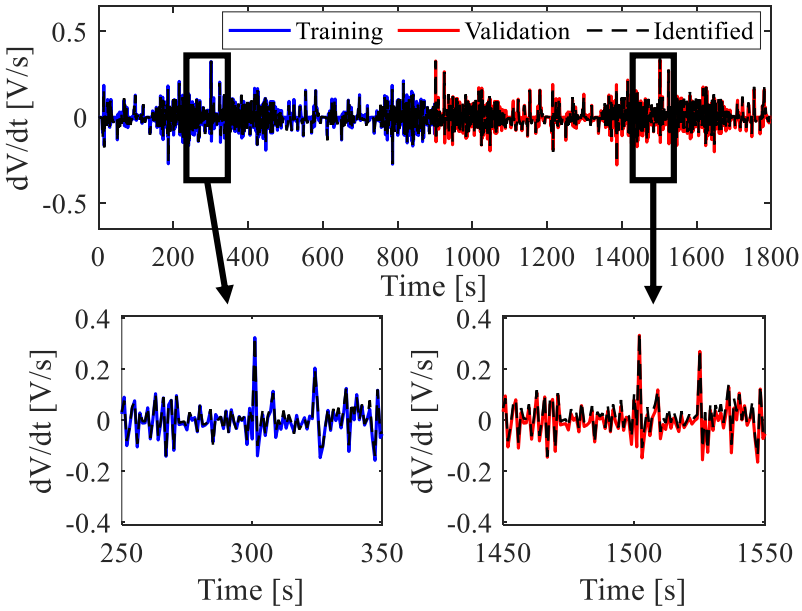


Figure 3.8. Validating Model II with the US06 Highway Driving signal.

limitations in terms of predictability and generalizability. These challenges arise due to several factors:

- **Lack of connection to battery physics**: The model does not include any information from the battery, which limits its interpretability and predictability.
- **Selection of sparsification parameters**: Using maximum likelihood from training data alone can lead to overfitting and reduced generalizability, particularly when training data growth.
- **Library selection issues**: Choosing inappropriate terms for the library may fail to capture the correct system dynamics, while including too many terms leads to an ill-conditioned model with correlated elements.
- **Sensitivity to noise and initial conditions**: The model's performance degrades under noise, and it heavily relies on initial values.

In the next chapter, we address these limitations by enhancing our methodology:

- **Physics-informed library construction**: We will select terms for the library based on battery physics, ensuring the model captures the correct dynamics and improves generalizability.

- **Refined sparsification tuning**: Instead of relying solely on AIC-inspired cost function, we propose a novel multi-objective cost function that considers training error, validation error, and model complexity, striking a balance between accuracy and generalizability.

These improvements will bring the model closer to real-world applications, increasing its reliability and interpretability for tasks like battery modeling.

CHAPTER 4

PHYSICS-INSPIRED MODELING

This chapter presents a foundation for physics-informed modeling of Li-ion batteries. Our approach is based on analysis of the governing equations of the DFN (Doyle-Fuller-Newman) model [27] and identifying the general forms that appear in the solution of these equations. Then, we use these terms as the core library terms rather than the general terms discussed in Chapter 3. Next, we develop discrete models for both voltage and SOC dynamics. We define our nonlinear tractable data-driven model by improving the SINDy from Chapter 3 for discrete systems and propose a cost function to enhance generalizability and operating range. These improvements increase the model's accuracy for more aggressive signals and broader operating ranges.

4.1 Interpretable nonlinear tractable data-driven model

In this section, we present our modeling approach. In section 4.1.1, we summarize a sparsification method for discovering nonlinear interpretable data-driven models. In section 4.1.2, we describe our approach to tune sparsification parameters through a hyperparameters reformulation.

4.1.1 Sparse data-driven modeling

Here, we present a nonlinear tractable data-driven modeling approach for energy storage systems. The model is created by sparsifying a library of candidate terms. The sparsification is achieved by adapting the SINDyC algorithm [92]. The SINDyC algorithm assumes that the system dynamic has the following form:

$$\mathbf{x}[k+1] = \mathbf{f}\left(\mathbf{x}[k], \mathbf{u}[k]\right), \tag{4.1}$$

where $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{u} \in \mathbb{R}^q$ are the state and input, respectively. $\mathbf{f}\left(\mathbf{x}, \mathbf{u}\right)$ represents the governing dynamics, and we assume that a few active terms can represent it. The sparse regression technique aims to find these active terms from a library of candidate

linear and nonlinear functions (also called the system's observables). To evaluate the library, we collect m snapshots of the state $\mathbf{x}$ and input $\mathbf{u}$ in time as follows:

$$\mathbf{X} \stackrel{\text{def}}{=} X[k,m] \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{x}[k] & \mathbf{x}[k+1] & \cdots & \mathbf{x}[k+m-1] \end{bmatrix}^T, \tag{4.2}$$

$$\mathbf{U} \stackrel{\text{def}}{=} U[k,m] \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{u}[k] & \mathbf{u}[k+1] & \cdots & \mathbf{u}[k+m-1] \end{bmatrix}^T. \tag{4.3}$$

The algorithm aims at identifying the dominant library terms that represent the system dynamics (4.1) as in:

$$\mathbf{X}' = \Theta(\mathbf{X},\mathbf{U})\Xi, \tag{4.4}$$

where $\mathbf{X}' \stackrel{\text{def}}{=} X[k+1,m]$ is a shifted temporal matrix of $\mathbf{X}$, and Ξ is a sparse vector of coefficients corresponding to elements of the library, and $\Theta(\mathbf{X}, \mathbf{U})$ is the candidate library that has the following form:

$$\Theta(\mathbf{X},\mathbf{U}) = \begin{bmatrix} | & | & & | & | & & | & \\ \mathbf{X} & \mathbf{X}^2 & \cdots & \sin(\mathbf{X}) & \exp(\mathbf{X}) & \cdots & \mathbf{U} & \cdots \\ | & | & & | & | & & | & \end{bmatrix}. \tag{4.5}$$

The sparse coefficients of (4.4), Ξ, are determined by minimizing the following cost function that captures the trade-off model accuracy and complexity:

$$\Xi^* = \arg\min_{\Xi} \|\mathbf{X}' - \Theta(\mathbf{X},\mathbf{U})\Xi\|_2 + P(\Xi). \tag{4.6}$$

In (4.6), $P(\Xi)$ is the regularizer to promote sparsity. By defining $P(\Xi) = \xi_{\text{th}} \|\Xi\|_0 + \lambda \|\Xi\|_2$, we arrive at:

$$\Xi^* = \arg\min_{\Xi} \|\mathbf{X}' - \Theta(\mathbf{X},\mathbf{U})\Xi\|_2 + \xi_{\text{th}} \|\Xi\|_0 + \lambda \|\Xi\|_2, \tag{4.7}$$

where hyperparameters ξ_{th} and λ are the threshold and regularization parameters, respectively. These parameters balance accuracy and complexity. The 0-norm nullifies components in Ξ whose absolute values are less than a threshold ξ_{th} to increase parsimony. In addition,λ is the regularization parameter that regulates the coefficients. We can balance the model's accuracy and complexity by the combination of these terms. We repeat the regression until the nonzero coefficients of Ξ converge. This algorithm is known as sequentially thresholded ridge regression (STRidge) [106, 112] and works better than other regression methods like LASSO and sequentially thresholded least-squares (STLS) due to correlated features present in the electrochemical processes as some of them are governed by PDEs.

4.1.2 Optimizing the hyperparameters

Choosing the threshold and regularization parameter is not trivial since different values lead to different sparse coefficients and models. Increasing ξ_{th} will eliminate more terms and produce a sparser model, and increasing λ will reduce the magnitude of the coefficients. A common approach to determine the threshold ξ_{th} and the regularization parameter λ is minimizing an Akaike information criterion (AIC)-inspired cost function on the training data [104]. Here, we propose reformulating the regularization problem to include these parameters as hyperparameters and optimizing them using an additional validation dataset.

Fig. 4.1 shows the procedure for choosing the hyperparameters with training and validation datasets. Selecting hyperparameters involves two steps that balance model complexity and accuracy. In the first step, we define priory sets of hyperparameters and use training data with known input/output to build the sparse models for all sets of hyperparameters by solving (4.7). The second step involves minimizing a cost function (4.8) comprised of the root mean square error (RMSE) of the training data,

RMSE of the validation data, and the number of terms as a measure of the model complexity.

$$(\lambda, \xi_{th}) = \min . \mathcal{J}(\Xi) := \rho_1 E_t(x, \hat{x}) + \rho_2 E_v(x, \hat{x}) + \rho_3 K \tag{4.8}$$

where ρ_1 and ρ_2 are the penalties on the training and validation data's modeling errors, respectively, and $E_t(x, \hat{x})$ and $E_v(x, \hat{x})$ are the RMSE of training and validation data, respectively. Therefore, errors from both the training and validation datasets are incorporated, capturing in-sample and extra-sample errors. Including extra-sample error helps the model avoid overfitting to the training data. Additionally, we incorporate the number of nonzero terms in the sparse coefficient vector, Ξ, into the cost function as a measure of model complexity. The number of nonzero terms in the sparse vector of coefficients, Ξ, is represented by $K \ll \rho$, with ρ_3 serving as a penalty on model complexity.

Also, RMSE is given by:

$$E(x, \hat{x}) = \sqrt{\frac{1}{m+1} \sum_{k=1}^{m+1} (x[k] - \hat{x}[k])^2}, \tag{4.9}$$

where $\hat{x}$ is the predicted output from the model.

The model identifications were performed in the normalized coordinates to eliminate dissimilarity in the library elements scales affecting the regression. We normalized the library by dividing each element by its corresponding column's absolute maximum value.

4.2 Tractable Data-driven Model of Li-ion Batteries

4.2.1 Data Generation

We used the DFN model to generate the voltage from the input current. The Python Battery Mathematical Modelling (PyBaMM) [113] package was used for the

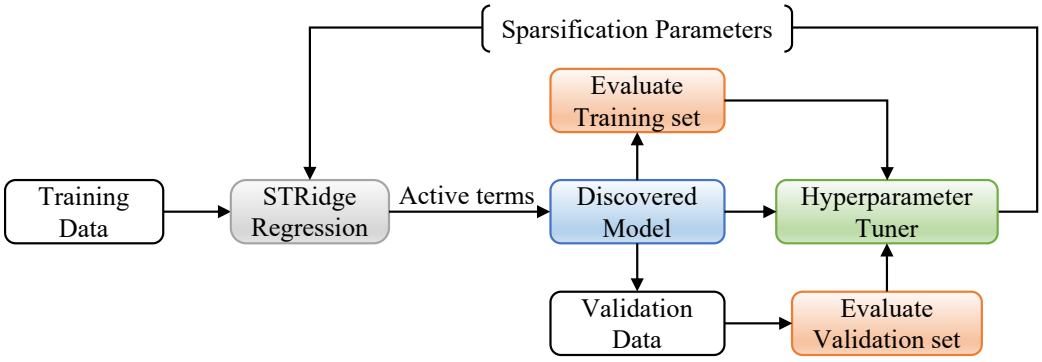


Figure 4.1. Schematic of the proposed data-driven modeling technique: The algorithm utilizes STRidge regression on the training data to identify the minimum number of active terms. Then, the hyperparameter tuner uses a grid search approach to minimize a cost comprising of the training and validation errors and the number of active terms (a measure of sparsity) to determine the sparsification parameters and, in turn, discovering the most appropriate model representing the process.

battery simulation. We modeled a 21700 cylindrical Li-ion cell with material NMC811 for the simulation [114].

We seek the sparse equation of the LiB by using only available data in the following form:

$$V[k+1] = \Theta(V[k], \mathrm{SOC}[k], I[k])\Xi_1, \tag{4.10}$$

$$\mathrm{SOC}[k+1] = \Theta(V[k], \mathrm{SOC}[k], I[k])\Xi_2. \tag{4.11}$$

where $V[k]$ and $I[k]$ represent the voltage and the electrical current, respectively, and SOC[k] is the negative electrode state of charge at time step k. The model's outputs are the voltage and SOC, and electrical current is the input, as shown in Fig. 4.2. The $\widehat{V}$ and $\widehat{\mathrm{SOC}}$ are the predicted outputs from the proposed model. Furthermore, the $\widetilde{V}$ and $\widetilde{\mathrm{SOC}}$ are the errors between the simulated data and identified data. We predict the voltage and SOC using their initial values and electrical current profile.

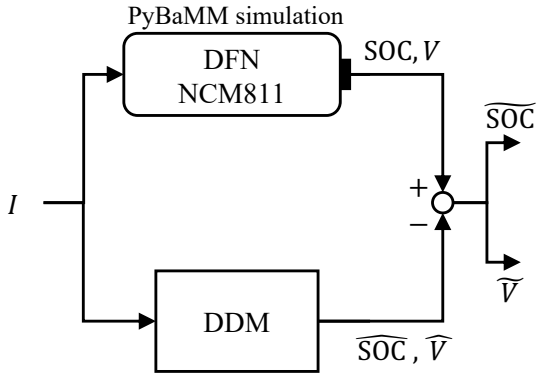


Figure 4.2. Block diagram of the nonlinear data-driven model.

4.3 Selection of Library Terms

The battery data can be used directly for predicting voltage and SOC through the library $\Theta(V, \mathrm{SOC}, I)$, which enables real-time estimation. While using generic nonlinear terms can lead to very small $\widetilde{V}$, the data will be prone to fitting into an incorrect nonlinear model, resulting in poor performance in unseen scenarios. Therefore, we use functions associated with the physics of the LiB in our library instead of adding many terms that are not based on the battery's physics.

4.3.1 Physics-informed Library

Here, we use the DFN physical model to build the library. The DFN model consists of two electrically separated porous electrodes (cathode and anode) and a separator. The Li-ions are transported by diffusion inside the active particles in the solid phase. In the liquid phase, the Li-ions traverse the electrolyte interface via Butler-Volmer kinetics to the other electrodes, mostly related to the charge transfer. The ions dissolved in the electrolyte pass through the separator to the opposite electrode. Internal battery dynamics are intercalation, electrochemical kinetics, and diffusion, and they are expressed by combinations of ordinary differential equations (ODEs) and partial differential equations (PDEs).

It is noted that we only described some equations for building library terms, and complete model equations with boundary conditions are given in [6, 27]. One of the processes inside the battery is Lithium transportation, which relates to Lithium concentrations. Lithium concentration in solid phase $c_s^{\pm}$ is radially symmetric diffusion in spherical coordinate as follows:

$$\frac{\partial c_s^{\pm}}{\partial t}(x,r,t) = \frac{1}{r^2}\frac{\partial}{\partial r}\left[D_s^{\pm} r^2 \frac{\partial c_s^{\pm}}{\partial r}(x,r,t)\right], \tag{4.12}$$

where r and x are the radial and longitudinal directions, respectively. The superscript "$\pm$" presents the positive and negative electrode, and $D_s^{\pm}$ is the diffusion coefficient. The Lithium concentration in electrolyte, c_e, is governed by Fick's law and molar flux $j_n^{\pm}$ as follows:

$$\frac{\partial c_e}{\partial t}(x,t) = \frac{\partial}{\partial x}\left[D_e \frac{\partial c_e}{\partial x}(x,t)\right] + \frac{(1-t_c^0)a^{\pm}}{\varepsilon_e F} j_n^{\pm}(x,t), \tag{4.13}$$

where F is the Faraday's constant, ε_e is the volume fraction of the electrolyte, t_c^0 is the transference number and $a^{\pm}$ is the specific interfacial surface area. Moreover, the diffusion coefficient D_e is a function of electrolyte concentration. Lithium transportations are diffusion processes, which are PDEs ((4.12) and (4.13)) and have the exponential function and zero-order Bessel function, which can be expressed as trigonometric terms with Fourier transformation; therefore, trigonometric and exponential terms are in their solution and is included in our physics-inspired library.

The molar flux is governed by the Butler-Volmer equation:

$$j_n^{\pm}(x,t) = \frac{1}{F} i_0^{\pm}(x,t)\left[e^{\frac{\alpha_a F}{RT}\eta^{\pm}(x,t)} - e^{-\frac{\alpha_c F}{RT}\eta^{\pm}(x,t)}\right], \tag{4.14}$$

where R is the universal gas constant, T is the temperature, and α_a and α_c are anodic and cathodic charge transfer coefficients, respectively. Moreover, $c_{ss}^{\pm}$ is the Lithium

concentration at the solid particle surface. The Butler-Volmer equation (4.14) has hyperbolic sine functions in the solution, which are exponential terms and are added into the physics-inspired library. The $i_0^{\pm}$ is the exchange current density, and the intercalation overpotential $\eta^{\pm}$ is an arithmetical combination of the solid potential, the electric electrolyte potential, molar flux, and open circuit potential that depends on the solid material.

The next process is the electrolyte potential ϕ_e, which is a function of ionic current $i_e^{\pm}$ and c_e:

$$\frac{\partial \phi_e}{\partial x}(x,t) = \frac{i_e^{\pm}(x,t)}{\kappa} - \frac{2RT}{F}(1-t_c^0) \times \left(1 + \frac{d\ln f_{c/a}}{d\ln c_e}(x,t)\right) \frac{\partial \ln c_e}{\partial x}(x,t), \tag{4.15}$$

where $f_{c/a}$ is mean molar activity coefficient in electrolyte and κ is the electolyte conductivity. The ionic current is a function of the molar flux, derived from the conservation of charge. The ionic current is a function of the molar flux, derived from the conservation of charge. Note that in the electrolyte, all the electrical current is $(i_e^{\pm})$. The equation for the ionic current is given by:

$$\frac{\partial i_e^{\pm}}{\partial x}(x,t) = a_s F j_n^{\pm}(x,t). \tag{4.16}$$

The specific interfacial area is a_s. Using Kirckoff's law the solid potential $\phi_s^{\pm}$ is given by:

$$\frac{\partial \phi_s^{\pm}}{\partial x}(x,t) = \frac{i_e^{\pm}(x,t) - I(t)}{\sigma^{\pm}}, \tag{4.17}$$

where, $\sigma^{\pm}$ is the effective electronic conductivity of the electrode. Voltage is the difference in the solid potential between the two ends of the electrode as follows:

$$V(t) = \phi_s^{+}(0^{+},t) - \phi_s^{-}(0^{-},t). \tag{4.18}$$

Finally, SOC is a function of the solid concentration (see (3.6)). However, according to the enhanced coulomb counting approach [15], we can estimate SOC with the integral of current in the following form:

$$SOC(t) = SOC(0) - \frac{1}{Q_n}\int_0^t \eta_c I d\tau, \tag{4.19}$$

where Q_n is the battery capacity and η_c is the coulomb efficacy.

In conclusion, to build the library of the candidate terms, we use the voltage, electrical current, SOC, and their exponential, trigonometric, and hyperbolic sinusoidal. In addition, we include the integral of the current to the library, and to consider its evolution, we append the double integral of the current to the library, too. Finally, we add a combination of polynomial terms up to order two to model other nonlinearities in the system. In this scenario, our library consists of 32 terms based on the internal processes of the LiBs, which are given by:

$$\Theta(\mathbf{V}, \mathbf{SOC}, \mathbf{I}) = \begin{bmatrix} | & | & | & & | & | & | & | & \\ V & V^2 & VI & \cdots & \sin(V) & \exp(\mathrm{SOC}) & \cos(I) & \int(I) & \cdots \\ | & | & | & & | & | & | & | & \end{bmatrix}. \tag{4.20}$$

Our physics-inspired model does not indicate a combination of the physical model (DFN) with a data-driven approach. The DFN model provides insight into the physical law of battery dynamics. Therefore, we include the DFN's solution forms inspired terms in the library Θ instead of using generic nonlinear terms to discover governing equations more connected to physics. This approach enabled a more generalizable model for improved performance with unseen data.

Table 4.1 presents the physics-inspired terms. To account for unmodeled dynamics of the DFN and other system nonlinearities, we combined simple terms (excluding trigonometric, exponential, and hyperbolic functions) up to the second order to construct our physics-inspired library.

Table 4.1. Terms for creating the physics-inspired library

Terms:	V	SOC	I	$\int I$	$\iint I$
$\sin(V)$	$\sin(I)$	$\sin(\text{SOC})$	$\cos(V)$	$\cos(I)$	$\cos(\text{SOC})$
$\exp(V)$	$\exp(I)$	$\exp(\text{SOC})$	$\sinh(V)$	$\sinh(I)$	$\sinh(\text{SOC})$

4.4 Dataset

The sparse model depends on the captured data; therefore, the input must be chosen carefully to have perfectly distributed data. First, we select the working space by choosing the upper and lower bounds of the electrical current. We pick the maximum value of the electrical current discharge 20 A or 4 C-rates, and the minimum value of the electrical current charge -10 A. The average value of the electrical current is around 5 A, and the duration is around 40 minutes with 50 ms sampling time [115] that results in approximately 50% SOC variation. The uniformly distributed random electrical current signal was employed to generate data.

Fig. 4.3a shows the stochastic current as the input for generating the training data set, and Fig. 4.3b illustrates the histogram of the input current and voltage corresponding to it. We sweep all the possible voltage values for any SOC from 90% to 40%. Fig. 4.3b represents the working space, and for any current, voltage, and SOC in this space, the identified model should be valid. This workspace covers high current C-rates, which is suitable for aggressive current profiles with abrupt changes.

For the validation data set, we use the current profile from the US-highway (US06) drive cycle input current profile [116] (see Fig. 4.4). The duration of the US06 drive cycle is 600 seconds with 50 ms sampling time, and around 10 US06 drive cycles were used to have 50% SOC variation. The maximum value of the current equals 16 A, which is more than 3 C-rates of the cell.

Similarly, the test data generated with EPA Urban Dynamometer Driving Schedule (UDDS) drive cycle current profile [116]. Fig. 4.5 shows the UDDS current signal

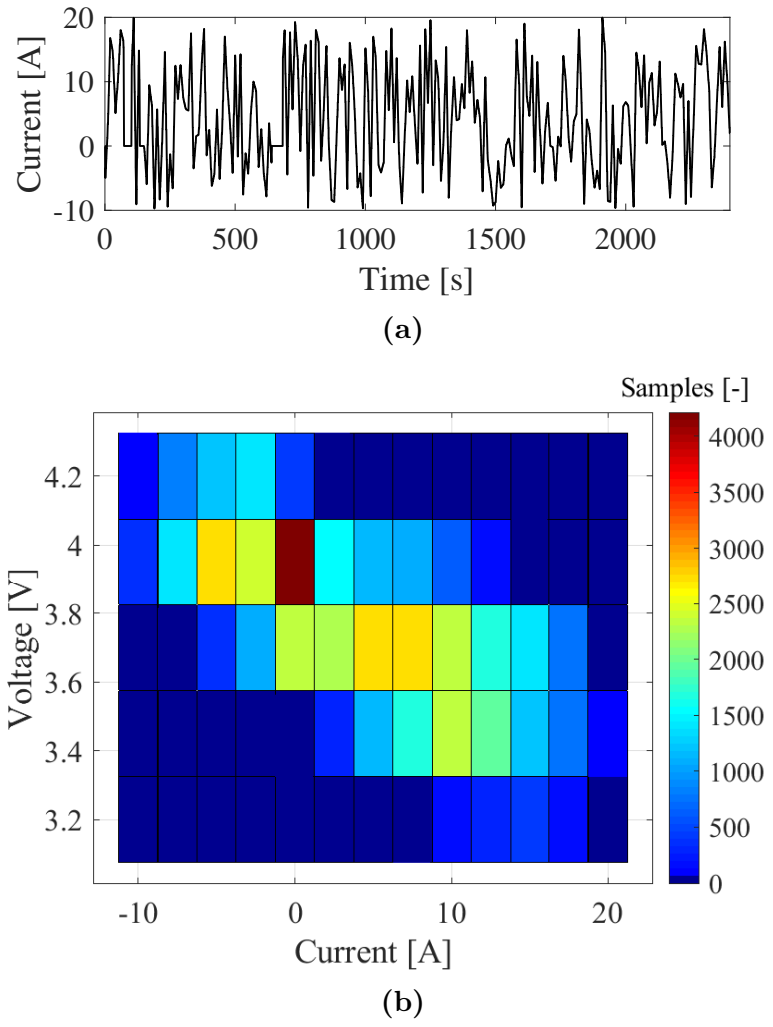


Figure 4.3. Distribution of the input current for the training data set. (a) Input current, and (b) Histogram of the current and voltage.

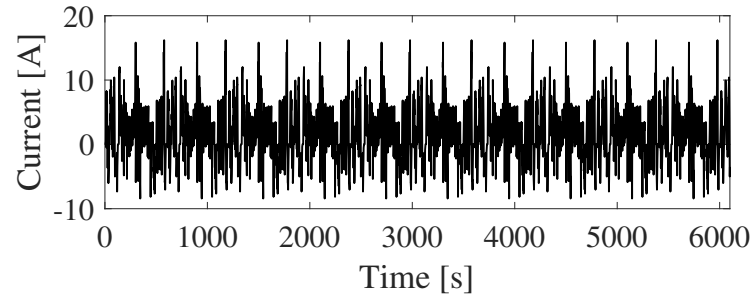


Figure 4.4. 10 US06 current profile cycles for validation data.

used to generate test data. The upper value of the current is around 16 A, and

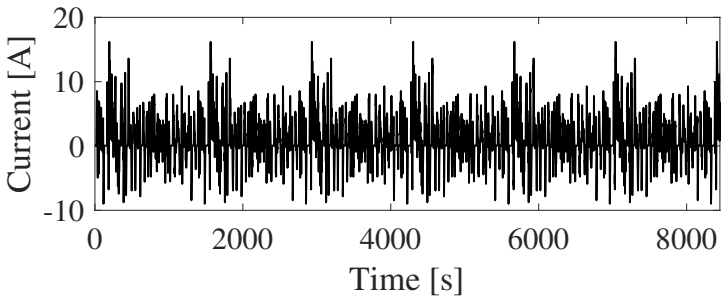


Figure 4.5. 6 UDDS current profile cycles for test data.

about 6 UDDS drive cycles were employed to have SOC variation from 90% to 40% discharge.

4.5 Simulation Results

In this section, we used the proposed nonlinear data-driven model of §4.2 to predict the voltage and SOC of the LiBs by using the electrical current as the input. The DFN was used to generate the output signals from the input currents of §4.4. After generating the training data, we use the AIC-inspired loss function (Discrete model of (3.13)) to find the regularization parameters λ and ξ_{th} in (4.7) for comparison with our proposed cost function.

We selected $\rho_1 = \rho_2 = 100$ and $\rho_3 = 0.1$ in the cost function. We assign equal weights to the training and validation RMSE to ensure a fair assessment of the model's performance. We used cross-validation, using only the initial value to predict the time series data to assess the RMSE of training data. The weight of the sparsity component is also modified to accommodate the scale difference between the RMSE and sparsity components. This enables us to discover governing equations that balance sparsity and accuracy for both voltage and SOC.

The exhaustive grid search approach was implemented to find these parameters. The observables were normalized to ensure the library terms have similar scales in the regression optimization. We normalized each column of the library term by dividing each element by the absolute maximum value of the corresponding column (observ-

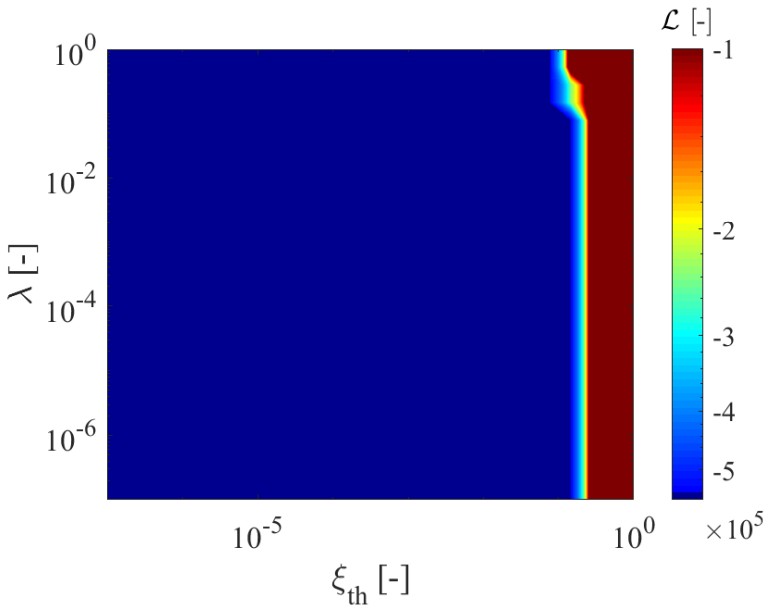


Figure 4.6. AIC-Loss function for SOC of the training data.

able). It is noted that the identified model was regressed onto the original coordinates to refrain from having non-physical terms in the prediction.

The sparse equation is decoupled to find the sparsification parameters for voltage and SOC separately. Fig. 4.6 represents the value of the AIC-inspired loss function of the SOC. For very small values of the regularization parameters, the model overfits the data and cannot omit less important terms. Furthermore, when ξ_{th} is very small, the model is not sparse and does not depend on λ anymore. It demonstrates that ξ_{th} plays a more important role than λ in the accuracy and sparsity of the model.

Here, the optimal value of the ξ_{th} and λ are calculated by solving the minimization problem (4.8). We use the validation data as shown in Fig. 4.1 to find the best model by selecting the optimal value for the sparsification parameters. We solved the problem once for the SOC only by assuming we know the values of the voltage and second for the voltage only by assuming we know the values of the SOC.

Fig. 4.7 illustrates the NRMSE of the US06 validation data for SOC. In contrast to Fig. 4.6 that the identified model has small training NRMSE for the very small value of ξ_{th}, for validation, the model has high validation NRMSE due to overfitting

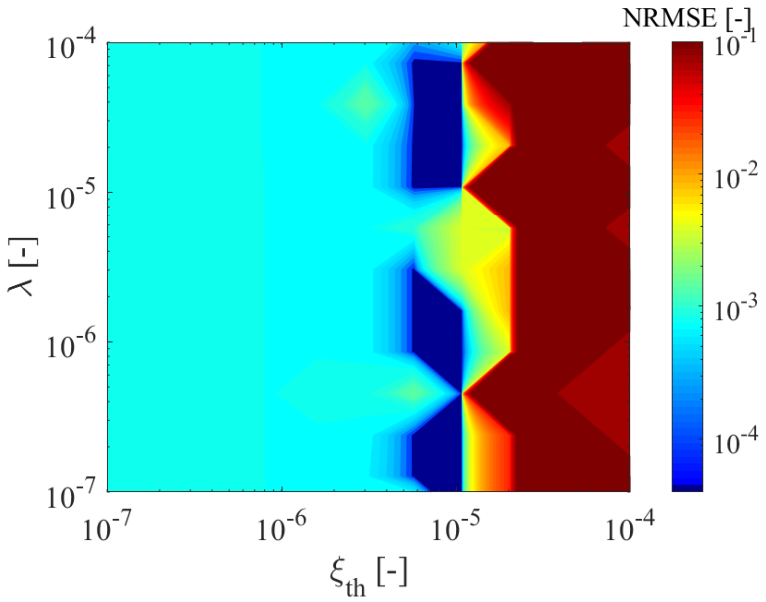


Figure 4.7. NRMSE for SOC of the validation data.

Table 4.2. Optimal parameters information.

Parameters	Value for SOC	Value for Voltage
λ	1.08×10^{-5}	7.50×10^{-4}
ξ_{th}	5.74×10^{-6}	2.15×10^{-2}
Number of active terms	19	33

the training data. This means that we have diminishing returns for training data, and unnecessary terms appear in the model that does not change the accuracy and overfits the training data. It is noted that in some parts of the red area in Fig. 4.7, the NRMSE is higher than the upper bound in the color bar (10^{-1}) to distinguish the smaller NRMSE better.

The values of the sparsification parameters are presented in Table 4.2. We calculate the optimal values for SOC and voltage separately by solving the minimization problem (4.8). The optimal values from Table 4.2 are used to find the sparse model from the training data set, and the performance of the sparse model is evaluated by validation and test data.

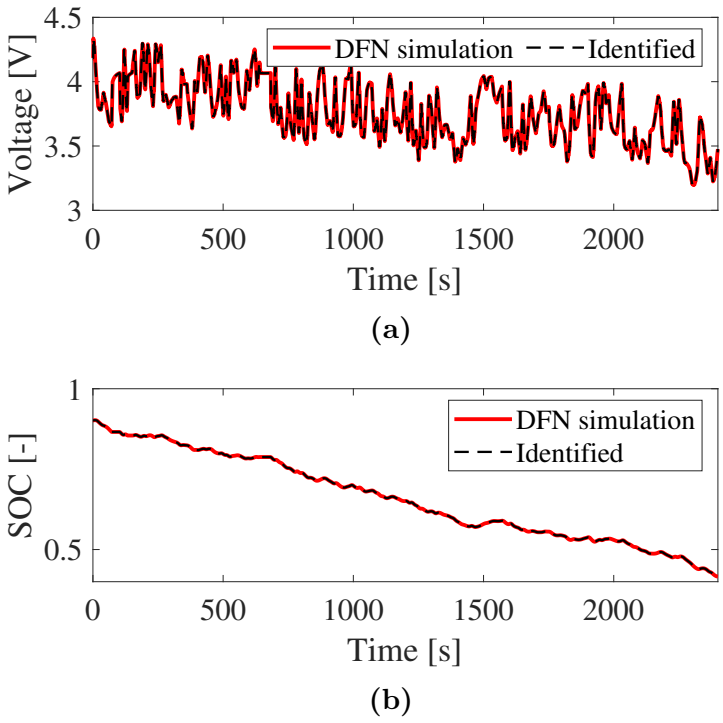


Figure 4.8. Identifying the system with the training data set. (a) Training voltage, and (b) Training SOC.

Fig. 4.8 depicts the identified model from the training data set compared to the simulation outputs. The voltage and SOC are generated from the training's current profile in Fig. 4.3a. The NRMSE of the voltage and SOC are 3.2×10^{-4} and 10^{-8}, respectively. The number of data points is 48000, which equals 2400 seconds with 50 ms sampling rate.

We used the US06 drive cycle current profile (Fig. 4.4) to evaluate the applicability of the identified model. The same coefficients from training data are used to generate the identified model. Fig. 4.9 shows the performance of the identified model for the US06 validation data, where the NRMSE of voltage and SOC are 6.1×10^{-3} and 2.2×10^{-5}. The identified model predicts the voltage and SOC simultaneously by only using the current profile.

After training the model using the stochastic input and US06 drive cycles, we check the model's performance with unseen test data. Fig. 4.10 displays the robustness of

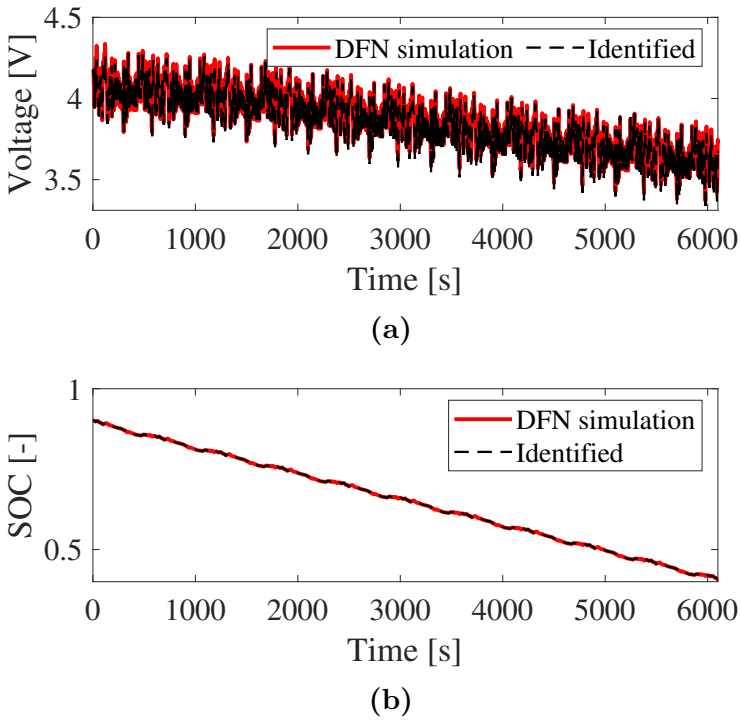


Figure 4.9. Validating the identified model with US06 drive cycle. (a) US06 validation voltage, and (b) US06 validation SOC.

the identified model to the UDDS test data. Fig. 4.10a shows that the model can track the voltage of the very aggressive UDDS current profile that yields NRMSE of 6.3×10^{-3}. Furthermore, Fig. 4.10b displays the identified model predicted the SOC for 8450 seconds with NRMSE of 2.8×10^{-3}. Furthermore, we can use the identified voltage and known input current to calculate the generated battery's power, which is an important parameter for electric and hybrid vehicles.

4.6 Summary

In this chapter, we presented a physics-inspired, data-driven model of LiBs using measurable data. We estimate the voltage and SOC by using only current as an input of the system. For generating simulation data, we used the DFN model from the PyBaMM package with the parameter set of the 21700 cylindrical Li-ion cell with NMC811 material. To avoid overfitting the model and balance complexity and generalizability, we optimized the model by proposing a cost function comprising

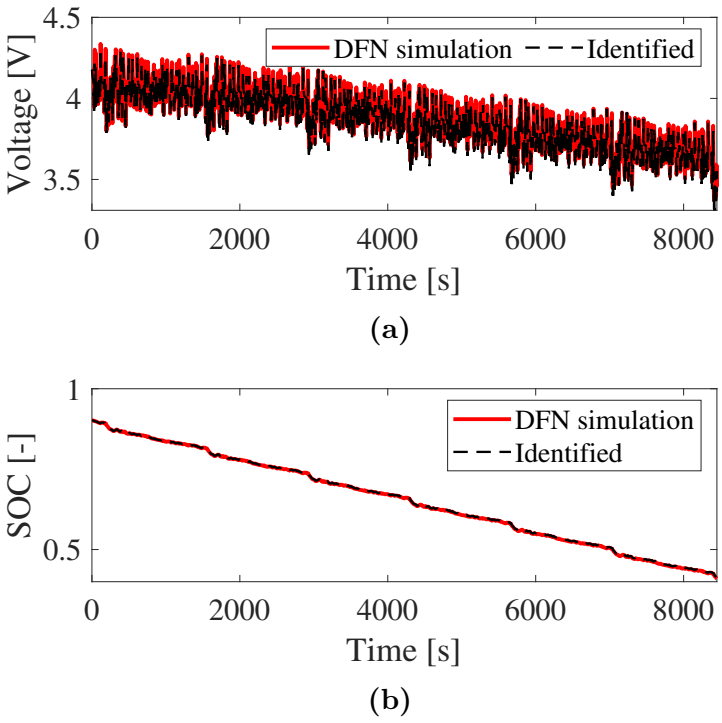


Figure 4.10. Testing the performance of the identified model with UDDS drive cycle. (a) UDDS test voltage, and (b) UDDS test SOC.

training, validation data sets, and the number of parameters as a measure of the model complexity.

The model's working range was defined to be between $2C$-rates charged to $4C$-rates discharged from 90% SOC to 40% SOC. The US06 drive cycle was used for the validation data with the current signal higher than $3C$-rates. To evaluate the model's performance with the unseen test, we employed the UDDS current profile with 50% SOC variation with the aggressive electrical current profile higher than $3C$-rates. The model achieved a small error (NRMSE = 10^{-8}) for SOC prediction of the training data set. Similarly, the model attained NRMSE of 3.2×10^{-4} for voltage prediction of the training data set. Additionally, the model yielded the NRMSE of 6.3×10^{-3} for both the validation and unseen test data set.

Despite its enhanced predictability and generalizability due to the physics-inspired library and improved cost function for sparsification tuning, the model is currently

accurate only with ideal, noise-free simulation data. Its performance degrades in real-world scenarios, particularly due to its reliance on an accurate initial value and its inability to update parameters based on observations.

In the next chapter, we address these challenges by augmenting the model with a joint unscented Kalman filter (JUKF) to mitigate the effects of noise from real-world experimental data. Furthermore, we develop a framework that allows SOC estimation without requiring precise knowledge of the initial value. This adaptive framework will use noisy voltage measurements to update the model parameters dynamically, ensuring robustness and convergence to the correct values even under uncertain conditions.

CHAPTER 5

AN ONLINE CO-ESTIMATION FRAMEWORK

In this chapter, we apply our proposed physics-inspired data-driven algorithm in chapter 4 to create a tractable data-driven model using measurement data. Due to the noisy nature of the measurement data, the model's performance can be compromised, similar to other machine learning techniques. To address this, we augment the model with a Joint Unscented Kalman filter (JUKF) to mitigate the effect of the measurement noise and enhance estimation. We develop two approaches: i) a model with JUKF to mitigate only the noise effect, and ii) a co-estimation model that uses voltage dynamics as a SOC-voltage map alongside SOC dynamics to enhance estimation and simultaneously mitigate noise.

5.1 Kalman Filter Augmentation

Here, we briefly introduce the discrete Kalman filter that forms the foundation of the noise mitigation in 5.1.1. The algorithm uses a joint unscented Kalman filter as discussed in 5.1.2. Then, in section 5.1.3, the joint unscented Kalman filter is added to the discovered dynamics to correct the measured noisy voltage data. We integrate the joint unscented Kalman filter with the tractable data-driven battery model to utilize voltage dynamics (4.10) as a SOC-voltage map, connecting it to the SOC dynamics (4.11) to provide a closed-loop SOC estimation framework.

5.1.1 Discrete Kalman filter

Let us consider a discrete linear model of the following form:

$$x_k = Ax_{k-1} + w_k, \tag{5.1}$$

$$y_k = Hx_k + v_k, \tag{5.2}$$

where x_k and y_k are the state and output measurements, respectively; A is a input state matrix, and H is the output state matrix. In (5.1), w_k and v_k are the process

and measurement noises, which are assumed to be Gaussian white noise, with the covariance matrix Q and R, respectively. The Kalman filter algorithm [117] consists of 2 steps: prediction and correction. The prediction step is given by:

$$\hat{x}_k^- = A\hat{x}_{k-1}, \tag{5.3}$$

$$P_k^- = AP_{k-1}A^T + Q, \tag{5.4}$$

where $\hat{x}_{k-1}$ and $\hat{x}_k^-$ are the posterior and prior estimates of the states at time instant $k-1$ and k, respectively. Similarly, P_{k-1} and P_k^- are the posterior and prior estimation error covariance matrix.

In the correction step, we rectify equations by introducing the Kalman gain as follows:

$$K_k = P_k^- H^T \left(HP_k^- H^T + R\right)^{-1}. \tag{5.5}$$

We update the posterior state estimate by using the error between the actual output and predicted output ($\hat{y}_k = H\hat{x}_k^-$) in the following way:

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - \hat{y}_k). \tag{5.6}$$

Finally, the posterior error covariance is given by:

$$P_k = (I - K_k H)P_k^-. \tag{5.7}$$

Fig. 5.1 displays the prediction and correction steps of the Kalman filter.

5.1.2 Joint unscented Kalman filter

The Kalman filter framework consists of prediction and correction steps. The prediction step estimates the state and uncertainty of the current state based on past observations. The correction step updates the state and uncertainty based on new ob-

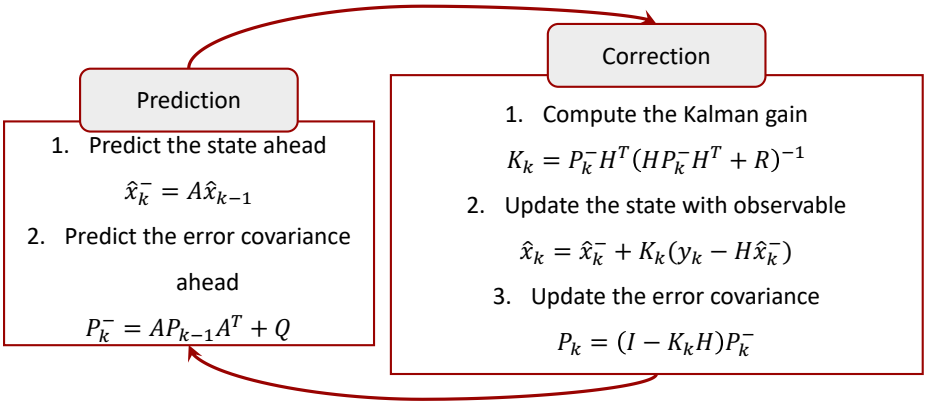


Figure 5.1. Kalman filter framework.

servations. The unscented Kalman filter (UKF) utilizes the Kalman filter framework for nonlinear systems such as Li-ion battery dynamics instead of traditional methods like the extended Kalman filter algorithm that linearizes the nonlinear system [118].

The discrete systems have two possible uncertainties: i) uncertainty in the system's parameter and ii) uncertainty in the state dynamics. Integrating the Joint UKF (JUKF) helps address uncertainty in both the nonlinear states and parameters concurrently, enhancing the algorithm's robustness. Let us consider a discrete nonlinear dynamical system with stationary dynamics for the parameter $\mathbf{n}$:

$$\mathbf{n}[k+1] = \mathbf{n}[k] + w_1[k], \tag{5.8}$$

$$\mathbf{x}[k+1] = f(\mathbf{x}, w_2[k]; \mathbf{n}[k]) \tag{5.9}$$

$$\mathbf{y}[k] = h(\mathbf{x}[k], v[k]; \mathbf{n}[k]), \tag{5.10}$$

where $w_1[k]$ and $w_2[k]$ are the process noises, which are assumed to be Gaussian white noise, $\mathbf{x}[k]$ and $\mathbf{y}[k]$ represent the state and output in the k time step, respectively, and $v[k]$ is the zero mean measurement noise with covariance R. We note that the system dynamics f and h are known.

To estimate both states and parameters simultaneously, we combine these variables into a unified joint state vector $\mathbf{z} = \begin{bmatrix} \mathbf{x}^T & \mathbf{n}^T \end{bmatrix}^T$ with dimension of L. The

state-space equations can be represented more concisely and conveniently using this joint state vector as follows:

$$\mathbf{z}[k+1] = F(\mathbf{z}[k]) + \mathbf{w}[k], \tag{5.11}$$

$$\mathbf{y}[k] = H(\mathbf{z}[k]) + v[k], \tag{5.12}$$

where $\mathbf{w} = \begin{bmatrix} {w_2}^T & w_1^T \end{bmatrix}^T$ is a vector of zero mean and covariance Q. In what follows, we use the following notation for an arbitrary parameter (z), $\hat{\mathbf{z}}[k]$ denotes the estimate of $\mathbf{z}[k]$; $\hat{\mathbf{z}}^-[k]$ denotes prior estimate of $\mathbf{z}[k]$. Starting with initial values for the augmented state, $\hat{\mathbf{z}}$, and covariance $\mathrm{P}_{\mathbf{z}}$ and, employing the unscented transform to generate $2L+1$ sigma points $\mathcal{X}_i$ to capture statistics of $\mathbf{y}$ in the following manner:

$$\begin{aligned} &\mathcal{X}[k-1] = \\ &\begin{bmatrix} \hat{\mathbf{z}}[k-1] & \hat{\mathbf{z}}[k-1] + \sqrt{(L+\gamma)\,\mathrm{P}_{\mathbf{z}}} & \hat{\mathbf{z}}[k-1] - \sqrt{(L+\gamma)\,\mathrm{P}_{\mathbf{z}}} \end{bmatrix}, \end{aligned} \tag{5.13}$$

where $\gamma = \alpha^2(L+\kappa) - L$ is a scaling parameter, α controls the spread of sigma points around $\hat{\mathbf{z}}$, and is commonly assigned to small value such as 10^{-3} [118]. The constant κ is the second scaling parameter and is typically set to $3 - L$ [119].

We generate the sigma points using the nonlinear state function in the following manner:

$$\mathcal{X}_i^-[k] = F(\mathcal{X}_i[k-1]) \tag{5.14}$$

Now, we can determine the prior estimate of the state $\hat{\mathbf{z}}^-[k]$ and the corresponding error covariance matrix $\mathrm{P}_{\mathbf{z}}^-[k]$ as follows:

$$\hat{\mathbf{z}}^-[k] = \sum_{i=0}^{2L} W_{m,i}\mathcal{X}_i^-[k], \tag{5.15}$$

$$\mathrm{P}_{\mathbf{z}}^-[k] = \sum_{i=0}^{2L} W_{c,i}\left(\mathcal{X}_i^-[k] - \hat{\mathbf{z}}^-[k]\right)\left(\mathcal{X}_i^-[k] - \hat{\mathbf{z}}^-[k]\right)^T + Q, \tag{5.16}$$

where factors $W_{m,i}$ and $W_{c,i}$ serve as the weighting coefficients and are expressed as:

$$W_{m,0} = \frac{\gamma}{L+\gamma}, \tag{5.17}$$

$$W_{c,0} = \frac{\gamma}{L+\gamma} + 1 - \alpha^2 + \beta, \tag{5.18}$$

$$W_{m,i} = W_{c,i} = \frac{1}{2(L+\gamma)}, \quad i = 1, \ldots, 2L+1 \tag{5.19}$$

where β integrates prior information about the state distribution and optimally is 2 for Gaussian distribution [118].

The propagated sigma points (5.14) are utilized for generating the measurement sigma points:

$$\mathcal{Y}_i^-[k] = H(\mathcal{X}_i^-[k]). \tag{5.20}$$

Through the utilization of the weighting factor, a priory measurement estimate is determined as follows:

$$\hat{\mathbf{y}}^-[k] = \sum_{i=0}^{2L} W_{m,i} \mathcal{Y}_i^-[k]. \tag{5.21}$$

Now, we can compute the measurement covariance $\mathrm{P}_{\mathbf{y}}$ and cross-covariance $\mathrm{P}_{\mathbf{zy}}$, respectively, in the following manner:

$$\mathrm{P}_{\mathbf{y}}[k] = \sum_{i=0}^{2L} W_{c,i} \left(\mathcal{Y}_i^-[k] - \hat{\mathbf{y}}^-[k]\right) \left(\mathcal{Y}_i^-[k] - \hat{\mathbf{y}}^-[k]\right)^T + R, \tag{5.22}$$

$$\mathrm{P}_{\mathbf{zy}}[k] = \sum_{i=0}^{2L} W_{c,i} \left(\mathcal{X}_i^-[k] - \hat{\mathbf{z}}^-[k]\right) \left(\mathcal{Y}_i^-[k] - \hat{\mathbf{y}}^-[k]\right)^T, \tag{5.23}$$

The Kalman gain matrix for correction is carried out as follows:

$$\mathcal{K}[k] = \mathrm{P}_{\mathbf{zy}}[k] \mathrm{P}_{\mathbf{y}}^{-1}[k]. \tag{5.24}$$

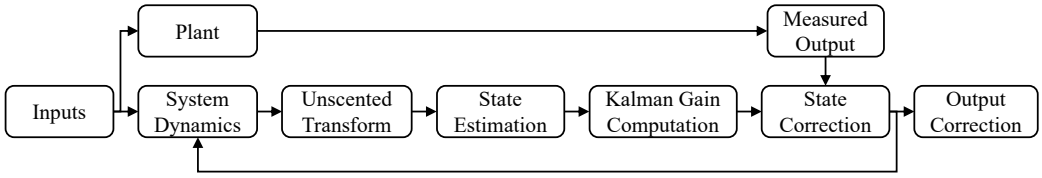


Figure 5.2. Unscented Kalman filter block diagram.

Following the computation of the Kalman gain filter, the posterior state estimate is determined as follows:

$$\hat{\mathbf{z}}[k] = \hat{\mathbf{z}}^{-}[k] + \mathcal{K}[k](\mathbf{y}[k] - \hat{\mathbf{y}}^{-}[k]) \tag{5.25}$$

Ultimately, the correction of the estimation error covariance is performed as follows:

$$\mathrm{P}_{\mathbf{z}}[k] = \mathrm{P}_{\mathbf{z}}^{-}[k] - \mathcal{K}[k]\mathrm{P}_{\mathbf{y}}[k]\mathcal{K}^{T}[k] \tag{5.26}$$

Fig. 5.2 displays the UKF framework for updating the state and correcting measurement using the unscented transformation for estimating the state. Algorithm 1 presents the Unscented Kalman filter procedure.

Algorithm 1 Unscented Kalman Filter (UKF)

Input: Initial state estimate $\hat{\mathbf{z}}[0]$, initial covariance $\mathbf{P_z}[0]$, process noise covariance Q, measurement noise covariance R, nonlinear functions F (state transition) and H (measurement), and measurements $\mathbf{y}[k]$
Output: Estimated state $\hat{\mathbf{z}}[k]$, covariance $\mathbf{P_z}[k]$
Initialize parameters: α, κ, β, and compute $\gamma = \alpha^2(L+\kappa) - L$
for each time step $k = 1, 2, \ldots$ **do**

1. Generate sigma points:

$$\mathcal{X}[k-1] = \left[\hat{\mathbf{z}}[k-1], \hat{\mathbf{z}}[k-1] + \sqrt{(L+\gamma)\mathbf{P_z}[k-1]}, \hat{\mathbf{z}}[k-1] - \sqrt{(L+\gamma)\mathbf{P_z}[k-1]}\right]$$

2. Propagate sigma points through the system model:

$$\mathcal{X}_i^-[k] = F(\mathcal{X}_i[k-1]) \; \forall i$$

3. Predict the state and covariance:

$$\hat{\mathbf{z}}^-[k] = \sum_{i=0}^{2L} W_{m,i}\mathcal{X}_i^-[k]$$

$$\mathbf{P_z}^-[k] = \sum_{i=0}^{2L} W_{c,i}\left(\mathcal{X}_i^-[k] - \hat{\mathbf{z}}^-[k]\right)\left(\mathcal{X}_i^-[k] - \hat{\mathbf{z}}^-[k]\right)^T + Q$$

4. Propagate sigma points through the measurement model:

$$\mathcal{Y}_i^-[k] = H(\mathcal{X}_i^-[k])$$

5. Predict the measurement:

$$\hat{\mathbf{y}}^-[k] = \sum_{i=0}^{2L} W_{m,i}\mathcal{Y}_i^-[k]$$

6. Compute measurement covariance and cross-covariance:

$$\mathbf{P_y}[k] = \sum_{i=0}^{2L} W_{c,i}\left(\mathcal{Y}_i^-[k] - \hat{\mathbf{y}}^-[k]\right)\left(\mathcal{Y}_i^-[k] - \hat{\mathbf{y}}^-[k]\right)^T + R$$

$$\mathbf{P_{zy}}[k] = \sum_{i=0}^{2L} W_{c,i}\left(\mathcal{X}_i^-[k] - \hat{\mathbf{z}}^-[k]\right)\left(\mathcal{Y}_i^-[k] - \hat{\mathbf{y}}^-[k]\right)^T$$

7. Compute Kalman gain:

$$\mathcal{K}[k] = \mathbf{P_{zy}}[k]\mathbf{P_y}^{-1}[k]$$

8. Update the state estimate:

$$\hat{\mathbf{z}}[k] = \hat{\mathbf{z}}^-[k] + \mathcal{K}[k]\left(\mathbf{y}[k] - \hat{\mathbf{y}}^-[k]\right)$$

9. Update the error covariance:

$$\mathbf{P_z}[k] = \mathbf{P_z}^-[k] - \mathcal{K}[k]\mathbf{P_y}[k]\mathcal{K}^T[k]$$

end for

5.1.3 SOC and voltage co-estimation framework

As described previously, we aim to find the identified nonlinear tractable data-driven model (4.10) and (4.11) with the JUKF of section 5.1.2 to ensure convergence of SOC values in the presence of noise. Using (4.11) with the wrong initial state leads to inaccurate predictions. Another issue with the model is the noisy voltage data, which deteriorates the accuracy of the model. In what follows, we describe our proposed approach to address these issues.

We treat this noise as uncertainties w_V and w_Ξ in the discovered dynamics for voltage (4.10) and in the sparse vector of coefficients corresponding to the voltage

dynamics, respectively. In contrast, the SOC is not readily measurable. Therefore, the estimated SOC can be inaccurate and needs to be constantly updated, and this inaccuracy will introduce uncertainty w_S to SOC dynamics (4.11). To address these, we incorporate the JUKF into our model to mitigate the effect of noise from measured voltage data and converge the estimated SOC with inaccurate initial conditions to its actual value. Fig. 5.3 displays the proposed model with JUKF where $\widehat{V}$ and $\widehat{\text{SOC}}$ are the predicted output. The nonlinear tractable data-driven model with uncertainty is formulated as follows:

$$\Xi_{r,1}[k+1] = \Xi_{r,1}[k] + w_{\Xi}[k], \tag{5.27}$$

$$V[k+1] = \theta_{r,1}[k]\Xi_{r,1}[k] + w_V[k], \tag{5.28}$$

$$\text{SOC}[k+1] = \theta_{r,2}[k]\Xi_{r,2}[k] + w_S[k], \tag{5.29}$$

$$V_o[k] = V[k] + v_V[k], \tag{5.30}$$

where $\Xi_{r,1}$ and $\Xi_{r,2}$ are the nonzero terms of the sparse vector of coefficients for voltage Ξ_1 and SOC Ξ_2, respectively. Additionally, $\theta_{r,1}[k]$ and $\theta_{r,2}[k]$ refer to the terms in the k^{th} row of $\Theta(\mathbf{V}, \text{SOC}, \mathbf{I})$ corresponding to the $\Xi_{r,1}$ and $\Xi_{r,2}$, respectively; $v[k]$ is the measurement noise, and V_o is the noisy measured voltage.

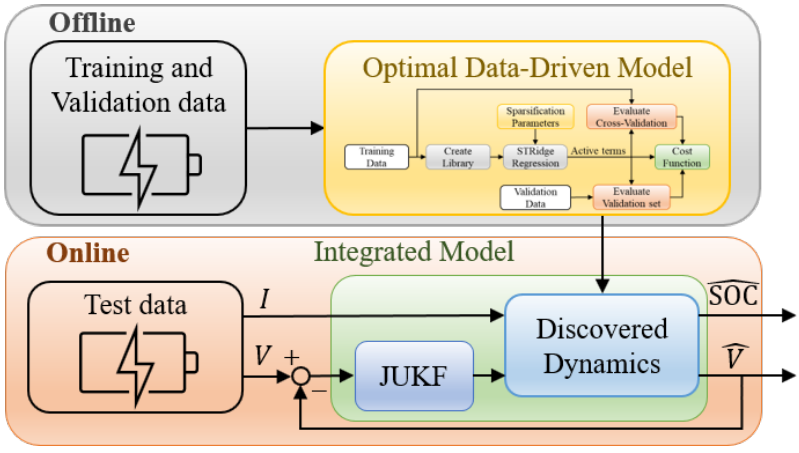


Figure 5.3. Block diagram of the augmented battery model with JUKF.

Noise Mitigation

For noise mitigation, we assume the process noise w_S in SOC dynamics is zero, and the initial value of the SOC is known. Therefore, the JUKF only corrects the voltage data, and refined voltage data is used for the SOC dynamics. We concatenate the states V and $\Xi_{r,1}$ into a single joint state vector, $\mathbf{z} = \begin{bmatrix} V & \Xi_{r,1} \end{bmatrix}^T$ to refine the voltage data.

Co-estimation Framework

Similar to section 5.1.2, V and SOC in (5.28) and (5.29) are the states $\mathbf{x}$ in (5.9), $\Xi_{r,1}$ is the parameters in (5.8), and V_o is the output in (5.10). Hence, we introduce the augmented state $\mathbf{z} = \begin{bmatrix} V & \text{SOC} & \Xi_{r,1} \end{bmatrix}^T$ for utilization in the JUKF, and measured voltage data V_o serves as the model output. The JUKF updates the state $\mathbf{z}$ to estimate the nonzero sparse coefficients for voltage, SOC dynamics, and voltage dynamics based on the measured voltage data. Consequently, the estimated SOC with an incorrect initial value converges to the estimated SOC with the correct initial value. Simultaneously, the voltage data is refined by mitigating the effect of measurement noise. Although, in practical applications, the electrical current is noisy too, our calculations account for this by considering the effect of electrical current noise as process noise. Therefore, the voltage dynamics is used as the SOC-voltage map, and refining voltage data and its parameter, $\Xi_{r,1}$ leads to correcting the SOC-voltage map operando, contributing to the convergence of the estimated SOC to its actual value by connecting it to the SOC dynamics.

5.1.4 Simulation Results on Noise Mitigation

We used the same simulation data from section 4.5 to evaluate the performance of the model in noise mitigation.

The Gaussian noise with a mean magnitude equal to 5% of the voltage's variance is added to the voltage data in all simulations. The process noise covariance matrix Q and measurement noise covariance R should be chosen appropriately. Although the measurement noise covariance can be estimated based on the sensor, assigning the process noise covariance is challenging. To exclude dissimilarity in the scales of the library elements, we normalize all the library terms by dividing each element by the absolute maximum of the corresponding column in the library. Furthermore, the identified model reverted to its original coordinates to avoid having non-physical errors in the prediction.

Fig. 5.4 shows the span of the noisy training data. The current is uniformly distributed to capture the possible values of the voltage and SOC in the operating range of the battery.

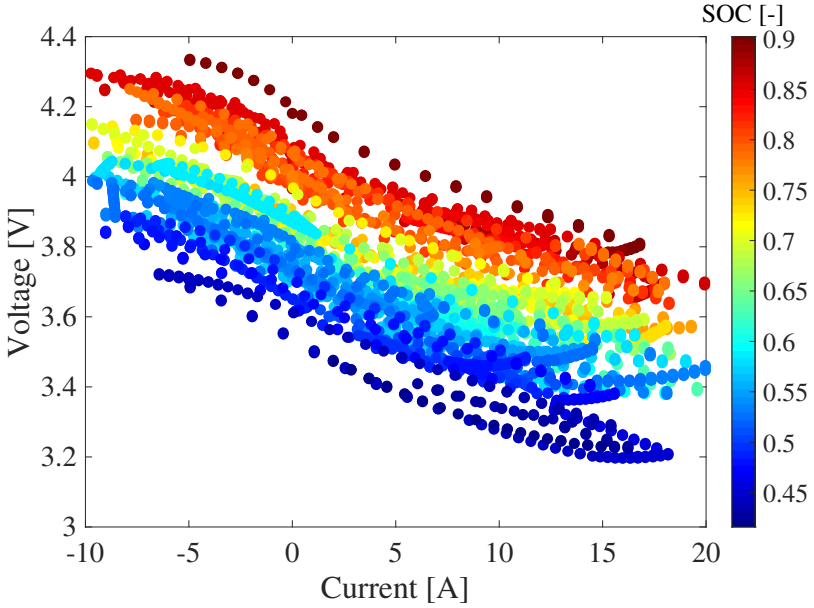
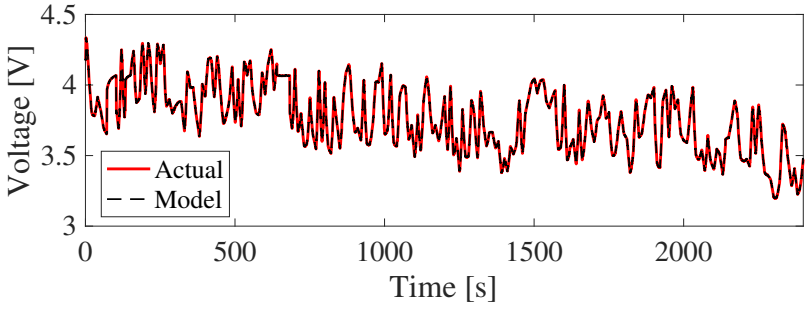


Figure 5.4. Noisy training data span.

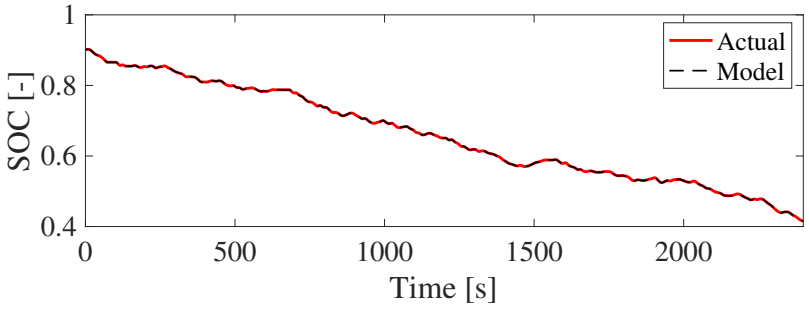
Due to the noise, we retrain the model using the noisy stochastic training data and noisy US06 validation data. Table 5.1 presents the optimal values of the sparsification parameters.

Table 5.1. Model parameters for identification.

Parameters	Voltage	SOC
ξ_{th}	0.0909	5.7362×10^{-6}
λ	0.0051	1.0826×10^{-5}
Number of active terms	15	29



(a)



(b)

Figure 5.5. Performance of the identified model on the training data set. (a) Training voltage with noise, and (b) Training SOC.

Fig. 5.5 shows the plot of the model for the training data set. The model achieved the voltage and SOC NRMSE values of 10^{-3} and 1.0077×10^{-7}, respectively. Note that the NRMSE of the generated model with noisy data is higher than that of the model without noisy data, as presented in Section 4.5. The error is approximately 10 times larger, highlighting the impact of noise on the model's performance.

Here, we evaluate the model's performance on the validation data. The model is identified with the training data and predicts the voltage and SOC concurrently by only using the current profile. The model is identified in noisy voltage data, and we

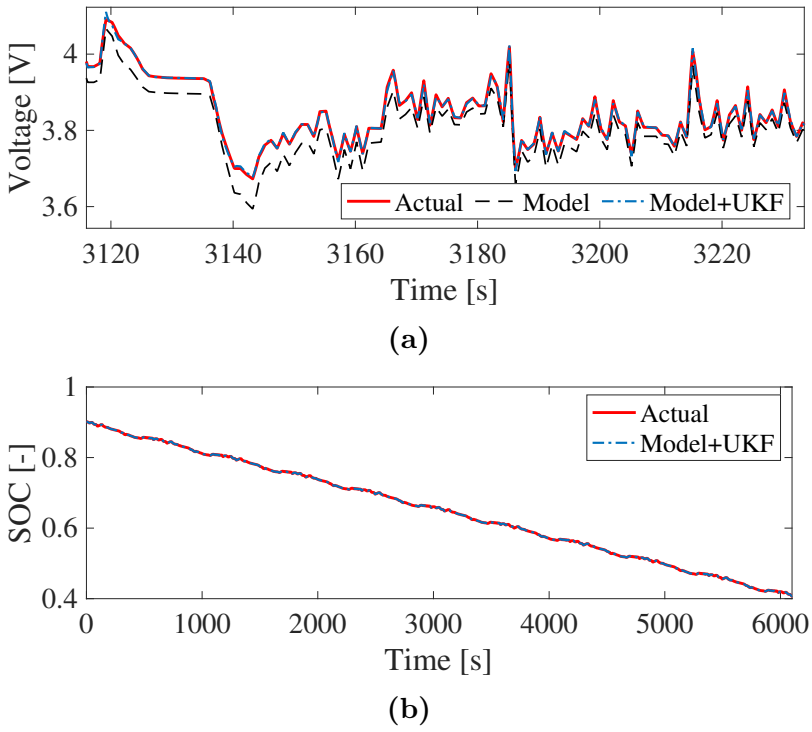


Figure 5.6. Model validation results on noisy US06 cycles. (a) Zoomed view of US06 voltage, and (b) US06 SOC.

remove the effect of the noise by updating the sparse vector of coefficients for voltage as described in section 5.1.3. Finally, the data-driven model utilizes the updated voltage for SOC estimation. Fig. 5.6 illustrates the accuracy of the identified model with and without the Kalman filter. The NRMSE of 1.1×10^{-3} and 5.0963×10^{-5} were achieved for voltage and SOC prediction with the unscented Kalman filter, respectively.

Finally, we utilize test data to check the generalizability of the identified model. Fig. 5.7 displays the identified model compared to the UDDS test data. The model with filter can track the noisy voltage and SOC where the NRMSE are 9.0568×10^{-4} and 1.1585×10^{-4}, respectively. The model is general and can track the output of the aggressive current profile higher than 3C-rates with noisy data. In addition, we evaluate the model's performance on the UDDS test data using only an unscented Kalman filter (without updating the coefficients) and without any Kalman filter.

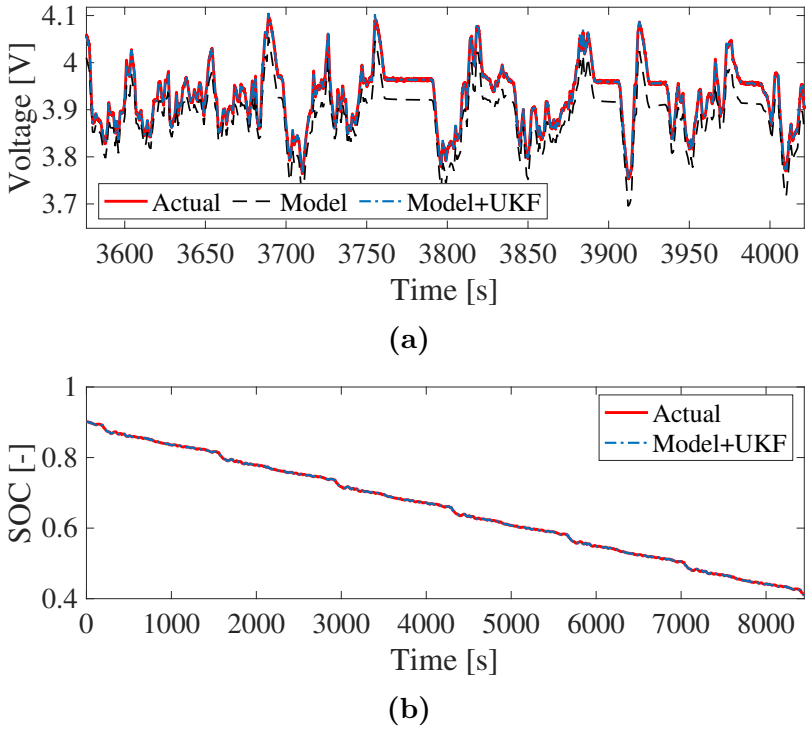


Figure 5.7. Model test results on the unseen noisy UDDS cycles. (a) Zoomed view of UDDS voltage, and (b) UDDS SOC.

The model achieves an NRMSE of 2.3×10^{-3} and 10^{-2}, respectively, for voltage data, both of which are higher than the NRMSE achieved using the JUKF. These results highlight the critical role of the Kalman filter and coefficient updates in mitigating noise effects.

5.2 Summary

In this chapter, we augment the physics-inspired data-driven algorithm introduced in Chapter 4 to develop a tractable data-driven model for battery dynamics using noisy data. Similar to other machine learning models, the model's accuracy can be compromised due to the noise. To address this issue, we integrate a JUKF into the model to mitigate the effect of measurement noise and improve state estimation. We propose two approaches: the first only addresses noise reduction from voltage data, while the second introduces a co-estimation framework that updates both voltage and

SOC dynamics, enhancing estimation performance in uncertainties, such as unknown initial conditions.

The discrete Kalman filter is introduced to provide the foundation, followed by the development of the JUKF, which is applied to nonlinear systems like Li-ion batteries. The JUKF is augmented into our model to estimate both the states and system parameters simultaneously, providing a robust solution for managing uncertainties in the system.

We also introduce a co-estimation framework that utilizes voltage dynamics as a SOC-voltage map, enabling simultaneous estimation of SOC and mitigation of noisy voltage data. This framework helps ensure accurate SOC estimation even with unknown initial conditions. The proposed model with JUKF is then demonstrated to enhance prediction accuracy in noise mitigation from noisy simulated data. The model yielded training and validation errors less than 1.1×10^{-3} and 5.1×10^{-5} for voltage and SOC prediction. We evaluated the model generalizability with the unseen 6.2 drive cycles UDDS test data, where NRMSE of 9.0568×10^{-4} was obtained for prediction.

CHAPTER 6

EXPERIMENTS

This section explains the experiment procedure, the battery material, and the data collection and analysis. The experimental data are used to build the tractable data-driven model.

6.1 Battery cell and test procedure

The quality and accuracy of the identified model depend on several parameters, including the dataset used for the identification process. The data should be rich enough to be distributed over the entire design space. We designed electrical current profiles to capture battery behavior in the expected operating range. The experiments were conducted on a 21700 cylindrical Li-ion cell with graphite anode and NMC811 (Nickel-Manganese-Cobalt) cathode with the main characteristics listed in Table 6.1. The cell has a nominal capacity of 4.8 Ah. The charge and discharge cutoff voltages were 4.2 V and 2.5 V, respectively. The maximum continuous charging current is 1.44 A (0.3 C) with the 50 mA charge cutoff current.

All experiments were conducted using the MACCOR 4200 as illustrated in Fig. 6.1. The Model 4200 allows for configuration with multiple current range channels, covering up to 200 A, with an accuracy of ± 30 ηA for each channel. These channels

Table 6.1. NMC811 cell characteristics.

Parameters	**Value**
Manufacturer	LG Chem
Cathode material	$LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$
Anode material	Graphite
Nominal capacity	4.8 Ah
Nominal voltage	3.6 V
Charging/discharging cut-off voltage	4.2 V/2.5 V
Charging/discharging current	1.44 A/4.8 A
Cut-off current	50 mA
Charge/discharge operating temperature	0 to 45°C/-20 to 55°C

Figure 6.1. Maccor model 4200 and thermal chamber.

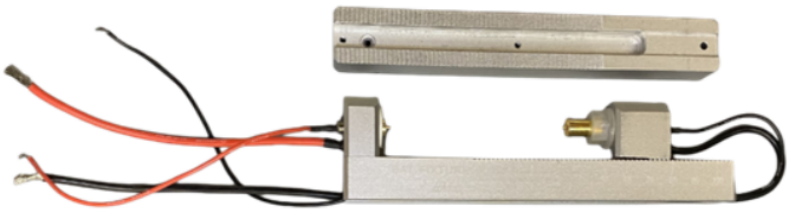


Figure 6.2. Cylindrical cell holder.

can also be set up for voltage ranges of -2 V to 8 V or 0 V to 5 V as standard. The accuracy of measured current and voltage on these channels is ±0.02% of the Full Scale Range (FSR). Electrical current serves as the input to the battery tester, while voltage, temperature, and battery capacity are the measurable outputs.

The cell is placed in the programmable Thermal Test Chamber MSK-TE906-80L, which has an 80-liter capacity (see Fig. 6.1). The chamber features a programmable working temperature ranging from -70°C to 150°C with the maximum rate of ±5°C/min and temperature accuracy of ±1°C. Fig 6.2 shows a double self-locking aluminum alloy CNC four-wire battery holder fixture, which is used to hold the battery and connect it to the MACCOR battery cycler.

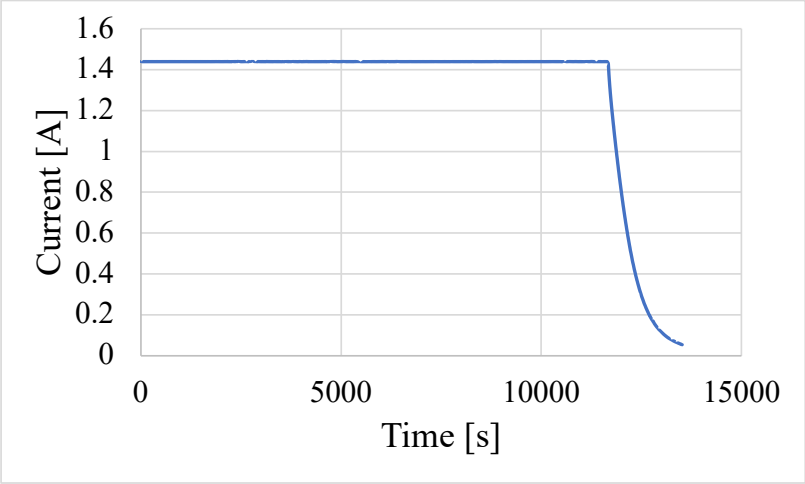


Figure 6.3. Electrical current profile for CCCV charge method.

In the experiment, we cycle the battery 3 times with the CCCV method. The CCCV charge method starts with a constant current, and when the battery reaches its cut-off voltage, the charging current drops to keep the battery's voltage constant. The charging current keeps coming down until it reaches the cut-off (end condition) value. The same procedure happens for the CCCV discharge method.

According to the cell's specification, the capacity of the cell is 4.8 Ah. The maximum constant current for the charge is 0.3 C rates, which is 1.44 A. The cut-off voltage is 4.2 V, and the cut-off current is 50 mA. Fig. 6.3 displays the CCCV charge profile for the cell. The constant current 1.44 A is employed to the battery until it reaches the 4.2 V cut-off voltage and drops until it reaches the 50 mA. At that point, we can state that the battery is fully charged. Fig. 6.4 illustrates the voltage response of the corresponding CCCV curves. When the battery reaches 4.2 V, there is no increase in the voltage, and the battery stays at 4.2 V. For discharge, the maximum constant current is 14.4 A (3 C rates). The cut-off voltage is 2.5 V, and the end of the discharge current is the same as a charge current (50 mA). To avoid damaging the cell, the battery's voltage does not exceed 4.25 V for charge and does not fall behind 2 V for discharge.

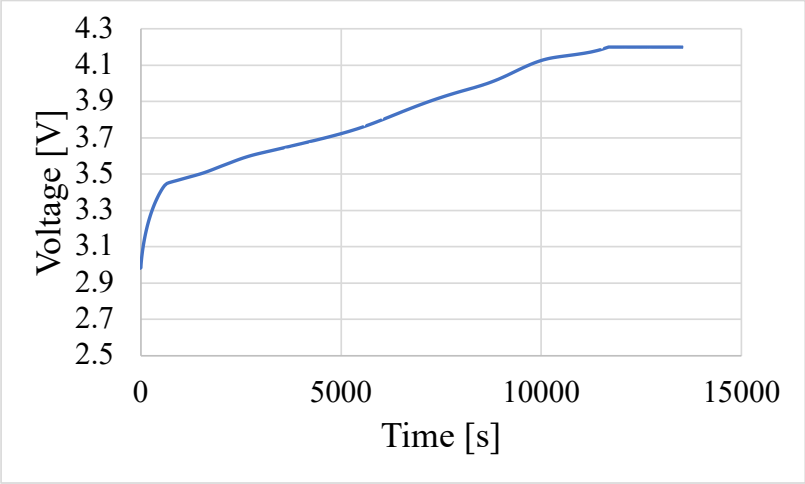


Figure 6.4. CCCV charge voltage response.

All the experiments were performed at 10°C, 25°C, and 40°C in a thermal test chamber. The experiments were performed on a fresh cell. We charge and discharge the cell with the CCCV method with a 0.3 C-rate to a maximum voltage of 4.2 V and a discharge rate of 1 C to a minimum voltage of 2.5 V. The CCCV charging method begins with a constant current. Once the battery reaches its cut-off voltage, the charging current decreases to maintain a constant voltage. The current continues to decrease until it reaches the specified cut-off (end condition) value. A similar process occurs during the CCCV discharge method. Fig. 6.3 displays the CCCV electrical current charge profile for the cell, and Fig. 6.4 illustrates the voltage response of the corresponding CCCV method. The cell underwent 15 charge-discharge cycles before the modeling experiments. In what follows, we explain the experiments to collect data for modeling.

6.1.1 Drive Cycles

In this section, we explain the drive cycles used in this study for training, validation, and testing of the models.

The first cycle is an equivalent current profile of the aggressive US-highway (US06) driving cycle [116, 120] used for validation studies. The US06 drive cycle simulates the condition of driving an electric vehicle at high speed and aggressively. Figure 6.5

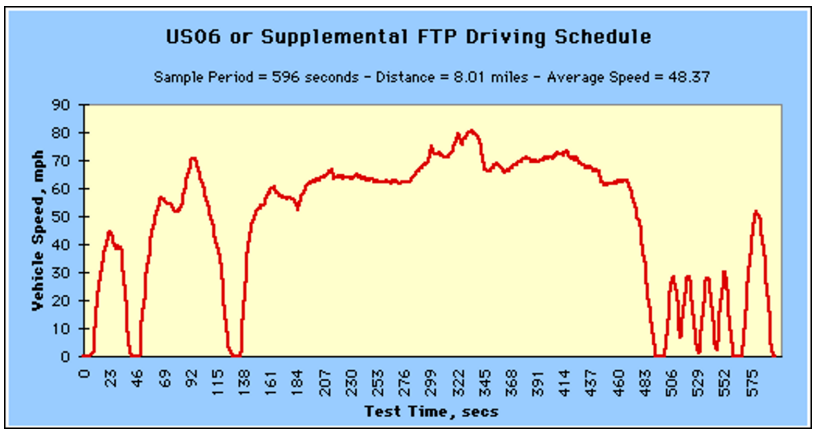


Figure 6.5. The US06 driving schedule.

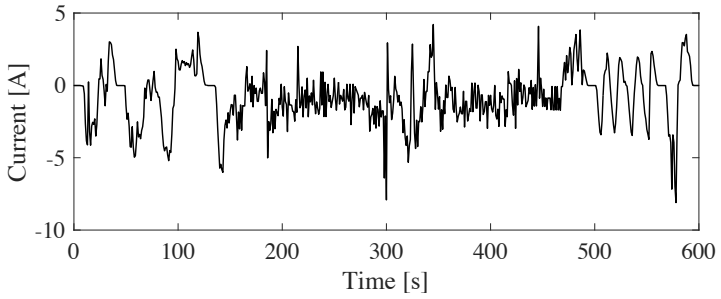


Figure 6.6. Current profile based on the US06 drive cycle.

illustrates the velocity profile of the US06 drive cycle [120], characterized by periods of high acceleration and deceleration. The corresponding electrical current profile [116], derived from the US06 velocity data, was used to generate the validation dataset, as shown in Fig. 6.6.

The second profile is the equivalent current of the standard city drive cycle, or UDDS (Urban Dynamometer Driving Schedule) [116, 120]. Fig. 6.7 shows the UDDS driving profile [120], and Fig. 6.8 shows the current profile derived from the UDDS cycle. This current profile was used for model evaluations (test data).

The last cycle is a customized drive cycle generated using random inputs. The upper bound and lower bound of these inputs were defined as the maximum and minimum electrical currents of the last two profiles (US06 and UDDS). Therefore, we set the limits to 5 A and 8.5 A for charge and discharge. The random signals were

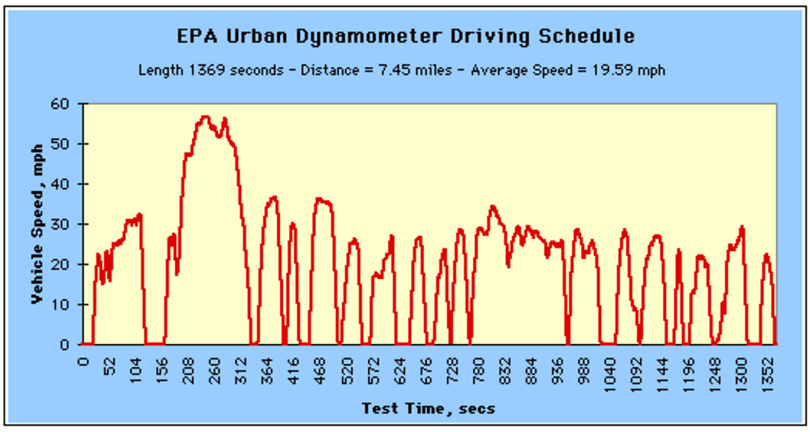


Figure 6.7. The UDDS profile.

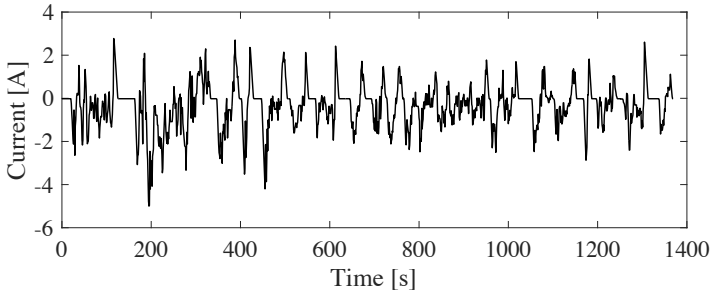


Figure 6.8. Current profile based on the UDDS drive cycle.

generated from a uniform distribution of these limits at each sampling time to cover the span of the design space. This dataset was used to train the models.

The current profiles were applied to the fully charged cell after resting in the thermal chamber for 2 hours. The sampling time was 50 ms for more accurate modeling [115]. We repeated the drive cycle profiles for validation and test data until the voltage reached 2.5 V. Then, the cell was fully discharged with the constant voltage (CV) for a more accurate depth of discharge (DoD) estimation. All experiments were repeated to ensure the repeatability of the data. A summary of the experiments is shown in Fig. 6.9

6.2 SOC calculation method

The SOC of commercial cells cannot be measured directly; however, we can use charge and discharge capacities and the initial value of the SOC from the experi-

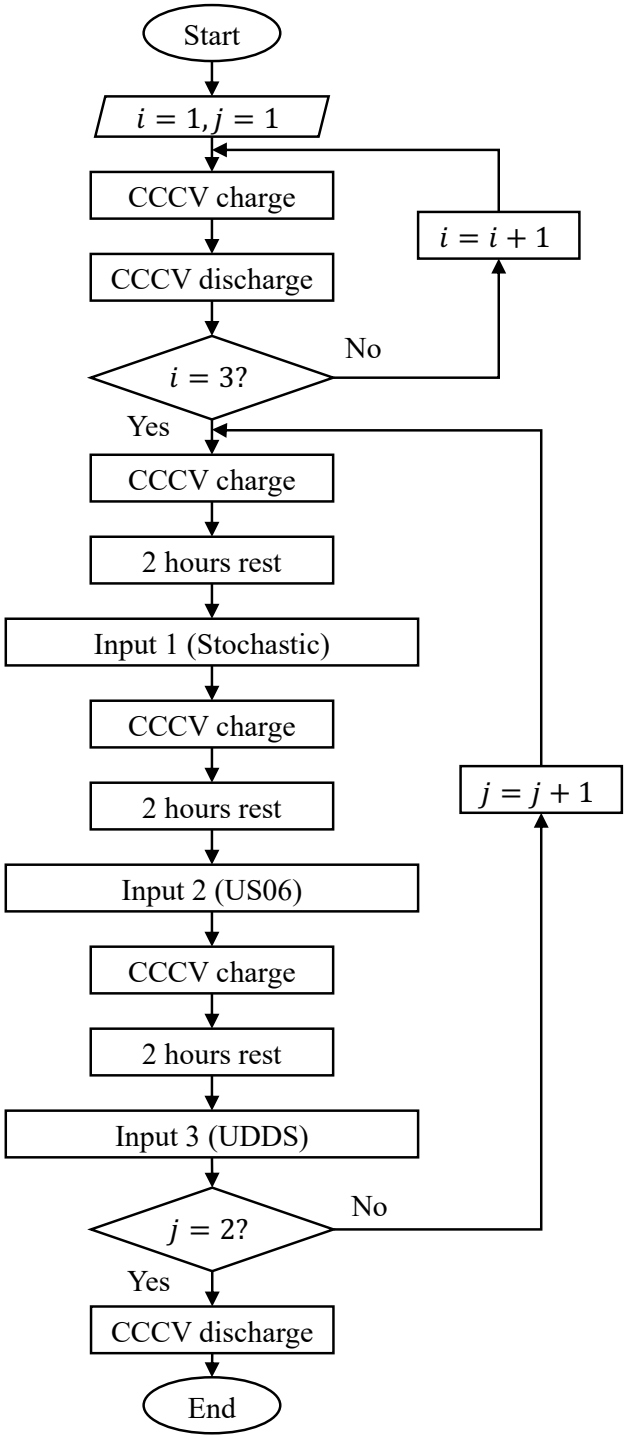


Figure 6.9. The procedure of the experiment.

ment to estimate the SOC. We estimated the SOC through enhanced coulomb counting (4.19) and used it as the reference SOC. To access the SOC's initial value, we fully discharged the battery (0% SOC) and then fully charged the battery; therefore, the initial value of the SOC for the experiment was considered 100%. We calculated the coulomb efficacy η_c offline by using the ratio of the battery's capacity at full charge and discharge. The discharge capacity of the battery is measured after completing the cycles (e.g., training data cycles); therefore, this method is not suitable for real-time SOC estimation.

6.3 Experimental results

Here, we present the experimental results on a 21700 cylindrical cell with material NMC811. Fig. 6.10 displays the custom, US06, and UDDS current profiles, measured voltage, temperature, and capacity for generating data, followed by charging and rest time at 10°C. Similarly, Fig. 6.11 and Fig. 6.12 display the collected data for constructing the model at 25°C and 40°C, respectively.

Now, we provide a detailed view of the experimental data, particularly for developing a model at 25°C. Fig. 6.13a shows the custom electrical current profile and voltage measurement applied to the cell. The current ranged from -8.2 A (discharge) to 4.6 A (Charge/ regenerative). The cell started at 100% SOC with a voltage of 4.2 V. As shown in the figure, after reaching the minimum voltage of 2.5 V (at about 9820 seconds), the battery was discharged using the CV method to a fully discharged state (50 mA cutoff current). A similar procedure was followed for the other current profiles. Fig. 6.13b shows the US06 current profile and voltage measurement. We repeated US06 until we reached the minimum voltage. Fig. 6.13c shows the UDDS current profile and measured voltage. The current profile was repeated until the cell reached its minimum voltage, followed by a CV, as shown in the figure.

Fig. 6.14 shows the span of the measured training data, which defines our design space. The current is uniformly distributed to capture the possible values of the

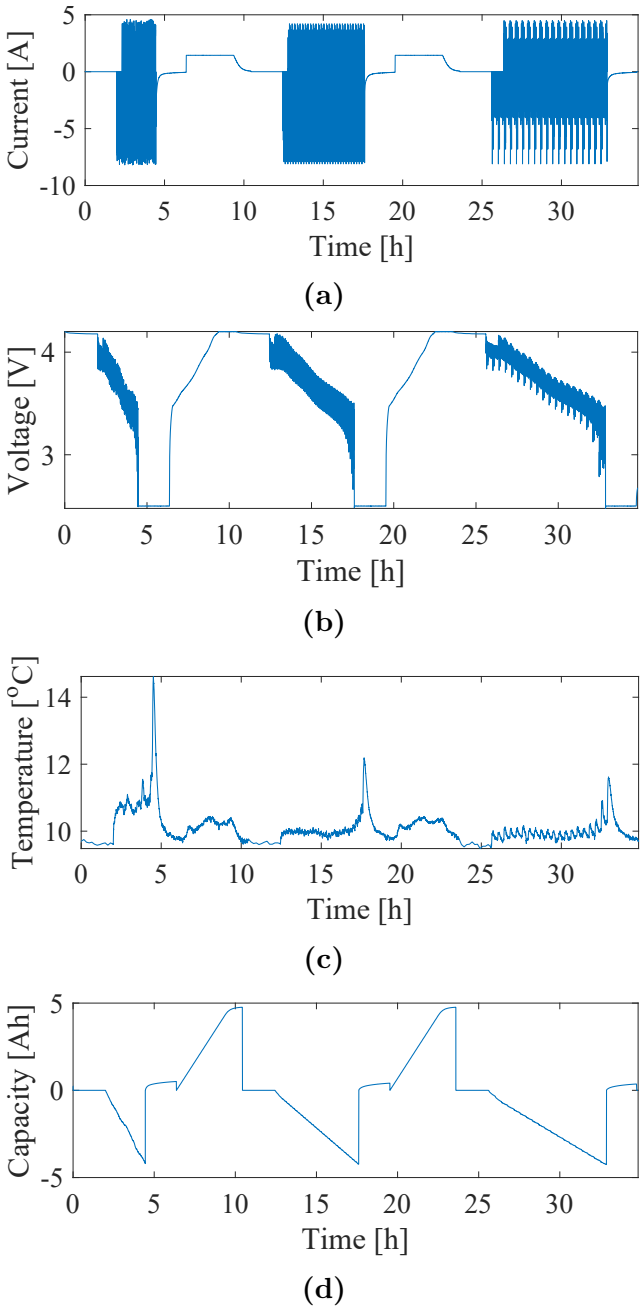


Figure 6.10. Collected experimental data at 10°C. (a) electrical current, (b) voltage, (c) temperature, and (d) cell capacity.

voltage and SOC in the operating range of the battery. The identified model should be valid for any point within this region.

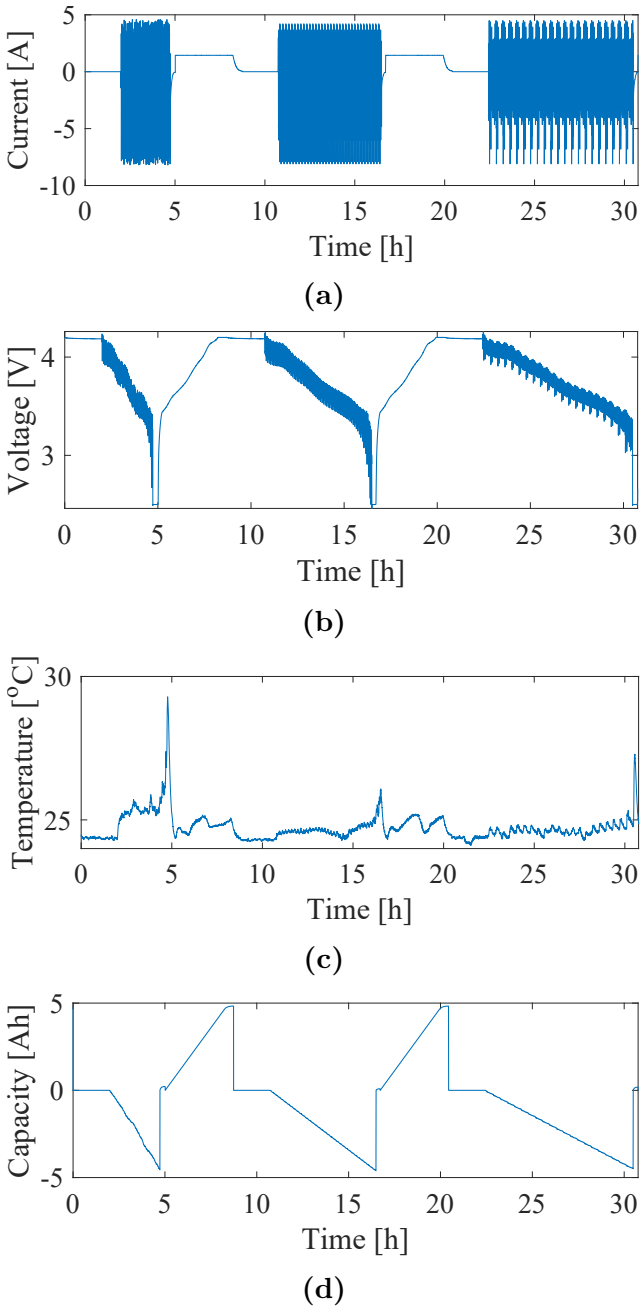


Figure 6.11. Collected experimental data at 25°C. (a) electrical current, (b) voltage, (c) temperature, and (d) cell capacity.

6.4 Summary

This chapter outlined the experimental setup conducted on an NMC811 Li-ion cell. We provided a detailed description of the cell's chemical composition and electrical

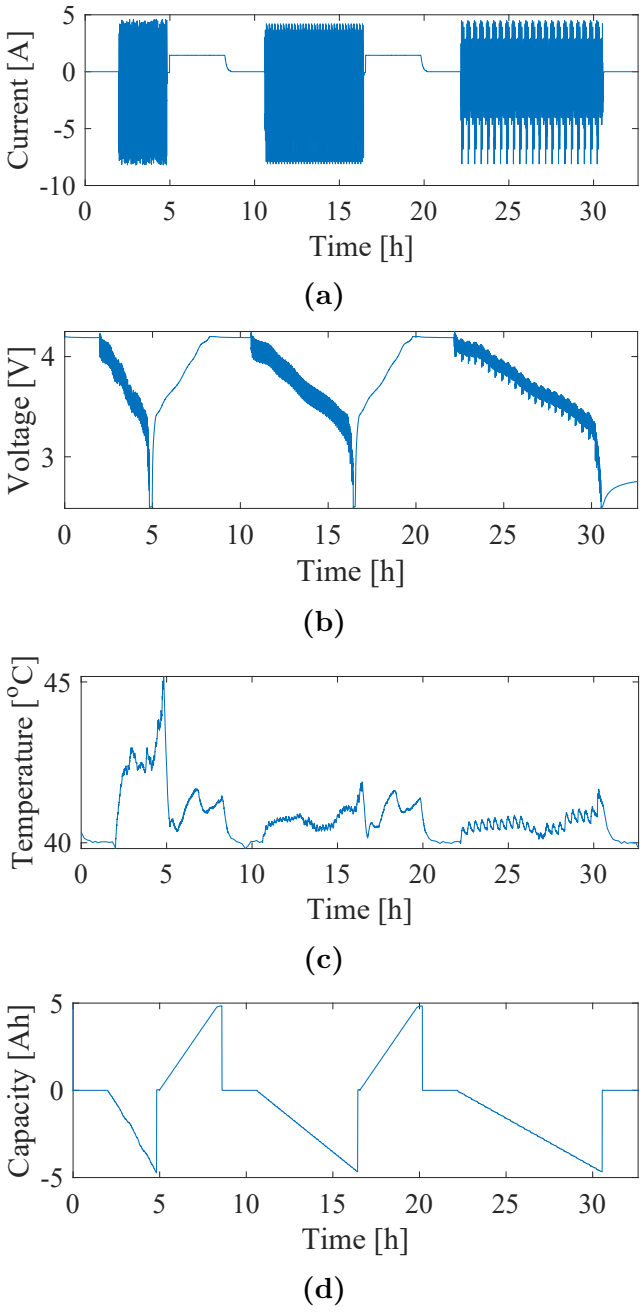


Figure 6.12. Collected experimental data at 40°C. (a) electrical current, (b) voltage, (c) temperature, and (d) cell capacity.

specifications, along with an overview of the MACCOR 4200 battery tester and the MSK-TE906-80L thermal chamber used to control testing conditions.

The experimental procedure for battery cycling was explained, with distinct datasets collected for model training, validation, and testing. A stochastic custom current pro-

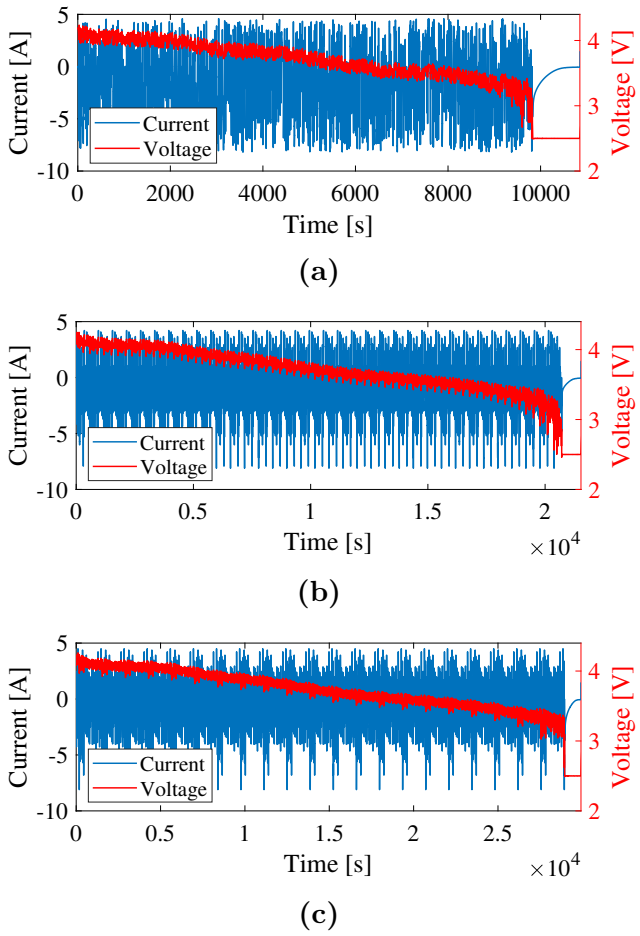


Figure 6.13. Collected experimental data. (a) Custom input followed by CV, (b) US06 input followed by CV, and (c) UDDS input followed by CV.

file, representing aggressive driving conditions, was used for training. The highway driving profile was applied for validation, and the urban driving schedule served as test data, allowing for a comprehensive evaluation of the model across real-world scenarios. Experiments were conducted under a range of thermal conditions: 25°C as the baseline, 10°C to represent cold conditions, and 40°C for hot environments.

The design space of the data-driven model, which spans the full operational range of the cell, was established to ensure coverage of all possible data points within the cell's operating limits. The SOC was estimated using enhanced coulomb counting, with the initial SOC set to 100% after a full charge cycle. This comprehensive dataset

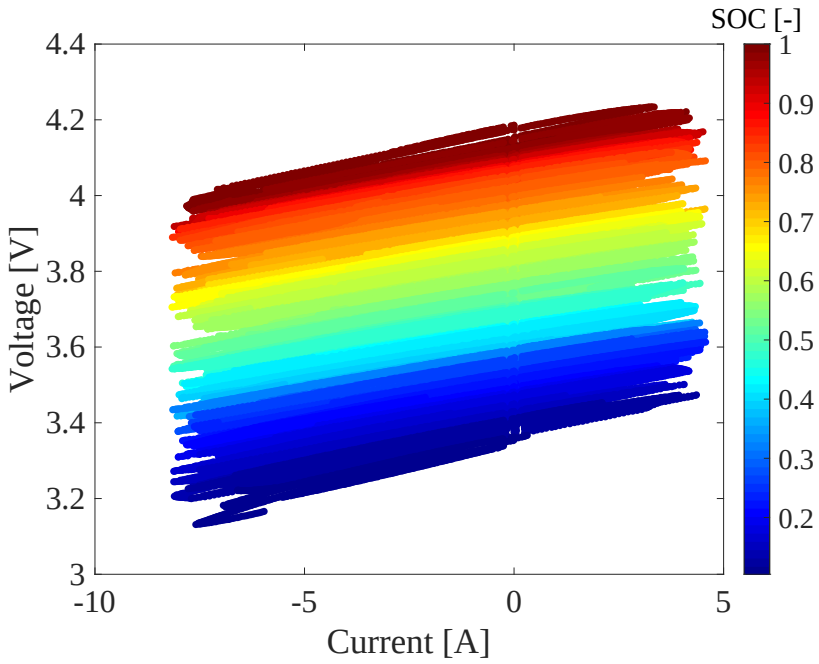


Figure 6.14. Design space of the model.

provides the foundation for accurate model development, covering the entire voltage and SOC ranges to support robust SOC estimation and model validation.

CHAPTER 7

RESULTS AND DISCUSSION

In this chapter, we present the results of applying our proposed physics-inspired, data-driven model enhanced with a joint unscented Kalman filter. The process begins by optimizing the model's sparsification parameters using the novel cost function with experimental data. Next, we assess the model's performance in mitigating noise from measured voltage data and improving SOC estimation accuracy. We also explore the model's adaptability to various operating conditions.

We then demonstrate the effectiveness of the co-estimation framework, which allows the model to adapt its parameters under noisy and uncertain conditions online, even when the initial state is unknown. Lastly, we showcase the model's robustness and accuracy across different operating environments by refining its coefficients. Note that since we are using experimental data across the entire battery range, RMSE was used instead of NRMSE, as the latter becomes very large in the low SOC region. The following sections provide a detailed analysis of the specific results and performance of the proposed model.

7.1 Sparsification Parameters Optimization Results

In this section, we use the experimental data to find a nonlinear tractable data-driven model of LiBs. As discussed in section 4.1.2, the primary step after selecting the library is optimizing the hyperparameters (threshold ξ_{th} and regularization λ), which is achieved by minimizing (4.8).

For both SOC and voltage, we select $\rho_1 = \rho_2 = 100$ and $\rho_3 = 0.1$. Training and validation RMSE are equally weighted to ensure a balanced model performance on both data sets. In addition, we adjust the weights to account for the scale difference between the RMSE and sparsity components. As a result, we can create a novel evaluation metric that effectively considers both accuracy and sparsity.

Table 7.1. Optimal hyperparameters λ, ξ_{th}, and model's number of terms.

Parameters	**SOC**	**Voltage**
λ	2.40×10^{-7}	5.74×10^{-6}
ξ_{th}	9.24×10^{-4}	1.17×10^{-2}
Sparsity	9	9

Following the steps shown in Fig. 4.1, we conduct STRidge regression on the training data to find a set of models. Each model is generated using a pair of sparsification parameters λ and ξ_{th} in an extended range of parameter space (e.g., $\lambda,\ \xi_{th} \in [10^{-8}, 1]$). Subsequently, we assess each model's performance by evaluating the error of the training data by cross-validation (using only the initial value to predict the time series data) and validation data error. We calculate the cost function (4.8) for each model using both the cross-validation and validation error, along with the number of nonzero terms in each model's sparse vector of coefficients. Finally, the sparsification parameter pair that results in the lowest cost is selected as the optimal sparsification parameter set, and we use it in STRidge regression to identify the active terms and their coefficients from the training data.

Fig. 7.1a displays the cost function value versus the number of active (nonzero) terms in the SOC sparse coefficients, assuming voltage data is known. We note that while different values of the hyperparameters can result in the same number of terms, we present only the set with the minimum cost (4.8) for each number of terms in the figure. Assuming the SOC data is known, we identify the optimal hyperparameters for voltage prediction. The corresponding voltage prediction cost of the identified model with respect to the number of nonzero terms is shown in Fig. 7.1b.

Table 7.1 shows the optimal values of the hyperparameters λ, ξ_{th}, and sparsity for the selected model. Among 32 functions in the library, the sparse model requires 9 functions to estimate SOC and voltage.

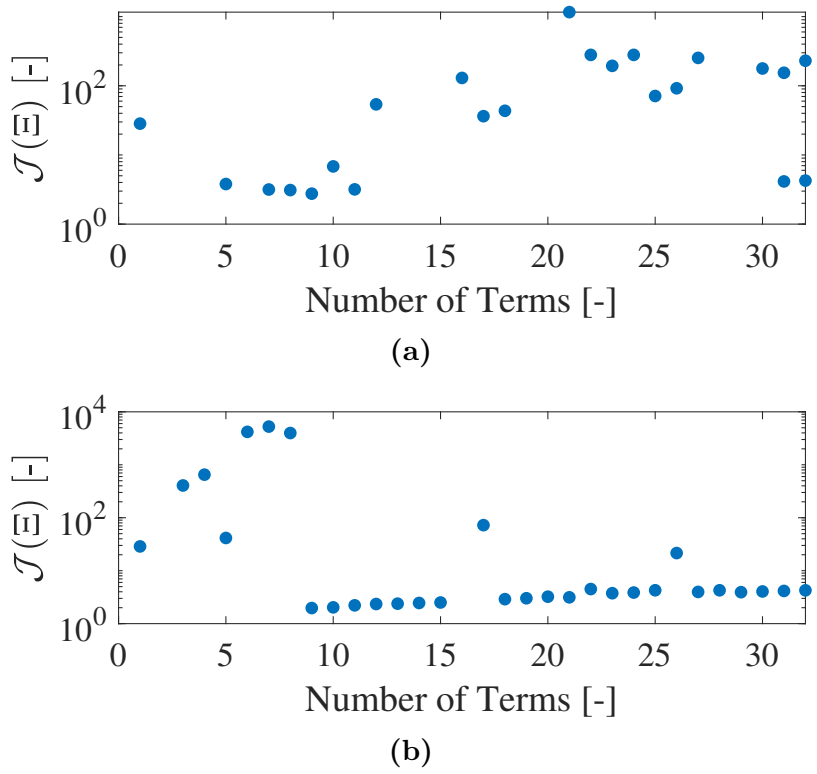


Figure 7.1. Cost function as a function of model complexity (Number of Terms). (a) SOC dynamic with the minimum cost with 9 terms. (b) Voltage dynamic with the minimum cost with 9 terms.

Small values of λ and ξ_{th} yield the sparse model incorporating additional terms that result in diminishing returns. Adding more terms does not necessarily result in a better model, as it may overfit the training data. While more terms may lead to lower RMSE of the training data, the resulting model degrades the validation error. Usually, these terms are unnecessary and provide a model that is not general and has a significant error for unseen data. Additionally, using only RMSE can provide us with an accurate model which is not sparse. For voltage prediction, utilizing 15 terms offers the lowest validation RMSE of 0.0078; however, many are unnecessary, and adding them increases the model complexity. Instead, we select 9 terms with a validation RMSE of 0.0086 using our proposed cost function. Fig. 7.2a and Fig. 7.2b display the SOC and voltage prediction of the training data. This approach allows for creating an accurate tractable data-driven model while avoiding overfitting by including only

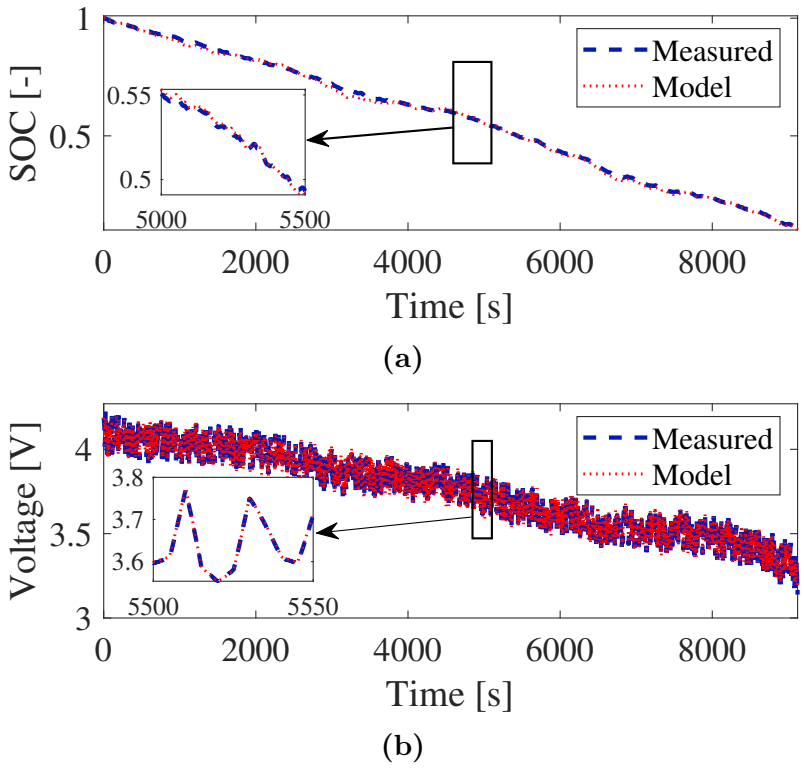


Figure 7.2. Model performance on the training data. (a) SOC dynamics. (b) Voltage dynamics.

active functions. We evaluate the model's performance with validation and test data sets to ensure the model is generalizable to new data.

7.2 Model in The Presence of Noise

Here, we show the result of using a JUKF (section 5.1.2) in improving the performance of the nonlinear model subject to the noisy voltage measurement. We mitigate the noise from voltage data with suitable Q for validation and test data sets, while the measurement noise covariance R is computed based on the covariance of the training voltage measurements.

Fig. 7.3 illustrates the model predictions on the validation data, where the model achieves an RMSE of 0.0099 for SOC prediction. After calculating the sparse vector of coefficients Ξ_1 and Ξ_2 with training data, the identified model predicted voltage and SOC concurrently using the input current profile. We employed the JUKF to rectify

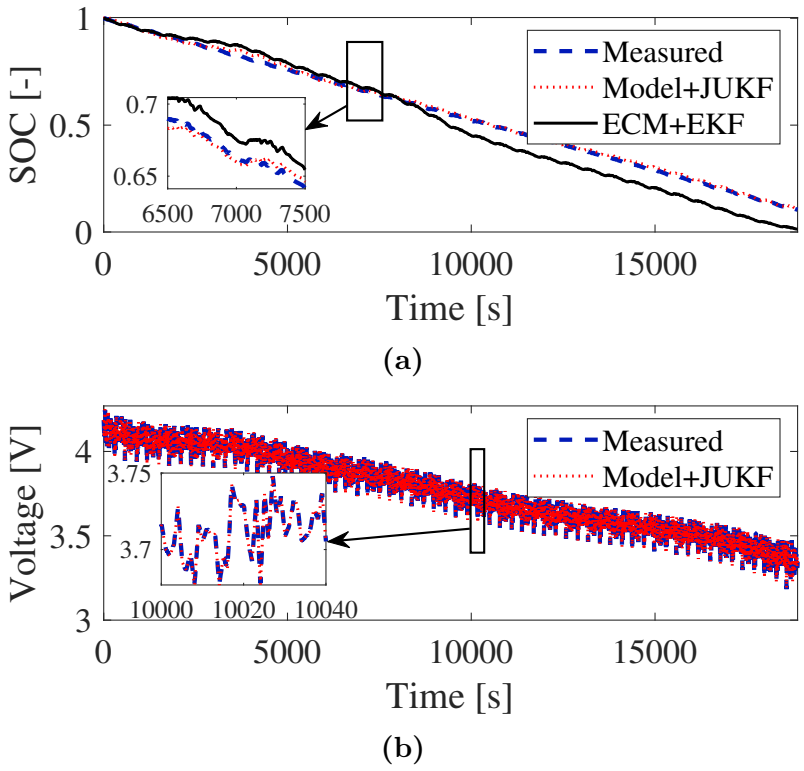


Figure 7.3. Model validation results on US06 drive cycles. (a) US06 SOC. (b) US06 voltage.

the measured voltage data and use the corrected value to predict SOC, reducing the RMSE from 0.0713 to 0.0099 for SOC prediction. The model shows high accuracy in correcting voltage with improvement in RMSE from 0.0579 V to 0.0008 V.

Next, we compare our data-driven model to an established and commonly used method, ECM with EKF. We tuned ECM parameters using stochastic training data. Similar to our method, we used the validation data's current, voltage, and initial value of SOC to predict the time series SOC data and correct the measured voltage data. Fig. 7.3a illustrates the superior performance of the proposed approach (RMSE: 0.0099) compared to the ECM with EKF (RMSE: 0.0629). Moreover, ECM with EKF corrected the voltage data with an RMSE of 0.0155, significantly higher than our proposed model's RMSE of 0.0008. Therefore, the proposed modeling approach is a more effective and accurate alternative approach for SOC prediction and voltage correction.

Table 7.2. Identified voltage model and updated coefficients with JUKF.

Terms	Initial	After JUKF
V	0.9474	0.9416
SOC	-0.0819	-0.0791
I	0.0677	0.0668
$\int I$	-0.0675	-0.0658
SOC^2	-0.1146	-0.1081
$\exp(V)$	-0.0906	-0.0918
$\exp(\mathrm{SOC})$	0.2292	0.2373
$\cos(\mathrm{SOC})$	-0.0516	-0.0437
$\sinh(V)$	0.0952	0.0873

Table 7.2 displays the identified functions with their coefficients from the training data and the updated coefficients obtained using JUKF for voltage dynamics in the US06 dataset. The parameters were slightly adjusted to refine the noisy voltage data.

To verify the generalizability of the identified model, we utilize test data. Fig. 7.4 shows the results of the identified models on the unseen UDDS test data, which was applied to the cell for approximately 7.5 hours. The RMSE of the SOC prediction with JUKF is 0.0126, which shows the accuracy of the prediction in unseen scenarios. Furthermore, the model achieved an RMSE of 0.0006 V for correcting the voltage. The model's ability to accurately predict the voltage and SOC on unseen test data demonstrates its robustness and practical applicability.

The model's RMSE on the highway (US06) and urban (UDDS) driving scenarios demonstrate that the model is suitable for real-world driving situations. The model also calculates the battery's power, a critical parameter for electric and hybrid vehicles, by combining the estimated/corrected voltage and electrical current.

7.2.1 Noise mitigation evaluation on different operating conditions

To verify the model's accuracy, we conducted the same experiments outlined in section 6.1 at 40°C. We used the governing equation (9 terms for both voltage and

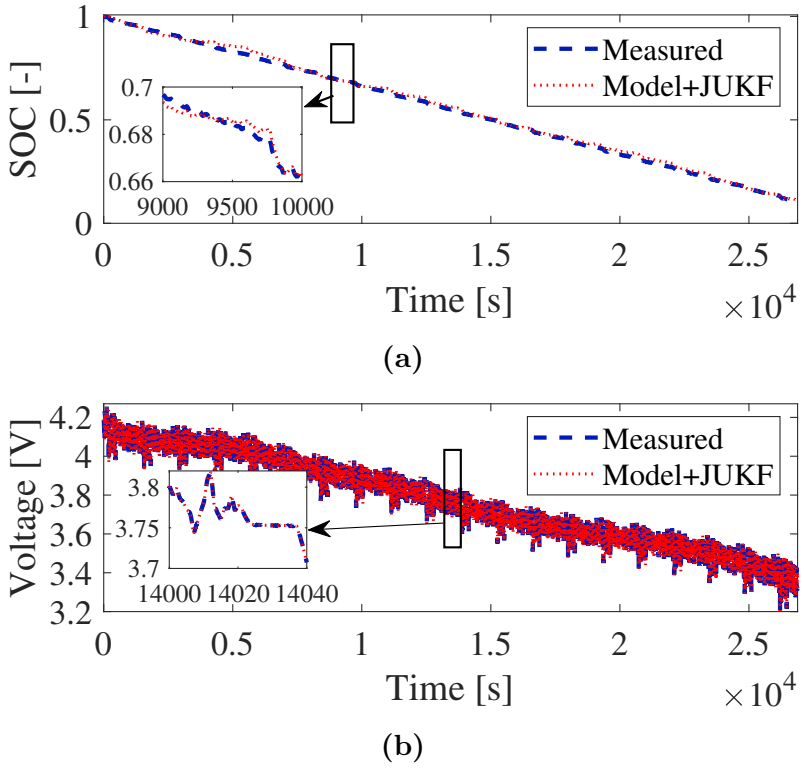


Figure 7.4. Model test results on the unseen UDDS drive cycles. (a) UDDS SOC. (b) UDDS voltage.

SOC according to Fig. 7.1) identified from 25°C data and updated their coefficients using the training data at 40°C. We achieved an RMSE of 0.0021 and 0.0088 for voltage and SOC prediction from the training data, respectively. Similarly, we evaluated the model's performance on the US06 data as indicated in Fig. 7.5. The model was augmented with JUKF to correct voltage data, resulting in RMSE of 0.0006 and RMSE of 0.015 for SOC prediction. Finally, we evaluated the model with JUKF performance on the UDDS data to ensure its generalizability and robustness. We utilized the same coefficients obtained from the 40°C training data and achieved an RMSE of 0.0165 for SOC prediction and an RMSE of 0.0005 for correcting voltage data. Therefore, we discovered the active functions using our proposed cost function (4.8) from training and validation data at 25°C. We updated the coefficients corresponding to the discovered active functions using the stochastic training data at 40°C and confirmed the model's accuracy by achieving an RMSE of 0.015 for US06

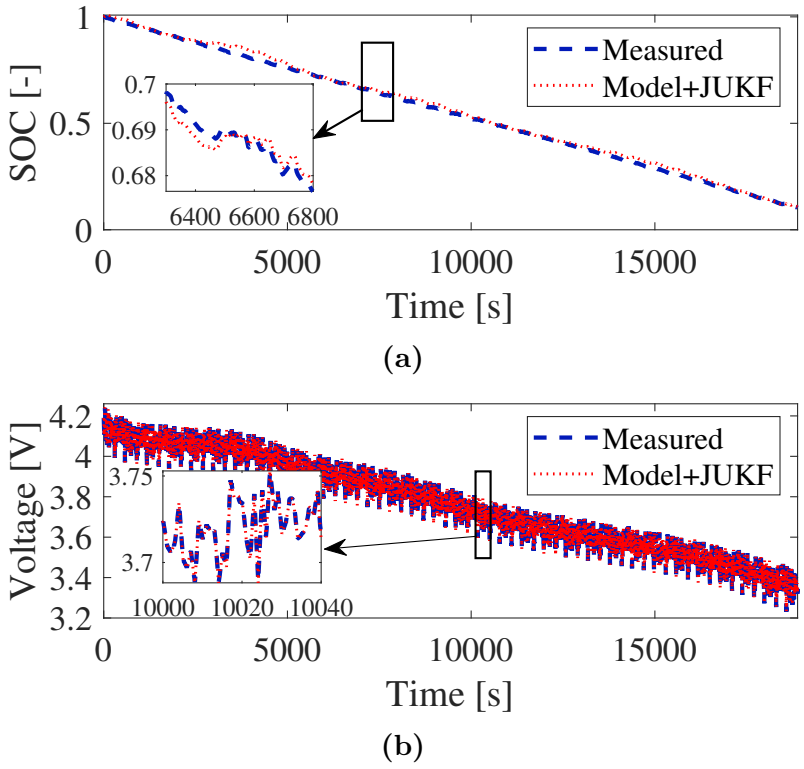


Figure 7.5. Model results on US06 drive cycles at 40°C. (a) US06 SOC. (b) US06 voltage.

SOC prediction. The results show that the model was able to generalize well and capture the underlying dynamics of the system under different operating conditions.

Hence, we can implement the proposed method to construct tractable data-driven models for other LiBs to predict voltage and SOC. It is worth reiterating that the model needs remarkably less data than other machine learning techniques, and it helps us train the model quickly. In addition, reformulating the cost function in the SINDy algorithm for tuning sparsification parameters helps us select a robust and generalizable model with high accuracy in unseen data and different operating conditions. The traditional SINDy cost function, AIC-inspired, utilizes the error of the training data and the number of active terms, resulting in a model with only one term for SOC predictions. The model identified with the AIC-inspired cost function has only one term (Integral of current from enhanced coulomb counting); however,

it performs poorly in US06 and UDDS current profiles. Therefore, the proposed cost function, which includes the error of training and validation data sets, as well as the number of active terms, is necessary to attain a more accurate model.

We note that while the procedure to discover the library terms can be considered mostly an offline calculation, the coefficients can be adjusted in real-time, allowing the model to adapt to changes in the operating environment and aging of the cell.

7.3 SOC Estimation and Voltage Correction Results

This section evaluates the proposed co-estimation algorithm on the validation and test data.

Fig. 7.6a illustrates the SOC estimation for the validation data. The identified model simultaneously estimates the SOC and voltage using the electrical current profile. The initial value of the SOC is set at 1 (fully charged). However, we use 0.8 SOC (20% uncertainty) as the initial value for SOC estimation. The figure shows that the algorithm adjusted the voltage signal to account for the measurement noise. Also, the SOC converges with its true value. We note that the SOC correction happens at every step, and once the SOC reaches its true value, the algorithm does not adjust it anymore. The model with JUKF outperforms the original model as it updates the parameters and states using the measured voltage data information.

As a next step, we compared our method with a commonly used technique: ECM with EKF. We tuned the ECM parameters using our training data and the EKF parameters using the validation data. We then tested our model on the unseen test data and compared the results with those of the ECM with EKF. Fig. 7.6b displays the SOC estimation error for the validation data. The proposed model converges toward the actual SOC compared to the ECM with EKF, showcasing the robustness of our proposed approach, particularly at low SOC levels. The SOC validation RMSE of the model after convergence is 0.0102, highlighting its high accuracy. The results showed that our method significantly outperformed the ECM with EKF.

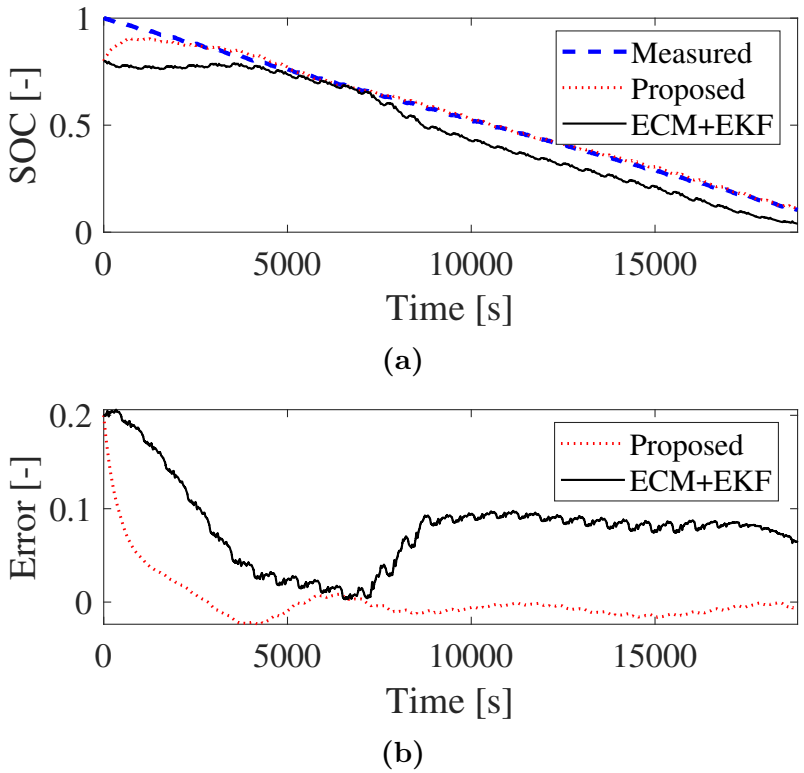


Figure 7.6. SOC validation results on US06 drive cycles. (a) SOC estimation result. (b) SOC error curve.

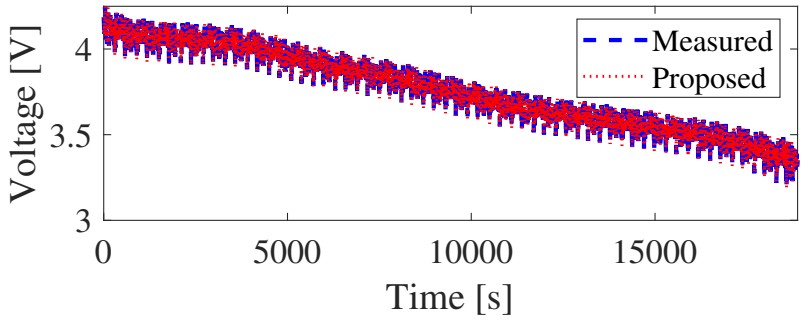


Figure 7.7. Voltage validation results on US06 drive cycles.

Fig. 7.7 presents the voltage correction result under the US06 validation data. The JUKF effectively handles model uncertainties, achieving an RMSE of 0.0008 V. As can be observed, the uncertainty in SOC does not affect the performance of the proposed model for correcting the voltage data, which is used as a SOC-voltage map.

The model's performance was further tested on the unseen UDDS test data set. Fig. 7.8 and Fig. 7.9 display the SOC estimation and voltage correction results, re-

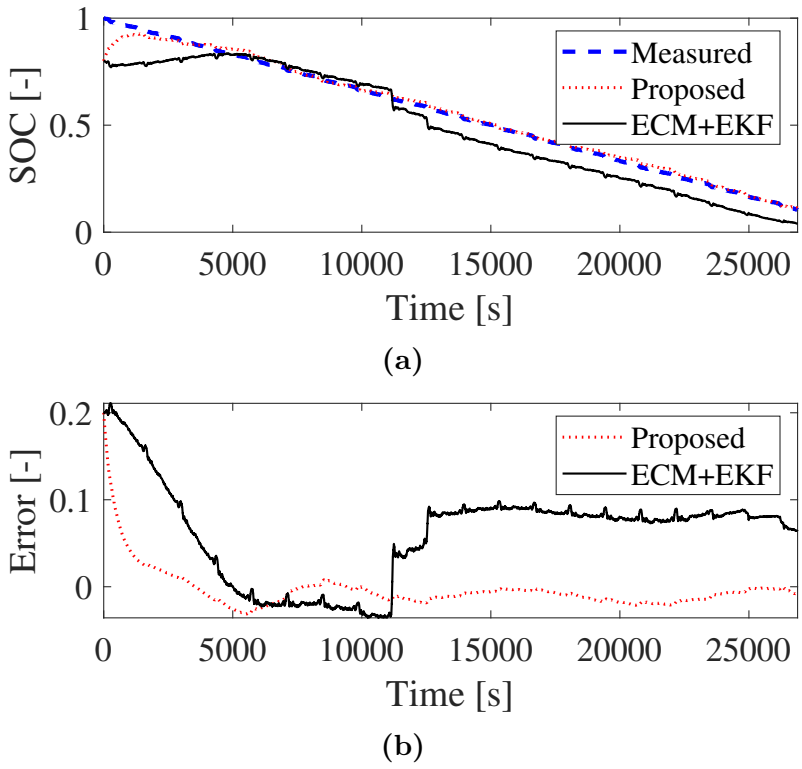


Figure 7.8. SOC results on UDDS test started with a 20% initial SOC error. (a) SOC estimation result. (b) SOC error curve.

spectively. Similar to the validation dataset, we considered 20% uncertainty in the SOC initial value. As shown in Fig. 7.8b, the proposed model converges to the actual SOC. However, the ECM with EKF converges slowly and tends to diverge around 50% SOC, maintaining a stable overall error. In addition, the model has an RMSE of 0.0130 after convergence, highlighting the model's generalizability in unseen scenarios. Furthermore, the model achieves an RMSE of 0.0006 V for correcting UDDS voltage data, demonstrating high accuracy for refining the SOC-voltage map during operation.

The RMSE and mean absolute error (MAE) for SOC prediction after convergence and voltage correction for US06 validation and UDDS test data are given in Table. 7.3. We used the Maccor model 4200 battery tester with a $\pm 0.02\%$ full-scale range (FSR) voltage accuracy. Hence, the experiment's voltage measurement accuracy is approximately ± 0.0004 V, in close agreement with the achieved MAE from our model. The

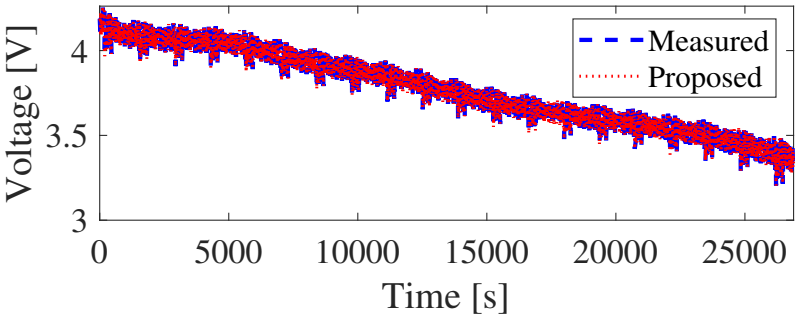


Figure 7.9. Voltage results on UDDS test data.

Table 7.3. SOC estimation after convergence and voltage correction results.

		RMSE	MAE
US06	SOC	0.0102	0.0085
	Voltage	0.0008 V	0.0005 V
UDDS	SOC	0.0130	0.0108
	Voltage	0.0006 V	0.0004 V

results indicated that the discovered governing equations are robust and can be used confidently.

7.3.1 Evaluation on Different Temperature States

Here, we evaluate the performance of our model under varying operating conditions. The experiments described in section 6.1 were repeated at temperatures of 10°C and 40°C to verify the model's accuracy. The model identified at 25°C is used as the baseline, meaning we only update the coefficients of the governing equations, which contain 9 terms each for both voltage and SOC, as shown in Fig. 7.1. In this scenario, the coefficients are updated using the least squares error method rather than STRidge since the governing equation is already known from the 25°C data.

At 10°C, the model achieved an RMSE of 0.0062V for voltage prediction and 0.0061 for SOC prediction using the training data. Additionally, when tested with a 20% error in the initial SOC value, the model attained an RMSE of 0.0283 for SOC

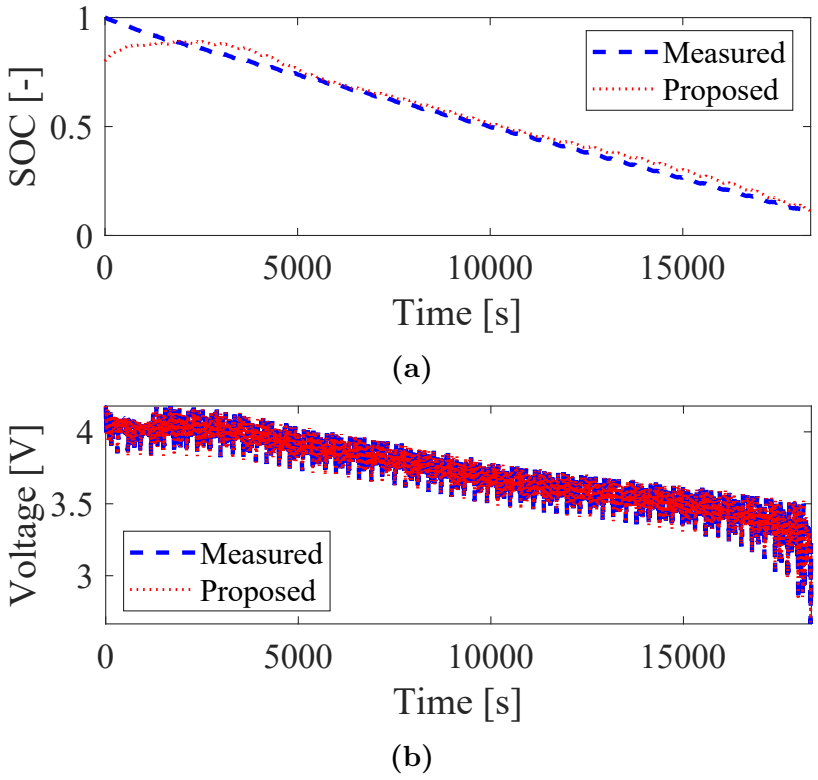


Figure 7.10. Model results on US06 at 10°C. (a) SOC estimation. (b) Voltage correction.

estimation after convergence and an RMSE of 0.0026 V for voltage correction during the US06 drive cycle, as shown in Fig. 7.10.

Finally, we evaluated the model's performance using UDDS data, starting with an 80% initial SOC (representing a 20% error). The model successfully converged to the actual SOC value with an RMSE of 0.0325 after convergence and an RMSE of 0.0019 V for voltage correction, as shown in Fig. 7.11. It is worth noting that at lower temperatures, the battery's internal resistance increases. To prevent the voltage from exceeding the cell limit, we reduced the charging current (regenerative) at the start of the cycles.

A similar approach was followed for 40°C, where cross-validation on the training data yielded an RMSE of 0.0083 for SOC prediction and 0.0021 V for voltage prediction when adjusting the coefficients. For the US06 drive cycle, as shown in Fig, 7.12a, the SOC with a 20% initial error converged to its true value with an RMSE of 0.0144

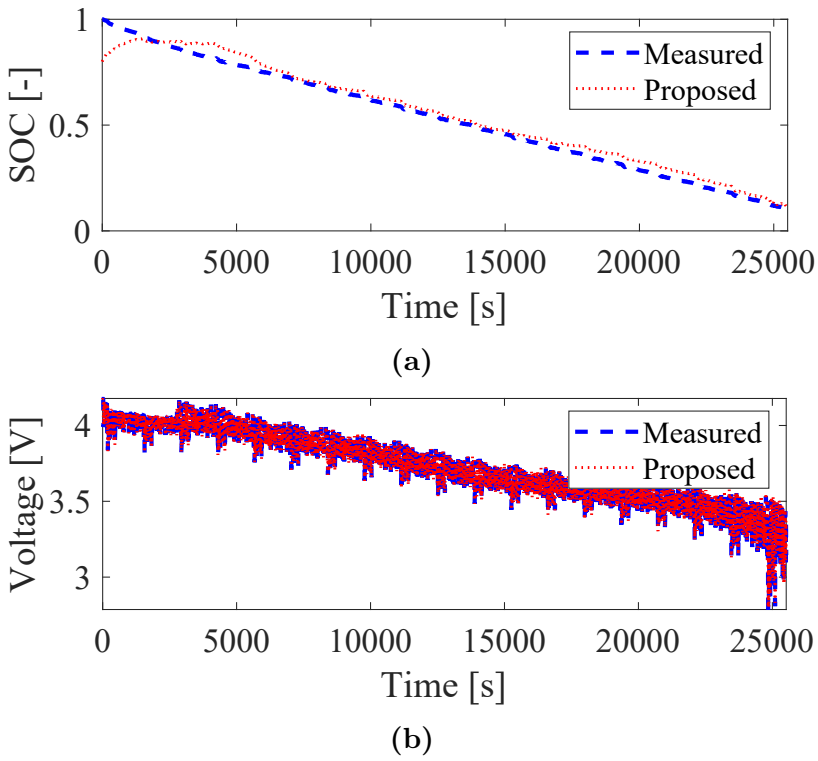


Figure 7.11. Model results on UDDS at 10°C. (a) SOC estimation. (b) Voltage correction.

after convergence. Fig. 7.12b displays that the JUKF effectively mitigated the impact of measurement noise, achieving an RMSE of 0.0006 V for the voltage data.

Fig.7.13a shows the model's accuracy in correcting the SOC with a 20% initial error for UDDS data, resulting in an RMSE of 0.0158 after convergence. Furthermore, Fig.7.13b illustrates the robustness of JUKF in handling noise, achieving an RMSE of 0.0005 V in voltage correction.

These results demonstrate the model's generalizability and applicability across different operating conditions. Only the coefficients of the governing equation identified at 25°C were updated using training data at different temperatures, proving that the underlying dynamics were correctly captured at 25°C. This same methodology can be extended to other temperatures or battery health states.

Table 7.4 shows the identified functions with their coefficients for both voltage and SOC dynamics at 10°C, 25°C, and 40°C. The coefficients for 10°C and 40°C were

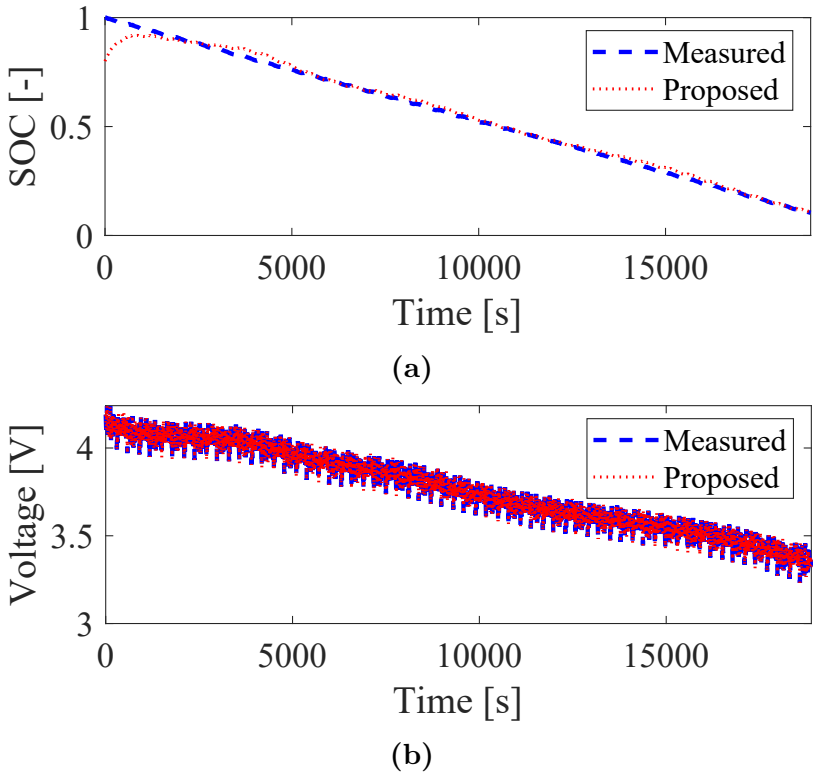


Figure 7.12. Model results on US06 at 40°C. (a) SOC estimation. (b) Voltage correction.

retrained using the identified functions at 25°C with our proposed multi-objective cost function.

Fig. 7.14 and Fig. 7.15 present the sparse vector of coefficients for all temperatures, normalized to the coefficients at 25°C. These figures illustrate the trend of the coefficients across different temperatures for the voltage and SOC models, respectively. Voltage dynamics are primarily influenced by the term V, as voltage in discrete dynamics largely depends on its previous value and a few additional terms. Similarly, in SOC dynamics, the value of SOC is highly dependent on its previous value. Terms like $\exp(V)$ show the exponential dependence of voltage on charge transfer, while $\cos(\text{SOC})$ reflects the periodic behavior related to Li-ion concentration variations in the battery. Hyperbolic terms such as $\sinh(V)$ capture the influence of Li-ion intercalation in the Butler-Volmer equation.

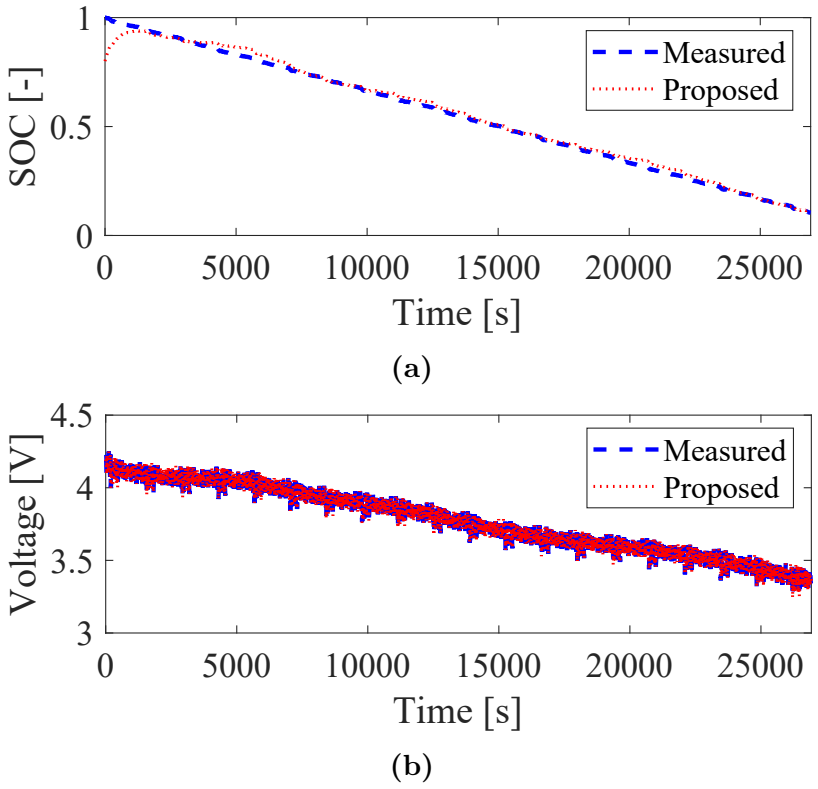


Figure 7.13. Model results on UDDS at 40°C. (a) SOC estimation. (b) Voltage correction.

Furthermore, Fig. 7.16 and Fig. 7.17 display the value of each term (function with their coefficients) during CCCV charge method at all temperatures, for voltage and SOC model, respectively, which illustrate V and SOC terms are primarily governed by the battery's state and are less sensitive to temperature changes. A reduction in the absolute values of trigonometric, exponential, and hyperbolic terms in voltage dynamics with increasing temperature indicates the temperature dependence of the diffusion process and Li-ion intercalation. For voltage dynamics, the State of Charge (SOC) plays a crucial role, as SOC is a function of the solid concentration, which significantly impacts the terminal voltage, as shown in Fig.7.16. Similarly, voltage affects SOC, as illustrated in Fig.7.17. These figures indicate that both terms are influenced by temperature, while the hyperbolic terms in the SOC model remain relatively independent of temperature. The hyperbolic term is associated with the charge transfer process, which primarily affects the terminal voltage.

Table 7.4. Identified SOC and voltage dynamics at 10°C, 25°C, and 40°C.

SOC				Voltage			
Terms	10°C	25°C	40°C	Terms	10°C	25°C	40°C
V	0.0234	0.0204	0.0162	V	0.9083	0.9620	0.9653
SOC	1.0216	1.0235	1.0261	SOC	-0.2457	-0.0819	-0.0730
V^2	0.00110	-0.0012	-0.0067	I	0.0961	0.0704	0.0636
SOC^2	0.0399	0.0420	0.0473	$\int I$	-1.9159	-1.4056	-1.2699
$\exp(V)$	0.0214	0.0207	0.0222	SOC^2	-0.3418	-0.1146	-0.1023
$\exp(\text{SOC})$	-0.0366	-0.0994	-0.0416	$\exp(V)$	-0.1107	-0.0338	-0.0294
$\cos(\text{SOC})$	0.0175	0.0173	0.0196	$\exp(\text{SOC})$	0.2523	0.0843	0.0751
$\sinh(V)$	-0.0542	-0.0479	-0.0407	$\cos(\text{SOC})$	-0.1554	-0.0523	-0.0468
$\sinh(\text{SOC})$	0.0132	0.0124	0.0148	$\sinh(V)$	0.2502	0.0827	0.0728

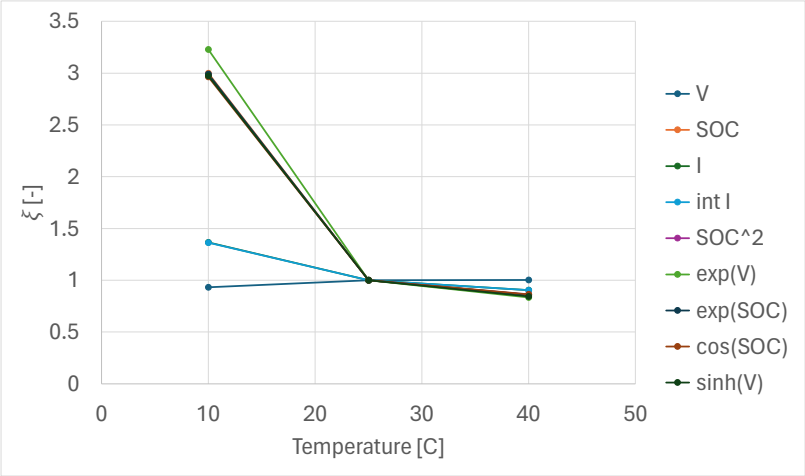


Figure 7.14. Trend of the normalized sparse vector of coefficients for voltage model.

7.4 Performance and Computational Cost Analysis

In this section, we compare our proposed physics-inspired model with the standard sparse identification method (SINDy). In the SINDy algorithm, we utilized a polynomial library with terms up to the third order and selected optimal sparsification parameters (λ and ξ_{th}) using an AIC-inspired cost function. Table 7.5 presents the results of the basic SINDy method alongside our proposed algorithm, which we have named Augmented SINDy (ASINDy).

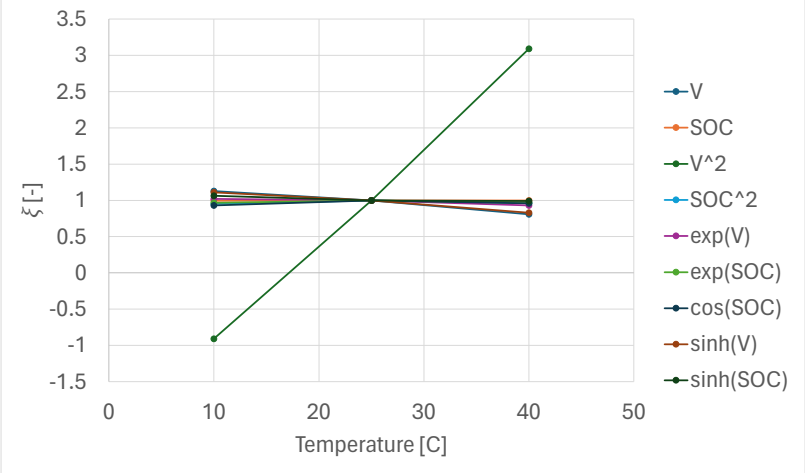


Figure 7.15. Trend of the normalized sparse vector of coefficients for SOC model.

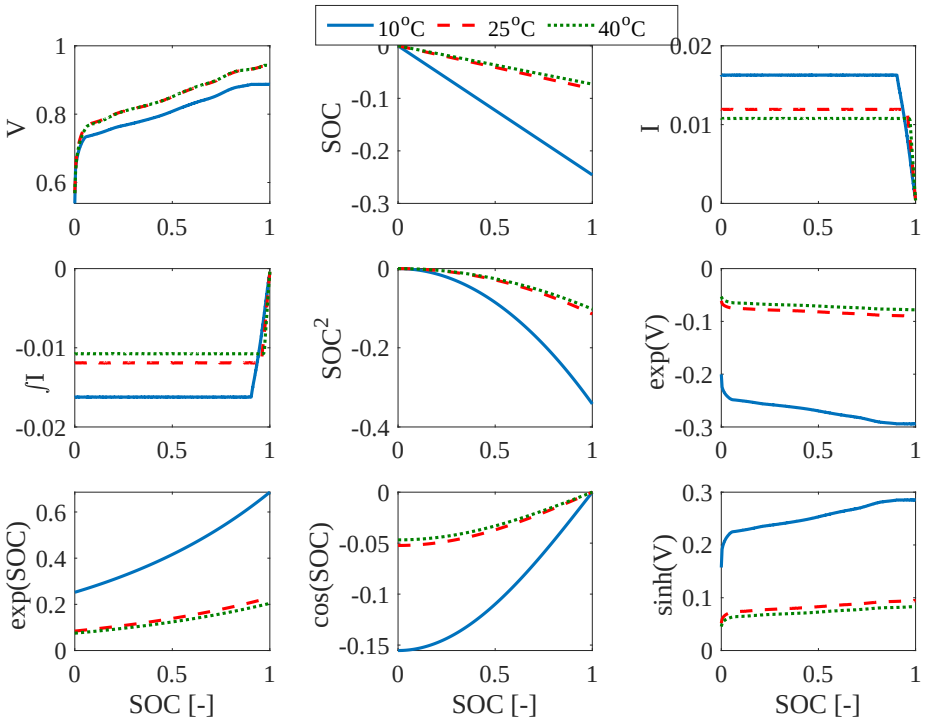


Figure 7.16. Voltage model terms for CCCV charge at different temperatures.

The results demonstrate that the generic SINDy approach requires significantly more terms than ASINDy to capture the dynamics while also yielding a higher error. This underscores the importance of designing a physics-inspired library and optimizing sparsification parameters through a multi-objective cost function. The ASINDy model achieves a sparser representation with higher accuracy, effectively capturing the

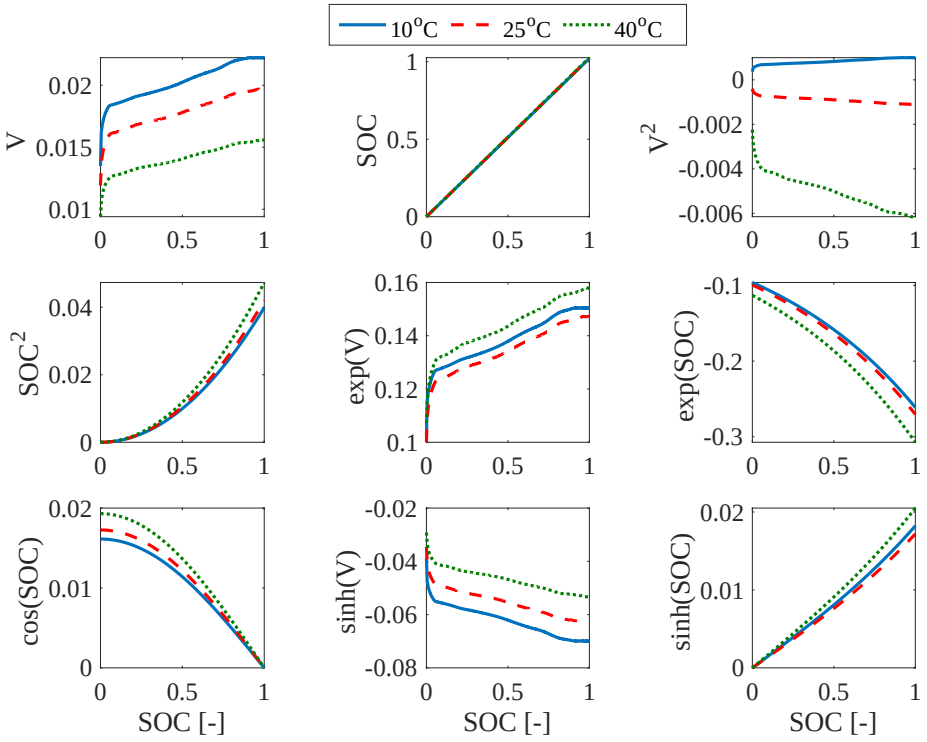


Figure 7.17. SOC model terms for CCCV charge at different temperatures.

Table 7.5. Comparision between SINDy and our proposed ASINDy models.

Criteria	**SINDy**	**ASINDy**
Number of terms for SOC dynamics	42	9
Number of terms for voltage dynamics	41	9
SOC prediction RSME on US06 data	3.9%	1%
Voltage prediction RSME on US06 data	17 mV	0.8 mV
SOC prediction RSME on UDDS data	4.4%	1.3%
Voltage prediction RSME on UDDS data	18 mV	0.6 mV
Robust to noisy measured data	No	Yes
Need initial value	Yes	No

governing equations of the battery and highlighting the advantages of our proposed enhancements.

Finally, we compare the computational cost of our proposed ASINDy method with a mechanistic model (DFN) and an ECM combined with EKF. Table 7.6 summarizes the CPU computational time results for the UDDS data. The DFN model, known

Table 7.6. Comparison of computational time for different models.

Model	Computational Time [s]
DFN model	195815
ECM + EKF	12
ASINDy	259

as one of the most accurate mechanistic models, requires detailed information about battery internal parameters. However, its significantly higher computational time limits its application in real-time systems. Conversely, the ECM with EKF is the fastest model but lacks accuracy, particularly in low SOC regions. Our proposed ASINDy method strikes an effective balance, offering computational efficiency with high accuracy in SOC estimation. This makes it a suitable solution for real-time and precise SOC estimation, particularly enhancing the performance and reliability of electric vehicles.

7.5 Discussion

We proposed a data-driven approach to co-estimate SOC and Voltage values and their dynamics. The approach is a significant step from other existing approaches in the literature. Our algorithm works in two steps: 1) Discovering the dynamics as in (4.10) and (4.11) and 2) using a joint co-estimation to remove noise from the voltage and estimate the correct SOC values. While the discovered SOC dynamics in (4.11) can accurately predict the SOC dynamics, it requires the correct initial values. Otherwise, its performance degrades and makes it unattractive for practical applications. A common method for estimating initial values of SOC is using SOC-OCV maps [121], which require a significant amount of training data from a control group of batteries. Cell-to-cell variations and changes in the cells limit the accuracy of SOC-OCV mappings, and their performance degrades over time. Our approach, on the other hand, learns such mapping using terminal voltage instead of OCV and adapts to the new conditions of the batteries. We showed the SOC converges to its

correct values through experimental studies. The proposed method can be integrated into the battery management system (BMS) to correct voltage and estimate SOC concurrently. The model requires less data for training than other machine learning algorithms, and due to its simplicity, it can be implemented in BMS. In addition, accurate SOC estimation is crucial for precise state of power (SOP) estimation in batteries, as it indicates the available energy within the battery, ensuring efficient and safe battery operation.

The other significant contribution of this study is the presented sparse identification algorithm. Integrating the joint unscented Kalman filter to join different dynamics for an improved closed-loop model resulted in an adaptive modeling technique and filtering out the measurement noise in voltage. We showed the accuracy of the method in determining the SOC dynamics, and the steady-state RMSE error was 0.013. This was achieved by replacing the common approach in obtaining the sparse nonlinear dynamics using an AIC-inspired cost function with a new cost function (4.8). We note that using the AIC-inspired cost function resulted in only one term (compared to 9 terms using (4.8)), and the RMSE of the test data was significantly higher (RMSE = 0.301). These results point to a limitation of the AIC-inspired cost function as it tries to estimate the model's prediction error using only "in-sample" errors. While such a cost function is effective in some scenarios, it does not yield an acceptable generalizable error (test data) for the complex dynamics of Li-ion batteries. On the other hand, our proposed formulation resembles the cross-validation method as it uses "extra-sample" (validation data) to balance the model's complexity and accuracy. As a result, the model is more generalizable (smaller test errors) when tested using simulation studies [122] or experimental battery data.

We also presented a methodological approach for picking the initial library terms using domain knowledge. While the exact analytical solutions of the battery dynamics are not tractable, their solution form provided a guideline to create the library terms.

The process of identifying library terms and their coefficients was conducted offline; however, the coefficients can be adapted online. This adaptability enables the model to perform well under various operating conditions by updating the SOC-voltage map. Moreover, the proposed approach can be applied to other complex dynamical systems to construct a tractable data-driven model. It can provide insight into the dynamic behavior of the system, aiding in system control or optimization. Overall, the proposed approach is an effective and efficient way to create a tractable data-driven model and gain valuable insights into the system's behavior.

7.6 Summary

In this chapter, we applied our physics-inspired data-driven model, augmented with JUKF, for online SOC estimation. Noisy experimental data from an NMC811 cell was used to construct and evaluate the model. After constructing the physics-inspired library, the model's sparsification parameters were tuned using the proposed multi-objective cost function, which considered training error, validation error, and sparsity. With the optimal parameters, we discovered the governing equations for both SOC and voltage dynamics, each with only nine terms. This sparsity highlights the model's generalizability and its ability to capture correct dynamics without overfitting to training data.

Key findings include:

Noise Mitigation: The JUKF was used to mitigate noise in the measured voltage data. The JUKF improved the accuracy and stability of SOC prediction. The experimental results from US06 current profiles (validation data) demonstrated that by using the noisy voltage measurements, the augmented model with JUKF improves the RMSE from 0.0579 V and 0.0713 to 0.0008 V and 0.0099 for voltage and SOC, respectively.

The model accurately predicted time series SOC in US06 validation data with an RMSE of 0.0099, compared to 0.0629 RMSE obtained with the ECM with EKF. This

indicates a significant enhancement in the accuracy of our model. Furthermore, the model's accuracy was verified through experiments conducted at 40°C, utilizing the governing equation identified from 25°C data. The coefficients were updated with stochastic training data at 40°C, confirming model accuracy with an RMSE 0.015 for US06 SOC prediction.

Co-estimation Framework: In addition, JUKF was expanded to utilize the measured voltage data to reduce the uncertainty in SOC dynamics, especially due to inaccurate SOC initial values. It created a model that updates state variables and model parameters dynamically with the measured data. This improved the accuracy of the model and enabled an accurate SOC prediction.

Experimental results showed the convergence of SOC to the actual values with an RMSE of 0.0130 for SOC estimation and an RMSE of 0.0004 V for voltage. We showed that while ECM with EKF has the ability to converge to the actual SOC values in mid-SOC levels, it fails in the low-SOC regions, and in 40% SOC, the relative error is about 22.2%, which is equal to $\approx$ 9% SOC, and with almost constant offset error, this value becomes more significant at lower SOC levels. For example, at 10% SOC, the relative error reaches 61.5%.

The model's ability to adapt to different operating conditions, from low to high temperatures, was tested using experimental data at 10°C and 40°C. Only the coefficients of the identified governing equation at 25°C were updated with the stochastic training data at other temperatures, and the model achieved accurate SOC predictions under varying temperatures. At both 10°C and 40°C, SOC converged to its correct value, with unknown initial conditions, with an RMSE less than 0.0330 after convergence.

Additionally, the model effectively mitigated the effect of noise in the voltage measurements, achieving an RMSE of 0.0026 V for voltage correction at 10°C and 0.0006 V at 40°C. These results demonstrate the model's accuracy, transferability, and

robustness across a wide range of operating conditions, from cold to hot environments, making it highly promising for real-world applications.

CHAPTER 8

SUMMARY AND CONCLUSION

In this chapter, we summarize the key findings of the research, focusing on the development of a physics-inspired, data-driven modeling technique and its application for SOC estimation. We discuss the major contributions of this study, including a detailed study of the battery's electrochemistry to determine physics-inspired terms, the development of a cost function to achieve a parsimonious and generalizable model, using a joint unscented Kalman filter to mitigate the effect of measurement noise and develop a co-estimation framework for a novel SOC estimation. A discussion of potential applications and the conclusion follows this.

8.1 Summary

This dissertation has made significant contributions to the field of sparse data-driven modeling and its application to discovering governing equations of Li-ion batteries for SOC estimation. Sparse data-driven modeling has shown advantages over other machine learning methods by using less data and providing explicit control-oriented models. However, most of the applications of this method have been limited to examples with known closed-form models.

The first part of the research was the application of state-of-the-art algorithms to battery modeling and showed their limitations. Specifically, we noted three limitations of these models: 1) lack of connection to physics, 2) reliance on a simple cost function using only training data, and 3) their sensitivity to initial values and measurement noise. We tackled these issues and provided a novel data-driven framework and its application for advanced battery management systems.

First, the previous research overlooked the importance of domain expertise in modeling complex systems and relied on generic library terms for modeling. As these models are nonlinear, there are several nonlinear models that can fit training data and then fail when tested with unseen data. As an example, developing models around

polynomials requires higher-order polynomial terms, which results in correlated features, leading to ill-conditioned problems and overfitting. To avoid such issues, we build a model around physics-inspired library terms based on a detailed study of the high-fidelity Doyle-Fuller-Newman electrochemical model of batteries. This model captures the main internal processes contributing to the state-of-charge and voltage dynamics of batteries based on the porous electrode theory. However, using such models in battery management systems is not practical since they include several partial differential equations (PDEs) and require over 30 internal battery parameters. Our analysis showed the presence of specific forms in the solution of these PDEs, and we used them in our baseline library. This approach enabled a more generalizable data-driven discovery of the governing equations of Li-ion batteries, regardless of the sparsification parameters.

Another limitation of the other sparse data-driven modelings was using a cost function based on the Akaike Information Criterion (AIC). This approach aims to minimize the predictive error by solely using the training data, and it shows significant limitations in modeling commercial batteries. Specifically, the AIC-inspired cost function is based on the maximum likelihood of the training data, which is susceptible to overfitting as the size of the training dataset increases. The AIC-inspired cost function suggests only one term for modeling SOC dynamics, which is similar to the coulomb counting method and does not capture the complex nature of SOC dynamics in other scenarios. To fix this issue, we created a multi-objective cost function and adopted a method from other machine learning techniques to use separate training and validation datasets, as well as sparsity, to determine the optimal number of terms. This approach resulted in a more complex but realistic governing equation of SOC dynamics and showed very accurate performance on unseen data. This approach is generic and can be used in sparse data-driven modeling of complex systems while mitigating the poor generalizability of other approaches. The algorithm fine-

tunes model sparsification parameters with the multi-objective cost function. The resulting model is a trade-off between the model's complexity and accuracy, hence reducing the susceptibility to overfitting. The model showed superior performance in cross-validation studies, especially in more challenging scenarios with highly nonlinear battery dynamics at low SOC levels.

Measurement noise degrades the performance of data-driven models as these models are highly sensitive to the quality of the measured data. We proposed incorporating a joint unscented Kalman filter to enhance SOC and voltage predictions. The JUKF minimizes the impact of noise on the measured voltage signals. We first evaluated the impact of JUKF on simulated data with Gaussian noise, which effectively removed noise from voltage measurements. As shown in Chapter 5, the same training data with and without noise can result in significantly different models, with noise deteriorating model performance.

A challenge with developing accurate and generalizable data-driven models is the quality of the data and the applicability of the modeling techniques on real-world data. While there are many studies with experimental data on Li-ion batteries, they mostly focus on simple charge-discharge cycles. Such data is very limited in modeling as they are not sufficiently rich, and one would need a copious amount of such data for modeling and dealing with the unbalanced data. Here, we analyzed the operating range of the batteries and developed a custom driving cycle to excite the system in these operating ranges and used them to model the batteries. We showed the efficiency and other advantages of this approach by conducting comprehensive experiments on a Li-ion cell. Results demonstrated that the model for both voltage and SOC dynamics required only nine terms, showcasing its interpretability and sparsity. Comparisons with a common approach used in battery management systems (ECM+EKF) showed the limitation of such approaches in accurate SOC prediction, particularly at low

SOC levels. On the other hand, our model effectively mitigated noise effects, reducing RMSE from 0.0713 to 0.0099 in SOC prediction for the highway drive cycle.

In earlier chapters, we developed a control-oriented and physics-informed data-driven modeling technique with low noise sensitivity. This technique can be used to model other complex systems; however, it needs a known initial value. In some applications, such as battery management systems, this value is available through a standard practice known as calibration. In the (re-)calibration step, the initial value of SOC is adjusted after a full charge or discharge to eliminate cumulative errors. In section 5.1.3, we introduce a co-estimation framework to eliminate the cumulative errors and estimate the SOC values without the need for recalibrations. This is achieved by deriving the nonlinear voltage-SOC dynamics instead of using tabulated voltage-soc maps after extensive experimental studies, which have been used in other studies and methods. Our novel framework simultaneously estimates SOC and adapts model parameters, allowing for reliable performance in uncertain conditions. This co-estimation framework continually adjusts parameters in the voltage dynamics to account for noisy measurements while updating the SOc values in SOC dynamics. We demonstrated the robustness and accuracy of the approach by initiating the algorithm with 20% charge difference in the initial SOC estimate. The SOC converged to its true value. After convergence, the error remained below 0.013 for the rest of the cycle.

The model's transferability across different operating conditions was further validated through experiments at 10°C and 40°C. These tests confirmed the model's robustness and adaptability across varied environments. By adjusting only the coefficients in the governing equations, the model retained high accuracy, achieving an RMSE below 0.033 after convergence for SOC estimation with a 20% error in initial values. This adaptability underscores the model's potential for real-world applications facing temperature fluctuations.

In conclusion, the developed modeling framework and detailed battery dynamics studies approach provide a reliable and accurate tool for real-time SOC estimation and voltage correction of Li-ion batteries. We evaluated the accuracy of our approach on the standard city driving cycle and compared it to the commonly used ECM+EKF. We conducted these experiments on a commonly used chemistry in electric vehicles (NMC811), and our experimental results showed the convergence of SOC to the actual values. Accurate SOC estimation is critical for several tasks of battery management system such a ensuring safety, regulating the maximum charge and discharge, and developing advanced fast charging algorithms. The currently used SOC estimation algorithms have significant errors in the low and high SOC regions. This has forced the manufacturers to reduce vehicle efficiency in high SOC values by limiting regenerative braking and reducing the vehicle's range due to high error levels in low SOCs, contributing to the range anxiety of EV drivers. By using a nonlinear form for SOC dynamics as opposed to linear models, we addressed these limitations and demonstrated the model's performance. Further, our approach allows for developing advanced fast-charging strategies via designing optimal controllers on the data-driven model.

A significant contribution of this research was the development of a tractable data-driven model that extends beyond the typical input-output relationships found in traditional machine learning methods. This approach discovers the governing equations of complex dynamical systems. The interpretability of this model eliminates the need for in-situ measurement of internal parameters, unlike mechanistic models, allowing for broader and more accessible applications. The model's performance proved promising for real-world applications. Moreover, this technique needs less data than other machine learning methods, such as neural networks. Therefore, the model can be trained quickly while the battery's dynamics change due to aging or environmental conditions. Finally, this methodology can be applied to other complex dynamical

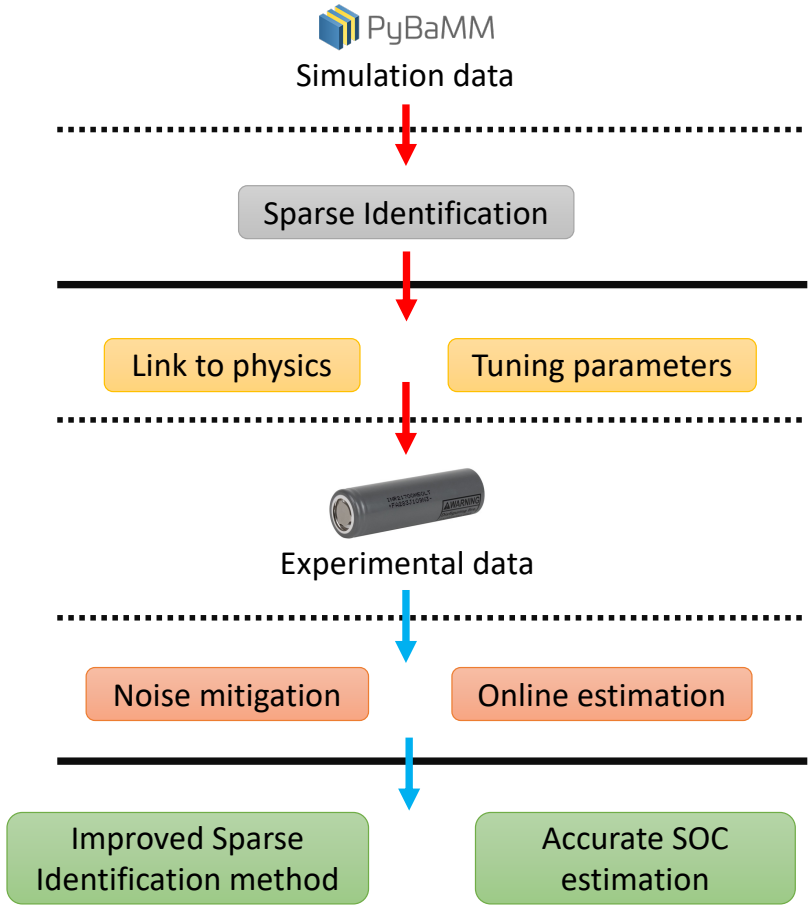


Figure 8.1. Summary of the proposed tractable data-driven method.

systems to uncover underlying equations, offering insights for system control and optimization. Figure 8.1 provides a comprehensive visualization of the proposed algorithm developed in this dissertation and its key contributions. The diagram showcases the advancements achieved, highlighting their impact on both battery modeling and data-driven fields.

8.2 Future Work

This research established an algorithm for accurate SOC estimation based on interpretable data-driven models, and several areas for future work remain to be explored. Although the model demonstrates generalizability across different driving scenarios and temperatures, further evaluation under a broader range of environmen-

tal conditions, such as extreme cold (below 0°C), is needed. Examining variations in model coefficients under these conditions can provide insights into which model terms are most affected by temperature and, in turn, reveal the internal battery processes linked to these changes.

Second, future work will apply this method to different cell chemistries, such as sodium-ion, an alternative technology with dynamics still being characterized. Given the adaptable nature of data-driven modeling, the model could be readily transferred to other battery chemistries, enabling evaluation of its performance and generalizability across various chemistries. Additionally, incorporating ensemble machine learning techniques into the data-driven model could further improve accuracy, expanding the modeling approach to form a more comprehensive framework by capturing additional parameters, such as temperature. Since temperature significantly impacts battery safety and degradation, accurately predicting it would assist thermal management systems in effectively preventing overheating. For example, data aggregation using statistical methods like bootstrapping could combine battery data across different temperatures, producing a single model for SOC estimation without a need for a look-up table while capturing temperature dynamics.

Another critical direction for future research is enhancing the model's ability to account for battery aging. As batteries degrade, their internal dynamics and SOC behavior shift. Developing a mechanism within the model to adjust for aging effects, either through periodic coefficient recalibration or self-updating, could extend the model's utility over the entire battery lifecycle.

Overall, advancing this research in these areas would contribute to more adaptable, resilient, and insightful battery management solutions. As the demand for accurate SOC estimation in electric vehicles and energy storage systems continues to grow, expanding these capabilities is essential for enhancing battery performance, longevity, and safety.

REFERENCES CITED

[1] Gholamreza Karimi and Xianguo Li. Thermal management of lithium-ion batteries for electric vehicles. Int. J. Energy Res., 37(1):13–24, 2013.

[2] Omidreza Ahmadzadeh, Renato Rodriguez, Jonah Getz, Samy Panneerselvam, and Damoon Soudbakhsh. The impact of lightweighting and battery technologies on the sustainability of electric vehicles: A comprehensive life cycle assessment. Environmental Impact Assessment Review, 110:107668, 2025.

[3] Omidreza Ahmadzadeh, Deepti Tewari, Judith Jeevarajan, and Damoon Soudbakhsh. Real-time internal short circuit detection in li-ion battery modules during field use. In 2024 American Control Conference (ACC), pages 3480–3485. IEEE, 2024.

[4] Omidreza Ahmadzadeh, Yan Wang, and Damoon Soudbakhsh. A data-driven framework for learning governing equations of li-ion batteries and co-estimating voltage and state-of-charge. Journal of Energy Storage, 84:110743, 2024.

[5] Gregory L Plett. Battery management systems, Volume I: Battery modeling. Artech House, 2015.

[6] Nalin A Chaturvedi, Reinhardt Klein, Jake Christensen, Jasim Ahmed, and Aleksandar Kojic. Algorithms for advanced battery-management systems. IEEE Control Syst., 30(3):49–68, 2010.

[7] Saehong Park, Dong Zhang, Reinhardt Klein, and Scott Moura. Estimation of cyclable lithium for li-ion battery state-of-health monitoring. In ACC 21, pages 3094–3101. IEEE, 2021.

[8] Thomas Stuart, Fang Fang, Xiaopeng Wang, Cyrus Ashtiani, and Ahmad Pesaran. A modular battery management system for hevs. sAE Transactions, pages 777–785, 2002.

[9] Gregory L Plett. Battery management systems, Volume II: Equivalent-circuit methods. Artech House, 2015.

[10] Dong Zhang, Satadru Dey, Luis D Couto, and Scott J Moura. Battery adaptive observer for a single-particle model with intercalation-induced stress. IEEE Trans Control Syst Technol, 28(4):1363–1377, 2019.

[11] Xuebing Han, Minggao Ouyang, Languang Lu, and Jianqiu Li. Simplification of physics-based electrochemical model for lithium ion battery on electric vehicle. Part I: Diffusion simplification and single particle model. J. Power Sources, 278:802–813, 2015.

[12] Scott J Moura, Federico Bribiesca Argomedo, Reinhardt Klein, Anahita Mirtabatabaei, and Miroslav Krstic. Battery state estimation for a single particle model with electrolyte dynamics. IEEE Trans Control Syst Technol, 25(2):453–468, 2017.

[13] Saehong Park, Dylan Kato, Zach Gima, Reinhardt Klein, and Scott Moura. Optimal experimental design for parameterization of an electrochemical lithium-ion battery model. J. Electrochem. Soc., 165(7):A1309, 2018.

[14] Madeleine Ecker, Thi Kim Dung Tran, Philipp Dechent, Stefan Käbitz, Alexander Warnecke, and Dirk Uwe Sauer. Parameterization of a Physico-Chemical Model of a Lithium-Ion Battery: I. Determination of Parameters. J. Electrochem. Soc., 162(9):A1836–A1848, 2015.

[15] Kong Soon Ng, Chin-Sien Moo, Yi-Ping Chen, and Yao-Ching Hsieh. Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. Applied Energy, 86(9):1506–1511, 2009.

[16] Yasser Ghoulam, Tedjani Mesbahi, Peter Wilson, Sylvain Durand, Andrew Lewis, Christophe Lallement, and Christopher Vagg. Lithium-ion battery parameter identification for hybrid and electric vehicles using drive cycle data. Energies, 15(11):4005, 2022.

[17] Yuejiu Zheng, Minggao Ouyang, Xuebing Han, Languang Lu, and Jianqiu Li. Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles. Journal of Power Sources, 377:161–188, 2018.

[18] Yinjiao Xing, Wei He, Michael Pecht, and Kwok Leung Tsui. State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures. Applied Energy, 113:106–115, 2014.

[19] Paulo Kemper and Dongsuk Kum. Extended single particle model of li-ion batteries towards high current applications. In IEEE Veh. Power Propuls. Conf. VPPC 2013, pages 1–6, 2013.

[20] Tanvir R Tanim, Christopher D Rahn, and Chao-Yang Wang. A temperature dependent, single particle, lithium ion cell model including electrolyte diffusion. J Dyn Syst Meas, 137(1), 2015.

[21] Karen E Thomas, John Newman, and Robert M Darling. Mathematical modeling of lithium batteries. In Advances in Lithium-Ion Batteries, pages 345–392. Springer, 2002.

[22] Ulrike Krewer, Fridolin Röder, Eranda Harinath, Richard D Braatz, Benjamin Bedürftig, and Rolf Findeisen. Dynamic models of li-ion batteries for diagnosis and operation: a review and perspective. Journal of the electrochemical society, 165(16):A3656, 2018.

[23] Marc Doyle. Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymer/Insertion Cell. Journal of The Electrochemical Society, 140(6):1526, 1993.

[24] WB Gu and CY Wang. Thermal-electrochemical modeling of battery systems. Journal of The Electrochemical Society, 147(8):2910, 2000.

[25] P Ramadass, Bala Haran, Parthasarathy M Gomadam, Ralph White, and Branko N Popov. Development of first principles capacity fade model for li-ion cells. Journal of the Electrochemical Society, 151(2):A196, 2004.

[26] Pankaj Arora, Marc Doyle, and Ralph E White. Mathematical modeling of the lithium deposition overcharge reaction in lithium-ion batteries using carbon-based negative electrodes. Journal of The Electrochemical Society, 146(10):3543, 1999.

[27] Joel C Forman, Scott J Moura, Jeffrey L Stein, and Hosam K Fathy. Genetic identification and fisher identifiability analysis of the doyle–fuller–newman model from experimental cycling of a lifepo4 cell. Journal of Power Sources, 210:263–275, 2012.

[28] Mehrdad Mastali Majdabadi, Siamak Farhad, Mohammad Farkhondeh, Roydon A Fraser, and Michael Fowler. Simplified electrochemical multi-particle model for lifepo4 cathodes in lithium-ion batteries. Journal of Power Sources, 275:633–643, 2015.

[29] Huazhen Fang, Yebin Wang, Zafer Sahinoglu, Toshihiro Wada, and Satoshi Hara. State of charge estimation for lithium-ion batteries: An adaptive approach. Control Engineering Practice, 25:45–54, 2014.

[30] Adam Smiley and Gregory L Plett. An adaptive physics-based reduced-order model of an aged lithium-ion cell, selected using an interacting multiple-model kalman filter. Journal of Energy Storage, 19:120–134, 2018.

[31] Xiaosong Hu, Shengbo Li, and Huei Peng. A comparative study of equivalent circuit models for li-ion batteries. Journal of Power Sources, 198:359–367, 2012.

[32] Xinfan Lin, Youngki Kim, Shankar Mohan, Jason B Siegel, and Anna G Stefanopoulou. Modeling and estimation for advanced battery management. Annual Review of Control, Robotics, and Autonomous Systems, 2:393–426, 2019.

[33] Hamed Hossein Afshari, Mina Attari, Ryan Ahmed, Ali Delbari, Saeid Habibi, and Tina Shoa. Reliable state of charge and state of health estimation using the smooth variable structure filter. Control Engineering Practice, 77:1–14, August 2018.

[34] Zuolu Wang, Guojin Feng, Dong Zhen, Fengshou Gu, and Andrew Ball. A review on online state of charge and state of health estimation for lithium-ion batteries in electric vehicles. Energy Reports, 7:5141–5161, November 2021.

[35] Rui Xiong, Linlin Li, and Jinpeng Tian. Towards a smarter battery management system: A critical review on battery state of health monitoring methods. Journal of Power Sources, 405:18–29, November 2018.

[36] Jan Philipp Schmidt and Ellen Ivers-Tiffée. Pulse-fitting–a novel method for the evaluation of pulse measurements, demonstrated for the low frequency behavior of lithium-ion cells. Journal of Power Sources, 315:316–323, 2016.

[37] Ping Shen, Minggao Ouyang, Languang Lu, Jianqiu Li, and Xuning Feng. The co-estimation of state of charge, state of health, and state of function for lithium-ion batteries in electric vehicles. IEEE Transactions on vehicular technology, 67(1):92–103, 2017.

[38] Hongwen He, Rui Xiong, and Jinxin Fan. Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach. energies, 4(4):582–598, 2011.

[39] Chris Gould, Jiabin Wang, Dave Stone, and Martin Foster. Ev/hev li-ion battery modelling and state-of-function determination. In International Symposium on Power Electronics Power Electronics, Electrical Drives, Automation and Motion, pages 353–358. IEEE, 2012.

[40] Cheng Zhang, Walid Allafi, Quang Dinh, Pedro Ascencio, and James Marco. Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. Energy, 142:678–688, 2018.

[41] Abel Sancarlos, Morgan Cameron, Andreas Abel, Elias Cueto, Jean-Louis Duval, and Francisco Chinesta. From rom of electrochemistry to ai-based battery digital and hybrid twin. Archives of Computational Methods in Engineering, 28(3):979–1015, 2021.

[42] Prashant Shrivastava, Tey Kok Soon, Mohd Yamani Idna Bin Idris, and Saad Mekhilef. Overview of model-based online state-of-charge estimation using kalman filter family for lithium-ion batteries. Renewable and Sustainable Energy Reviews, 113:109233, 2019.

[43] Caihao Weng, Jing Sun, and Huei Peng. An open-circuit-voltage model of lithium-ion batteries for effective incremental capacity analysis. In Dynamic systems and control conference, volume 56123, page V001T05A002. American Society of Mechanical Engineers, 2013.

[44] Yuan Zou, Xiaosong Hu, Hongmin Ma, and Shengbo Eben Li. 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.

[45] Jiamiao Xie, Xingyu Wei, Xiqiao Bo, Peng Zhang, Pengyun Chen, Wenqian Hao, and Meini Yuan. State of charge estimation of lithium-ion battery based on extended kalman filter algorithm. Frontiers in Energy Research, 11:1180881, 2023.

[46] Qizhe Lin, Xiaoqi Li, Bicheng Tu, Junwei Cao, Ming Zhang, and Jiawei Xiang. Stable and accurate estimation of soc using exogenous kalman filter for lithium-ion batteries. Sensors, 23(1):467, 2023.

[47] Bin Yao, Yongxiang Cai, Wei Liu, Yang Wang, Xin Chen, Qiangqiang Liao, Zaiguo Fu, and Zhiyuan Cheng. State-of-charge estimation for lithium-ion batteries based on modified unscented kalman filter using improved parameter identification. International Journal of Electrochemical Science, 19(5):100574, 2024.

[48] Jörg Illig. Physically based impedance modelling of lithium-ion cells, volume 27. KIT Scientific Publishing, 2014.

[49] Mohsen Derakhshan, Elham Sahraei, and Damoon Soudbakhsh. Detecting mechanical indentation from the time constants of li-ion batteries. Cell Reports Physical Science, 3(11):101102, 2022.

[50] Francesco Ciucci. Modeling electrochemical impedance spectroscopy. Current Opinion in Electrochemistry, 13:132–139, 2019.

[51] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. Statistics and computing, 14(3):199–222, 2004.

[52] Meru A Patil, Piyush Tagade, Krishnan S Hariharan, Subramanya M Kolake, Taewon Song, Taejung Yeo, and Seokgwang Doo. A novel multistage support vector machine based approach for li ion battery remaining useful life estimation. Applied energy, 159:285–297, 2015.

[53] Xuning Feng, Caihao Weng, Xiangming He, Xuebing Han, Languang Lu, Dongsheng Ren, and Minggao Ouyang. Online state-of-health estimation for li-ion battery using partial charging segment based on support vector machine. IEEE Transactions on Vehicular Technology, 68(9):8583–8592, 2019.

[54] Dickshon NT How, Mahammad A Hannan, Molla S Hossain Lipu, Khairul SM Sahari, Pin Jern Ker, and Kashem M Muttaqi. State-of-charge estimation of li-ion battery in electric vehicles: A deep neural network approach. IEEE Transactions on Industry Applications, 56(5):5565–5574, 2020.

[55] Di Zhu, Jeffrey Joseph Campbell, and Gyouho Cho. Battery voltage prediction using neural networks. In 2021 IEEE Transportation Electrification Conference & Expo (ITEC), pages 807–812. IEEE, 2021.

[56] Rajendra P Joshi, Jesse Eickholt, Liling Li, Marco Fornari, Veronica Barone, and Juan E Peralta. Machine learning the voltage of electrode materials in metal-ion batteries. ACS Appl. Mater. Interfaces, 11(20):18494–18503, 2019.

[57] Shuangqi Li, Hongwen He, Chang Su, and Pengfei Zhao. Data driven battery modeling and management method with aging phenomenon considered. Appl. Energy, 275:115340, 2020.

[58] Carlos Vidal, Pawel Malysz, Mina Naguib, Ali Emadi, and Phillip J Kollmeyer. Estimating battery state of charge using recurrent and non-recurrent neural networks. Journal of Energy Storage, 47:103660, 2022.

[59] Ruxiu Zhao, Phillip J Kollmeyer, Robert D Lorenz, and Thomas M Jahns. A compact unified methodology via a recurrent neural network for accurate modeling of lithium-ion battery voltage and state-of-charge. In 2017 IEEE Energy Conversion Congress and Exposition (ECCE), pages 5234–5241. IEEE, 2017.

[60] Der-Tsai Lee, Shaw-Ji Shiah, Chien-Ming Lee, and Ying-Chung Wang. State-of-charge estimation for electric scooters by using learning mechanisms. IEEE transactions on Vehicular Technology, 56(2):544–556, 2007.

[61] Ephrem Chemali, Phillip J Kollmeyer, Matthias Preindl, and Ali Emadi. State-of-charge estimation of li-ion batteries using deep neural networks: A machine learning approach. Journal of Power Sources, 400:242–255, 2018.

[62] Xiaosong Hu, Shengbo Eben Li, and Yalian Yang. Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation in Electric Vehicles. IEEE Transactions on Transportation Electrification, 2(2):140–149, June 2016.

[63] Hao Tu, Scott Moura, and Huazhen Fang. Integrating electrochemical modeling with machine learning for lithium-ion batteries. In 2021 American Control Conference (ACC), pages 4401–4407, 2021.

[64] Todd R Ferguson and Martin Z Bazant. Phase transformation dynamics in porous battery electrodes. Electrochimica Acta, 146:89–97, 2014.

[65] Zahra Nozarijouybari and Hosam K. Fathy. Machine learning for battery systems applications: Progress, challenges, and opportunities. J. Power Sources, 601:234272, May 2024.

[66] Tahmineh Raoofi and Melih Yildiz. Comprehensive review of battery state estimation strategies using machine learning for battery management systems of aircraft propulsion batteries. Journal of Energy Storage, 59:106486, 2023.

[67] Elias Galiounas, Tom G Tranter, Rhodri E Owen, James B Robinson, Paul R Shearing, and Dan JL Brett. Battery state-of-charge estimation using machine learning analysis of ultrasonic signatures. Energy and AI, 10:100188, 2022.

[68] Hao Tu, Scott Moura, Yebin Wang, and Huazhen Fang. Integrating physics-based modeling with machine learning for lithium-ion batteries. Applied Energy, 329:120289, 2023.

[69] Zhenhua Cui, Le Kang, Liwei Li, Licheng Wang, and Kai Wang. A combined state-of-charge estimation method for lithium-ion battery using an improved bgru network and ukf. Energy, 259:124933, 2022.

[70] Shaosen Su, Wei Li, Jianhui Mou, Akhil Garg, Liang Gao, and Jie Liu. A hybrid battery equivalent circuit model, deep learning, and transfer learning for battery state monitoring. IEEE Transactions on Transportation Electrification, 9(1):1113–1127, 2022.

[71] Qian-Kun Wang, Yi-Jun He, Jia-Ni Shen, Zi-Feng Ma, and Guo-Bin Zhong. A unified modeling framework for lithium-ion batteries: An artificial neural network based thermal coupled equivalent circuit model approach. Energy, 138:118–132, 2017.

[72] Manashita Borah, Shida Jiang, Junzhe Shi, and Scott Moura. Nonlinear fractional dynamics integrated physics-informed neural network model for lifepo 4 batteries in electric vehicles. In 2024 American Control Conference (ACC), pages 1429–1434. IEEE, 2024.

[73] TO Ting, Ka Lok Man, Eng Gee Lim, and Mark Leach. Tuning of kalman filter parameters via genetic algorithm for state-of-charge estimation in battery management system. The Scientific World Journal, 2014(1):176052, 2014.

[74] Chunsheng Wang, Ripeng Li, Yuan Cao, and Mutian Li. A hybrid model for state of charge estimation of lithium-ion batteries utilizing improved adaptive extended kalman filter and long short-term memory neural network. Journal of Power Sources, 620:235272, 2024.

[75] Manashita Borah, Qiao Wang, Scott Moura, Dirk Uwe Sauer, and Weihan Li. Synergizing physics and machine learning for advanced battery management. Communications Engineering, 3(1):134, 2024.

[76] Saehong Park, Dong Zhang, and Scott Moura. Hybrid electrochemical modeling with recurrent neural networks for li-ion batteries. In 2017 American control conference (ACC), pages 3777–3782. IEEE, 2017.

[77] Fei Feng, Sangli Teng, Kailong Liu, Jiale Xie, Yi Xie, Bo Liu, and Kexin Li. Co-estimation of lithium-ion battery state of charge and state of temperature based on a hybrid electrochemical-thermal-neural-network model. Journal of Power Sources, 455:227935, 2020.

[78] Tobias Hofmann, Jacob Hamar, Marcel Rogge, Christoph Zoerr, Simon Erhard, and Jan Philipp Schmidt. Physics-informed neural networks for state of health estimation in lithium-ion batteries. Journal of the Electrochemical Society, 170(9):090524, 2023.

[79] Chenyu Xue, Bo Jiang, Jiangong Zhu, Xuezhe Wei, and Haifeng Dai. An enhanced single-particle model using a physics-informed neural network considering electrolyte dynamics for lithium-ion batteries. Batteries, 9(10):511, 2023.

[80] Damoon Soudbakhsh, Anuradha M Annaswamy, Yan Wang, Steven L Brunton, Joseph Gaudio, Heather Hussain, Draguna Vrabie, Jan Drgona, and Dimitar Filev. Data-driven control: Theory and applications. In ACC'2023, pages 1922–1939. IEEE, 2023.

[81] Alfredo Vellido, José David Martín-Guerrero, and Paulo JG Lisboa. Making machine learning models interpretable. In ESANN, volume 12, pages 163–172. Bruges, Belgium, 2012.

[82] Or Biran and Courtenay Cotton. Explanation and justification in machine learning: A survey. In IJCAI-17 workshop on explainable AI (XAI), volume 8, pages 8–13, 2017.

[83] Simon Vollert, Martin Atzmueller, and Andreas Theissler. Interpretable machine learning: A brief survey from the predictive maintenance perspective. In 2021 26th IEEE international conference on emerging technologies and factory automation (ETFA), pages 01–08. IEEE, 2021.

[84] Jichen Zhu, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G Michael Youngblood. Explainable ai for designers: A human-centered perspective on mixed-initiative co-creation. In 2018 IEEE Conference on Computational Intelligence and Games (CIG), pages 1–8. IEEE, 2018.

[85] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information fusion, 58:82–115, 2020.

[86] Jer-Nan Juang and Richard S Pappa. Effects of noise on modal parameters identified by the eigensystem realization algorithm. J Guid Control Dyn, 9(3):294–303, 1986.

[87] Juan M Caicedo. Practical guidelines for the natural excitation technique (NExT) and the eigensystem realization algorithm (ERA) for modal identification using ambient vibration. Exp. Tech., 35(4):52–58, 2011.

[88] Jer-Nan Juang, Minh Phan, Lucas G Horta, and Richard W Longman. Identification of observer/Kalman filter Markov parameters-Theory and experiments. J Guid Control Dyn, 16(2):320–329, 1993.

[89] M. Phan, L. G. Horta, J.-N. Juang, and R. W. Longman. Improvement of Observer/Kalman Filter Identification (OKID) by Residual Whitening. Journal of Vibration and Acoustics, 117(2):232–239, April 1995.

[90] Tony F Chan. An improved algorithm for computing the singular value decomposition. ACM Trans. Math. Software, 8(1):72–83, 1982.

[91] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech., 656:5–28, 2010.

[92] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. PNAS, 113(15):3932–3937, 2016.

[93] Jennifer Annoni and Peter Seiler. A method to construct reduced-order parameter-varying models. INT J ROBUST NONLIN, 27(4):582–597, 2017.

[94] Kevin K Chen, Jonathan H Tu, and Clarence W Rowley. Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses. J. Nonlinear Sci., 22(6):887–915, 2012.

[95] Urban Fasel, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. Proceedings of the Royal Society A, 478(2260):20210904, 2022.

[96] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Sparse identification of nonlinear dynamics with control (SINDYc). IFAC-PapersOnLine, 49(18):710–715, 2016.

[97] Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proceedings of the Royal Society A, 474(2219):20180335, 2018.

[98] Kadierdan Kaheman, J Nathan Kutz, and Steven L Brunton. Sindy-pi: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. Proceedings of the Royal Society A, 476(2242):20200279, 2020.

[99] Renato Rodriguez, Omidreza Ahmadzadeh, Yan Wang, and Damoon Soudbakhsh. Data-driven discovery of lithium-ion battery state of charge dynamics. Journal of Dynamic Systems, Measurement, and Control, 146(1):011101, 2024.

[100] Alan A Kaptanoglu, Jared L Callaham, Aleksandr Aravkin, Christopher J Hansen, and Steven L Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. Physical Review Fluids, 6(9):094401, 2021.

[101] Kadierdan Kaheman, Steven L Brunton, and J Nathan Kutz. Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. Machine Learning: Science and Technology, 3(1):015031, 2022.

[102] Jingyi Wang, Jesús Moreira, Yankai Cao, and Bhushan Gopaluni. Time-variant digital twin modeling through the kalman-generalized sparse identification of nonlinear dynamics. In 2022 American Control Conference (ACC), pages 5217–5222. IEEE, 2022.

[103] Omidreza Ahmadzadeh, Yan Wang, and Damoon Soudbakhsh. Sparse modeling of energy storage systems in presence of noise. IFAC-PapersOnLine, 56(2):3764–3769, 2023.

[104] Niall M Mangan, J Nathan Kutz, Steven L Brunton, and Joshua L Proctor. Model selection for dynamical systems via sparse regression and information criteria. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2204):20170009, 2017.

[105] Tanvir R. Tanim, Christopher D. Rahn, and Chao-Yang Wang. State of charge estimation of a lithium ion cell based on a temperature dependent and electrolyte enhanced single particle model. Energy, 80:731–739, February 2015.

[106] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. Sci. Adv., 3(4):e1602614, 2017.

[107] Samuel Rudy, Alessandro Alla, Steven L Brunton, and J Nathan Kutz. Data-driven identification of parametric partial differential equations. SIAM J. Appl. Dyn., 18(2):643–660, 2019.

[108] Aoxue Chen and Guang Lin. Robust data-driven discovery of partial differential equations with time-dependent coefficients. arXiv preprint arXiv:2102.01432, 2021.

[109] John Newman. Fortran Programs. http://www.cchem.berkeley.edu/jsngrp /fortran.html, 2008.

[110] Stephen Boyd, Neal Parikh, and Eric Chu. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Now Publishers Inc, 2011.

[111] M Matti Maricq, Diane H Podsiadlik, and Richard E Chase. Gasoline vehicle particle size distributions: Comparison of steady state, ftp, and us06 measurements. Environ. Sci. Technol, 33(12):2007–2015, 1999.

[112] Omidreza Ahmadzadeh, Renato Rodriguez, and Damoon Soudbakhsh. Modeling of li-ion batteries for real-time analysis and control: A data-driven approach. In 2022 American Control Conference (ACC), pages 392–397. IEEE, 2022.

[113] Valentin Sulzer, Scott G Marquis, Robert Timms, Martin Robinson, and S Jon Chapman. Python battery mathematical modelling (pybamm). Journal of Open Research Software, 9(1), 2021.

[114] Chang-Hui Chen, Ferran Brosa Planella, Kieran O'regan, Dominika Gastol, W Dhammika Widanage, and Emma Kendrick. Development of experimental techniques for parameterization of multi-scale lithium-ion battery models. Journal of The Electrochemical Society, 167(8):080534, 2020.

[115] Renato Rodriguez, Omidreza Ahmadzadeh, Yan Wang, and Damoon Soudbakhsh. Discovering governing equations of li-ion batteries pertaining state of charge using input-output data. In 2023 American Control Conference (ACC). IEEE, 2023.

[116] Jorn M Reniers, Grietus Mulder, and David A Howey. Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. Journal of The Electrochemical Society, 166(14):A3189, 2019.

[117] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.

[118] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373), pages 153–158. Ieee, 2000.

[119] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter. Kalman filtering and neural networks, pages 221–280, 2001.

[120] OAR US EPA. Dynamometer Drive Schedules. https://www.epa.gov/vehicle-and-fuel-emissions-testing/dynamometer-drive-schedules, September 2015.

[121] Hongwen He, Rui Xiong, Xiaowei Zhang, Fengchun Sun, and JinXin Fan. 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(4):1461–1469, 2011.

[122] Omidreza Ahmadzadeh, Renato Rodriguez, Yan Wang, and Damoon Soudbakhsh. A physics-inspired machine learning nonlinear model of li-ion batteries. In 2023 American Control Conference (ACC), pages 3087–3092, 2023.