

Perspective

Selecting the appropriate features in battery lifetime predictions

Alexis Geslin,^{1,2} Bruis van Vlijmen,^{1,2} Xiao Cui,^{1,2} Arjun Bhargava,³ Patrick A. Asinger,⁴ Richard D. Braatz,⁴ and William C. Chueh^{1,2,*}

SUMMARY

Data-driven models are being developed to predict battery lifetime because of their ability to capture complex aging phenomena. In this perspective, we demonstrate that it is critical to consider the use cases when developing prediction models. Specifically, model features need to be classified to differentiate whether or not they encode cycling conditions, which are sometimes used to artificially increase the diversity in battery lifetime. Many use cases require the prediction of cell-to-cell variability between identically cycled cells, such as production quality control. Developing models for such prediction tasks thus requires features that do not rely on cycling conditions. Using the dataset published by Severson et al. in 2019 as an example, we show that features encoding cycling conditions boost model accuracy because they predict the protocol-to-protocol variability. However, models based on these features are less transferable when deployed on identically cycled cells. Our analysis underscores the concept of using the right features for the right prediction task. We encourage researchers to consider the usage scenarios that they are developing models for and whether or not to include cycling conditions in their models in order to avoid data leakage. Equally important, benchmarking model performance should be carried out between models developed for the same use case.

INTRODUCTION

Optimizing rechargeable batteries is a challenging and necessary task as energy storage is deployed to decarbonize transportation and the electricity grid. The battery design space is large, spanning chemistry, architecture, manufacturing processes, and usage conditions. Besides, battery lifetime is long, and its evaluation is time and resource consuming, generally lasting several months to years, even under accelerated testing conditions. Over the past few decades, there have been substantial efforts to shorten battery lifetime evaluation, such as developing accelerated testing protocols (typically under elevated temperatures)^{1,2} and electrochemical methods (such as high-precision coulometry).³ Physics-based models^{4–7} and semi-empirical models^{8,9} have also been widely used to capture battery degradation trajectories. More recently, substantial progress has been made in data-driven approaches for battery lifetime prediction, typically involving machine learning (ML) methods on large datasets.^{10,11}

There are many use cases for battery lifetime prediction ML models, and the model development is specific to each use case. In [Figure 1](#), we provide an overview of

CONTEXT & SCALE

Machine learning plays a critical role in accelerating materials research and optimizing battery performance. In recent years, scientists and engineers have developed accurate data-driven models to predict the lifetime of Li-ion batteries. The versatility of such models and their ability to shorten battery testing have drawn substantial attention from industry and academia. Feature engineering has been instrumental in achieving promising model performance. In this perspective paper, we encourage researchers to determine whether they should include features that encode cycling conditions depending on their use case.

In model training datasets, battery cell aging is related to both intrinsic variability between cells and induced variability between cycling conditions. We demonstrate that for certain applications, such as production quality control, models should only use features that capture cell-to-cell variability. In these cases, features that encode information on cycling conditions should be avoided. Through several analyses, we show that when this constraint is not upheld, model performance is artificially inflated. Using the right features for the right task is essential to building data-driven models suitable for real use cases.

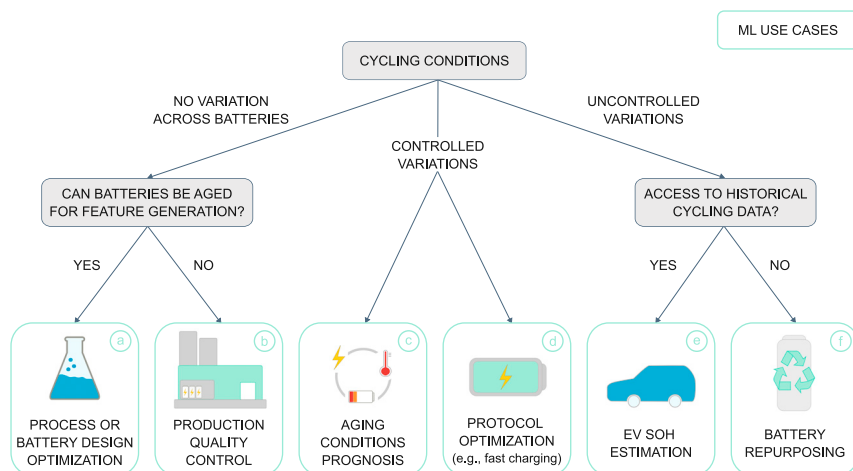


Figure 1. Machine learning algorithms for battery research can be deployed for various use cases
This flowchart illustrates the correspondence between available data and use cases.

several typical use cases and classify them based on whether the cycling conditions are varied when carrying out the cycling experiments. For example, a battery engineer who aims to optimize the cell design will cycle batteries identically for a fair comparison (Figure 1 scenario a).^{12,13} Alternatively, a model sorting short-lasting batteries out of a production line needs to rely on minimal testing as batteries cannot be cycled more than a few times (Figure 1 scenario b).¹⁴ In another scenario, a battery engineer aims to determine the impact of usage conditions, such as the depth of discharge and the charging current, on battery lifetime.^{1,15,16} For such a use case, an experiment with intentionally varied aging conditions is needed to build a prognosis model (Figure 1 scenario c). Such a model could be further integrated into more complex architectures to optimize over a large protocol parameter space (Figure 1 scenario d).^{17,18} In another use case, an electric vehicle (EV) engineer aims to integrate a prediction model in the battery management system (BMS) to estimate the battery state of health (SOH).^{10,19,20} Contrary to the previous scenario, cells are cycled variably during driving, and the BMS can access historical data for predictions (Figure 1 scenario e). Lastly, for battery repurposing, models do not have access to historical cycling data (Figure 1 scenario f).²¹ These use cases are diverse and involve cycling conditions that are either constant or variable.

Differences in battery aging trajectories are due to both cell-to-cell variability and differences in external aging factors, such as cycling rate and temperature. On the one hand, several studies quantified the intrinsic cell-to-cell variability.^{14,22–26} On the other hand, varying the cycling conditions is an artificial yet convenient way to enhance the diversity of the dataset and decrease the number of cells that need to be tested. This approach of diversifying the dataset via varied cycling conditions is analogous to accelerating aging via elevated temperature cycling, with the goal being to decrease the resources (number of cells and time) required.

In 2019, Severson et al.¹¹ demonstrated that ML, combined with a large dataset, is effective for predicting battery lifetime by employing data-driven feature engineering. The authors achieved accurate early lifetime predictions for 124 commercial lithium iron phosphate (LFP)/graphite cells using observations from the first 100 aging cycles. The cells were charged under different fast-charging protocols, but the discharge protocol was identical for all cells. Thus, the charging data explicitly encoded the cycling conditions, whereas the discharging data (for a given cycle)

¹Department of Materials Science & Engineering, Stanford University, Stanford, CA 94305, USA

²Applied Energy Division, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

³Toyota Research Institute, Los Altos, CA 94022, USA

⁴Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*Correspondence: wchueh@stanford.edu

<https://doi.org/10.1016/j.joule.2023.07.021>

did not. Following this work, several battery datasets have been published.^{25,27–31} Features inspired by Severson's work have been used to create models that can be transferred to different datasets^{32–34} or different chemistries.³⁵ Other feature engineering methods^{36,37} and more complex ML algorithms, such as neural networks, have also been deployed for lifetime and SOH estimation.^{20,38–40} These complex architectures can be used to improve lifetime prediction accuracy⁴¹ or to enable predictions with fewer aging cycles.⁴² Both of these objectives have been used as metrics to benchmark model performance. Although ML approaches for battery lifetime predictions are powerful, they also have pitfalls.⁴³ Among them are “non-legitimate features,”⁴⁴ which are those linked or correlated with the outcome. In the ML community, the term “data leakage” is used when information about the target is contained in the training data but may not be accessible in real testing conditions. This results in the model performance being artificially inflated during training but then deteriorating when the trained model is deployed for real use cases.

In this perspective, we analyze the Severson dataset to demonstrate that the community needs to pay attention to which features to use, depending on the usage scenario. Using features developed by Greenbank and Howey³⁷ extracted on (1) the charging data, (2) the discharging data, and (3) the entire charge-discharge data, we show that models relying on features based on the charging data have substantially better prediction accuracies than models using features from the discharging data. This is because these former models directly capture protocol-to-protocol variations. In fact, we show that models using no aging data at all and only the cycling conditions give good performance. However, when deployed on identically cycled cells, models based on the charging data do not maintain the same level of prediction accuracy. More generally, we show that feeding information about the aging conditions into a lifetime prediction ML model will bias the model to learn the protocol-to-protocol variations instead of learning the cell-to-cell variations. Thus, for use cases aiming to detect cell-to-cell variability, features encoding cycling conditions are inappropriate and should be avoided to prevent data leakage.

METHODOLOGY

In the Severson dataset, the authors artificially amplified the variability between cells by changing the charging protocols (two constant current [CC] steps varied across cells, followed by a CC, constant voltage [CCCV] step). As a result, the cycle life (defined as the cycle number when the capacity reaches 80% of nominal value) varied between 148 and 2,237 cycles. The discharge conditions, on the other hand, were kept constant for all cells, providing a common diagnostic across cycling conditions. In [Figure 2](#), we define a feature classification scheme, specific to this dataset, reflecting how these features are derived:

- “Class 0” features: solely based on cycling protocol parameters (no actual battery cycling data)
- “Class 1” features: derived from the charging data during aging, which encode charging protocols
- “Class 2” features: derived from the discharging data during aging

This classification scheme is generalizable to other battery datasets. Specifically, class 1 features are those that encode aging conditions either through cycling or through calendar aging.¹⁶ Class 2 features rely solely on regions of the cycling curves that are kept constant across all cells. This can be a diagnostic or check-up cycle,

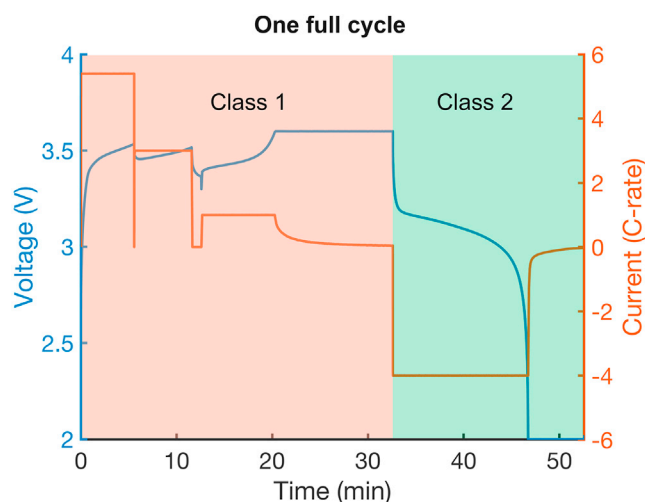


Figure 2. Voltage and current versus time profiles from one full charge and discharge cycle in the Severson dataset

Marked are the temporal regions from which class 1 and class 2 features are generated.

performed at regularly spaced intervals, to probe the state of degradation.^{16,25,45} Including a reset cycle at the start of such check-up cycles is essential to erase any explicit information about aging conditions. In the Severson dataset, there is no reset cycle between the charge and the discharge, but the CCCV at the end of charge is kept constant across all cells and mitigates explicit data leakage to the discharge data.

The purpose of this perspective is to assess the importance of the feature classes for lifetime predictions. Thus, we need features that are derived equivalently across charging and discharging curves (e.g., using the same feature extraction routine). For this reason, we employ features developed by Greenbank and Howey.³⁷ Briefly, these features quantify the fraction of time a given time series of interest (voltage, current, etc.) spends in a given window (e.g., 3.5–3.6 V) during a given time interval (10-cycle window in our study). See [supplemental information](#) for a complete explanation of the featurization. The code used for this analysis is publicly available on Github at <https://github.com/geslina>. For simplicity, models using features based on the charging data, the discharging data, or the full cycling data are referred to as “charge” (class 1), “discharge” (class 2), or “full” (class 1) model, respectively. Finally, we derive the features at different points in the early cycles of the cells. This allows us to study how the information carried by such features evolves during the early aging of the cells.

The ML task is to predict the battery lifetime and to compare the impact of incorporating features encoding cycling conditions directly or indirectly (class 0 and class 1). We employed a regularized linear model (elastic net) and an ensemble model (random forest regressor), the latter being able to capture nonlinear correlations. Most of the discussion is based on the results from the random forest model, which provides better accuracy. To prevent overfitting and to optimize model hyperparameters, a 10-fold cross-validation was carried out systematically using the GridSearchCV class from Python package SKlearn⁴⁶ (see [supplemental information](#) for details). An 80–20 train-test split was used. Because we observe that the prediction accuracy depends on train-test splits, all of the analyses conducted here are repeated over 10 random train-test splits. To evaluate the usefulness of an ML

approach, we employed a “dummy regressor” from the SKlearn library⁴⁶ as a baseline. For a given training set, such a model always predicts the same output value (in this study, the mean of the training set labels) regardless of the input features.

RESULTS AND DISCUSSION

Class 0 model is a decent baseline model without the need for any aging data

Using solely the cycling protocols and no battery aging data (e.g., only class 0 features), we show good but artificially inflated “early-prediction” model performance, compared with the dummy regressor. We stress that this is not an early-prediction model but a model that predicts cycle life as a function of cycling conditions. Table 1 reports the mean absolute percentage error (MAPE) for both the linear model and the random forest model. These results show that by solely using cycling conditions and no aging data, lifetime can be predicted with 26.4% error, a significant improvement over the 41.2% error of the dummy regressor. Importantly, this confirms that cycling conditions are predictive, even without aging data, as should be expected.^{15,47}

Class 1 features carry more information than class 2 features

Next, we compare ML performance using class 1 and class 2 features (Figure 3) as a function of the number of early cycles used as inputs. The errors of the charge and full models are equivalent and substantially better than the discharge model. Importantly, the charge models work equally well using only features from cycle 1 vs. features from cycle 150. The low prediction errors of these models are explained by the fact that the charge features encode both the protocol-to-protocol and the cell-to-cell variability. This makes the charge models even more accurate than the class 0 model. By contrast, the discharge models do not rely on any cycling conditions information. Thus, their predictive power is limited if only the first tens of cycles are employed for the featurization, as shown in Figure 3. However, their accuracy improves steadily with increasing early aging data, approaching the other two models’ accuracy after over 100 cycles, consistent with Severson et al.¹¹ Additionally, Figure S4 demonstrates that all models’ errors are approximately equivalent if cycling information is manually included into the features. These results confirm that a model that includes features that encode the intentionally varied cycling conditions will have an artificially inflated early-prediction accuracy. On the contrary, predicting lifetime without knowledge of the cycling conditions is a considerably more challenging task.

Class 2 models are more appropriate to study cell-to-cell variability

Finally, we compare charge and discharge models when deployed on a subset of cells from the dataset that were cycled identically. This special case examines how we can predict the lifetime of cells in which variation arises purely from intrinsic differences between cells, with one use case being production quality control. Although the Severson dataset does not contain many protocols with a large number

Table 1. Train and test mean absolute percentage errors (MAPEs) for the baseline models

Model	Train error		Test error	
	MAPE	(SD)	MAPE	(SD)
Dummy regressor	40.4%	(1.3%)	41.2%	(7.4%)
Class 0 (elastic net)	29.6%	(1.2%)	29.7%	(4.4%)
Class 0 (random forest)	16.9%	(1.5%)	26.4%	(4.3%)

The mean MAPE and SD are calculated across 10 train-test splits.

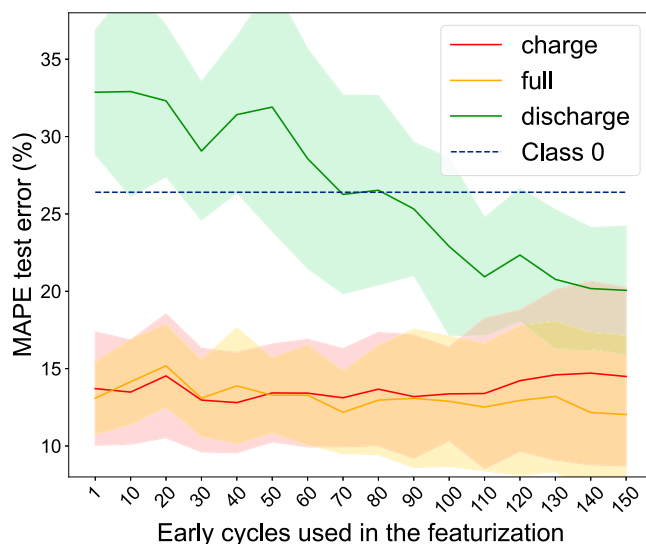


Figure 3. Evolution of the accuracies of random forest models based on how much aging data are included as inputs to the models

A class 0 model relies solely on cycling conditions and thus does not need any aging data. The shaded areas represent the values within 1 SD, calculated across the 10 train-test splits results.

of cell repeats, there are 5 protocols with at least 6 repeats (Figure S5). For each of these protocols, we train and test a charge model and discharge model on the entire dataset, excluding the identically cycled cells of that protocol. We then deploy these models on the retained cells and compare the models errors (Table 2). Protocols “5.0C-(67%)-4.0C” and “5.3C-(54%)-4.0C” (rows 1 and 4 in Table 2) show that the discharge model performs similarly if not outperforms the charge model on identically cycled cells. This trend is opposite to the model trained on the full dataset, which includes cells from many charging conditions. Class 1 features are less capable of capturing intrinsic cell-to-cell variability than class 2 features because their predictive power relies strongly on the charging protocol encoded in the data (which is not varied in this test subset of the data). Additionally, both models deployed on cells from protocols “5.6C-(19%)-4.6C” and “5.6C-(36%)-4.3C” (rows 2 and 3 in Table 2) show a significant error reduction, irrespective of the model. As mentioned earlier, the test performance is highly dependent on the train-test split. These two protocols both have a narrow distribution and a mean close to that of the entire dataset (802 cycles) as shown in the parity plots in Figure S6. Such protocols may lead to “easier” predictions because these cells are similar to the train set cells. Such protocols provide little information regarding the ability of a model to extrapolate for datapoints outside of the train set. By contrast, in the case of protocol 5.0C-(67%)-4.0C (row 1), which has a higher spread in cycle life distribution, a model relying on the cell-to-cell variability (“discharge model”) would perform better at extrapolating. Lastly, protocol “4.8C-(80%)-4.8C” (row 5 in Table 2) has the largest number of repeats but also contains 3 cells with a cycle life higher than 1,600 (more than two standard deviations [SDs] away from the mean). The errors of these models reported in Table 2 are dominated by the models’ ability to fit outliers (Figure S6). Thus, these outliers make the performance results for that protocol less representative.

We note that the Severson dataset is not optimal to study cell-to-cell variability. First, the protocol-to-protocol variability is more pronounced than the cell-to-cell variability because of the aggressive cycling conditions (Figure S4). Second, the dataset

Table 2. Errors when models are deployed on identically cycled cells

Protocol (CC1-(SOC%)-CC2)	Repeats	Model test error			Unseen identically cycled cells	
		Cycle life	Charge	Discharge	Charge	Discharge
		Mean (SD)	MAPE (SD)	MAPE (SD)	MAPE (SD)	MAPE (SD)
5.0C-(67%)-4.0C	6	1,116 (402)	13.3% (6.3%)	18.8% (4.7%)	35.9% (5.5%)	25.0% (2.4%)
5.6C-(19%)-4.6C	7	963 (138)	13.1% (5.4%)	19.2% (4.7%)	7.0% (2.2%)	10.2% (1.7%)
5.6C-(36%)-4.3C	8	931 (126)	15.3% (7.6%)	22.5% (6.0%)	8.1% (1.0%)	11.7% (2.5%)
5.3C-(54%)-4.0C	8	1,074 (121)	14.9% (7.2%)	22.7% (5.7%)	18.2% (2.4%)	18.0% (4.4%)
4.8C-(80%)-4.8C	9	1,032 (550)	17.3% (8.8%)	21.0% (7.4%)	13.4% (1.4%)	20.0% (2.3%)

5 protocols in the “Severson dataset” contain at least 6 repeats. The mean MAPE and SD are calculated across 10 train-test splits. The overall average lifetime in the entire dataset is 802 cycles with an SD of 378. Class 2 features are more appropriate to capture cell-to-cell variability.

does not contain many protocols with a large number of repeats (Figure S5). Dechent and coworkers^{48,49} developed statistical methods to estimate the number of cell repeats needed to capture the underlying cell-to-cell variability of a population. They showed that, even for simple models with 3 parameters or less, at least 9–13 repeats are needed to reliably fit the data. The ML models used here rely on many more parameters, and the Severson dataset does not contain more than 9 repeats for any protocol. This emphasizes the need to design datasets tailored for specific applications. For example, to study cell-to-cell variability, datasets in Baumhöfer et al.,¹⁴ Rumpf et al.,²² Dechent et al.,^{23,24} and Weng et al.²⁵ are more appropriate.

The analysis and results are reproducible using different sets of features

For completeness, we repeated our analysis using features originally derived by Severson et al.¹¹ Only 7 of the 20 original features were employed; not all features could be derived identically on the charge and discharge data because the charging protocol has three steps, whereas the discharge protocol has only one. Figure S7 shows similar model performance trends as observed in Figure 3 (which uses features developed by Greenbank and Howey³⁷). Errors are noticeably higher compared with Figure 3 (~+5% in error for the best-performing models), which is expected because the features are not optimized. This analysis confirms that cycling conditions information can bias models, independently of the featurization method.

CONCLUSIONS AND RECOMMENDATIONS

In recent years, the battery research community has deployed data-driven methods to predict battery lifetime. The dataset published by Severson et al.¹¹ among several others, is widely used for benchmarking model performance. In this dataset, charging conditions are varied to broaden the distribution of cycle life, whereas discharging conditions are kept constant across the cells. We demonstrate that a prediction model using only cycling conditions and no aging data can achieve decent (26.4% MAPE) predictions (compared with a dummy regressor, 41.2% MAPE) because cycling conditions strongly influence battery lifetime. More importantly, we show that models with features encoding the cycling conditions are more accurate than models that do not rely on cycling conditions. However, these models do not maintain the same level of performance when predicting intrinsic cell-to-cell variability among subsets of cells cycled identically.

Our results illustrate that ML models for lifetime prediction can be biased to learn the intentionally induced protocol-to-protocol variations in a dataset. In some use scenarios, such as prognosis predictions and cycling protocol optimization, this is

advantageous. However, in other scenarios, such as production quality control and chemistry/cell design optimization, the objective is to detect the intrinsic variability between the cells rather than how cycling conditions determine the lifetime. In such use cases, battery cells are cycled identically, making the prediction model task harder, as we show here.

We recommend that researchers carefully consider the use cases when developing lifetime prediction models and select the right features for the right prediction tasks to avoid data leakage. Importantly, benchmarking of lifetime prediction models should be carried out for models designed for the same use cases. Comparing model performance across use cases would be unfair as some models can rely on richer data encoding the protocol-to-protocol variability as demonstrated in this perspective.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.joule.2023.07.021>.

ACKNOWLEDGMENTS

This work was supported by the Toyota Research Institute through the Accelerated Materials Design and Discovery program. The authors thank Dr. Brian Storey for the insightful discussion and framing of the problem. The authors would like to thank Adrian Yao, Jianbo Wang, Vivek Lam, Devi Ganapathi, Emma Kaeli, Stephen Kang, and Shijing Sun for their help reviewing the manuscript. The authors also thank the reviewers for their suggestions to improve this paper.

AUTHOR CONTRIBUTIONS

Conceptualization, A.G., B.v.V., X.C., R.D.B., and W.C.C.; software and analysis, A.G., B.v.V., and A.B.; visualization, A.G., B.v.V., X.C., A.B., and P.A.A.; writing – original draft, A.G., B.v.V., and X.C.; writing – review and editing, A.G., B.v.V., X.C., A.B., P.A.A., R.D.B., and W.C.C.; funding acquisition, W.C.C.; supervision, W.C.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Leng, F., Tan, C.M., and Pecht, M. (2015). Effect of temperature on the aging rate of Li ion battery operating above room temperature. *Sci. Rep.* 5, 12967. <https://doi.org/10.1038/srep12967>.
- Schmalstieg, J., Käbitz, S., Ecker, M., and Sauer, D.U. (2013). From accelerated aging tests to a lifetime prediction model: analyzing lithium-ion batteries. In *World Electric Vehicle Symposium and Exhibition*, pp. 1–12. <https://doi.org/10.1109/EVS.2013.6914753>.
- Smith, A.J., Burns, J.C., Trussler, S., and Dahn, J.R. (2010). Precision measurements of the coulombic efficiency of lithium-ion batteries and of electrode materials for lithium-ion batteries. *J. Electrochem. Soc.* 157, A196–A202. <https://doi.org/10.1149/1.3268129>.
- Ramadesigan, V., Chen, K., Burns, N.A., Boovaragavan, V., Braatz, R.D., and Subramanian, V.R. (2011). Parameter estimation and capacity fade analysis of lithium-ion batteries using reformulated models. *J. Electrochem. Soc.* 158, A1048. <https://doi.org/10.1149/1.3609926>.
- Atalay, S., Sheikh, M., Mariani, A., Merla, Y., Bower, E., and Widanage, W.D. (2020). Theory of battery ageing in a lithium-ion battery: capacity fade, nonlinear ageing and lifetime prediction. *J. Power Sources* 478, 229026. <https://doi.org/10.1016/j.jpowsour.2020.229026>.
- O’Kane, S.E.J., Ai, W., Madabattula, G., Alonso-Alvarez, D., Timms, R., Sulzer, V., Edge, J.S., Wu, B., Offer, G.J., and Marinescu, M. (2022). Lithium-ion battery degradation: how to model it. *Phys. Chem. Chem. Phys.* 24, 7909–7922. <https://doi.org/10.1039/D2CP00417H>.
- Pozzato, G., Lee, S.B., and Onori, S. (2021). Modeling degradation of lithium-ion batteries for second-life applications: preliminary results. In *IEEE Conference on Control Technology and Applications (CCTA)*, pp. 826–831. <https://doi.org/10.1109/CCTA48906.2021.9659267>.
- Ecker, M., Gerschler, J.B., Vogel, J., Käbitz, S., Hust, F., Dechent, P., and Sauer, D.U. (2012). Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data. *J. Power Sources* 215, 248–257. <https://doi.org/10.1016/j.jpowsour.2012.05.012>.
- Gaspar, P., Gering, K., Dufek, E., and Smith, K. (2021). Challenging practices of algebraic battery life models through statistical validation and model identification via machine-learning. *J. Electrochem. Soc.* 168,

020502. <https://doi.org/10.1149/1945-7111/abde1>.
10. Richardson, R.R., Osborne, M.A., and Howey, D.A. (2017). Gaussian process regression for forecasting battery state of health. *J. Power Sources* 357, 209–219. <https://doi.org/10.1016/j.jpowsour.2017.05.004>.
 11. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* 4, 383–391. <https://doi.org/10.1038/s41560-019-0356-8>.
 12. Kim, S.C., Oyakhire, S.T., Athanitis, C., Wang, J., Zhang, Z., Zhang, W., Boyle, D.T., Kim, M.S., Yu, Z., Gao, X., et al. (2023). Data-driven electrolyte design for lithium metal anodes. *Proc. Natl. Acad. Sci. USA* 120, e2214357120. <https://doi.org/10.1073/pnas.2214357120>.
 13. Drakopoulos, S.X., Gholamipour-Shirazi, A., MacDonald, P., Parini, R.C., Reynolds, C.D., Burnett, D.L., Pye, B., O'Regan, K.B., Wang, G., Whitehead, T.M., et al. (2021). Formulation and manufacturing optimization of lithium-ion graphite-based electrodes via machine learning. *J. Phys. Sci.* 2, 100683. <https://doi.org/10.1016/j.xcrp.2021.100683>.
 14. Baumhöfer, T., Brühl, M., Rothgang, S., and Sauer, D.U. (2014). Production caused variation in capacity aging trend and correlation to initial cell performance. *J. Power Sources* 247, 332–338. <https://doi.org/10.1016/j.jpowsour.2013.08.108>.
 15. Ning, G., Haran, B., and Popov, B.N. (2003). Capacity fade study of lithium-ion batteries cycled at high discharge rates. *J. Power Sources* 117, 160–169. [https://doi.org/10.1016/S0378-7753\(03\)00029-6](https://doi.org/10.1016/S0378-7753(03)00029-6).
 16. Wildfeuer, L., Karger, A., Aygül, D., Wassiliadis, N., Jossen, A., and Lienkamp, M. (2023). Experimental degradation study of a commercial lithium-ion battery. *J. Power Sources* 560, 232498. <https://doi.org/10.1016/j.jpowsour.2022.232498>.
 17. Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* 578, 397–402. <https://doi.org/10.1038/s41586-020-1994-5>.
 18. Jiang, B., Gent, W.E., Mohr, F., Das, S., Berliner, M.D., Forsuelo, M., Zhao, H., Attia, P.M., Grover, A., Herring, P.K., et al. (2021). Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols. *Joule* 5, 3187–3203. <https://doi.org/10.1016/j.joule.2021.10.010>.
 19. Richardson, R.R., Birk, C.R., Osborne, M.A., and Howey, D.A. (2019). Gaussian process regression for in situ capacity estimation of lithium-ion batteries. *IEEE Trans. Ind. Inform.* 15, 127–138. <https://doi.org/10.1109/TII.2018.2794997>.
 20. Li, W., SenGupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D.U. (2021). Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *J. Power Sources* 482, 228863. <https://doi.org/10.1016/j.jpowsour.2020.228863>.
 21. Braco, E., San Martin, I., Sanchis, P., Ursúa, A., and Stroe, D. (2022). Health indicator selection for state of health estimation of second-life lithium-ion batteries under extended ageing. *J. Energy Storage* 55, 105366. <https://doi.org/10.1016/j.est.2022.105366>.
 22. Rumpf, K., Naumann, M., and Jossen, A. (2017). Experimental investigation of parametric cell-to-cell variation and correlation based on 1100 commercial lithium-ion cells. *J. Energy Storage* 14, 224–243. <https://doi.org/10.1016/j.est.2017.09.010>.
 23. Dechent, P., Willenberg, L., Eckert, M., and Sauer, D.U. (2021). Time series data on cycle tests on Samsung NR18650-35E. <https://doi.org/10.18154/RWTH-2021-07965>.
 24. Dechent, P., Lehner, S., and Sauer, D.U. (2021). Time series data on cycle tests on Samsung NR18650-15L1. <https://doi.org/10.18154/RWTH-2021-08740>.
 25. Weng, A., Mohtat, P., Attia, P.M., Sulzer, V., Lee, S., Less, G., and Stefanopoulou, A. (2021). Predicting the impact of formation protocols on battery lifetime immediately after manufacturing. *Joule* 5, 2971–2992. <https://doi.org/10.1016/j.joule.2021.09.015>.
 26. Zhou, Z., and Howey, D.A. (2022). Bayesian hierarchical modelling for battery lifetime early prediction. <https://doi.org/10.48550/arXiv.2211.05697>.
 27. dos Reis, G., Strange, C., Yadav, M., and Li, S. (2021). Lithium-ion battery data and where to find it. *Energy AI* 5, 100081. <https://doi.org/10.1016/j.egyai.2021.100081>.
 28. Preger, Y., Barkholtz, H.M., Fresquez, A., Campbell, D.L., Juba, B.W., Román-Kustas, J., Ferreira, S.R., and Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *J. Electrochem. Soc.* 167, 120532. <https://doi.org/10.1149/1945-7111/abae37>.
 29. Juarez-Robles, D., Jeevarajan, J.A., and Mukherjee, P.P. (2020). Degradation-safety analytics in lithium-ion cells: Part I. Aging under charge/discharge cycling. *J. Electrochem. Soc.* 167, 160510. <https://doi.org/10.1149/1945-7111/abc8c0>.
 30. Juarez-Robles, D., Azam, S., Jeevarajan, J.A., and Mukherjee, P.P. (2021). Degradation-safety analytics in lithium-ion cells and modules: Part III. Aging and safety of pouch format cells. *J. Electrochem. Soc.* 168, 110501. <https://doi.org/10.1149/1945-7111/ac30af>.
 31. van Vlijmen, B., Lam, V., Asinger, P.A., Cui, X., Ganapathi, D., Sun, S., Herring, P.K., Gopal, C.B., Geise, N., Deng, H.D., et al. (2023). Interpretable data-driven modeling reveals complexity of battery aging. Preprint at ChemRxiv. <https://doi.org/10.26434/chemrxiv-2023-zdl2n-v2>.
 32. Roman, D., Saxena, S., Robu, V., Pecht, M., and Flynn, D. (2021). Machine learning pipeline for battery state-of-health estimation. *Nat. Mach. Intell.* 3, 447–456. <https://doi.org/10.1038/s42256-021-00312-3>.
 33. Jones, P.K., Stimming, U., and Lee, A.A. (2022). Impedance-based forecasting of lithium-ion battery performance amid uneven usage. *Nat. Commun.* 13, 4806. <https://doi.org/10.1038/s41467-022-32422-w>.
 34. Zhu, J., Wang, Y., Huang, Y., Bhushan Gopaluni, R.B., Cao, Y., Heere, M., Mühlbauer, M.J., Mereacre, L., Dai, H., Liu, X., et al. (2022). Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nat. Commun.* 13, 2261. <https://doi.org/10.1038/s41467-022-29837-w>.
 35. Paulson, N.H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S.J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. *J. Power Sources* 527, 231127. <https://doi.org/10.1016/j.jpowsour.2022.231127>.
 36. Zhang, Y., Tang, Q., Zhang, Y., Wang, J., Stimming, U., and Lee, A.A. (2020). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nat. Commun.* 11, 1706. <https://doi.org/10.1038/s41467-020-15235-7>.
 37. Greenbank, S., and Howey, D. (2022). Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life. *IEEE Trans. Ind. Inform.* 18, 2965–2973. <https://doi.org/10.1109/TII.2021.3106593>.
 38. Xu, F., Yang, F., Fei, Z., Huang, Z., and Tsui, K.-L. (2021). Life prediction of lithium-ion batteries based on stacked denoising autoencoders. *Reliab. Eng. Syst. Saf.* 208, 107396. <https://doi.org/10.1016/j.res.2020.107396>.
 39. Li, W., Zhang, H., van Vlijmen, B., Dechent, P., and Sauer, D.U. (2022). Forecasting battery capacity and power degradation with multi-task learning. *Energy Storage Mater.* 53, 453–466. <https://doi.org/10.1016/j.ensm.2022.09.013>.
 40. Su, L., Wu, M., Li, Z., and Zhang, J. (2021). Cycle life prediction of lithium-ion batteries based on data-driven methods. *eTransportation* 10, 100137. <https://doi.org/10.1016/j.etrans.2021.100137>.
 41. Fei, Z., Yang, F., Tsui, K.-L., Li, L., and Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. *Energy* 225, 120205. <https://doi.org/10.1016/j.energy.2021.120205>.
 42. Strange, C., and dos Reis, G. (2021). Prediction of future capacity and internal resistance of Li-ion cells from one cycle of input data. *Energy AI* 5, 100097. <https://doi.org/10.1016/j.egyai.2021.100097>.
 43. Kaufman, S., Rosset, S., Perlich, C., and Stitelman, O. (dec 2012). Leakage in data mining: formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data* 6, 1–21. <https://doi.org/10.1145/2382577.2382579>.
 44. Kapoor, S., and Narayanan, A. (2022). Leakage and the reproducibility crisis in ML-based science. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2207.07048>.
 45. Eldesoky, A., Bauer, M., Bond, T., Kowalski, N., Corsten, J., Rathore, D., Dressler, R., and Dahn, J.R. (2022). Long-term study on the impact of

- depth of discharge, C-rate, voltage, and temperature on the lifetime of single-crystal NMC811/artificial graphite pouch cells. *J. Electrochem. Soc.* *169*, 100531. <https://doi.org/10.1149/1945-7111/ac99a6>.
46. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* *12*, 2825–2830.
47. Celik, B., Sandt, R., dos Santos, L.C.P., and Spatschek, R. (2022). Prediction of battery cycle life using early-cycle data, machine learning and data management. *Batteries* *8*, 266. <https://doi.org/10.3390/batteries8120266>.
48. Dechent, P., Greenbank, S., Hildenbrand, F., Jbabdi, S., Sauer, D.U., and Howey, D.A. (2021). Estimation of Li-ion degradation test sample sizes required to understand cell-to-cell variability. *Batteries Supercaps* *4*, 1821–1829. <https://doi.org/10.1002/batt.202100148>.
49. Strange, C., Allerhand, M., Dechent, P., and dos Reis, G. (2022). Automatic method for the estimation of Li-ion degradation test sample sizes required to understand cell-to-cell variability. *Energy AI* *9*, 100174. <https://doi.org/10.1016/j.egyai.2022.100174>.