# High-Fidelity Data-Driven Modeling of Battery Energy Storage for Grid Applications

Erfan Kharazian, Amarsagar Reddy Ramapuram Matavalam
*Arizona State University*
Tempe, AZ, USA
{ekharazi,amar.sagar}@asu.edu

Hyungjin Choi, Ujjwol Tamrakar
*Sandia National Laboratories*
Albuquerque, NM, USA
{hchoi, utamrak}@sandia.gov

***Abstract*—Aggressive use of battery energy storage systems (BESSs) for grid services like frequency regulation can accelerate battery capacity fading when trying to achieve control objectives without accounting for degradation. To mitigate this, control strategies must integrate models that accurately capture degradation to balance battery health and optimal performance. While physics-based models offer accurate representation of these dynamics, they are often computationally intensive for real-time use. This paper proposes developing surrogate models using dynamic mode decomposition, a data-driven method that simplifies nonlinear battery behavior into equivalent linear representation from time-series data. We focus on learning multi-scale dynamics, including fast (e.g., voltage) and slow (e.g., degradation) dynamics, enabling efficient real-time control. The resulting models will enable the integration of degradation dynamics into control framework, improving techno-economic performance**

***Index Terms*—energy storage systems, energy storage control, control hardware in the loop, real-time simulation**

## I. Introduction

THE role of battery energy storage systems (BESSs) is becoming increasingly significant across a wide range of applications for flexible energy management in modern electricity grids. This includes enhancing grid reliability and resiliency, facilitating load shifting, and providing ancillary services. Understanding and applying the electrochemical dynamics of BESSs is crucial for optimizing performance and making informed decisions regarding their utilization in the grid. In this context, accurate battery models are essential for improving efficiency, reducing costs, and enhancing the reliability of energy systems. In this work, we propose a novel data-driven approach that effectively reduces computational costs while enabling the seamless integration of models into existing battery control and optimization frameworks.

Utility-scale BESSs represent substantial investments intended for long-term use. However, aggressive utilization of BESSs for grid services, such as frequency regulation, can lead to significant degradation that adversely affects their lifespan. Consequently, it is essential to consider the effects of specific grid services on battery degradation and longevity during operation. To achieve the desired techno-economic targets, it is vital to optimize grid services and controls in relation to service requirements and degradation [1]. Conservative control strategies often restrict the ability of BESSs to fulfill the technical requirements of grid services [2], particularly in applications involving grid services dispatched over short time intervals, such as frequency and voltage regulation or fast-frequency response [3], [4].

To accurately capture the degradation dynamics of BESSs, a physics-based model, such as the Doyle-Fuller-Newman (DFN) model, is typically required. This model is often defined by a set of highly non-linear differential algebraic equations (DAEs) or partial differential equations (PDEs). However, utilizing detailed electrochemical models for optimal BESS control presents significant challenges due to their complexity [5]. In contrast, empirical and reduced-order models, such as the equivalent circuit model (ECM) or energy reservoir model, can alleviate computational burdens but often lead to inaccurate analyses and are constrained to specific conditions that do not extrapolate well to real-world grid application profiles [1], [5]. Furthermore, these models require domain expertise in battery electrochemistry for their development and application, making it challenging for engineers in adjacent fields, such as power systems, to utilize them effectively. Recently, data-driven approaches, including neural network methods, have gained popularity for battery modeling and control [6], [7]. However, the performance of these models is not guaranteed due to their heuristic nature, and they lack the physical insight typically provided by analytical models because of their black-box representation.

In this paper, we propose a novel data-driven framework for constructing a surrogate linear model of complex physics-based battery models. Specifically, we leverage the extended dynamic mode decomposition algorithm to optimally learn a finite-dimensional invariant subspace using time-series battery data, where the linear models are effectively embedded to represent the original physics-based models. The proposed modeling framework offers several advantages:

- The method offers simpler, accurate models compared to nonlinear ones, facilitating efficient predictions, control, and analysis of battery dynamics;
- We build surrogate models using battery measurement or simulated data, eliminating the need for domain expertise in battery chemistry;
- Our proposed modeling approach, in contrast to typical neural network methods, enables analytical interpretability of time-series data through linear representation, rather than relying on conventional black-box techniques.

979-8-3315-8138-1

We demonstrate the effectiveness of the proposed framework for accurately predicting the dynamics of original physics-based electrochemical battery models. This framework produces computationally efficient surrogate models that capture the degradation dynamics in BESSs, facilitating their application in real-world scenarios, such as frequency regulation.

The organization of the subsequent sections is as follows: In Section II, we present our proposed data-driven modeling approach, which focuses on creating surrogate model blocks to effectively represent internal states, voltage, SoC, and capacity loss resulting from degradation. In Section III, we detail our data collection strategies, which involve simulations conducted with PyBaMM, a widely used open-source battery simulation package, by using input current profiles sourced from the New England ISO regulation market. Next, Section IV illustrates the effectiveness of our method in predicting battery dynamics and capacity loss under various operating profiles linked to the New England ISO regulation market signal. Lastly, in Section V we conclude by summarizing our proposed method and discussing potential future directions.

## II. Data-Driven Battery Linear Surrogate Model

In this section, we outline our proposed methodology for developing surrogate models for batteries. We begin by presenting the individual linear surrogate model components. Subsequently, we introduce a data-driven algorithm that employs extended dynamic mode decomposition to learn each component using time-series battery data.

### A. Building Blocks of Linear Surrogate Battery Model

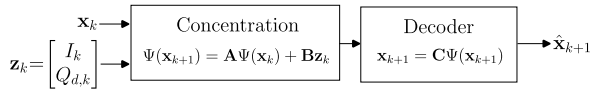
This work presents three distinct components that serve as linear surrogate models: 1) lithium-ion concentrations, 2) voltage and SoC, and 3) capacity loss due to degradation, as illustrated in Fig. 1. As illustrated in the figure, our objective is to develop linear dynamical models for each dynamic state that accurately represent the original battery dynamics.

The first component, depicted in Fig. 1a, formulates a linear model to characterize the evolution of lithium-ion concentration, which is influenced by the current control input $I$ and the capacity loss $Q_d$ in an affine manner. In this model, the battery internal states $\mathbf{x}$ are transformed into a new coordinate system denoted by $\Psi$. This system comprises basis functions of the states, represented as $\psi_i(\mathbf{x})$ for $i = 1, \ldots, N$, and is defined as follows:

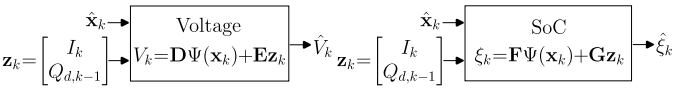
$$\Psi(\mathbf{x}) := [\psi_1(\mathbf{x}), \ldots, \psi_N(\mathbf{x})]^\top, \tag{1}$$

where the original nonlinear dynamics of the battery are effectively represented as linear system dynamics. The decoder block is responsible for converting the lifted states $\Psi$ back into the original states $\mathbf{x}$. The predicted states are denoted as $\hat{\mathbf{x}}$.

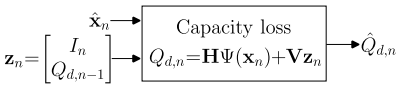
The subsequent components, illustrated in Figs. 1b and Fig. 1c, represent surrogate models for voltage, SoC, and capacity loss, respectively, with $I$ and $Q_d$ serving as inputs collected in $\mathbf{z}$. It is important to emphasize that these components do not capture dynamic behavior; instead, they establish algebraic relationships that map the predicted states $\hat{\mathbf{x}}$ and




(a) Surrogate block for lithium-ion concentration




(b) Surrogate blocks for voltage and SoC




(c) Surrogate block for capacity loss

Fig. 1: Proposed surrogate model blocks for predicting lithium-ion concentration, voltage, and capacity loss.

inputs and $\mathbf{z}$ to the corresponding predictions of voltage, SoC and capacity loss, denoted by $\hat{V}$, $\hat{\xi}$, and $\hat{Q}_d$, respectively. The learning of these mappings occurs separately, as indicated by the distinct notations for time $k$ and $n$, where $k$ represents time with a fast time step (e.g., seconds) and $n$ denotes time with a longer time step (e.g., minutes or hours). This distinction arises from the differing time scales, with voltage and SoC evolving at a significantly faster rate than capacity loss.

The linear system models in Fig. 1 offer significant advantages for integration into battery control and optimization frameworks, primarily due to their well-established theoretical foundations and practical applications of linear system theory. In the following section, we will outline an approach for identifying the lifted states and linear matrices within each surrogate model block using time-series battery data.

### B. Proposed Data-Driven Modeling Approach

In this section, we present a modeling approach based on extended dynamic mode decomposition (EDMD) [8]. The EDMD algorithm constructs a surrogate model in a lifted function space using time-series data, enabling a linear representation of original physics-based battery models. This linear model exhibits a strong connection to the Koopman operator, which analyzes the evolution of observables over time, permitting a linear characterization of nonlinear dynamics [9], [10]. By applying EDMD, we obtain a finite-dimensional approximation of the Koopman operator from observed data, which significantly broadens the applicability of the Koopman framework across various engineering disciplines.

To develop the surrogate model components outlined previously, we utilize the EDMD framework. Initially, we collect time-series data represented as $\{\mathbf{x}_k, I_k, V_k, \xi_k, Q_{d,k}\}_{k=1}^{T}$ with a time step of $\Delta t_k$. Here, $\mathbf{x}_k$ captures battery internal states, such as the ion concentrations, at time $k$, while $V_k$, $\xi_k$ and $Q_{d,k}$ correspond to the battery voltage, state of charge, and capacity loss, respectively. $I_k$ denotes the battery current control input at time $k$, and $T$ signifies the total duration of the collected time-series data. For capacity loss learning, we down-sample

the original time-series to a larger time step $\Delta t_n$ , resulting in $\{\mathbf{x}_n, I_n, V_n, Q_{d,n}\}_{n=1}^{N}$ with $N = T\Delta t_k/\Delta t_n$.

Subsequently, we address a series of regression problems formulated as follows:

$$\min_{\mathbf{A},\mathbf{B}} \sum_{k=1}^{T-1} |\Psi(\mathbf{x}_{k+1}) - \mathbf{A}\Psi(\mathbf{x}_k) - \mathbf{B}\mathbf{z}_k|^2 + \lambda\|[\mathbf{A},\mathbf{B}\|_2, \quad (2a)$$

$$\min_{\mathbf{D},\mathbf{E}} \sum_{k=1}^{T} |V_k - \mathbf{D}\Psi(\mathbf{x}_k) - \mathbf{E}\mathbf{z}_k|^2 + \lambda\|[\mathbf{D},\mathbf{E}\|_2, \quad (2b)$$

$$\min_{\mathbf{F},\mathbf{G}} \sum_{k=1}^{T} |\xi_k - \mathbf{F}\Psi(\mathbf{x}_k) - \mathbf{G}\mathbf{z}_k|^2 + \lambda\|[\mathbf{F},\mathbf{G}\|_2, \quad (2c)$$

$$\min_{\mathbf{H},\mathbf{V}} \sum_{n=1}^{N} |Q_{d,n} - \mathbf{H}\Psi(\mathbf{x}_n) - \mathbf{V}\mathbf{z}_n|^2 + \lambda\|[\mathbf{H},\mathbf{V}\|_2. \quad (2d)$$

In this context, $\mathbf{z} = [I, Q_d]^\top$. The solutions to equations (2a) through (2d) yield the linear system matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ pertinent to the concentration surrogate models. Notably, for a polynomial basis $\Psi$, $\mathbf{C}$ need not be learned, as the original states $\mathbf{x}$ are inherently embedded in $\Psi$. In this case, $\mathbf{C}$ is a constant matrix that contains zero entries, except for the first two diagonal entries, which are set to unity. Meanwhile, the resolutions of equations (2b)–(2d) provide the linear mapping matrices $\mathbf{D}$, $\mathbf{E}$, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{V}$. $\|\cdot\|$ with $\lambda > 0$ denotes L2 regularization term (i.e., Ridge) to reduce overfitting by promoting model robustness and generalizability.

## III. Battery Data Collection

We detail our data collection strategy, utilizing PyBaMM, an open-source software platform tailored for simulating physics-based battery models that encompass degradation mechanisms.

### A. Frequency Regulation Signal

We employ data from the New England regulation market. To create realistic regulation signals, we generate a synthetic regulation control signal based on the simulated 24-hour automatic generator control (AGC) setpoint data provided by NE-ISO [11]. We follow the numerical procedure described in [12], selecting representative 2-hour signals that encapsulate the majority of the dynamics associated with actual usage. These signals are classified into two distinct categories: one representing average usage and the other reflecting highly aggressive usage, as shown in Fig. 2, with standard deviation utilized as the metric for distinguishing between the average and aggressive inputs. Each signal is combined and subjected to random perturbations to generate slightly varied distributions for each day, resulting in multi-day duty cycles that consist of both average and aggressive segments.

### B. Physics-based Battery Model Simulation using PyBaMM

In this study, we used PyBaMM (Python Battery Mathematical Modelling) to simulate various electrochemical battery models [13]. Its flexibility accommodates different battery technologies, such as NCA and NMC, while accounting for thermal dynamics and degradation mechanisms. Since measurement data primarily stems from periodic cycling experiments, which may not capture all operating conditions, we use simulated data for our analysis. Nonetheless, the proposed method is applicable to experimental data if it is comprehensive and covers a wide range of conditions.

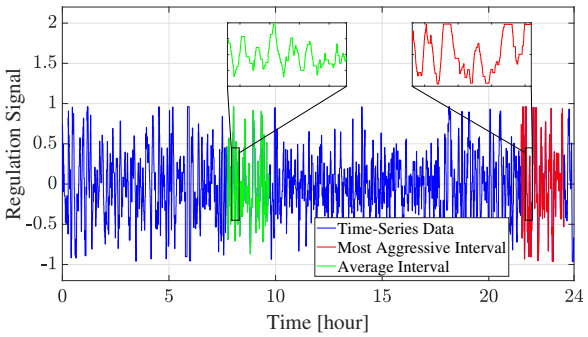



Fig. 2: Average and aggressive signals derived from the New England ISO market signal data.

Using previously developed control signals, we simulate a physics-based battery model in PyBaMM, focusing on current profiles at various C-rates. We select the single particle model with electrolyte (SPMe), which provides moderate accuracy at higher C-rates compared to the standard SPM [14]. SPMe builds upon the Doyle-Fuller-Newman (DFN) model, reducing computational complexity while accurately capturing electrochemical dynamics. By incorporating electrolyte dynamics, it enhances the SPM's fidelity, making it suitable for real-time control applications. However, the model still remains complex which may hinder its adoption in real-time control applications. For more details, see [5], [14].

We consider degradation mechanisms including SEI (Solid Electrolyte Interphase) layer growth and lithium plating as they are the major contributors to battery capacity decline in lithium ion batteries. The SEI is a passivation layer that develops on the anode surface during the cycling of lithium-ion batteries. Its ongoing growth consumes lithium ions and raises the internal resistance of the battery, which ultimately results in decreased capacity and efficiency, thereby impacting the battery's cycle life. On the other hand, lithium plating occurs when lithium metal is deposited on the anode surface rather than intercalating into the anode material. This phenomenon reduces the availability of lithium ions, leading to diminished capacity and an increased risk of safety hazards.

## IV. Numerical Demonstration

In this section, we demonstrate the proposed surrogate battery modeling approach to the SPMe model, which captures lithium-ion electrochemical dynamics along with the SEI growth and lithium plating as major degradation mechanisms.

### A. Synthetic Input Signal Scenarios

We utilize day-long NE-ISO regulation signal data to create two 2-hour segments corresponding to aggressive and average operating scenarios, respectively. These segments are combined to form 24-hour signal data, which is then randomly perturbed to generate regulation signal data for a three-week period for our study. We consider two operating conditions with different C-rates: 0.1 and 0.3 which corresponds to approximately 10 and 3.3 hours, respectively, for a complete charge or discharge.

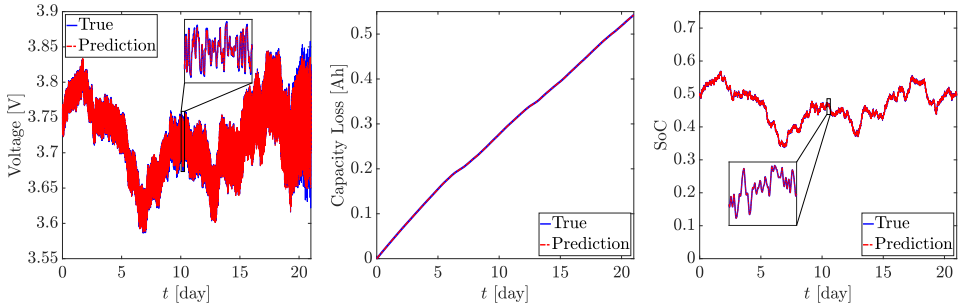



Fig. 3: Predicted voltage, capacity loss, and SoC for 0.1 C.

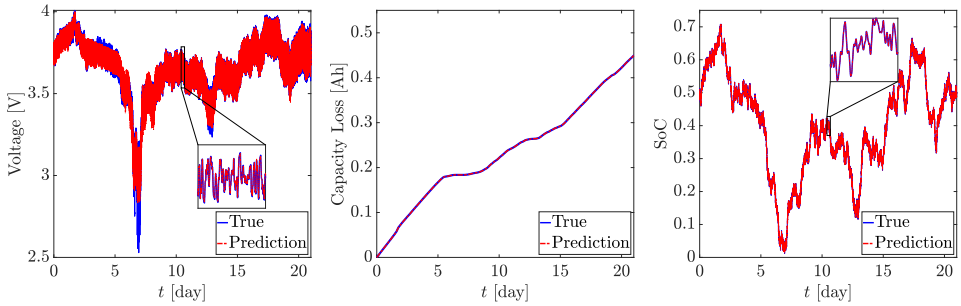



Fig. 4: Predicted voltage, capacity loss, and SoC for 0.3 C.

### *B. PyBamm Simulation*

For simulation using PyBaMM, we consider the SPMe model, which incorporates coupled degradation mechanisms, primarily focusing on SEI growth and lithium plating. The case study also considers different C-rates including 0.1 and 0.3, an initial SoC of 0.5, and upper and lower cut-off voltages of 4.2 V and 2.5 V, respectively. The nominal cell capacity is 5.0 Ah. All model parameters for SPMe and the degradation models in PyBaMM are adopted from [15].

The battery model states collected from the simulations include average negative and positive particle concentrations, battery terminal voltage, state of charge (SoC), and capacity loss. We set the simulation time step to $\Delta t_k = 4$ [s] to align with the update interval of the real-time signals in the New England ISO regulation market. For the capcity loss model, we downsample the data for the time step $\Delta t_n = 300$ [s].

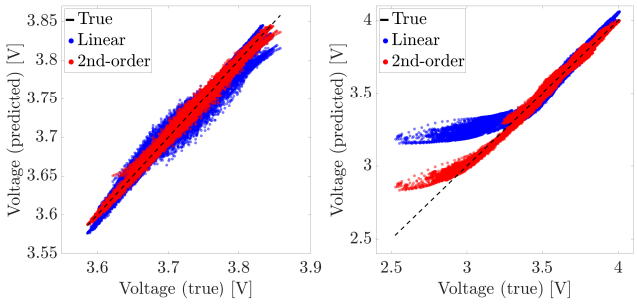
### *C. Surrogate Model Construction*

We apply the proposed approach for surrogate battery models described in Section II. The key contribution of the proposed surrogate modeling framework is its ability to predict two essential SPMe battery dynamics: the output terminal voltage (fast dynamics) and total capacity loss (slow dynamics).

Voltage, SoC, and total capacity loss predicted by our proposed method (marked in red) across both C-rates are illustrated in Figs. 3–4, as generated by the proposed surrogate model blocks (Fig. 1). In each case, basis functions are constructed using second-order polynomials. For reference, the trajectories simulated using the original SPMe model are overlaid in blue. To facilitate a detailed comparison between the true and predicted trajectories, magnified views of the voltage trajectories are provided in the insets. The close alignment between the surrogate model predictions and the SPMe simulations underscores the effectiveness of the proposed approach in faithfully capturing the dynamics of the original physics-based model.

We observe that the case with a 0.1 C-rate exhibits a greater total capacity loss of approximately 11% relative to the nominal cell capacity, compared to the case with a 0.3 C-rate, which reaches 9% by the end of the simulation. This trend is illustrated in the total capacity loss for different C-rates in Figs. 3–4. The observed behavior may be attributed to the fact that, over the same simulation period, the 0.3 C-rate scenario undergoes more significant charge-discharge cycles with larger depths of discharge. These deeper cycles could result in shorter exposure periods for the electrolyte to sustain side reactions at the anode surface, thereby partially suppressing SEI-forming side reactions and reducing degradation over time. Additionally, low C-rate operation may prolong the time spent at higher SoC, which is a key driver for capacity fade. In contrast, deeper discharge associated with 0.3 C-rate cycling could lower the average SoC, slightly mitigating degradation.

Additionally, the accuracy of the 2nd-order basis adopted in our method is validated against linear regression, as shown in Fig. 5 which illustrates scatter plots for true voltage vs. predicted voltages for both 0.1 and 0.3 C-rates. As the figure shows, under faster full charging and discharging conditions (i.e., 0.3 C-rate), the linear regression model fails to capture the voltage dynamics over a longer time horizon. In contrast, our surrogate model, which employs a 2nd-order polynomial




(a) 0.1 C-rate. (b) 0.3 C-rate.

Fig. 5: Scatter plots for true vs. predicted voltages.

TABLE I: Accuracy between Linear vs. 2nd-order basis

| C-Rate | Metric | Linear | 2nd-order | Improvement |
|---|---|---|---|---|
| 0.1 | RMSE | 0.005844 | 0.002505 | 57.1 % |
| | MAE | 0.004295 | 0.001660 | 61.3 % |
| | $R^2$ | 0.946424 | 0.989917 | 4.6 % |
| 0.3 | RMSE | 0.049367 | 0.016817 | 65.9 % |
| | MAE | 0.027624 | 0.009835 | 64.4 % |
| | $R^2$ | 0.898910 | 0.991308 | 10.3 % |

basis, successfully follows the true voltage responses. This performance is quantified using standard metrics including the root mean square error (RMSE), mean absolute error (MAE), and R-squared ($R^2$), as shown in Table I. The results show that the linear regression produces noticeably higher RMSE and MAE values and lower R-squared values compared to the 2nd-order polynomial basis. This implies that our Koopman-based surrogate model, which employs 2nd-order polynomial ridge regression, outperforms the linear regression model and demonstrates the strong generalization capability of the proposed method.

## V. Conclusions and Future Work

In this paper, we introduce a data-driven framework for battery degradation modeling using extended dynamic mode decomposition, which offers a linear representation of complex electrochemical models. This framework and its obtained preliminary results enhance computational efficiency for battery optimization and economic analysis, addressing the nonlinearity of high-fidelity models. We utilized the PyBaMM simulation package to collect time- series data from a physics-based Li-ion battery model with regulation signals from the New England ISO. Our numerical experiments show that the linear-operator-based surrogate model accurately reflects the original physics-based model and is suitable for optimization and control applications. We are also exploring its use in frequency regulation control problems with battery energy storage systems to balance performance and degradation costs.

## Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Electricity (OE), Energy Storage Division. This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan.

## References

[1] J. M. Reniers, G. Mulder, S. Ober-Blöbaum, and D. A. Howey, "Improving optimal control of grid-connected lithium-ion batteries through more accurate battery and degradation modelling," *Journal of Power Sources*, vol. 379, pp. 91–102, 2018.

[2] N. Collath, B. Tepe, S. Englberger, A. Jossen, and H. Hesse, "Aging aware operation of lithium-ion battery energy storage systems: A review," *Journal of Energy Storage*, vol. 55, 2022.

[3] D.-I. Stroe, M. Swierczynski, A.-I. Stroe, R. Laerke, P. C. Kjaer, and R. Teodorescu, "Degradation behavior of lithium-ion batteries based on lifetime models and field measured frequency regulation mission profile," *IEEE Transactions on Industry Applications*, vol. 52, no. 6, pp. 5009–5018, 2016.

[4] Y. Shi, B. Xu, Y. Tan, D. Kirschen, and B. Zhang, "Optimal battery control under cycle aging mechanisms in pay for performance settings," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2324–2339, 2019.

[5] D. M. Rosewater, D. A. Copp, T. A. Nguyen, R. H. Byrne, and S. Santoso, "Battery energy storage models for optimal control," *IEEE Access*, vol. 7, pp. 178 357–178 391, 2019.

[6] R. R. Ardeshiri, B. Balagopal, A. Alsabbagh, C. Ma, and M.-Y. Chow, "Machine learning approaches in battery management systems: State of the art: Remaining useful life and fault detection," in *2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES)*, 2020, pp. 61–66.

[7] F. Heinrich, P. Klapper, and M. Pruckner, "A comprehensive study on battery electric modeling approaches based on machine learning," *Energy Informatics*, vol. 4, no. S3, sep 2021.

[8] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the koopman operator: Extending dynamic mode decomposition," *Journal of Nonlinear Science*, vol. 25, no. 6, p. 1307–1346, Jun. 2015.

[9] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. New York: Springer-Verlag, 1994.

[10] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dynamics*, vol. 41, no. 1-3, pp. 309–325, 2005.

[11] New England ISO. New England ISO Market Operations ISO Express. [Online]. Available: https://www.iso-ne.com/markets-operations/iso-express

[12] D. Rosewater and S. Ferreira, "Development of a frequency regulation duty-cycle for standardized energy storage performance testing," *Journal of Energy Storage*, vol. 7, pp. 286–294, 2016.

[13] V. Sulzer, S. G. Marquis, R. Timms, M. Robinson, and S. J. Chapman, "Python Battery Mathematical Modelling (PyBaMM)," *Journal of Open Research Software*, vol. 9, no. 1, p. 14, 2021.

[14] S. G. Marquis, V. Sulzer, R. Timms, C. P. Please, and S. J. Chapman, "An asymptotic derivation of a single particle model with electrolyte," *Journal of The Electrochemical Society*, vol. 166, no. 15, p. A3693, Nov 2019.

[15] S. E. J. O'Kane, W. Ai, G. Madabattula, D. Alonso-Alvarez, R. Timms, V. Sulzer, J. S. Edge, B. Wu, G. J. Offer, and M. Marinescu, "Lithium-ion battery degradation: how to model it," *Physical Chemistry Chemical Physics*, vol. 24, no. 13, p. 7909–7922, 2022.