

MIT Open Access Articles

Cycle Life Prediction for Lithium-ion Batteries: Machine Learning and More

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: J. Schaeffer et al., "Cycle Life Prediction for Lithium-ion Batteries: Machine Learning and More," 2024 American Control Conference (ACC), Toronto, ON, Canada, 2024, pp. 763-768.

Published Version: 10.23919/acc60939.2024.10644790

Publisher: IEEE

Permanent Link: <https://hdl.handle.net/1721.1/157671>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: <http://creativecommons.org/licenses/by-nc-sa/4.0/>



Cycle Life Prediction for Lithium-ion Batteries: Machine Learning and More

Joachim Schaeffer^{1,†}, Giacomo Galuppini^{2,†}, Jinwook Rhyu³, Patrick A. Asinger⁴, Robin Droop⁵, Rolf Findeisen⁶, and Richard D. Braatz^{7,*}, *IEEE Fellow*

Abstract—Batteries are dynamic systems with complicated nonlinear aging, highly dependent on cell design, chemistry, manufacturing, and operational conditions. Prediction of battery cycle life and estimation of aging states is important to accelerate battery R&D, testing, and to further the understanding of how batteries degrade. Beyond testing, battery management systems rely on real-time models and onboard diagnostics and prognostics for safe operation. Estimating the state of health and remaining useful life of a battery is important to optimize performance and use resources optimally.

This tutorial begins with an overview of first-principles, machine learning, and hybrid battery models. Then, a typical pipeline for the development of interpretable machine learning models is explained and showcased for cycle life prediction from laboratory testing data. We highlight the challenges of machine learning models, motivating the incorporation of physics in hybrid modeling approaches, which are needed to decipher the aging trajectory of batteries but require more data and further work on the physics of battery degradation. The tutorial closes with a discussion on generalization and further research directions.

I. INTRODUCTION

Energy storage is vital for the transition to a sustainable future. In particular, electrochemical energy storage devices are essential for applications that require high energy- and power density, such as electric vehicles, portable electronic devices, electric vertical takeoff and landing aircraft, grid and mobile storage, and many more. This article focuses on Lithium-Ion-Batteries (LIBs), currently the most prominent and widely used type of electrochemical battery. However, many of the approaches that we present in this tutorial are also applicable to batteries beyond lithium-ion, such as sodium-ion or solid-state batteries.

LIB cells consist of two porous electrodes and a separator immersed in a liquid electrolyte [1]. Lithium ions deinter-

calate from the porous cathode, pass through the separator, and intercalate into the anode while charging. The required electric work is supplied by an external circuit. This process is reversed during discharging when the battery powers an external device. The cathode usually consists of a metal oxide, such as Lithium-Iron-Phosphate (LFP), Nickel-Cobalt-Aluminium (NCA), or Nickel-Manganese-Cobalt (NMC), while almost all LIBs use graphite as anode material, sometimes with a small fraction of silicon to improve the anode energy density [2]. Unfortunately, during cycling and also while resting, batteries degrade. The complex interplay of many different degradation mechanisms manifests in degradation modes, such as Loss of Lithium Inventory (LLI) and Loss of Active Material (LAM), leading to capacity fade and power fade [3], [4]. The challenges of understanding the interplay of degradation mechanisms led to a rise of machine learning methods for forecasting the battery cycle life (e.g., [5], [6], [7], [8]). Machine learning-based battery modeling has attracted a lot of research in the last few years. One example of successful forecasting of battery cycle life in research is the optimization of battery fast charging [9]. However, care must be taken regarding data leakage [10]. Furthermore, many machine learning models for LIBs suffer from limited generalizability due to a lack of openly available training data [11], [12]. Many different cell chemistries [13] and diverse operational conditions in real-world battery usage make it impossible to generate enough datasets to encapsulate all conditions that can occur in operation. Predicting cycle life accurately is important; however, predicting only the capacity is not enough for making informed decisions in R&D and for second life [14]. Advancing the understanding of degradation, for example, by analyzing degradation modes from diagnostic or cycling data, is needed [15]. Furthermore, obtaining optimal performance while ensuring safety is important to use resources optimally and to safeguard life and property [9], [16], [17], [18]. For this task, the battery management system estimates State-Of-Charge (SOC), State-Of-Health (SOH), and ideally also End-Of-Life (EOL), which can be framed in the context of model-based diagnostics and prognostics [19].

This tutorial is structured as follows. The next section gives an overview of state-of-the-art first-principles, machine learning, and hybrid battery modeling approaches (middle layer, Fig. 1). Subsequently, battery cycle life prediction is showcased (top layer, Fig. 1), highlighting recent improvements and limitations motivating hybrid models. The last part of the tutorial discusses the importance of diagnostic data to

¹ Technische Universität Darmstadt, Darmstadt, Germany. joachim.schaeffer@iat.tu-darmstadt.de

² University of Pavia, Pavia, PV, Italy.

giacomo.galuppini@unipv.it

³ Massachusetts Institute of Technology, Cambridge, MA, USA. jrrhyu@mit.edu

⁴ Massachusetts Institute of Technology, Cambridge, MA, USA. pasinger@mit.edu

⁵ Technische Universität München, München, Germany. r.droop@tum.de

⁶ Technische Universität Darmstadt, Darmstadt, Germany. rolf.findeisen@iat.tu-darmstadt.de

⁷ Massachusetts Institute of Technology, Cambridge, MA, USA. braatz@mit.edu

* Corresponding Author

† These two authors contributed equally.

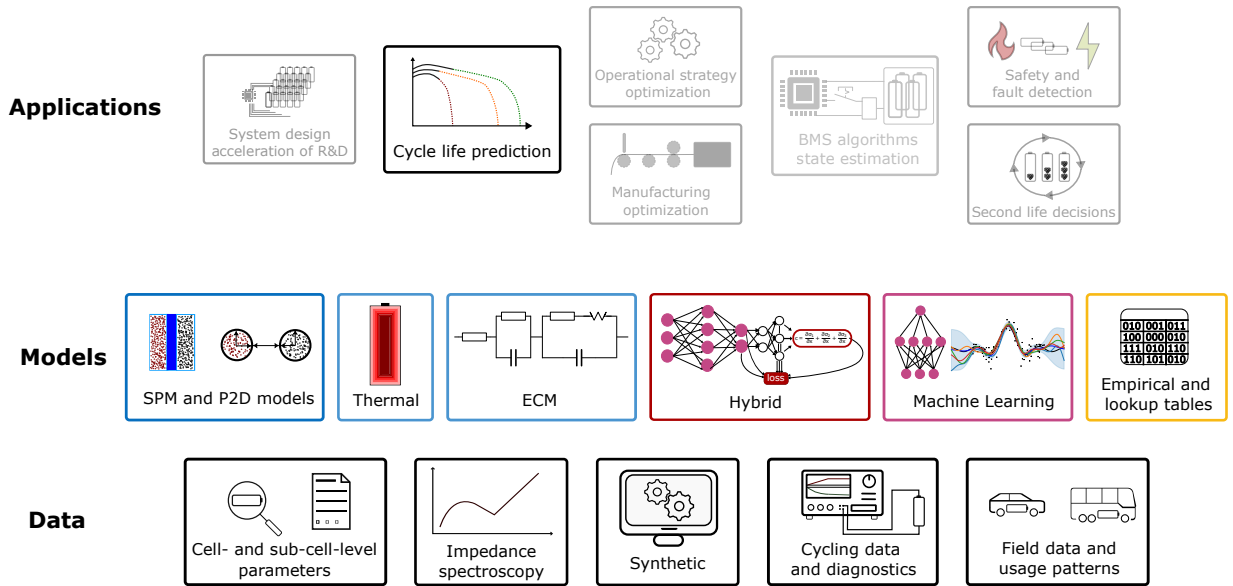


Fig. 1. Data, model, and application overview for battery system modeling.

understand degradation further and improve generalizability.

II. MODELING

First-Principles Models describe short-term battery dynamics by using laws for reaction, diffusion, and conduction processes such as Butler-Volmer kinetics [20] and coupled ion-electron transfer kinetics [21], yielding well-established electrochemical models (e.g., Porous Electrode Theory (PET) [22], Single Particle Model (SPM), kinetic Monte Carlo (kMC) model [23], and Multiphase PET (MPET) [24]). Despite their complexity, with resulting models consisting of hundreds of Differential and Algebraic Equations (DAEs), several efficient implementations are nowadays available to cope with strict computing time requirements [25]. However, first-principles modeling of degradation mechanisms and their interactions is a very active research area, with a focus on thermal effects, Solid-Electrolyte Interphase (SEI) layer growth and cracking, lithium plating, loss of active material, etc. [3], [26], [27], [28], [29]. Such physics-based models can provide a mechanistic framework (e.g., [30], [31]) to capture the influence of degradation physics on electrochemical behavior and provide good extrapolation capability, as well as transferability to a wide range of chemistries [32]. However, their design is rather challenging, as it may involve degradation modes that are not yet well understood. A further challenge is represented by parameter identifiability issues, which may be due to the structure of the model itself, as well as by the informativity of the experimental dataset available for parameter identification [33], [34], [35].

Machine Learning Models gained popularity in recent years for LIBs because there is no need to deal with complex physics and challenges related to parameter identifiability. For example, Severson et al. [5] proposed a machine learning approach for predicting cycle life by combining feature engineering with a regularized linear model. Particularly,

the features based on ΔQ (i.e., the difference of discharge capacity between two cycles over voltage, Fig. 3a) showed excellent predictive performance on a large dataset of 124 commercial LFP cells under controlled cycling conditions. This approach demonstrated the value of integrating domain knowledge into the design of features for LIBs. Building on [5], the Battery Evaluation and Early Prediction (BEEP) open-source Python library generates a structured feature matrix based on the provided cycling data [6]. Alternatively, the fused LASSO, an interpretable linear model, predicts the cycle life directly from ΔQ [8]. A different approach is to use histogram-based features (e.g., [36]) to extend the prediction of cycle life to the capacity fade trajectory, which is important for prognostics. Machine learning approaches have also successfully been applied to Electrochemical Impedance Spectroscopy (EIS) data [37], [38]. Jones et al. [38] developed a capacity forecasting model based on EIS data that can be applied to cells with different cycling histories. Furthermore, several studies have also used acoustic data as a non-invasive method. Stress waves can capture physical phenomena such as bubble generation, electrode expansion, and crack formations [39], which affect degradation and performance. Although data-driven approaches can result in accurate and rapid predictions, training purely data-driven models may require a huge amount of data, and the resulting models may yield little physical insight into degradation and often struggle with generalization beyond the intended application or cell type.

Hybrid Models combine mechanistic and machine learning models, enabling physical insights while keeping the flexibility of machine learning models [14]. As an example, the work of Ramadesigan et al. [40] estimated the cycle life by using a data-driven model to predict the evolution behaviors of effective transport and kinetic parameters where the values at each cycle were estimated from a PET model. In this case,

the short-term battery dynamics are still governed by first-principles laws, while the long-term dynamics are described by a data-driven model. While promising, this approach may suffer from practical identifiability issues during parameter estimation. Furthermore, the trend of unidentifiable parameters over the cycle number yields little insight. To overcome this issue, [33] and [34] carry out a nonlinear identifiability analysis for PET and MPET models, respectively. Once identifiability issues are properly addressed (e.g., by using Maximum-a-Posteriori estimation), the procedure can be safely applied to design hybrid models. The updated methodology is applied in [41] to regress kinetic and transport coefficients of a PET model over discharge curves of 95 NCA cells from a disassembled Tesla Model 3, all cycled at least to 80% of nominal capacity. The hybrid model developed in this way can be used to predict the battery discharge dynamics from pristine to EOL conditions. Another hybrid modeling approach is to use a data-driven model to learn the prediction error of a previously trained first-principles model [14], increasing the flexibility and prediction accuracy of the first-principles model and improving generalizability and physical insights over a purely data-driven model. A different approach is to constrain a data-driven model to fulfill specific first-principles laws. The methodology proposed in [42] regularizes the training loss of a neural network with the desired physical laws, leading to the design of a Physics-Informed Neural Network (PINN). This approach has three advantages: (1) increased generalizability, (2) reduced overfitting, and (3) reduced amount of training data. Last, to overcome the need for large training datasets, data-driven models can also be trained on a combination of real and synthetic data (e.g., [12]). The latter can be generated by previously trained first-principles models. This approach to *transfer learning* [43] has been applied successfully for battery state and health estimation [44], [45], [46]. The next section describes an application of machine learning modeling.

III. TUTORIAL: INTERPRETABLE & ACCURATE DATA-DRIVEN CYCLE LIFE PREDICTION

Ideally, the first step of building any battery modeling effort should be defining the application precisely. Next, suitable models can be selected, which then determine the data that are needed. Figure 1 is a schematic of this top-down selection and decision process. Many possible paths are suited for a specific modeling task. Unfortunately, openly available battery data are scarce [11], and battery cycling experiments are expensive. Consequently, research groups wanting to explore reasonable combinations of applications and models oftentimes have no other choice than using a dataset that is not optimally suited for the investigated application, risking data leakage and overly optimistic prediction accuracies [10].

Here, we revisit the problem of battery cycle life prediction from early laboratory cycling data as posed in [5] and restrict the main part of the tutorial to machine learning approaches. The high-level motivation of the cycle life prediction explored in [5] is to shorten laboratory battery cycling tests

and accelerate the optimization of fast charging protocols as demonstrated in [9].

A. Data Generation

Generally, data for cycle life prediction can range from electrode level to modules and system data. With respect to the controllability of the interfering variables, the data can be classified into laboratory, synthetic, and field data [47] (Fig. 1). This tutorial focuses on cell-level laboratory time series cycling data. The selection of cell chemistry(-ies) and form factor(s), and the cycling experimental design (i.e., charging current, discharging current, Depth-of-Discharge (DoD) and Depth-of-Charge (DoC), temperature, rest times, and diagnostic cycles) are crucial design choices which depend on the research question [5], [9], [15]. In turn, the number of cells needed depends on the range of the experimental conditions. The dataset of Severson et al. [5] contains 124 cells (Fig. 2), but *only* varies the fast charging protocols between cells. All cycles are full cycles, discharge conditions are constant at 4C, the cells are cycled at a constant temperature, and all cells are cylindrical LFP cells from the same manufacturer and production batch.

B. Feature Design

Features condense and transform input data, helping the model to learn the most relevant relationships between input and output data. Domain knowledge is important to extract suitable and interpretable features. The ΔQ_{100-10} feature (i.e., the difference of discharge capacity over voltage between cycles 100 and 10, Fig. 3a) was developed in [5], to capture cycle-to-cycle evolutionary changes in capacity and internal resistance, before significant capacity fade occurred (for more details see [5]). $\Delta Q \in \mathbb{R}^{124 \times 1000}$ is a wide matrix, where each row corresponds to a cell and each column to a discrete voltage between the voltage limits of the experiment.

Latent variable methods such as Partial Least Squares (PLS) or Principal Component Analysis (PCA) construct lower dimensional representations of the data and perform regression in this latent space [48], allowing to directly use ΔQ as input. Alternatively, nonlinear transformation $f: \mathbb{R}^p \rightarrow \mathbb{R}$, can be applied to yield scalar features. Examples are the variance and mean feature used in [5].

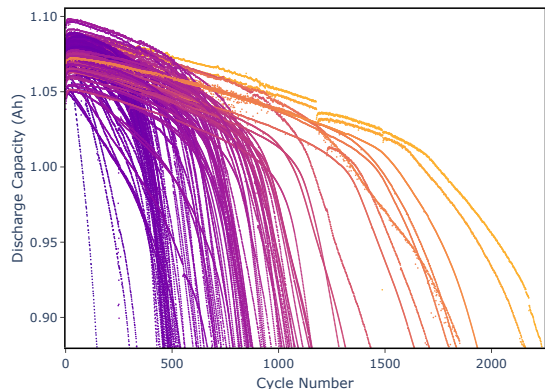


Fig. 2. Discharge capacity fade over cycle number for LFP data [5].

A different approach is to accumulate the time spent in certain operating regions (e.g., defined by current, temperature, and SOC or voltage) and use histogram features [36]. This approach is also applicable to field data. However, on the LFP dataset of [5], this approach risks yielding overly optimistic prediction performances due to implicit data leakage if the charging data is used as input [10].

Feature scaling is needed for features with different units or if the absolute value of the features is very different, and we want the model to treat features as equally important a priori. Feature transformations, $g: \mathbb{R} \rightarrow \mathbb{R}$, such as the logarithm, root, or reciprocal function, can remove skewness in feature distributions and improve the correlation of the feature with the response variable. Feature selection is used to improve model interpretability and generalization. For example, the Pearson correlation coefficient, or the maximum correlation coefficient for nonlinear correlations, can indicate the relevance of the features [49] and can also be used for feature selection.

C. Feature Selection with Regularized Regression

Consider a model of the form $\hat{y} = \Phi\theta$, where \hat{y} is the vector of predicted outputs (i.e., cycle life), Φ is the predictor aka feature matrix, and θ is the vector of model coefficients. The model is linear in the parameters, but its predictor matrix Φ can contain arbitrary nonlinear transformations of the original input¹. Assume n observations are available, and consider a total of p predictors (including all transformations of original inputs). If $n \gg p$, and the identifiability condition $\text{rank}(\Phi) = p$ holds, an Ordinary Least Squares (OLS) approach can effectively estimate the model coefficients. However, for most process data – batteries in particular – these conditions are usually hard to meet [49]. Typical process data are characterized by strong multicollinearity between model inputs, resulting in identifiability issues [50]. The feature engineering approach may also introduce further collinearity as a side effect. Regularization is an effective way to reduce identifiability issues and perform automatic selection of relevant predictors while at the same time providing an estimate of the model coefficients [50].

The Elastic Net (EN) is a regularized identification algorithm that combines the advantages of Ridge Regression (RR) and LASSO [50]. The EN estimate of the model coefficients is

$$\hat{\theta}^{\text{EN}} = \arg \min_{\theta} \|\mathbf{y} - \Phi\theta\|_2^2 + \lambda \left(\frac{1-\alpha}{2} \|\theta\|_2 + \alpha \|\theta\|_1 \right)$$

where λ and α are hyperparameters. Specifically, the 1-norm regularization associated with the LASSO allows feature selection by zeroing the parameters, i.e., discard predictors. The 2-norm regularization from RR improves feature selection by consistently pushing towards zero the parameters of all collinear features [50]. Rigorous Cross-Validation (CV) should be used to select λ and α . CV estimates the expected out-of-sample prediction error by splitting the training data

¹The EN model (Fig. 3 and [5]) uses scalar discharge features, with $\text{Var}(\Delta Q_{100-10})$ being the single most predictive feature.

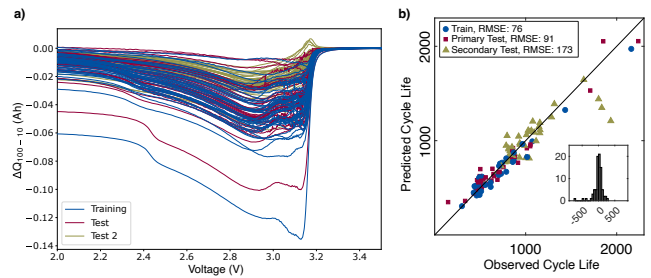


Fig. 3. a) ΔQ_{100-10} , modified from [8]. b) Predictions based on an EN model using discharge features, modified from [5]. Training (41 cells), primary (43 cells), and secondary (40 cells) test sets are color-coded.

into subsets, repeatedly training the model on all but the holdout set, and evaluating the model performance based on the holdout dataset [49], [50]. The estimated best hyperparameter values are those resulting in the model with the lowest average prediction error on the holdout data. Finally, the model is trained on all available training data using the best hyperparameter values found. Note that battery data are often characterized by an underlying group structure due to, e.g., batched production processes or in-situ operating conditions. To avoid fitting possible system biases, *grouped* CV [49] should be used, which splits data at the *group* level, ensuring that all of the data belonging to a group is assigned to the same subset.

D. Model Reliability, Predictive Performance, and Interpretability

Many metrics are available to evaluate a model’s predictive performance. Commonly used examples are Mean Square Error (MSE) and Root Mean Square Error (RMSE), R^2 , Average Absolute Percentage Error (AAPE) [50]. The use of (possibly multiple) *test* datasets is fundamental to avoid a biased evaluation of the predictive performance of the model. Moreover, this approach allows an even comparison of the performance achieved by different models (e.g., models based on different candidate predictors) [50]. The analysis of model residuals $\varepsilon = \mathbf{y} - \Phi\theta$ gives indications regarding the reliability of the model structure. Residuals resulting from a “good” model should be zero mean, normally distributed, and independent of \mathbf{y} . The analysis involves both visual inspection of the residual histogram and Q-Q plots, as well as statistical tests such as the χ^2 test [50]. Another fundamental issue is the characterization of the uncertainty affecting both the estimated parameters and the model prediction. The former aspect, which can be addressed by, for example, estimating the variance of model parameters and confidence intervals/hyperellipsoids, gives clues about the reliability of the model. An example is the case of confidence intervals spanning over positive and negative values for a parameter. In this case, the predictor can possibly be discarded from the model, as it is not even possible to significantly know the sign of its contribution to the model prediction [50]. The second aspect is instead important when decisions must be taken based on model predictions (e.g., timely replacement of cells in biomedical applications or design of safe charging

protocols). A final caveat is related to model interpretation. The fact that any particular specific predictor allows for good predictive performance does not imply the existence of a direct, causal relation between the predictor and the target variable. Therefore, any cause-effect inference inspired by the model structure or choice of predictors should be supported by domain expertise and physics-based arguments [49].

E. Battery Cycle Life Prediction Case Studies

Severson et al. [5] designed features based on ΔQ and subsequently used an EN for determining the regression coefficients. Figure 3 shows the resulting predictions on the training, primary, and secondary test data. The RMSE error is higher for the secondary test dataset; in particular, the model struggles to accurately predict the cycle life of the longer-lived cells in the secondary test dataset (Fig. 3b). The cells in the secondary test dataset were cycled approximately ten months after the cells in the training and test set and thus experienced higher calendar aging. Furthermore, few cells with a cycle life > 1200 cycles are contained in the training dataset. Recently, Schaeffer et al. [8] showcased a different approach, using the fused LASSO – a generalization of the LASSO particularly useful for discrete measurement data over a continuous domain – to directly learn from ΔQ . The resulting model is highly interpretable due to piecewise constant regression coefficients and yields similar prediction performance to the discharge model on the cells with cycle life < 1200 cycles.

IV. DISCUSSION ON GENERALIZATION AND CONCLUSION

Despite many recent advances, a key challenge of machine learning models for battery cycle life prediction is the risk of data leakage, which means that models will underperform in the target application compared to their in-house results. Careful experimental planning and model design are necessary to ensure the right features are constructed for the intended application. Furthermore, machine learning models risk overfitting to the specific experimental conditions. For example, while predictive models trained on *accelerated cycle aging data* may hasten research cycles and quality control, their generalization to real-world operation would likely be poor due to their ignorance of non-cycling aging (i.e., calendar aging). Given that many battery deployments spend most of their time idle (e.g., EVs), the influence of calendar aging must be accounted for in most practical lifetime prediction applications. The exclusion of calendar aging in lifetime predictions may lead to overly optimistic lifetime projections. Insights from machine learning models based on *high C-rate* data are limited; even the most interpretable models, such as the fused LASSO, can only support musings about underlying physics and degradation reactions. Therefore, diagnostic cycles that quantify degradation in rich and complementary ways are necessary. For example, (1) performing slow (e.g., C/25) charge and/or discharge to obtain high fidelity pseudo-OCV data to conduct differential

voltage analysis to quantify degradation modes such as LLI and cathode and anode LAM, and (2) performing pulse tests with various magnitudes and directions at multiple states of charge and temperatures to capture resistance changes. The complementary pseudo-equilibrium and dynamic characterizations help to paint a more detailed picture of cell degradation. [15] demonstrates this kind of experimental design and analysis and opens many new research avenues, such as hybrid models, to improve robustness, generalization, and physical understanding. Diverse cycling data combined with rich diagnostics is essential for improving the understanding of battery degradation and thus improving battery design and management.

To conclude, this tutorial showcases different battery models and recommends a top-down decision-making process, referring to the layer in Fig. 1, for battery modeling R&D. The tutorial section showcased battery cycle life prediction, including recent advances on the dataset in [5]. Furthermore, we discussed limitations and directions for future research to improve the understanding of degradation and, in turn, build better batteries.

REFERENCES

- [1] A. Yoshino, “The birth of the lithium-ion battery,” *Angewandte Chemie International Edition*, vol. 51, no. 24, pp. 5798–5800, 2012.
- [2] N. Nitta, F. Wu, J. T. Lee, and G. Yushin, “Li-ion battery materials: Present and future,” *Materials Today*, vol. 18, no. 5, pp. 252–264, 2015.
- [3] C. R. Birkel, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, “Degradation diagnostics for lithium ion cells,” *Journal of Power Sources*, vol. 341, pp. 373–386, 2017.
- [4] J. S. Edge, S. O’Kane, R. Prosser, N. D. Kirkaldy, A. N. Patel, A. Hales, A. Ghosh, W. Ai, J. Chen, J. Yang, S. Li, M.-C. Pang, L. Bravo Diaz, A. Tomaszewska, M. W. Marzook, K. N. Radhakrishnan, H. Wang, Y. Patel, B. Wu, and G. J. Offer, “Lithium ion battery degradation: What you need to know,” *Physical Chemistry Chemical Physics*, vol. 23, no. 14, pp. 8200–8221, 2021.
- [5] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggadakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz, “Data-driven prediction of battery cycle life before capacity degradation,” *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.
- [6] P. Herring, C. B. Gopal, M. Aykol, J. H. Montoya, A. Anapolsky, P. M. Attia, W. Gent, J. S. Hummelshøj, L. Hung, H.-K. Kwon, P. Moore, D. Schweigert, K. A. Severson, S. Suram, Z. Yang, R. D. Braatz, and B. D. Storey, “BEEP: A Python library for battery evaluation and early prediction,” *SoftwareX*, vol. 11, p. 100506, 2020.
- [7] V. Sulzer, P. Mohtat, S. Lee, J. B. Siegel, and A. G. Stefanopoulou, “Promise and challenges of a data-driven approach for battery lifetime prognostics,” in *Proceedings of the American Control Conference*, 2021, pp. 4427–4433.
- [8] J. Schaeffer, E. Lenz, W. C. Chueh, M. Z. Bazant, R. Findeisen, and R. D. Braatz, “Interpretation of high-dimensional linear regression: Effects of nullspace and regularization demonstrated on battery data,” *Computers & Chemical Engineering*, vol. 180, p. 108471, 2024.
- [9] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, and W. C. Chueh, “Closed-loop optimization of fast-charging protocols for batteries with machine learning,” *Nature*, vol. 578, no. 7795, pp. 397–402, 2020.
- [10] A. Geslin, B. van Vlijmen, X. Cui, A. Bhargava, P. A. Asinger, R. D. Braatz, and W. C. Chueh, “Selecting the appropriate features in battery lifetime predictions,” *Joule*, vol. 7, pp. P1956–1965, 2023.
- [11] L. Ward, S. Babinec, E. J. Dufek, D. A. Howey, V. Viswanathan, M. Aykol, D. A. Beck, B. Blaiszik, B.-R. Chen, G. Crabtree, S. Clark, V. De Angelis, P. Dechent, M. Dubarry, E. E. Eggleton, D. P. Finegan, I. Foster, C. B. Gopal, P. K. Herring, V. W. Hu, N. H. Paulson, Y. Preger, D. Uwe-Sauer, K. Smith, S. W. Snyder, S. Sripad, T. R.

- Tanim, and L. Teo, "Principles of the battery data genome," *Joule*, vol. 6, no. 10, pp. 2253–2271, 2022.
- [12] M. Dubarry and D. Beck, "Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis," *Journal of Power Sources*, vol. 479, p. 228806, 2020.
- [13] Y. Preger, H. M. Barkholtz, A. Fresquez, D. L. Campbell, B. W. Juba, J. Romàn-Kustas, S. R. Ferreira, and B. Chalamala, "Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions," *Journal of The Electrochemical Society*, vol. 167, no. 12, p. 120532, 2020.
- [14] M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. van Vlijmen, M. D. Berliner, M. Z. Bazant, R. D. Braatz, W. C. Chueh, and B. D. Storey, "Perspective—Combining physics and machine learning to predict battery lifetime," *Journal of The Electrochemical Society*, vol. 168, no. 3, p. 030525, 2021.
- [15] B. V. Vlijmen, P. A. Asinger, V. Lam, X. Cui, S. Sun, P. K. Herring, C. Balaji, N. Geise, H. D. Deng, H. L. Thaman, S. Dongmin, A. Trewartha, A. Anapolsky, B. D. Storey, E. William, R. D. Braatz, and W. C. Chueh, "Interpretable data-driven modeling reveals complexity of battery aging," pp. 1–19, 2023, chemRxiv preprint, <https://arxiv.org/abs/2309.00564>.
- [16] K. Liu, Y. Liu, D. Lin, A. Pei, and Y. Cui, "Materials for lithium-ion battery safety," *Science Advances*, vol. 4, no. 6, p. eaas9820, 2018.
- [17] G. Galuppini, M. D. Berliner, H. Lian, D. Zhuang, M. Z. Bazant, and R. D. Braatz, "Efficient computation of safe, fast charging protocols for multiphase lithium-ion batteries: A lithium iron phosphate case study," *Journal of Power Sources*, vol. 580, p. 233272, 2023.
- [18] —, "Efficient computation of robust, safe, fast charging protocols for lithium-ion batteries," *Control Engineering Practice*, vol. 145, p. 105856, 2024.
- [19] U. Krewer, F. Röder, E. Harinath, R. D. Braatz, B. Bedürftig, and R. Findeisen, "Review—Dynamic models of Li-ion batteries for diagnosis and operation: A review and perspective," *Journal of The Electrochemical Society*, vol. 165, no. 16, pp. A3656–A3673, 2018.
- [20] J. Butler, "Studies in heterogeneous equilibria. Part III. A kinetic theory of reversible oxidation potentials at inert electrodes," *Transactions of the Faraday Society*, vol. 19, no. March, pp. 734–739, 1924.
- [21] D. Fraggedakis, M. McEldrew, R. B. Smith, Y. Krishnan, Y. Zhang, P. Bai, W. C. Chueh, Y. Shao-Horn, and M. Z. Bazant, "Theory of coupled ion-electron transfer kinetics," *Electrochimica Acta*, vol. 367, p. 137432, 2021.
- [22] J. Newman and W. Tiedemann, "Porous-electrode theory with battery applications," *AIChE Journal*, vol. 21, no. 1, pp. 25–41, 1975.
- [23] F. Röder, R. D. Braatz, and U. Krewer, "Multi-scale simulation of heterogeneous surface film growth mechanisms in lithium-ion batteries," *Journal of The Electrochemical Society*, vol. 164, no. 11, pp. E3335–E3344, 2017.
- [24] R. B. Smith and M. Z. Bazant, "Multiphase porous electrode theory," *Journal of The Electrochemical Society*, vol. 164, no. 11, pp. E3291–E3310, 2017.
- [25] M. D. Berliner, D. A. Cogswell, M. Z. Bazant, and R. D. Braatz, "Methods—PETLION: Open-source software for millisecond-scale porous electrode theory-based lithium-ion battery simulations," *Journal of The Electrochemical Society*, vol. 168, no. 9, p. 090504, 2021.
- [26] M. B. Pinson and M. Z. Bazant, "Theory of SEI formation in rechargeable batteries: Capacity fade, accelerated aging and lifetime prediction," *Journal of The Electrochemical Society*, vol. 160, no. 2, pp. A243–A250, 2012.
- [27] J. M. Reniers, G. Mulder, and D. A. Howey, "Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries," *Journal of The Electrochemical Society*, vol. 166, no. 14, pp. A3189–A3200, 2019.
- [28] T. Gao, Y. Han, D. Fraggedakis, S. Das, T. Zhou, C.-N. Yeh, S. Xu, W. C. Chueh, J. Li, and M. Z. Bazant, "Interplay of lithium intercalation and plating on a single graphite particle," *Joule*, vol. 5, no. 2, pp. 393–414, 2021.
- [29] A. Jana, A. S. Mitra, S. Das, W. C. Chueh, M. Z. Bazant, and R. E. García, "Physics-based, reduced order degradation model of lithium-ion batteries," *Journal of Power Sources*, vol. 545, p. 231900, 2022.
- [30] M. El-Dalameh, M. Al-Greer, M. El-Dalameh, and I. Bashir, "Physics-based model informed smooth particle filter for remaining useful life prediction of lithium-ion battery," *Measurement*, vol. 214, p. 112838, 2023.
- [31] A. Downey, Y.-H. Lui, C. Hu, S. Laflamme, and S. Hu, "Physics-based prognostics of lithium-ion battery using non-linear least squares with dynamic bounds," *Reliability Engineering & System Safety*, vol. 182, pp. 1–12, 2019.
- [32] S. E. 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, pp. 7909–7922, 2022.
- [33] M. D. Berliner, H. Zhao, S. Das, M. Forsuelo, B. Jiang, W. H. Chueh, M. Z. Bazant, and R. D. Braatz, "Nonlinear identifiability analysis of the porous electrode theory model of lithium-ion batteries," *Journal of The Electrochemical Society*, vol. 168, no. 9, p. 090546, 2021.
- [34] G. Galuppini, M. D. Berliner, D. A. Cogswell, D. Zhuang, M. Z. Bazant, and R. D. Braatz, "Nonlinear identifiability analysis of multiphase porous electrode theory-based battery models: A lithium iron phosphate case study," *Journal of Power Sources*, vol. 573, p. 233009, 2023.
- [35] G. Pozzato and S. Onori, "Combining physics-based and machine learning methods to accelerate innovation in sustainable transportation and beyond: a control perspective," in *Proceedings of the American Control Conference*, 2023, pp. 640–653.
- [36] S. Greenbank and D. Howey, "Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 2965–2973, 2021.
- [37] J. Schaeffer, P. Gasper, E. Garcia-Tamayo, R. Gasper, M. Adachi, J. P. Gaviria-Cardona, S. Montoya-Bedoya, A. Bhutani, A. Schiek, R. Goodall, R. Findeisen, R. D. Braatz, and S. Engelke, "Machine learning benchmarks for the classification of equivalent circuit models from electrochemical impedance spectra," *Journal of The Electrochemical Society*, vol. 170, no. 6, p. 060512, 2023.
- [38] P. K. Jones, U. Stimming, and A. A. Lee, "Impedance-based forecasting of lithium-ion battery performance amid uneven usage," *Nature Communications*, vol. 13, no. 1, pp. 1–9, 2022.
- [39] K. Zhang, J. Yin, and Y. He, "Acoustic emission detection and analysis method for health status of lithium ion batteries," *Sensors*, vol. 21, no. 3, p. 712, 2021.
- [40] V. Ramadesigan, K. Chen, N. A. Burns, V. Boovaragavan, R. D. Braatz, and V. R. Subramanian, "Parameter estimation and capacity fade analysis of lithium-ion batteries using reformulated models," *Journal of The Electrochemical Society*, vol. 158, no. 9, pp. A1048 – A1054, 2011.
- [41] M. D. Berliner, "Simulating, Controlling, and Understanding Lithium-ion Battery Models," Ph.D. dissertation, Massachusetts Institute of Technology, 2023.
- [42] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- [43] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016.
- [44] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Applied Energy*, vol. 260, p. 114296, 2020.
- [45] C. Vidal, P. Kollmeyer, E. Chemali, and A. Emadi, "Li-ion battery state of charge estimation using long short-term memory recurrent neural network with transfer learning," in *IEEE Transportation Electrification Conference and Expo*, 2019, pp. 1–6.
- [46] N. H. Paulson, J. Kubal, L. Ward, S. Saxena, W. Lu, and S. J. Babinec, "Feature engineering for machine learning enabled early prediction of battery lifetime," *Journal of Power Sources*, vol. 527, p. 231127, 2022.
- [47] V. Sulzer, P. Mohtat, A. Aitio, S. Lee, Y. T. Yeh, F. Steinbacher, M. U. Khan, J. W. Lee, J. B. Siegel, A. G. Stefanopoulou, and D. A. Howey, "The challenge and opportunity of battery lifetime prediction from field data," *Joule*, vol. 5, no. 8, pp. 1934–1955, 2021.
- [48] J. Schaeffer and R. D. Braatz, "Latent Variable Method Demonstrator – Software for understanding multivariate data analytics algorithms," *Computers & Chemical Engineering*, vol. 167, p. 108014, 2022.
- [49] W. Sun and R. D. Braatz, "Smart process analytics for predictive modeling," *Computers & Chemical Engineering*, vol. 144, p. 107134, 2021.
- [50] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009, vol. 2.