Journal of The
Electrochemical Society

Long-Term Calendar Aging Across Commercial Lithium-Ion Cell Chemistries—Modeling and Early Prediction

To cite this article: Xiaofan Cui *et al* 2025 *J. Electrochem. Soc.* **172** 060521

Journal of The Electrochemical Society, 2025 **172** 060521
1945-7111/2025/172(6)/060521/15/$40.00

Long-Term Calendar Aging Across Commercial Lithium-Ion Cell Chemistries—Modeling and Early Prediction

Xiaofan Cui,[1,=] Florian Stroebl,[1,=] Maitri Uppaluri,[1,2,=] Vivek N. Lam,[2,3,=] William C. Chueh,[1,2,3,z] and Simona Onori[1,2,z]

[1]*Energy Sciences and Engineering, Stanford University, Stanford, CA, United States of America*
[2]*Applied Energy Division, SLAC National Accelerator Laboratory, Menlo Park, CA, United States of America*
[3]*Department of Materials Science and Engineering, Stanford University, Stanford, CA, United States of America*

Lithium-ion batteries are widely used in electric vehicles, stationary grid systems, and various consumer electronics today. In these cases, batteries remain idle for extended periods and are only in operation for a relatively small portion of their lifetime. Batteries degrade while at rest due to calendar aging induced mechanisms, and quantifying the effect of calendar aging is crucial for accurate battery capacity loss predictions. In this study, we formulate, validate and compare three different models to predict lithium-ion battery capacity fade from a long-term calendar aging dataset [V. N. Lam, X. Cui, F. Stroebl, M. Uppaluri, S. Onori, and W. C. Chueh, Joule, 9, 146 (2025)], namely semi-empirical, symbolic regression and data-driven models, are compared and validated using this dataset. We investigate each model's ability to interpolate and extrapolate aging behavior across different storage conditions and cell types testing their transferability. We observe that the three models accurately predict the end-of-life (EOL) of a cell. However, the data-driven model demonstrates a superior ability to predict the capacity fade trajectory for an unseen cell. Moreover, the data-driven model exhibits reasonable accuracy when predicting the EOL and capacity trajectories across various cell-types.
 [DOI: 10.1149/1945-7111/adde16]

Manuscript submitted December 2, 2024; revised manuscript received May 1, 2025. Published June 18, 2025.

List of Symbols

Category	Symbols and Descriptions
	General Parameters and Variables
Celltype	Cell type, categorical variable
SOC	State of Charge [%], categorical variable
T	Ambient Temperature [°C]
t	Time [weeks]
Q	Absolute discharge capacity [Ah]
EOL	End of life
	Vectors and Matrices
$\boldsymbol{T}$	Vector of temperatures
r	Length of $\boldsymbol{T}$
$\boldsymbol{t}$	Vector of t
$\boldsymbol{x}$	Aging Condition vector
$\boldsymbol{t}_{RPT}$	Vector of RPT times
$\boldsymbol{y}$	Aging Trajectory vector
	Aging Parameters and Measurements
Q_{rel}	Relative Capacity [-]
$t_{RPT,i}$	Time of the i-th RPT [weeks]
i	Counting index for RPTs
n	Number of RPTs
y_i	Scalar aging feature at i-th RPT
$Q_{rel,i}$	Relative Capacity at i-th RPT
Q_i	Absolute discharge capacity at i-th RPT
t_{EOL}	Time [weeks] of the RPT before $Q_{rel} < 0.9$ is surpassed
	Dataset Definitions
D	Data set used in this publication
D_{train}	Training subset of D
D_{test}	Testing subset of D
X	Space of all tested aging conditions
$X_{train/test}$	Input data set for training/testing
Y	Space of aging features
$Y_{train/test}$	Output data set for training/testing
m	Cumulative number of all tested cells
k	Number of repetitions of each aging condition
	Model Parameters and Predictions
f_θ	Parametric model function of SE model
$\boldsymbol{\theta}$	Parameter vector of SE model
Θ	Parameter space for SE model
$\hat{\boldsymbol{\theta}}$	Estimate of $\boldsymbol{\theta}$
$\hat{y}_i$	Prediction of y_i
$\hat{\boldsymbol{y}}$	Prediction of $\boldsymbol{y}$
$\hat{Q}_{rel}(T, t; \boldsymbol{\theta})$	T- and t-dependent SE model predicting Q_{rel}
$\hat{Q}_{rel}(t; \boldsymbol{\theta})$	t-dependent base model for SR predicting Q_{rel}
$\boldsymbol{\beta}$	Local parameter vector for SR
$\hat{\boldsymbol{\beta}}$	Estimate of $\boldsymbol{\beta}$
$\boldsymbol{\beta}(\hat{T}; \boldsymbol{\gamma})$	By SR identified parametric model functions for $\beta_0(T;\gamma)$, $\beta_1(T;\gamma)$
$\boldsymbol{\gamma}$	Parameter vector for global SR model
$\hat{\boldsymbol{\gamma}}$	Estimate of $\boldsymbol{\gamma}$
θ_j	j-th entry of $\boldsymbol{\theta}$
β_j	j-th element of $\boldsymbol{\beta}$
γ_j	j-th element of $\boldsymbol{\gamma}$
j	Counting index for parameter vectors
p	Total number of parameters
q	Number of input features
	Statistical Measures and Other Concepts
$L(\boldsymbol{y}, \hat{\boldsymbol{y}})$	Loss function for parameter estimation
$\hat{t}_{EOL}$	Prediction of t_{EOL}
RMSE	Root Mean Squared Error
R^2	Coefficient of determination

Across many applications, lithium-ion batteries may spend the majority of their lifetime at rest under no load. For instance, electric vehicles (EVs) are operated 10% of their time, and are predominantly idle.[1] Similarly, in stationary battery energy storage systems utilized for peak shaving, it has been observed that the aging effects

[=]Equal Contribution.
[z]E-mail: wchueh@stanford.edu; sonori@stanford.edu

attributed to inactivity are comparable to those caused by the charge-discharge cycles during operation.[2] The degradation of a battery over time under no applied current is termed "calendar aging". The emergence of electrified transport and renewable energy systems has highlighted the need to understand and mitigate the battery's calendar aging effects to accurately predict capacity loss during the battery's lifetime across a variety of conditions and cell chemistries.

Calendar aging of a cell is significantly impacted by its storage time and conditions (such as states-of-charge (SOC) and temperature).[3,4] Calendar aging data in the literature ranges from 3 months to 5 years, as summarized in the literature survey in Lam et al., which reported over 13 years of storage data.[5] To study the effects of calendar aging for longer periods, predictive models can be used to estimate the lifetime and the capacity fade trajectory of the battery.

In literature, several models have been established to predict calendar aging in lithium-ion batteries that vary in complexity and accuracy. There are mainly three distinct models: physics-based or electrochemical models,[6–10] algebraic semi-empirical models,[3,4,11–13] and data-driven machine-learning-based models.[14–16] Physics-based models can be used to predict capacity loss from calendar aging by modeling the aging mechanisms that lead to cell degradation during calendar aging. The most common degradation mechanism is the growth of the Solid Electrolyte Interphase (SEI) layer that leads to loss of cyclable lithium and increase in cell impedance. The development of electrochemical models for calendar aging was initiated by Ploehn et al. by simulating the diffusion-limited growth of the SEI layer.[6] The kinetic of the SEI layer growth was incorporated into the diffusion-limited model to form mixed-mode calendar aging models based on the SEI layer growth.[8,9] To include the SOC and temperature dependence on calendar aging, Kamyab et al. incorporated the Tafel relationship that describes the kinetically-limited SEI growth to depend on the storage SOC and the temperature of the cell.[9] Rahimian et al. added cathode degradation and electrolyte thermal decomposition equations to existing models to capture resistance growth and trends observed at high SOCs and temperatures.[10] Existing electrochemical models make assumptions about the dominant degradation mechanisms to model calendar aging in cells, and are typically validated using data from a single cell chemistry that was calendar aged for around 2 years.[9,10] In addition, solving these models in real-time can be computationally expensive, for predicting calendar aging behavior for over a decade.

Semi-empirical models employ simple algebraic equations with functional dependencies for stress factors such as temperature and SOC, making them ideal for predicting calendar aging trajectories in lithium-ion batteries in real-time applications. These models are based on well-established principles, mainly the Arrhenius law for temperature dependence and the power-law for storage time dependence.[3,11,17] Models for the SEI layer growth were initially characterized by the semi-infinite diffusion-limited growth using a square-root dependence of time ($t^{0.5}$) on capacity loss.[6–8] Later, studies showed that the time-dependence of calendar aging deviates from the square-root relationship, and instead a power-law (t^x) constitutes a more generalized expression for time-dependence.[5,18,19] Other functions, like the sigmoidal function proposed by Gering et al. are gaining popularity to describe the time dependence on calendar aging.[20] On the other hand, the temperature dependence on calendar aging is described by the Arrhenius relationship.[17,21,22] This relationship motivates accelerated aging studies by testing at higher temperatures to speed up cell degradation and extrapolate the model to predict performance at lower temperatures.[11,23–25] However, our recent study showed that the Arrhenius law does not explain temperature dependence under certain conditions when considering long-term (over 5 years) storage.[5] Some studies have gone beyond simple semi-empirical models by utilizing symbolic regression techniques to discover different functional forms for temperature-dependent calendar aging.[26,27]

Data-driven models are increasingly being recognized for their ability to serve as surrogates in simulating complex aging mechanisms and managing a range of aging scenarios. These models, particularly in the context of calendar aging, include diverse machine learning techniques such as Random Forest,[28,29] Gaussian Process Regression,[16,30] Artificial Neural Networks,[31] and

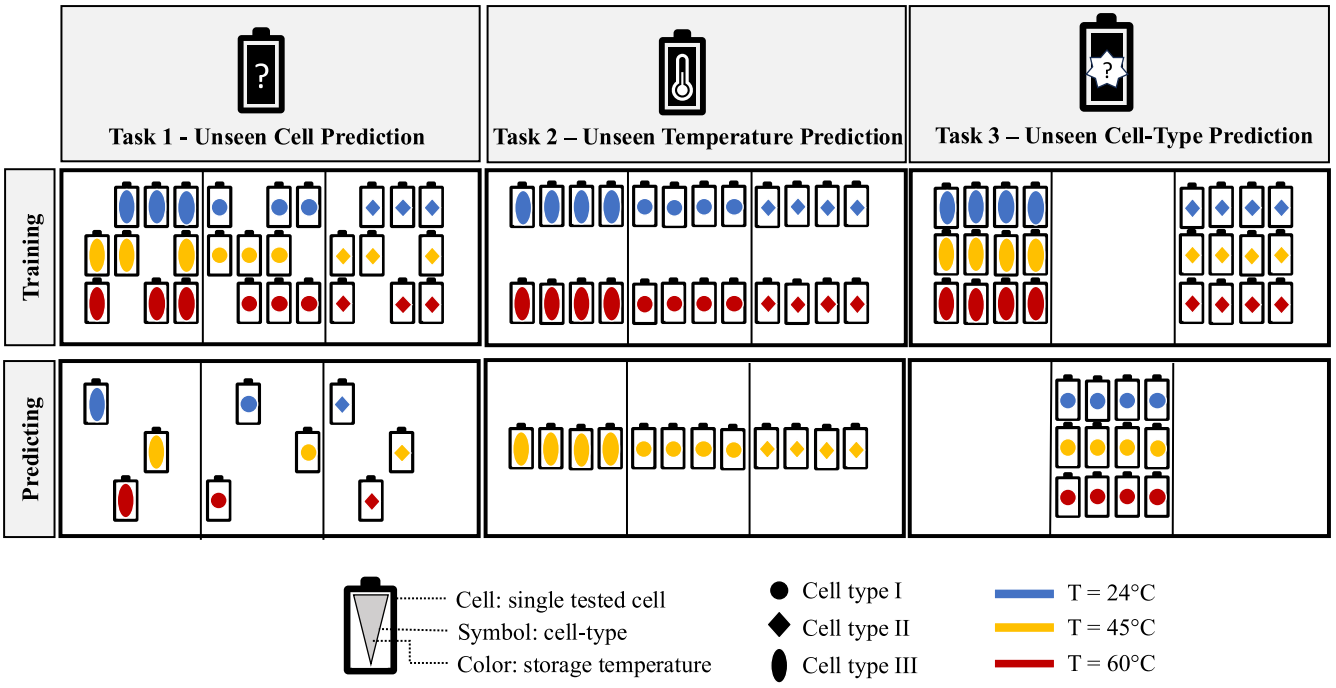

Figure 1. Early Aging Prediction Tasks considered in this paper: Task 1 - Predict the behavior of an unseen cell, Task 2 - Predict the behavior at an unseen storage temperature, Task 3 - Predict the behavior of an unseen cell type. The three tasks result in specific train/test splits. The training data is used to train Semi-Empirical (SE), Symbolic Regression (SR), and Data-driven (DD model) models for each task. Their prediction performance is evaluated by End-of-Life and trajectory-based prediction errors using the unseen testing data.

XGBoost.[15] A common trait observed with these models is that most of these algorithms are validated on short-term calendar-aging dataset. This raises concerns about whether the proposed modeling and prediction methods can be effective for long-term calendar aging. In addition, for all the aforementioned models, the transferability across cell types-characterized by form factors and chemistries-has rarely been examined. Several studies demonstrated models that were validated with datasets that consisted of multiple cell types.[15,30,31] However, since the model was trained on all cell types, its validity for an unknown cell type can also be assessed to evaluate the model's transferability.

An accurate model for calendar aging enables battery prognostics, which involves predicting a battery's remaining life and future degradation trends according to historical health data. Accurate battery prognostics can significantly reduce testing time and costs, especially for calendar aging. Predictive models are used for remaining useful life (RUL) and capacity loss trajectory predictions.[32–35] Given initial calendar-aging data - specifically, the early aging trajectory up to a defined threshold (e.g., 97% of the initial capacity) - RUL predictions estimate the remaining time until end-of-life (EOL). In contrast, capacity loss trajectory prediction forecasts the entire capacity evolution over time until EOL.

In Lam et al.,[5] we demonstrated that even for the same cell type, degradation trends can diverge under identical storage conditions due to cell-to-cell variability. Additionally, temperature dependence does not always follow clear Arrhenius behavior, and aging behavior can vary even among chemically similar cells from the same manufacturer. Given these transferability challenges, this study evaluates three primary tasks, as illustrated in Fig. 1, to assess model performance in predicting calendar aging. *Task 1* focuses on unseen cell prediction, where cells are categorized by type, storage temperature, and storage SOC, and a subset of cells under the same testing conditions are excluded from training to test model generalization. *Task 2* addresses unseen temperature prediction by withholding one temperature condition during training, training the model on the remaining conditions, and evaluating its performance on the withheld temperature. *Task 3* examines unseen cell type prediction, where one cell type is excluded from training, and the model is tested on its ability to predict aging for the withheld type. In this work, we develop three models-semi-empirical (SE), symbolic regression (SR), and data-driven (DD)-using the battery calendar aging dataset from[5]-to be assessed on their ability to predict end-of-life (EOL) and capacity trajectories across different training and testing data combinations. As discussed in Ref. 5 the dataset deviates from Arrhenius behavior and power-law relationships for temperature and time dependence. Our goal is to identify the modeling approach that best captures these deviations for long-term calendar aging predictions.

Dataset Description

The calendar aging dataset presented in[5] is used to construct and validate the models for the tasks outlined in this paper. An overview of the dataset is provided in Fig. 2. This dataset encompasses calendar aging data from 232 cells over a 13-year period, representing eight commercially available Li-ion battery cell types from five manufacturers. The cells were tested under two different SOCs (50% and 100% SOC) and four temperatures (24 °C, 45 °C, 60 °C, or 85 °C). The 85 °C data is excluded from this study due to the cells reaching EOL rapidly, offering limited benefits for early prediction. Diagnostic cycles or reference performance tests (RPTs) were performed at regular intervals during the calendar aging process. These cycles included three low-rate (C/5 charge and discharge) cycles, three high-rate (C/5 charge, discharge at 1C, 2C, or 3C depending on the cell type), a C/5 capacity check cycle to observe any damage done by high-rate cycles, and a final C/5 charge to the storage SOC (Fig. 3). The capacity metric used in this work is the discharge capacity from the third low-rate full charge-discharge cycle.

The relative capacity is determined as

$$Q_{rel,i} = \frac{Q_i}{Q_0}, \quad [1]$$

with Q_0 being the initially measured capacity, and Q_i representing the capacity measured at the ith RPT during the calendar aging test. Each aging condition was repeated with up to $k \leqslant 6$ cells for a given cell type. The combination of tested cell types, aging conditions, and number of cells is shown in Table I. For further testing details refer to Ref. 5.

Model Description

End-of-life (EOL) and capacity trajectory prediction models.— The capacity fading trajectories are divided into disjoint training and testing subsets, denoted as $D_{\text{train}} = (X_{\text{train}}, Q_{rel,\text{train}})$ and $D_{\text{test}} = (X_{\text{test}}, Q_{rel,\text{test}})$ respectively, where $X_{\text{train}}, X_{\text{test}}$ represent the input subsets with aging conditions and RPT times, and $Q_{rel,\text{train}}$, $Q_{rel,\text{test}}$ represent their corresponding relative capacity output subsets. D_{train} and D_{test} are used for model training and testing, respectively.

In this work, we evaluate the performance of three different models in predicting the EOL (set to 90% SOH) and the capacity loss trajectory. The SEM and the SR models use the whole trajectory from the training cells to determine model parameters. On the other hand, the DD model uses data from an initial portion of the trajectory, from SOH = 100% to a specified SOH threshold (SOH_{th}) as input. The DD model aims to estimate the EOL and the full capacity loss trajectory, given the initial portion of the degradation data between SOH = 100% and SOH = SOH_{th}. Unless differently specified, SOHth = 98%. $Q_{rel,train}$ represents the input portion of the capacity fade curve, which is used to train each of the models, and $Q_{rel,test}$ represents the output portion of the curve. The DD model's input corresponds to 2% of the initial capacity trajectory, from 100% SOH to 98% SOH. The corresponding output portion of the data ranges from 98% SOH to 90% SOH. The DD model is trained on both the input and the output data.

The training set is used to identify the parameters for each model. A prediction algorithm is then employed to fine-tune the model parameters, ensuring that the predicted relative capacity for the specific input subsets X_{train} aligns as closely as possible with the target values $Q_{rel,\text{train}}$. The proportion in which the dataset is split into the training and testing subsets, D_{train} and D_{test}, varies based on the task. A mathematical representation of the train and test split is described in Appendix B.

Semi-Empirical Model

Capacity loss from calendar aging can be modeled using semi-empirical models that capture the effects of stress factors such as temperature and SOC. The semi-empirical model for relative capacity as a function of storage temperature, SOC, and time is given by:

$$\hat{Q}_{rel}(T, SOC, t) = 1 - f(T, SOC, t) \quad [2]$$

The time dependence of capacity fade due to calendar aging is described by the power-law. Initially, this relationship was derived from the square-root dependence ($t^{0.5}$) of time on capacity fade caused by SEI layer growth,[21] which was later generalized into a t^b relationship.[18] To incorporate the power-law as the time dependency on capacity fade, Eq. 2 can be written as:

$$\hat{Q}_{rel}(T, SOC, t) = 1 - f_1(T, SOC) \cdot t^b \quad [3]$$

where $f(T, SOC, t) = f_1(T, SOC) \cdot t^b$, with f_1 being a nonlinear function that captures the temperature dependence of capacity $\hat{Q}_{rel}$, and b represents the power-law exponent.

The Arrhenius law is commonly used in literature to describe the temperature dependence on calendar aging, while the power-law is

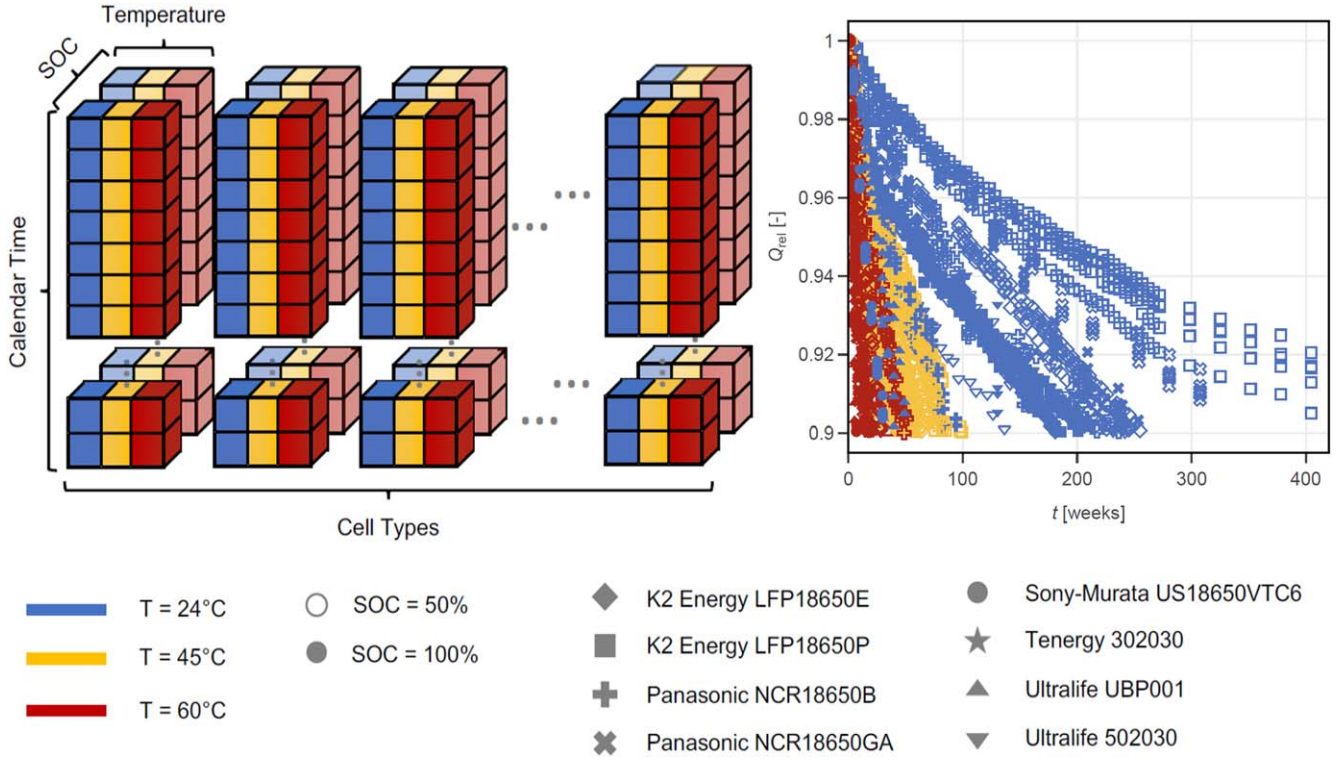

Figure 2. Calendar aging dataset. The dataset used in this study consists of calendar aging trajectories for eight different cell types, spanning three temperatures and two SOC levels. Each aging condition is replicated with up to six cells. Cell type and SOC are treated as categorical variables, while temperature and calendar time are considered continuous variables. The left part of the figure illustrates the dataset, where each cube represents data collected from multiple cells during a specific RPT. The right plot visualizes the relative capacity degradation Q_{rel}, over calendar time in weeks. From this data set, a task-specific train/test-split is constructed to evaluate model performance.

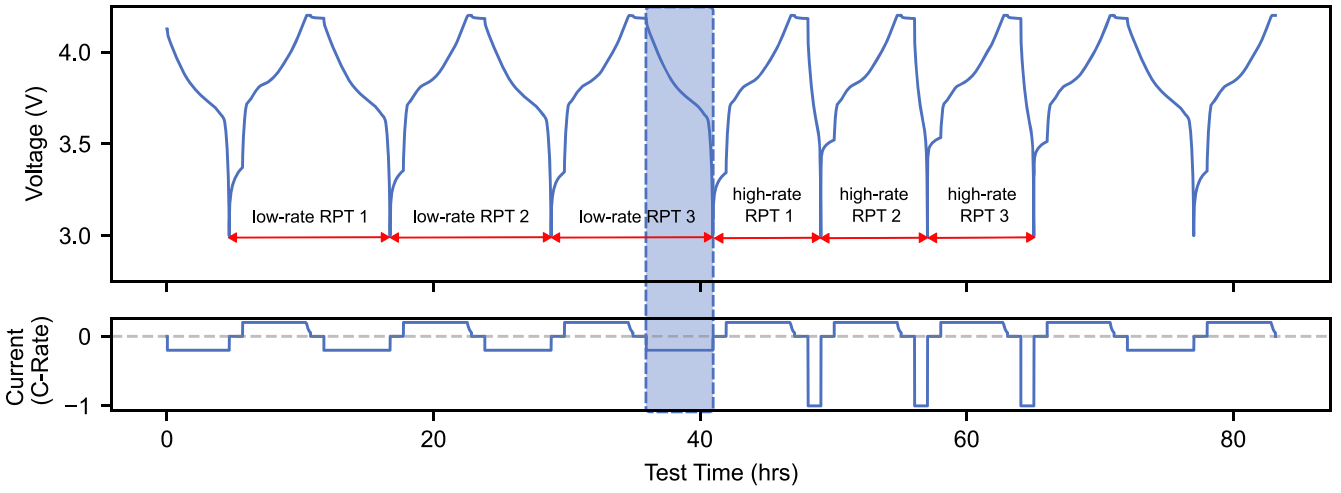

Figure 3. Current and voltage profile of the diagnostic cycles, with the highlighted region corresponding to the third low-rate RPT discharge.

used to describe the time-dependence.[11,18,21] Including this temperature dependence in the semi-empirical model results in:

$$\hat{Q}_{rel}(T, SOC, t) = 1 - f_2(SOC) \cdot a_0 \cdot exp\left(\frac{-E_a}{T}\right) \cdot t^b \quad [4]$$

where $f_1 = f_2(SOC) \cdot exp(\frac{-E_a}{T})$, a_0 represents the pre-exponent factor and E_a is the activation energy. In this work, the dataset used to build the model includes only two SOC values, which is insufficient for establishing a functional dependence on SOC. As a result, in the semi-empirical models developed here, SOC and battery cell type are treated as categorical variables. Therefore, a parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ is identified for each combination of cell type and SOC individually, capturing the specific characteristics of the aging behavior for each combination. Equation 4 can be rewritten in terms of the parameter vector $\boldsymbol{\theta}$:

$$\hat{Q}_{rel}(T, t; \boldsymbol{\theta}) = 1 - \theta_1 \cdot \exp\left(\frac{-\theta_2}{T}\right) \cdot t^{\theta_3} \quad [5]$$

where θ_1 is the pre-exponent factor, θ_2 corresponds to the activation energy in the Arrhenius law, and θ_3 is the power-law exponent. We assume that θ_1, θ_2 and θ_3 are independent of temperature and SOC.

Table I. Calendar Aging Data Set - aging conditions tested for each cell type, along with the number of cells tested under each condition.

Cell type	SOC [%]	T [°C]	#Cells
Ultralife 502030	50	24	4
	100	24	5
Ultralife UBP001	100	24	6
		60	6
Tenergy 302030	100	60	5
Sony-Murata US18650VTC6	100	24	6
		60	6
Panasonic NCR18650GA	50	24	6
		60	5
	100	24	6
		60	6
Panasonic NCR18650B	50	24	5
		45	3
		60	3
	100	24	5
		45	3
		60	3
K2 Energy LFP18650P	50	24	5
		45	5
		60	5
	100	24	5
		45	5
		60	5
K2 Energy LFP18650E	50	24	5
		45	5
		60	5
	100	24	4
		45	4
		60	5

The parameters are identified using the maximum likelihood estimator (MLE) based on experimental data. For Gaussian distributed errors, the MLE simplifies to a least-squares optimization problem, as shown in.[36]

Symbolic Regression Model

Symbolic Regression aims to find an algebraic expression that best fits a given data set. Unlike traditional regression methods, which rely on predefined model structures, SR seeks to uncover the underlying mathematical relationships within the data without being constrained by a specific model form. The objective in SR is to identify an expression, $g(x)$, that best fits a set of data points based on some error metric, without being limited to any predefined structure or form of $g(x)$.[37,38] In this work, we focus on identifying the best algebraic models that describe the relationship between temperature T and time t. We assume the model takes the form:

$$\hat{Q}_{rel}(t, T; \boldsymbol{\gamma}) = 1 - \beta_0(T; \boldsymbol{\gamma}) \cdot t^{\beta_1(T;\gamma)}, \quad [6]$$

with unknown functional forms for β_0 and β_1 expressed in terms of T and a generic vector of parameters $\boldsymbol{\gamma}$ to be identified.

In this work, we follow the procedure in[19] to calibrate (6):

1. Define a baseline model using the commonly adopted power-law calendar aging model:

$$\hat{Q}_{rel}(t; \hat{\boldsymbol{\beta}}) = 1 - \beta_0 \cdot t^{\beta_1}, \quad [7]$$

where $\hat{Q}_{rel}(t; \hat{\boldsymbol{\beta}})$ is the relative capacity of a battery cell stored for t weeks at ambient temperature T. $\hat{\boldsymbol{\beta}} = (\beta_0, \beta_1)$ are referred to as local model parameters, that are fitted to the data from a specific cell.

2. Identify (β_0, β_1) of the baseline model for each aging condition and cell type separately. For the baseline model, *SOC*, *Cell Type*, and *T* are considered categorical variables, resulting in a set of model parameters associated with each combination of *Cell Type*, *SOC*, and *T*. β_0 and β_1 determined for each cell can be included in a single vector $\hat{\boldsymbol{\beta}}$.
3. Utilize the identified parameters $\hat{\boldsymbol{\beta}}$ to determine a functional dependency of these parameters on T using SRs. In this work, we employ the PySR package to determine the functional forms of β_0 and β_1, given as $\beta_0 = \beta_0(T;\boldsymbol{\gamma})$ and $\beta_1 = \beta_1(T;\boldsymbol{\gamma})$.This results in a specific T- and t-dependent model function for each combination of categorical variables *Cell Type* and *SOC*. The model is called the global model, where the scalar local parameters of the base model $\hat{\boldsymbol{\beta}}$ are replaced by the prior identified equations:The models obtained for each *Cell Type* and *SOC* are summarized in Appendix A, where Table A.1 shows the expressions for β_0 and β_1 obtained for different cells tested using Task 1. In this work and for this dataset, depending on the test cell in Task 1, $\boldsymbol{\gamma}$ can be a vector containing either two values (γ_0, γ_1) three values $(\gamma_0, \gamma_1, \gamma_2)$ or four values $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$.

Data-Driven Model

In this work, we develop a *trajectory-to-trajectory model* to predict the relative capacity trajectory based on initial relative capacity data. Our model is data-driven, taking the initial portion of the capacity fade trajectory as input and predicting the remainder of the degradation trajectory until the end-of-life as the output. To tackle this trajectory-to-trajectory prediction challenge, we propose a data-driven prediction framework that employs an encoder-decoder architecture.

Various types of data-driven encoder-decoder models exist, including those based on Recurrent Neural Networks (RNN),[39] Long Short-Term Memory (LSTM) networks,[40] and Gate Recurrent Unit (GRU) networks,[41] among others. Notably, LSTM-based models have shown success in the trajectory-to-trajectory prediction of cycling aging.[42] In our study, the machine-learning architecture employs both a GRU network encoder and a GRU network decoder, as illustrated in Fig. 4.

We define a black-box data-driven (DD) model $\hat{Q}_{rel} = h(Q_{rel,early}, t)$, where $Q_{rel,early}$ represents the early-aging segment of the relative capacity data provided to the model, and $\hat{Q}_{rel}$ is the predicted remaining relative capacity trajectory. The variable t represents the calendar time. Each cell aging trajectory consists of N data pairs (t_1, q_1), (t_2, q_2), $\cdots$, (t_i, q_i), $\cdots$, (t_N, q_N), where N is the total number of data points for a given cell, t_i is the ith calendar time, and q_i is the ith relative capacity at the ith data point. The input to the model ($Q_{rel,early}$) consists of the first k data points, i.e., from (t_1, q_1), (t_2, q_2), ... (t_k, q_k), while the output ($\hat{Q}_{rel}$) is the predicted relative capacity from $(t_{k+1}, q_{k+1})...(t_N, q_N)$. The threshold for selecting the input data up to (t_k, q_k) can be determined by one of two methods: the *equal time threshold*, where all cells stop testing at the same time, or the *equal capacity threshold* method, where all cells stop testing after reaching the same relative capacity q_k. In the latter approach, the testing time is calculated by summing the time required for each cell to reach the capacity threshold q_k.

Implementing an equal-time threshold is commonly used in practice due to its simplicity and ease of planning and scheduling

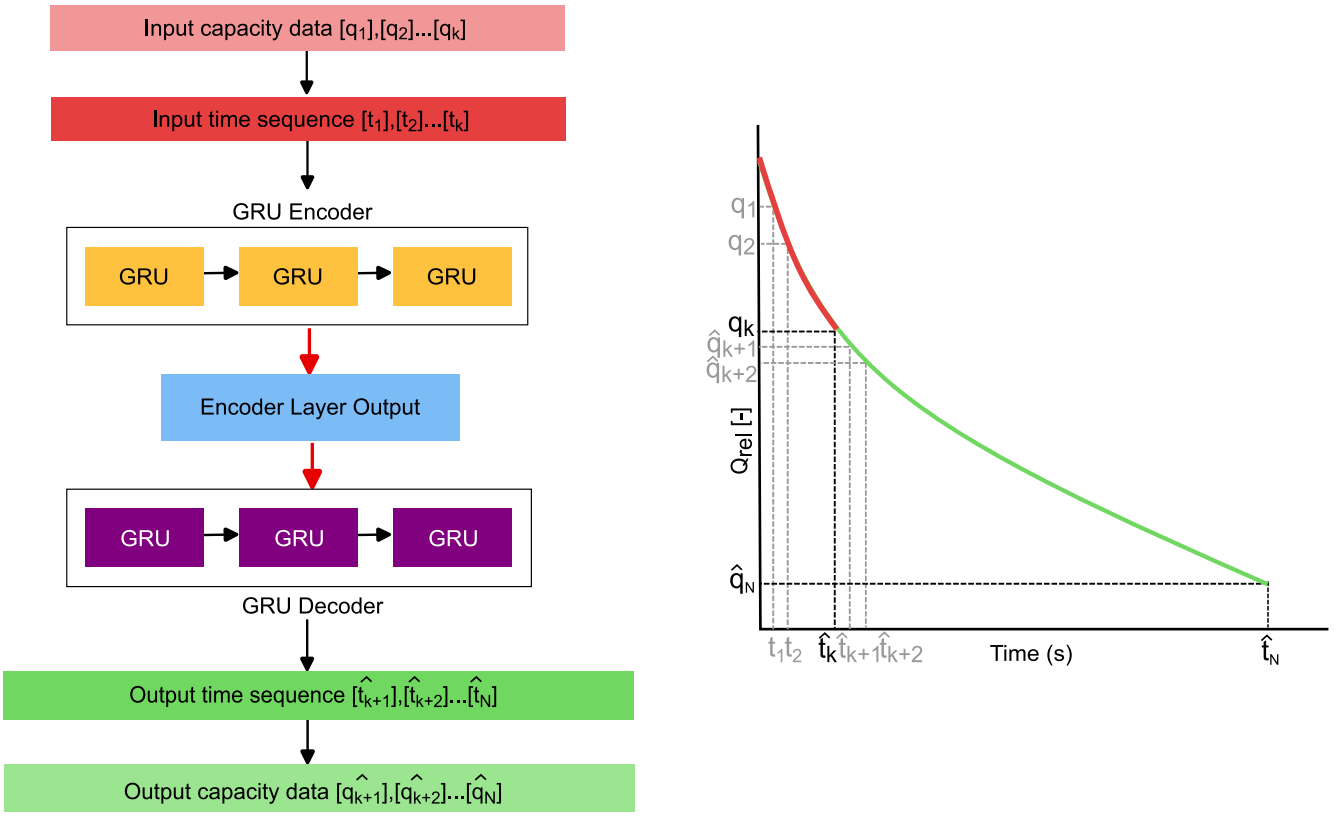

Figure 4. Schematic of the data-driven model framework, using the GRU-based encoder-decoder network. The time-sequence $[t_1, t_2, \ldots, t_k]$ is obtained from the input capacity data $[q_1, q_2, \ldots, q_k]$ and used as the input to the model to predict a time-sequence $[\hat{t}_{k+1}, \hat{t}_{k+2}, \ldots, \hat{t}_N]$ that corresponds to $[\hat{q}_{k+1}, \hat{q}_{k+2}, \ldots, \hat{q}_N]$.

test durations. However, as the data suggests, some cells degrade rapidly within a specific time period, while others may not show sufficient aging over the same duration. Data from these slower-aging cells might be inadequate for effectively training the model. As a result, testing cells to a specific capacity threshold proves to be a more effective approach for gathering data to train predictive lifetime models compared to using data up to a fixed amount of time. This idea is further supported by the comparisons in Appendix C, where the predictive model trained with equal-capacity-threshold data outperforms the model trained with equal-time-threshold data.

As shown in Ref. 5 the calendar aging rate varies significantly across different test conditions and cell types. For example, cells stored at 24 °C can maintain more than 90 % of their relative capacity after years of usage, while cells stored at 60 °C may degrade to 80% of their relative capacity within just a few weeks of testing. Additionally, cells from different manufacturers under the same aging conditions can exhibit widely varying longevity. As a result, using uniform time intervals for data discretization can create issues with grid resolution and computational efficiency. For slow-decaying cells, a coarser time discretization might be too broad, resulting in a loss of important information about intra-sample aging dynamics. Conversely, a finer discretization appropriate for fast-decaying cells can impose unnecessary computational burdens on slow-decaying cells.

Consequently, this paper uses an equal-capacity-threshold for trajectory prediction. We adopt a uniformly sampled sequence of relative capacities and obtain the corresponding time values through linear interpolation. Specifically, the time sequence $(t_1, t_2, \ldots, t_k)$ is provided as input to the DD model, which predicts the output time sequence $(\hat{t}_{k+1}, \hat{t}_{k+2}, \ldots, \hat{t}_N)$. Assuming a bijective piece-wise linear mapping $t(q)$ exists, we can estimate the relative capacities $(\hat{q}_{k+1}, \hat{q}_{k+2}, \ldots, \hat{q}_N)$. In this work, for all the tasks, t_k represents the time when $q_k = SOH_{th}\%$ SOH and t_N represents the threshold time (or the EOL). Data-driven models generally need large amounts of data to adequetly train the model for accurate predictability. We performed data augmentation and monotization to preprocess the data before implementing it in the model, to create a synthetic dataset to increase its size, and monotonized to remove the fluctuations in the capacity fade data. For more details about the data pre-processing, refer to Appendix B.

The structure of the network to predict the cell capacity trajectory is shown in Fig. 4. The DD model consists of an encoder and a decoder neural network. The encoder reads the input sequence and converts it into a context vector (encoder layer output), and the decoder which converts this context vector into an output sequence. The input of the encoder in this model is the initial portion of the lifetime trajectory $t_1, t_2, \ldots, t_k$. The output of the encoder is then fed into the decoder. The decoder outputs the predicted lifetime trajectory, $(\hat{t}_{k+1}, \hat{t}_{k+2}, \ldots, \hat{t}_N)$.

Both the encoder and decoder are built with multiple GRU cells. The encoder takes input sequences and processes them through two layers of GRU cells, each containing three hidden neurons. The output states of the encoder are then passed to the decoder as its initial states. The decoder also consists of two layers of GRU cells, each with three hidden neurons. It takes the encoder output and zero padding as input to generate output sequences, which are then combined with output states and passed through a dense layer with a linear activation function to produce the final predictions.[43] The GRU is a simplified version of the popular LSTM network, that employs a complex structure of three gates to control the flow of information.[44] Broadly, the GRU cell consists of an input layer, memory cells and an output layer. The detailed mechanism of a single GRU-based memory cell, that updates the information from the input to create the encoder layer output, and controls the output information in the decoder can be found in Appendix D.

End-of-life and capacity fade trajectory-based performance metrics.—Figure 5 illustrates the two performance metrics used for comparing the models; the EOL and the capacity trajectory.

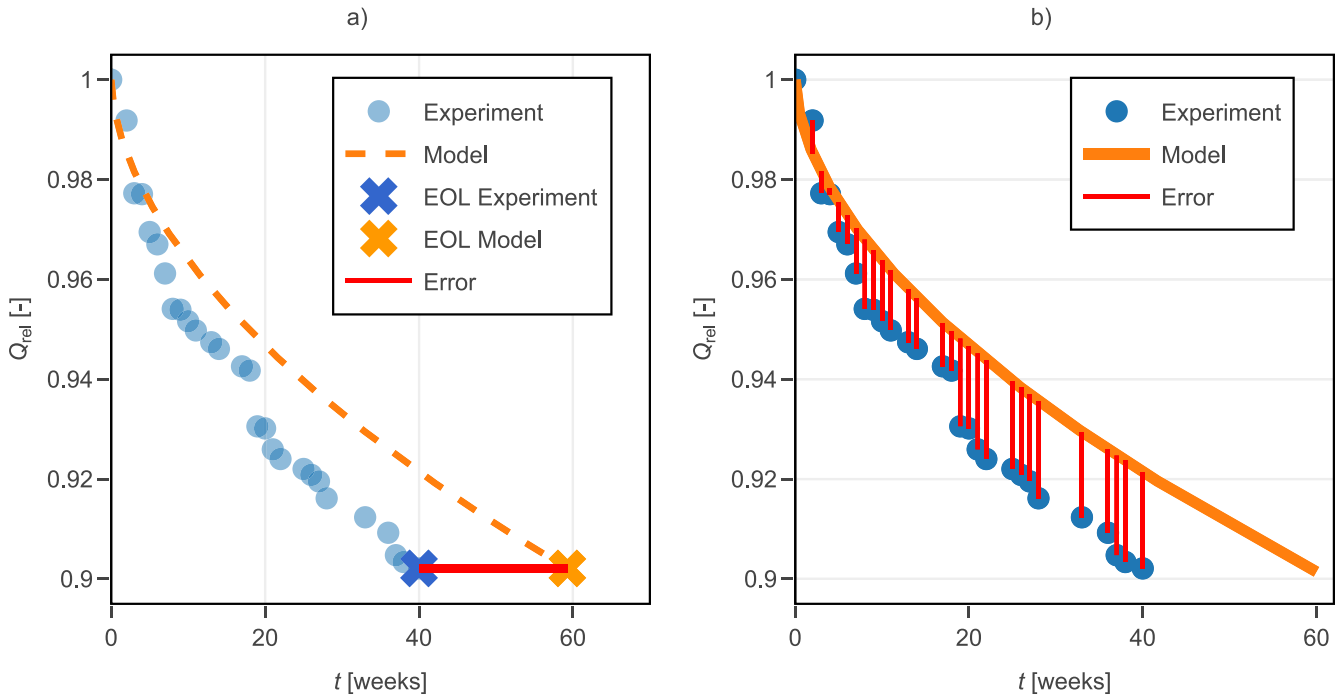

Figure 5. A schematic representation of the performance metrics for model comparison: (a) EOL is given in terms of weeks a battery cell lasts upon reaching the 90% SOH, and (b) Trajectory-based prediction.

Figure 5 illustrates these metrics to explain the difference between the two: Fig. 5a defines EOL based on the final point (in weeks) in the degradation trajectory, while Fig. 5b evaluates trajectory predictions.

We define t_{EOL} as the time when a cell's relative capacity reaches $Q_{\mathrm{rel,\,EOL}} = 90\%$, marking the end of life Fig. 5a. However, for some cells, particularly those aged at 24°C, relative capacity remains above 90% at the end of testing. To address this, we apply a power-law model, $Q_{\mathrm{rel}} = 1 - ab^t$ to extend (extrapolate) their aging trajectories until they reach 90% or $t_{\mathrm{EOL}} = t(Q_{rel} = 0.9)$. The extrapolated data are treated as synthetic and included in both training and testing to ensure model evaluation across all tasks.

The results are visualized in a parity plot with predicted $\hat{t}_{\mathrm{EOL}}$ vs. measured t_{EOL}, where greater distance to the diagonal line indicates worse prediction. The root mean square error, $RMSE_{EOL}$, is calculated from all N predictions for all cells for each of the tasks as follows:

$$RMSE_{\mathrm{EOL}}(t_{\mathrm{EOL}}, \hat{t}_{\mathrm{EOL}}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_{\mathrm{EOL,i}} - \hat{t}_{\mathrm{EOL,i}})^2}. \quad [8]$$

To evaluate the trajectory predictions of the models, we use the $RMSE_Q$ and the coefficient of determination (R^2) as statistical metrics. The $RMSE_Q$ quantifies the average discrepancies between the n observed values and the corresponding model predictions, giving greater weight to larger errors:

$$RMSE_Q(Q_{rel}, \hat{Q}_{rel}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{rel,i} - \hat{Q}_{rel,i})^2} \quad [9]$$

On the other hand, the coefficient of determination R^2 indicates how well the predicted values align with the actual data trends:

$$R^2(Q_{rel}, \hat{Q}_{rel}) = 1 - \frac{\sum_{i=1}^{n}(Q_{rel,i} - \hat{Q}_{rel})^2}{\sum_{i=1}^{n}(Q_{rel,i} - \bar{Q}_{rel})^2}, \quad [10]$$

where $\bar{Q}_{rel} = \frac{1}{n}\sum_{i=1}^{n} Q_{rel,i}$ is the mean of the observed relative capacities, n represents the total number of capacity evaluations, $Q_{rel,i}$ denotes the observed relative capacity values, and $\hat{Q}_{rel,i}$ are the corresponding values predicted by the model. In Fig. 5b, the observed values $Q_{rel,i}$ are depicted as experimental data points, while $\hat{Q}_{rel,i}$ represent the model trajectory predictions.

Results and Discussion

The capacity loss trajectory prediction tasks, described in Fig. 1, differ in the selection of data for training and testing. Throughout this paper, the SE and SR models use the entire trajectory from the traning cells, and unless stated otherwise, the initial capacity fade curve from $SOH = 100\%$ to $SOH_{th} = 98\%$ serves as the input to the DD model.

The first task, unseen cell prediction, involves classifying cells by type, storage temperature, and SOC, with a subset reserved for testing and excluded from training. The second task, unseen temperature prediction, trains the model on all but one temperature, which is then used for testing. The third task, unseen cell type prediction, follows a similar approach, omitting one cell type during training and using it exclusively for evaluation.

Task 1: unseen cell prediction.—Task 1 evaluates the models' ability to predict variations in individual cell behavior. To achieve this, the models are trained on a dataset that excludes one trajectory-i.e., one unseen cell-for each aging condition. The withheld trajectory serves as testing data, while the remaining cell data is used for training. The testing data is then used to assess the predictive performance of each model, SE, SR, and DD (introduced in Section 3).

The parameters of the SE model were identified using the nonlinear regression function *curve_fit* implemented as part of the Python package *SciPy*.[45] The Trust Region Reflective algorithm[46] was defined for least-squares optimization by setting the optimization method to "*trf*". The SR model functions were identified as follows:

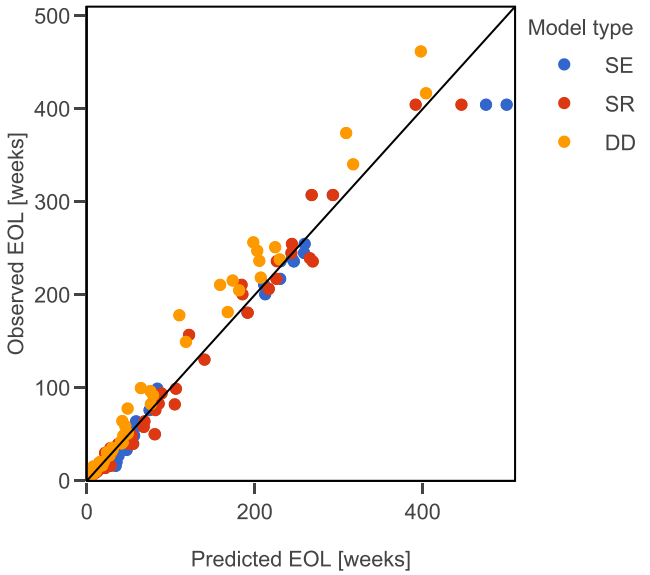

Figure 6. Unseen Cell Prediction - the semi-empirical model (SE), the symbolic regression model (SR) and the data-driven model (DD) are compared for EOL prediction using a parity plot, with the corresponding $RMSE_{EOL}$ values provided in Table II.

1. Estimate the parameters $\hat{\boldsymbol{\beta}} = (\beta_0, \beta_1)$ locally using the *curve_fit* implementation in *SciPy* with "*trf*"-optimizer.
2. Identify the functional dependence of $\boldsymbol{\beta}(T;\gamma)$ using the *PySR* symbolic regression implementation in Python.[47] To constrain complexity of the identifiable models, the operators for $\beta_0(T;\gamma)$ and $\beta_1(T;\gamma)$ were restricted to ("*", "/", "exp") and ("*", "/"), respectively. No nesting of multiple "exp" operators was allowed.
3. Re-identify the model parameters γ using the global optimization algorithm *differential_evolution* in *SciPy* for least-squares estimation.[45]

The DD model is implemented, trained, and tested using Python package Keras[48] and Tensorflow.[49] L_2 regularization with a weight of 10^{-6} is deployed to prevent overfitting. Considering the relatively small trainable parameters in DD model, a small weight was chosen. The Adam optimizer was selected for its adaptive learning rate properties, adjusting the learning rate for each parameter individually based on the first and second moments of the gradients. This allows for more efficient and stable convergence, especially in sparse gradients or noisy data. An exponential decay learning rate schedule starting at 0.0025 ensures the model can make significant updates initially while gradually reducing the learning rate to fine-tune the weights. A batch size of 10 of the training set balances training efficiency and model stability. Training for 1000 epochs provides sufficient time for the model to learn complex patterns, while an early stopping callback prevents overfitting by terminating training when performance ceases to improve. As shown in Fig. 4, the architecture includes two GRU layers, each with three neurons, to capture temporal dependencies, followed by a fully connected (dense) layer for output prediction. This design ensures the model effectively handles sequential data while maintaining a balance between complexity and computational efficiency. The choice of three neurons per GRU layer is sufficient to capture key temporal dependencies without overfitting, while also keeping computational demands manageable.

The EOL and trajectory-based performance metrics for all three models on the *Unseen cell prediction* task are shown in Fig. 6, Table II and Fig. 7. Table A.1 in Appendix A lists all identified SR models. Overall, all models perform well in both EOL-based and trajectory-based metrics. However, the SE and SR models perform

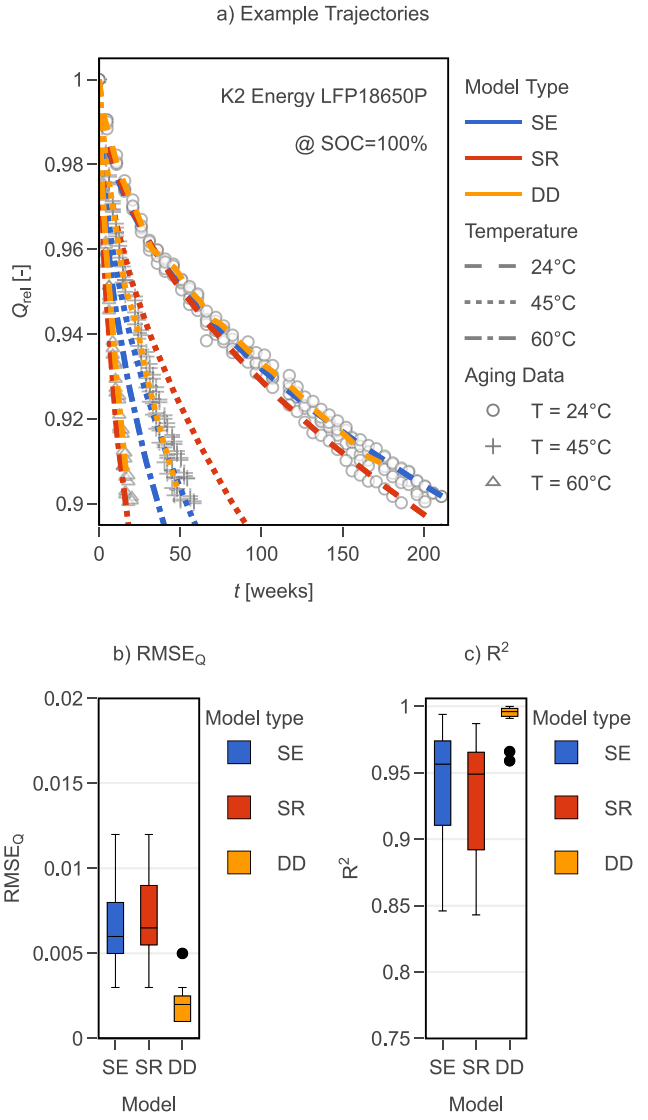

Figure 7. Unseen Cell Prediction - (a) trajectory predictions of SE, SR, and DD models for the example cell type *K2 Energy LFP18650P*, stored at SOC of 100%. (b) and (c) display the values of $RMSE_Q$ and the coefficient of determination (R^2) from the three models. Prediction errors over all the test cells are shown with a box plot. The box represents the middle 50% of prediction errors (25th to 75th percentile), with whiskers extending 1.5 times this range.

marginally better in EOL predictions, while the DD model exhibits a higher RMSE for EOL predictions.

The average RMSE across the dataset for the DD model is 0.0045, compared to 0.0075 for the SE model. The average R^2 for the DD model is higher, compared to the SE model. The advantage of the DD model over the algebraic models lies in its ability to leverage individual early aging data from cells, enabling the prediction of cell-specific aging trajectories, including cell-to-cell variability. In contrast, while the SE and SR models do not capture the cell-to-cell variability, they can predict the general trend of the aging trajectory for a given temperature and time.

The SE model assumes Arrhenius relationship for the temperature dependence, which, as shown in Ref. 5 diverges over longer timescales. This is observed in Fig. 7, where the model accuracy decreases at longer calendar aging times. The SR model contains unique expressions for each cell, which leads to increased error for trajectory predictions between the experimental data and model predictions for test cells. A key limitation of the SE and the SR

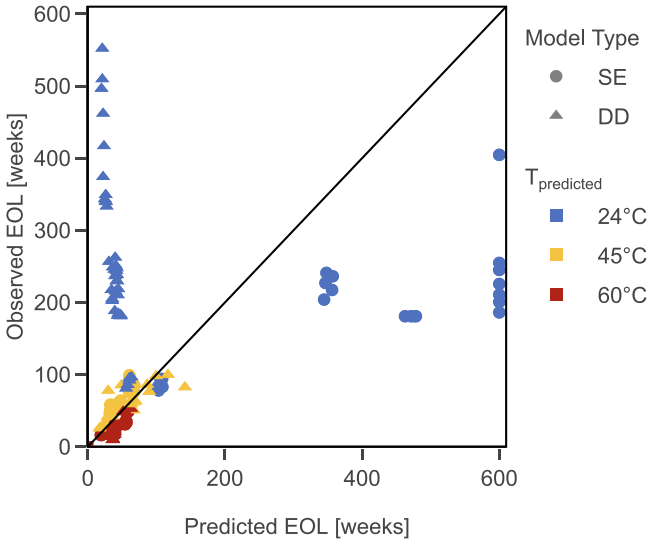

Figure 8. Unseen Temperature Prediction - the SE and DD models are compared for EOL prediction using a parity plot for each predicted temperature T_{pred} separately. The corresponding $RMSE_{EOL}$ metrics are shown in Table III.

approaches is that the model formulations are specific to a given cell-type operated at a particular SOC, making them non-transferable across different cells within a cell-type. In other words, the cell-to-cell variations are not captured by the SE and SR models. On the other hand, the DD model doesn't incorporate the SOC and temperature data for EOL and trajectory predictions. The dataset employed in this work encompasses variations in degradation trends, which might be overlooked by the DD model during training.

Task 2: unseen temperature prediction.—For this task, at least two temperatures are required, leading to three distinct cases based on the given conditions. First, extrapolation to lower temperatures uses data from 45 °C and 60 °C to predict the aging behavior at 24 °C. Second, interpolation uses data from 24 °C and 60 °C to forecast aging at 45 °C. Lastly, extrapolation to higher temperatures is conducted by applying data from 24 °C and 45 °C to predict aging at 60 °C. The SR derives expressions for temperature dependence $\beta_0(T;\gamma)$ and $\beta_1(T;\gamma)$ for a given cell-type. However, since functions for temperature dependence cannot be derived from only two temperature points, we opt not to evaluate the performance of this model for this specific case. Figure 8 shows the parity plot for the *unseen temperature prediciton task*, comparing the predictions for

Table II. RMSE in weeks for the EOL prediction of the SE, SR, and DD model for unseen cell prediction task.

	SE	SR	DD
$RMSE_{\text{EOL}}$ [weeks]	19.885113	15.143472	25.091912

Table III. $RMSE_{EOL}$ in weeks for the EOL prediction of the SE and DD model for each unseen temperature.

	SE	DD
$RMSE_{\text{EOL}}(T_{\text{predict}} = 24°\text{C})$ [weeks]	265.333410	250.740486
$RMSE_{\text{EOL}}(T_{\text{predict}} = 45°\text{C})$ [weeks]	17.909997	38.696998
$RMSE_{\text{EOL}}(T_{\text{predict}} = 60°\text{C})$ [weeks]	15.218209	15.575498

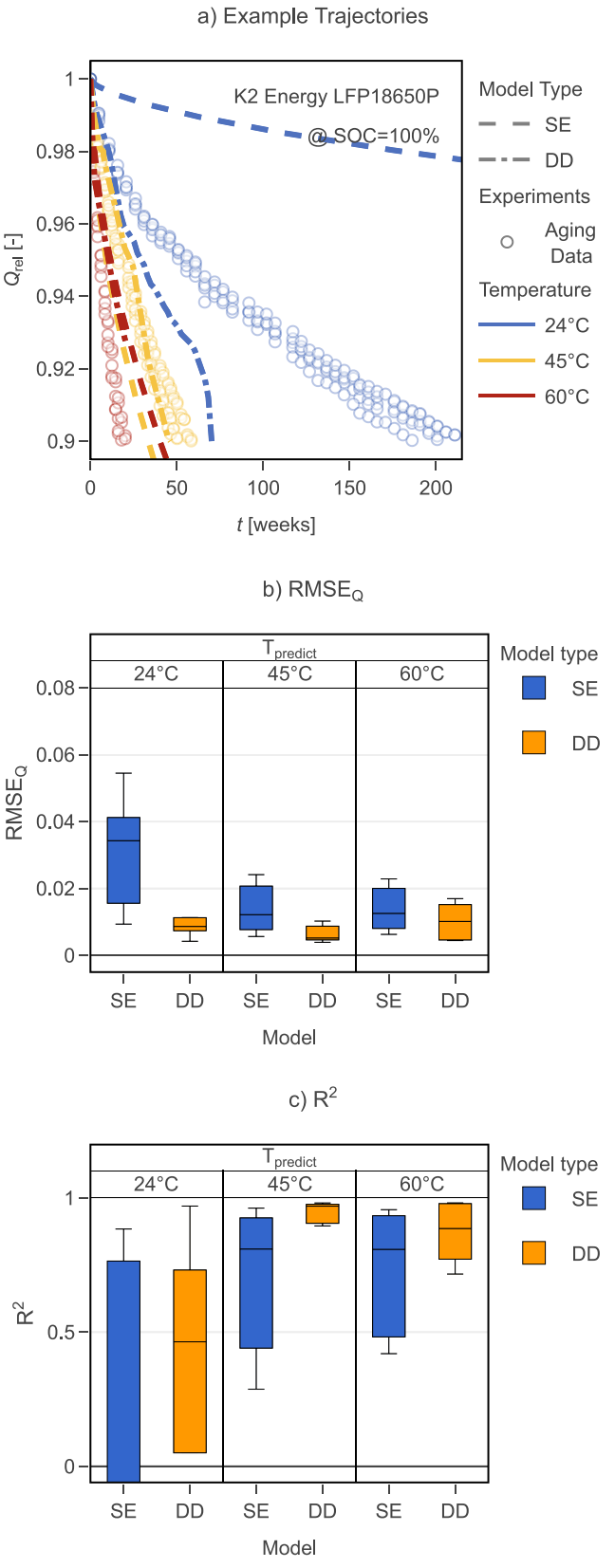

Figure 9. Unseen Temperature Prediction - (a) compares the fitted SE and DD models with experimental data for trajectory predictions, training on two temperatures and testing on the third temperature. (b) and (c) display the values of $RMSE_Q$ and the coefficient of determination (R^2) from the two models. The box plots show the variability of these metrics across different cell types and aging conditions.

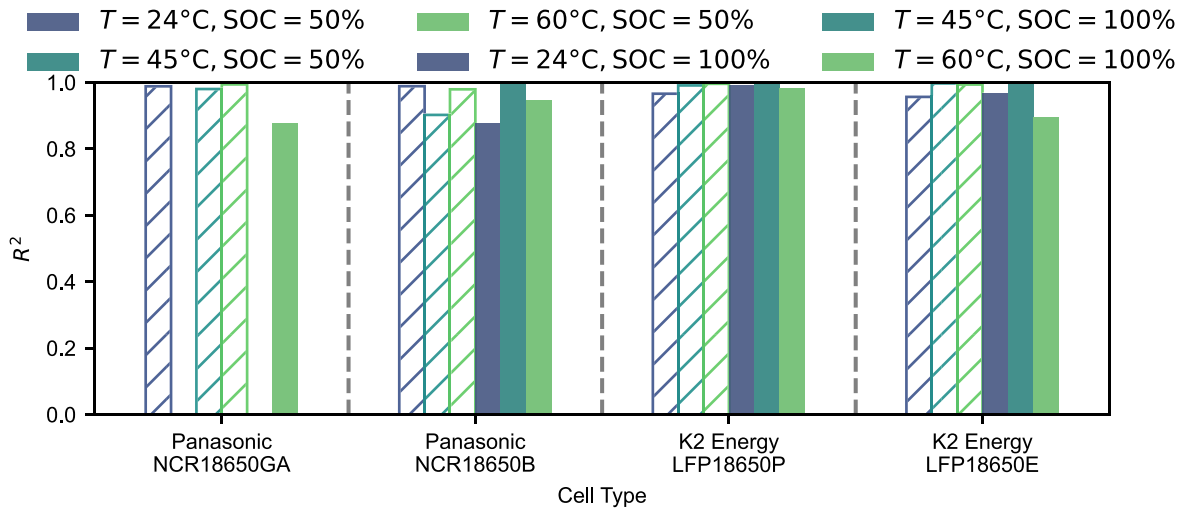

Figure 10. Task 3: Unseen cell type prediction results from the DD model (early prediction). The coefficient of determination (R^2) of trajectory predictions is evaluated across different cell types and aging conditions. This analysis examines how well the model generalizes to unseen cell types and whether it can accurately predict aging trajectories based on the available data from other cell types.

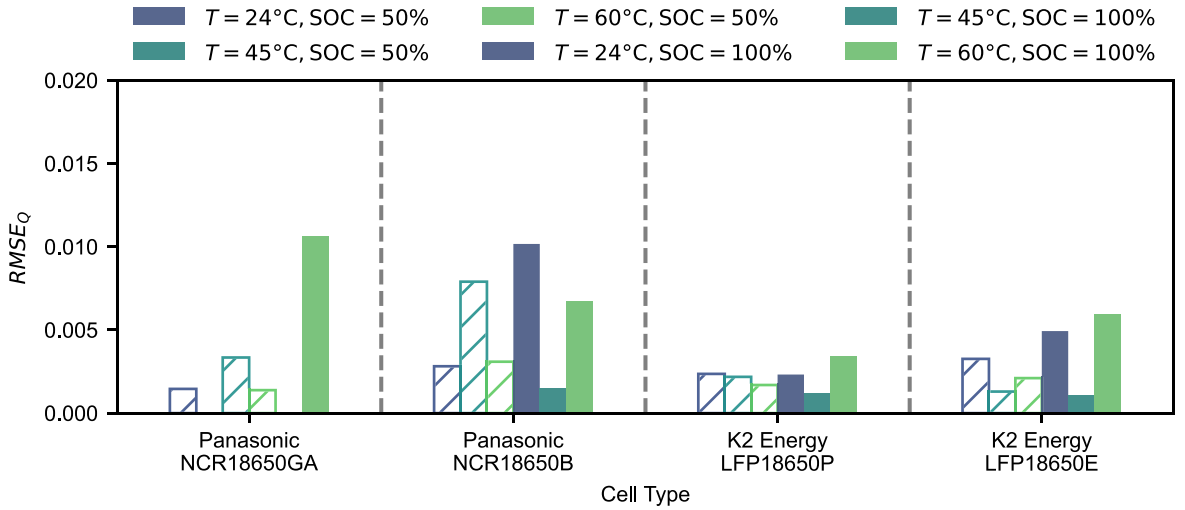

Figure 11. Unseen cell type prediction results from the DD model (early prediction). The $RMSE_Q$ of trajectory predictions is evaluated across different cell types and aging conditions.

$RMSE_{EOL}$ between the SE and the DD models. The $RMSE_{EOL}$ values are also summarized in Table III.

As observed from Fig. 9b, the prediction RMSE is lowest for 60 °C, with a slight increase for 45 °C. However, the RMSE for 24 °C predictions is significantly worse. This trend is also evident in the parity plot (Fig. 8) and R^2 values shown in Figs. 9a and c, respectively. The R^2 values for predictions at 45 °C and 60 °C are closer to 1, but the value for 24 °C can be extremely low, sometimes even negative. In the parity plot, predictions for 45 °C and 60 °C are closely clustered around the diagonal, while those for 24 °C are widely dispersed.

The underlying reason for the observed performance is that using data from 24 °C and 60 °C to predict at 45 °C is an interpolation with respect to temperature, while predicting for 24 °C and 60 °C involves extrapolation. As discussed in Ref. 5 the impact of temperature on calendar aging is complex and nonlinear, resulting in reduced prediction accuracy during extrapolation. Additionally, the accuracy for predicting at 24 °C is significantly lower than at 60 °C, highlighting that extrapolating from high-temperature data to predict lower temperatures is more challenging than extrapolating from low-temperature data for higher temperature predictions. Based on the RMSE and R^2 metrics, the DD demonstrates slightly better accuracy than the SE model for predicting various unseen temperature cases. The DD model predicts the trajectories for an unseen temperature well. The DD model was designed with a loss function for every point in the trajectory, ensuring a lower RMSE during trajectory predictions.

Task 3: unseen cell type prediction.—This task evaluates the model's transferability across four specific cell types: K2 Energy LFP18650E, Panasonic NCR18650B, Panasonic NCR18650GA, and K2 Energy LFP18650P. Data from all the three temperatures 24 °C, 45 °C and 60 °C) and SOCs (50% and 100%) are available for each of these cell types.

The methodology adopted for each experiment is consistent: one cell type is deliberately withheld from the training dataset, while the model is trained using the data from the remaining three cell types. This approach aims to assess the model's ability to generalize and predict behaviors across different cell types, ensuring that the model can effectively transfer knowledge learned from one cell type to another.

The cell type is a categorical variable that does not appear in the SE and SR models, as these models depend on numerical variables, specifically temperature and time. To address the cell-type dependency in capacity fading, two potential solutions can be considered

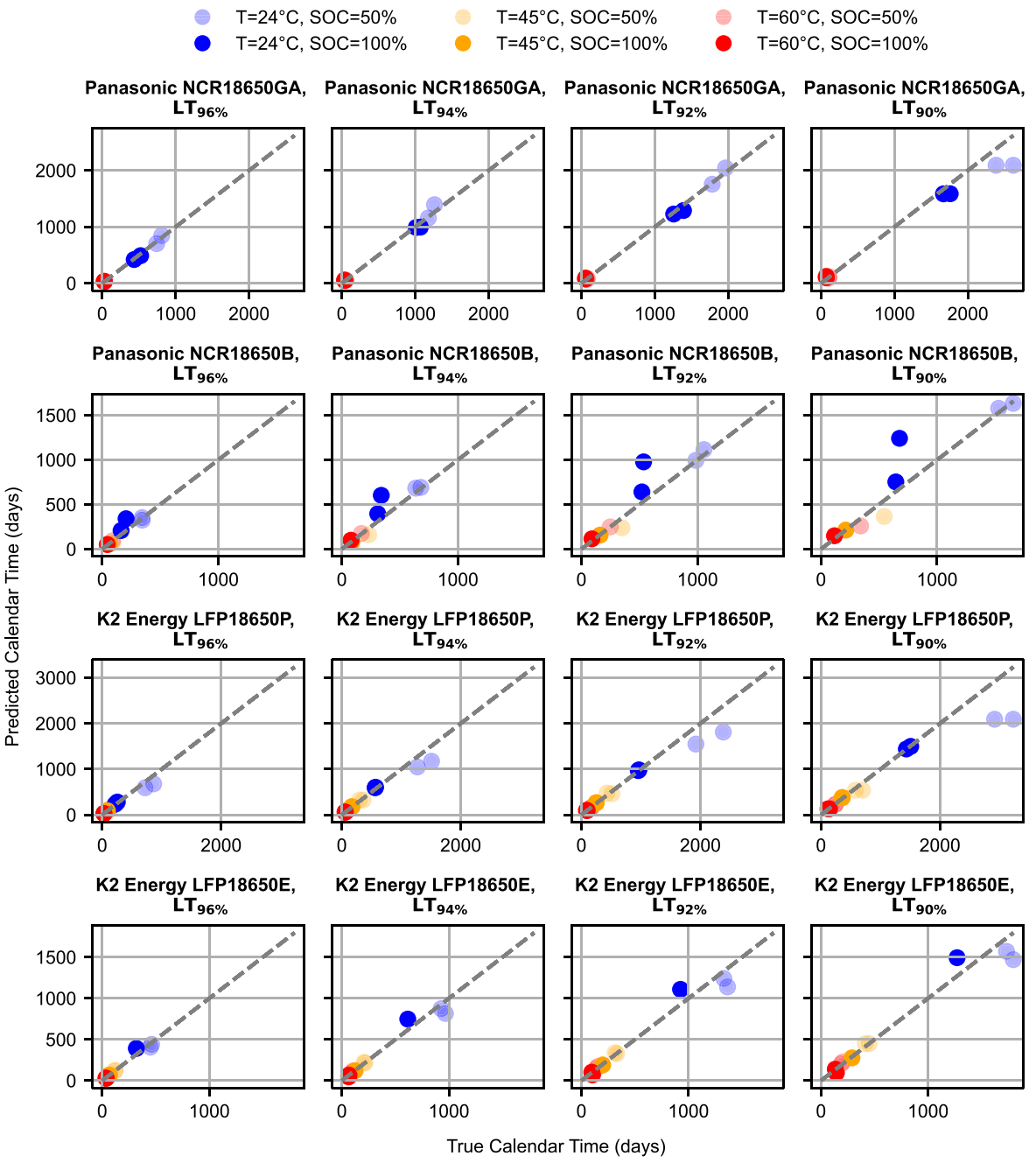

Figure 12. Task 3: Unseen cell type prediction. The DD model's EOL prediction capability is assessed using a parity plot. Each row corresponds to a different cell type, while each column displays the prediction errors of EOL at various capacity thresholds.

for the SE and SR models. The first approach involves encoding the cell-type variable as multiple binary indicators. For example, in a dataset with two cell types, one from K2 Energy and another from Panasonic, the SE model could be expanded as: $Q_{rel} = 1 - K\exp\left(-\frac{a_K}{T}\right)t^{b_K} - P\exp\left(-\frac{a_P}{T}\right)t^{b_P}$, where K is 1 and P is 0 if the cell type is from K2 Energy, and K is 0 and P is 1 if the cell type is from Panasonic. While this approach allows for incorporating cell type dependence, it faces a limitation: the lack of data for unseen cell types in the training set means that the model will output a trivial, constant prediction for the aging of these unseen cell types. The second approach is to exclude the cell type variable entirely from the model. While this simplification reduces complexity, it leads to inaccurate predictions. In this case, the model will predict the same aging curve for any cell type, given identical temperature and SOC conditions, which does not reflect the observed reality where different cell types exhibit distinct aging behavior. As demonstrated in Ref. 5 this simplification leads to inaccurate predictions, particularly for unseen cell types. Therefore, ror this task we evaluate only the results from the DD model.

We focus on assessing the DD model's performance using the four cell types. Test errors (RMSE and R^2) are shown in Figs. 10 and 11, respectively. Parity plots in Fig. 12 visualize the discrepancy between predicted and actual capacity fading, with columns representing different SOH thresholds. The x-axis corresponds to the withheld cell type, and the columns represent different SOH thresholds. As the SOH threshold decreases, accuracy declines due to the extrapolation challenge. The DD model predicts short-term aging well by learning early trajectory patterns but becomes less accurate for long-term predictions, where unexpected degradation

mechanisms may arise. The high R^2 value in Fig. 10 highlights the DD model's ability to transfer knowledge across different cell types. This transferability is due to the strong correlation between cell types within the dataset. For example, both the Panasonic NCR18650GA and Panasonic NCR18650B cells are NCA cells produced by Panasonic, while the K2 Energy LFP 18650P and K2 Energy LFP 18650E cells are LFP cells from K2 Energy. When one cell type is withheld for testing, the training set includes another cell of the same manufacturer and chemistry, aiding the DD model in transferring knowledge to the unseen cell type.

Comparing the results across three different scenarios, the SE and SR models predict reasonably well for Tasks 1 and 2. The DD model is an effective choice for trajectory predictions, but fail to capture the EOL during long-term calendar aging. As observed in Ref. 5 unexpected trends emerge at longer times that simpler models fail to predict. This behavior is chemistry-specific, and we observe that both the SE and the SR models are tailored to individual chemistries, making them non-transferable across different cell types.

Conclusions

This paper explores the modeling and prediction of our newly published long-term calendar aging dataset[5] which highlights several challenges such as time-varying nonlinearity, cell-to-cell variability, and cell-type dissimilarity. To address these challenges, we evaluate three models: semi-empirical, symbolic regression, and data-driven models. The data-driven model is based on a new trajectory-to-trajectory prediction framework. All models were evaluated for their ability to predict the EOL and capacity trajectories of unseen cells, unseen temperatures, and unseen cell-types. The data-driven (DD) model demonstrated good performance in predicting the degradation trajectory of unseen cells, effectively capturing nonlinearity and variability in cell aging trends. However, its accuracy in EOL prediction did not surpass that of the SE model. It is observed that the DD model, in some cases, tends to overfit the nonlinear aging behavior with multiple sigmoid functions, leading to singular local deviations near the EOL portion of the trajectories. In contrast, the SE model-due to its intrinsic structure of monomial nonlinearities-fitted more uniformly to the true degradation trends and exhibited a more homeomorphic relationship between the accuracies of EOL prediction and trajectory prediction. The SE model excelled in predicting aging at unseen high temperatures. Additionally, the data-driven model showed reasonable accuracy when predicting the capacity fade of a cell type using data from another cell type with the same chemistry and manufacturer, validating the feasibility of knowledge transfer across different cell types. This work raises two key research questions: (1) how to transfer knowledge from one cell type to another with different chemistry or manufacturer, and (2) whether calendar aging behaviors at low temperatures can be predicted using training data from higher temperature data.

Acknowledgments

This effort was sponsored in whole or in part by the United States Government (USG). The U.S. Government is authorized to reproduce and distribute original submissions for publication for Governmental purposes notwithstanding any copyright notation thereon.

Disclaimer

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the United States Government (USG).

Author Contributions

X.C., F.S., V.N.L., M.U. contributed to data analysis and algorithm development. X.C. and F.S. generated the figures. X.C., F.S., M.U. and S.O., wrote and revised the manuscript. S.O. and W. C.C supervised the work, and acquired funding for the project.

Declaration of Interest

X.C. is with the Department of Electrical and Computer Engineering, University of California Los Angeles, California, USA. F.S. is with the Institute for Sustainable Energy Systems, University of Applied Science HM, Munich, Germany. They were both affiliated with Stanford University at the time the research was conducted.

Appendix A: Symbolic Regression Model Functions

Appendix A summarizes the expressions obtained from the symbolic regression framework for Task 1 for different cell types and SOCs.

Table A·1. Task 1 - Identified model functions using SR.

Cell type	SOC [%]	SR Model
Ultralife 502030	50	$1 - \gamma_0 \cdot t^{\gamma_1}$
Ultralife 502030	100	$1 - \gamma_0 \cdot t^{\gamma_1}$
Ultralife UBP001	100	$1 - \gamma_0 e^{\gamma_1 T} \cdot t^{\gamma_2}$
Tenergy 302030	100	$1 - \gamma_0 \cdot t^{\gamma_1}$
Sony-Murata US18650VTC6	100	$1 - \gamma_0 T^{\gamma_1} \cdot t^{\gamma_2}$
Panasonic NCR18650GA	50	$1 - \gamma_0 e^{-\frac{\gamma_1}{T}} \cdot t^{\gamma_2}$
Panasonic NCR18650GA	100	$1 - \gamma_0 T^{\gamma_1} e^{-\gamma_2 T} \cdot t^{\gamma_3 T}$
Panasonic NCR18650B	50	$1 - \gamma_0 T \cdot t^{\gamma_1}$
Panasonic NCR18650B	100	$1 - \frac{e^{-\frac{\gamma_0}{T}}}{T} \cdot t^{\gamma_1}$
K2 Energy LFP18650P	50	$1 - \frac{\gamma_0 e^{\gamma_1 T}}{T} \cdot t^{\gamma_2}$
K2 Energy LFP18650P	100	$1 - \frac{\gamma_0 e^{\gamma_1 T}}{T^{\gamma_2}} \cdot t^{\gamma_3}$
K2 Energy LFP18650E	50	$1 - \frac{\gamma_0 e^{\gamma_1 T}}{T^{\gamma_2}} \cdot t^{\gamma_3}$
K2 Energy LFP18650E	100	$1 - \gamma_0 T \cdot t^{\gamma_1}$

Appendix B: Data Preprocessing for Data-Driven Model

The Data-Driven (DD) model requires a large training dataset to automatically extract the representative features. However, acquiring long-term trajectory data, especially for battery calendar aging experiments like those in this study, is time-consuming and labor-intensive. To address this challenge, we introduce data augmentation and monotonization techniques to enhance both the data and the performance of the DD model.

B.1. Data Augmentation.—The data augmentation uses the convex-hull method to create a synthetic training dataset. Figure B·1a shows the observed cell aging trajectories $Q_1(t)$ and $Q_2(t)$. The augmented cell aging trajectory $Q^a(t)$ is given by

$$Q^a(t) = wQ_1(t) + (1 - w)Q_2(t), \quad 0 < w < 1. \qquad [11]$$

This technique has previously found application in the computer science field.[50] However, to the best of the author's knowledge, this method has not been previously utilized to augment battery calendar aging datasets.

The real measured and augmented battery aging curves are compared in Figs. B·1(b), B·1(c). This convex-hull-based augmentation method is based on the assumption that the statistical distribution of a specific cell type should be continuous. In other words, if two cells originating from the same cell type and aging conditions display capacities Q_1 and Q_2, then we believe there is a non-zero probability that another cell from the same population will have a capacity between Q_1 and Q_2. The augmented trajectories are also as smooth as the real measured trajectories.

Equation 11 achieves this by bounding the derivatives of augmented trajectories by those of the real measured trajectories.

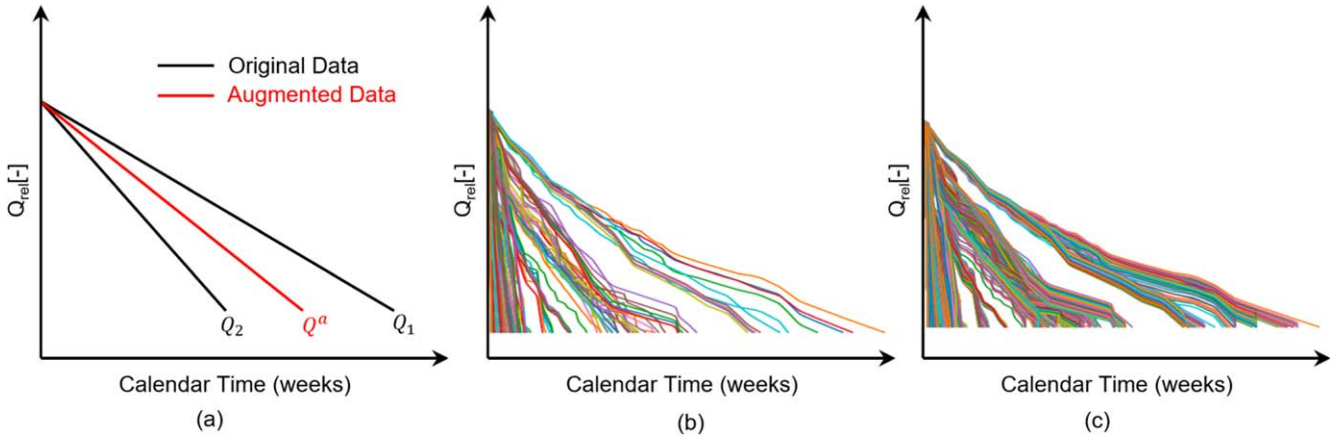

Figure B·1. Enhancing DD model performance with data augmentation: introducing the convex-hull method for synthetic battery aging trajectories to address the challenge of limited long-term data. (a) Convex hull of two aging trajectories; (b) Raw calendar time—relative capacity trajectories in the dataset; (c) Augmented calendar time—relative capacity trajectories in the dataset.

B.2. Data Monotonization.—The cell capacity during calendar aging is expected to exhibit a monotonic decrease. However, the experimental capacity data shows a non-monotonic decreasing trend. This leads to multiple lifetimes corresponding to a single SOH level. As the DD model necessitates function inversion, we introduce a monotonization approach that enables us to identify the maximum monotonic sub-sequence within a calendar-aging trajectory. The pseudo-code for the data monotization is provided in Algorithm 1.

Algorithm 1.Monotonize Arrays (Time vs Capacity)

function MONOTONIZEARRAYS (t (time), Q (capacity)
  $result_t \leftarrow []$ ▷List for time
  $result_Q \leftarrow []$ ▷List for capacity
  $i \leftarrow 0$
  while $i <$ length(Q) **do**
    $current_Q \leftarrow Q[i]$
    Append $t[i]$ to $result_t$ ▷Append time value
    Append $current_Q$ to $result_Q$ ▷Append capacity value
    while $i + 1 <$ length(Q) and $Q[i + 1] \geqslant current_Q$ **do**
      Increment i by 1
    Increment i by 1
Return $result_t$ (time), $result_Q$ (capacity)

B.3. Training Set and Testing Set Split.—In machine learning and statistical modeling, a common practice is splitting the available data set into two subsets: a training set and a testing set. This division allows models to be trained on one subset (training set) and evaluated on another unseen subset (testing set). Using the unseen subset, the model's prediction performance can be evaluated.

Let $D \subseteq X \times \mathbb{R}^+ \times Y$ be the set of all cell-specific aging conditions and corresponding trajectories. It is divided into a disjoint training and testing subset, denoted as $D_{\text{train}} = (X_{\text{train}}, Y_{\text{train}})$ and $D_{\text{test}} = (X_{\text{test}}, Y_{\text{test}})$ respectively, such that

$$D = D_{\text{train}} \cup D_{\text{test}},$$
$$D_{\text{train}} \cap D_{\text{test}} = \varnothing$$

with $X_{\text{train}}, X_{\text{test}} \subset X \times \mathbb{R}_+$ representing the input subsets with aging condition and RPT times, and $Y_{\text{train}}, Y_{\text{test}} \subset (Y^n)^m$ as its corresponding relative capacity output subsets. m is the sum of all tested cells. D_{train} and D_{test} are used for model training and testing, respectively. The proportion in which D is split into D_{train} and D_{test} varies based on the task.

Appendix C: Comparison Between Equal-Time-Threshold and Equal-Capacity-Threshold Prediction

The DD model takes the initial segment of the relative capacity trajectory as input and predicts the remaining trajectory as output. The input selection threshold can be determined using either an equal-time threshold, where all cells stop testing after the same duration, or an equal-capacity threshold, where all cells stop after reaching the same relative capacity. Figure C·1 compares trajectory prediction errors under equal-time and equal-capacity thresholds across different temperatures. The figure compares the prediction results from using both these thresholds for different testing times budgets, or different early-aging data, indicated by different times. In the figure, the x-axis represents different time-thresholds, and the y-axis represents the error in capacity predictions. For the case where the equal-capacity-threshold is used, the corresponding time (x-axis in the figure) is calculated as the mean of the corresponding times at which this capacity threshold is met for each cell type. When using the equal-time-threshold method, prediction accuracy improves with testing time but varies due to data stochasticity. When using the equal-capacity-threshold method, lower errors and faster convergence are achieved, highlighting its advantage in capturing diverse aging behaviors for improved prediction accuracy.

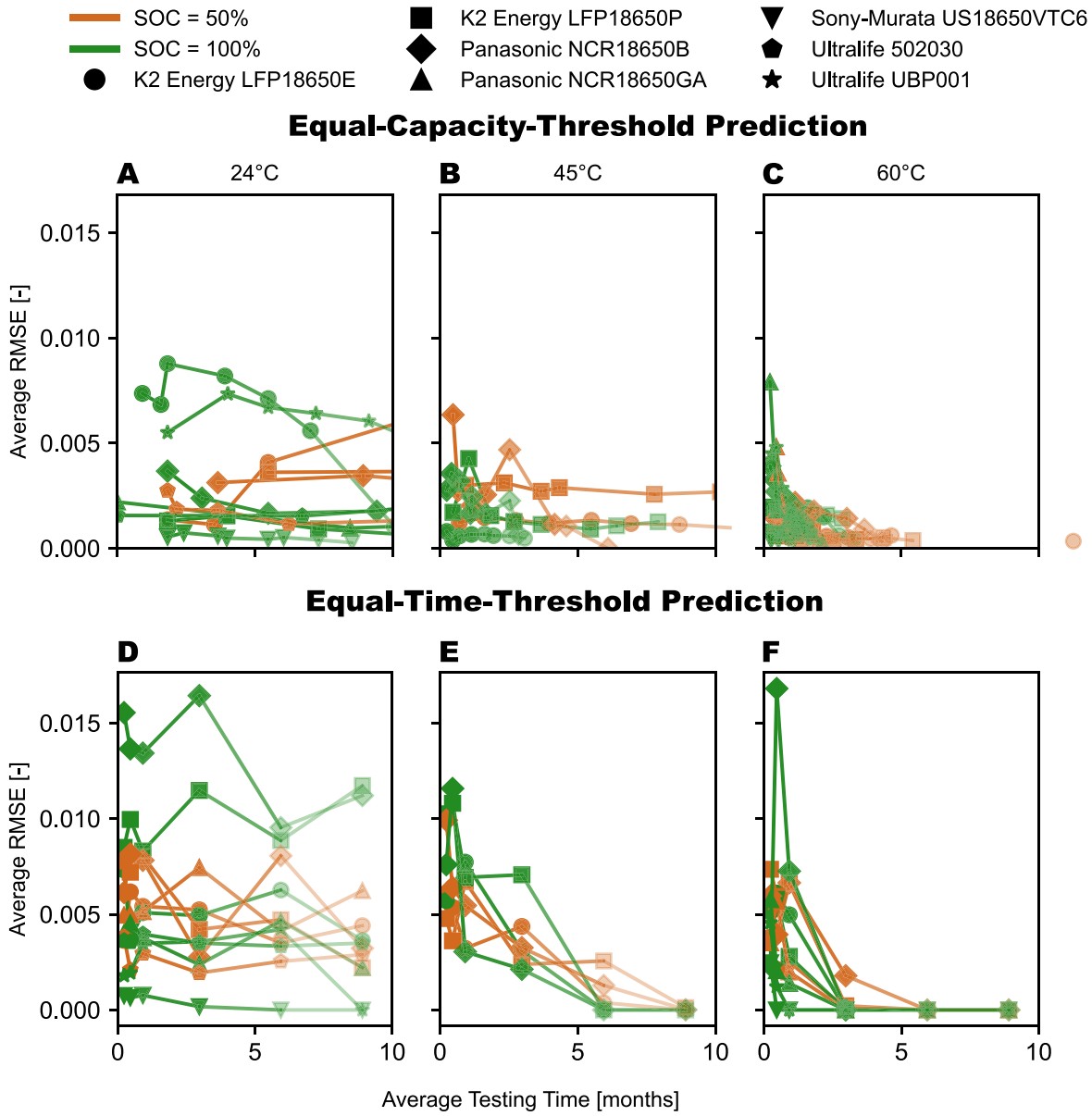

Figure C·1. Trajectory prediction error versus testing time budget at temperature 24 ˚C, 45 ˚C, 60 ˚C. The prediction error is quantified by the trajectory-wise average root mean square error (RMSE) between the predicted and ground-truth trajectories. (A), (B), (C) Trajectory prediction error using the equal capacity threshold. (D), (E), (F) Trajectory prediction error using the equal time threshold. Prediction errors generally follow a decaying trend with respect to the testing time budget, though variations may occur due to the stochastic nature of data-driven methods. For high-temperature cells that completely fail within the time threshold, no prediction is performed. Prediction errors are smaller and decrease more rapidly compared to the equal-time-threshold method in (D)-(F).

Appendix D: Mechanism of a Gate Recurrent Unit

The following equations describe how a Gate Recurrent Unit (GRU) neuron works:[51]

$$r[t] = \sigma(W_{ir}x[t] + b_{ir} + W_{hr}h[t-1] + b_{hr}) \quad [12]$$

$$z[t] = \sigma(W_{iz}x[t] + b_{iz} + W_{hz}h[t-1] + b_{hz}) \quad [13]$$

$$\hat{h}[t] = \tanh(W_{in}x[t] + b_{in} + r_t*(W_{hn}h[t-1] + b_{hn})) \quad [14]$$

$$h[t] = (1 - z[t])*\hat{h}[t] + z[t]*h[t-1] \quad [15]$$

where $x[t]$ is the input vector, $h[t]$ is the output vector, $\hat{h}[t]$ is the candidate activation vector, $z[t]$ is the update gate vector, $r[t]$ is the reset gate vector. σ is the sigmoid function that outputs a value between 0 and 1. tanh is the hyperbolic function. W and b are the weight and bias parameter matrices. Equation 12 computes the reset gate $r[t]$. The input $x[t]$ is subject to a linear transformation—comprising multiplication by a weight matrix and adding a bias vector. Concurrently, the previous state $h[t-1]$ also undergoes a linear transformation. These transformed inputs are then combined via the reset gate, and the aggregate is fed into a sigmoid activation function, which restricts the output to a range between 0 and 1, thereby dictating the extent of historical information to be retained or discarded. The update gate $z[t]$ is then determined as specified in Eq. 13. The computation mirrors the reset gate's, with distinct weights and bias parameters. As shown in Eq. 14, the candidate activation vector leverages the reset gate to merge the new memory content with past information. The previous state $h[t-1]$ is first linearly transformed by weight and bias vectors. Following this, the Hadamard product—element-wise multiplication—of the reset gate $r[t]$ with this transformed $h[t-1]$ is computed. The reset gate, a vector with elements ranging from 0 to 1, regulates the quantity of past information to be preserved; for instance, a zero value in $r[t]$ indicates a complete disregard for past data.

This Hadamard product, combined with the transformed input $x[t]$, is fed into a hyperbolic tangent activation function to produce the candidate activation. The last step involves the computation of $h[t]$, as depicted in Eq. 15. The update gate $z[t]$ decides the proportion of information to be obtained from the current candidate activation vector $h'[t]$ and the previous state $h[t-1]$.[52]

ORCID

Florian Stroebl https://orcid.org/0000-0002-9468-1956
Maitri Uppaluri https://orcid.org/0000-0002-6612-3009
Simona Onori https://orcid.org/0000-0002-6556-2608

References

1. J. Moody, E. Farr, M. Papagelis, and D. R. Keith, "The value of car ownership and use in the United States." *Nature Sustainability*, **4**, 769 (2021).
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*, **55**, 105634 (2022).
3. P. Keil, S. F. Schuster, Jörn Wilhelm, J. Travi, A. Hauser, R. C. Karl, and A. Jossen, "Calendar aging of lithium-ion batteries." *J. Electrochem. Soc.*, **163**, A1872 (2016).
4. M. Naumann, M. Schimpe, P. Keil, H. C. Hesse, and A. Jossen, "Analysis and modeling of calendar aging of a commercial $LiFePO_4$ /graphite cell." *Journal of Energy Storage*, **17**, 153 (2018).
5. V. N. Lam, X. Cui, F. Stroebl, M. Uppaluri, S. Onori, and W. C. Chueh, "A decade of insights: delving into calendar aging trends and implications." *Joule*, **9**, 1 (2025).
6. H. J. Ploehn, P. Ramadass, and R. E. White, "Solvent diffusion model for aging of lithium-ion battery cells." *J. Electrochem. Soc.*, **151**, A456 (2004).
7. P. Ramadass, B. Haran, P. M. Gomadam, R. White, and B. N. Popov, "Development of first principles capacity fade model for Li-ion cells." *J. Electrochem. Soc.*, **151**, A196 (2004).
8. M. Safari, M. Morcrette, A. Teyssot, and C. Delacourt, "Multimodal physics-based aging model for life prediction of Li-ion batteries." *J. Electrochem. Soc.*, **156**, A145 (2009).
9. N. Kamyab, J. W. Weidner, and R. E. White, "Mixed mode growth model for the solid electrolyte interface (SEI)." *J. Electrochem. Soc.*, **166**, A334 (2019).
10. S. Khaleghi Rahimian, M. M. Forouzan, S. Han, and Y. Tang, "A generalized physics-based calendar life model for Li-ion cells." *Electrochimica Acta*, **348**, 136343 (2020).
11. M. Ecker, J. B. Gerschler, J. Vogel, S. Käbitz, F. Hust, P. Dechent, and D. U. Sauer, "Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data." *Journal of Power Sources*, **215**, 248 (2012).
12. M. Schimpe, C. Piesch, H. C. Hesse, J. Pab, S. Ritter, and A. Jossen, "Power flow distribution strategy for improved power electronics energy efficiency in battery storage systems: Development and implementation in a utility-scale system." *Energies*, **11**, 533 (2018).
13. S. L. Hahn, M. Storch, R. Swaminathan, B. Obry, J. Bandlow, and K. P. Birke, "Quantitative validation of calendar aging models for lithium-ion batteries." *Journal of Power Sources*, **400**, 402 (2018).
14. E. J. Dufek, T. R. Tanim, B.-rong Chen, S. Kim, and E. J. Dufek, "Battery calendar aging and machine learning." *Joule*, **6**, 1363 (2022).
15. B. Celen et al., "Calendar ageing modelling using machine learning: an experimental investigation on lithium ion battery chemistries." *Open Research Europe*, **2**, 96 (2022).
16. K. Liu, T. R. Ashwin, X. Hu, M. Lucu, and W. D. Widanage, "An evaluation study of different modelling techniques for calendar ageing prediction of lithium-ion batteries." *Renew. Sustain. Energy Rev.*, **131**, 110017 (2020).
17. M. Schimpe, M. E. von Kuepach, M. Naumann, H. C. Hesse, K. Smith, and A. Jossen, "Comprehensive modeling of temperature-dependent degradation mechanisms in lithium iron phosphate batteries." *J. Electrochem. Soc.*, **165**, A181 (2018).
18. P. M. Attia, W. C. Chueh, and S. J. Harris, "Revisiting the $t^{0.5}$ dependence of SEI growth." *J. Electrochem. Soc.*, **167**, 090535 (2020).
19. P. Gasper, K. Gering, E. Dufek, and K. Smith, "Challenging practices of algebraic battery life models through statistical validation and model identification via machine-learning." *J. Electrochem. Soc.*, **168**, 020502 (2021).
20. K. L. Gering, "Novel method for evaluation and prediction of capacity loss metrics in Li-ion electrochemical cells." *Electrochimica Acta*, **228**, 636 (2017).
21. M. Broussely, S. Herreyre, P. Biensan, P. Kasztejna, K. Nechev, and R. J. Staniewicz, "Aging mechanism in Li ion cells and calendar life predictions." *J. Power Sources*, **97-98**, 13 (2001).
22. M. Ecker, N. Nieto, S. Käbitz, J. Schmalstieg, H. Blanke, A. Warnecke, and D. U. Sauer, "Calendar and cycle life study of Li(NiMnCo)O2-based 18650 lithium-ion batteries." *Journal of Power Sources*, **248**, 839 (2014).
23. J. Schmitt, A. Maheshwari, M. Heck, S. Lux, and M. Vetter, "Impedance change and capacity fade of lithium nickel manganese cobalt oxide-based batteries during calendar aging." *Journal of Power Sources*, **353**, 183 (2017).
24. A. Eddahech, O. Briat, and J. M. Vinassa, "Performance comparison of four lithium-ion battery technologies under calendar aging." *Energy*, **84**, 542 (2015).
25. S. Paarmann et al., "Short-term tests, long-term predictions – accelerating ageing characterisation of lithium-ion batteries." *Batteries & Supercaps*, **7**, e202300594 (2024).
26. P. Gasper, N. Collath, H. C. Hesse, A. Jossen, and K. Smith, "Machine-learning assisted identification of accurate battery lifetime models with uncertainty." *J. Electrochem. Soc.*, **169**, 080518 (2022).
27. K. Schofer, F. Laufer, J. Stadler, S. Hahn, G. Gaiselmann, A. Latz, K. P. Birke, and M.-benz Ag, "Machine learning-based lifetime prediction of lithium-ion cells." *Advanced Science*, 22006301 (2022).
28. S. Kim et al., "Accelerated battery life predictions through synergistic combination of physics-based models and machine learning." *Cell Reports Physical Science*, **3**, 101023 (2022).
29. K. S. R. Mawonou, A. Eddahech, D. Dumur, D. Beauvois, and E. Godoy, "State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking." *Journal of Power Sources*, **484**, 229154 (2021).
30. K. Liu, Y. Li, X. Hu, M. Lucu, and W. D. Widanage, "Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries." *IEEE Transactions on Industrial Informatics*, **16**, 3767 (2020).
31. M. Azkue, M. Lucu, E. Martinez-Laserna, and I. Aizpuru, "Calendar ageing model for li-ion batteries using transfer learning methods." *World Electric Vehicle Journal 2021, Vol. 12, Page 145*, **12**, 145 (2021).
32. M. S. Hossain Lipu, M. A. Hannan, A. Hussain, M. M. Hoque, P. J. Ker, M. H. M. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations." *Journal of Cleaner Production*, **205**, 115 (2018).
33. X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics." *Joule*, **4**, 310 (2020).
34. W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, "One-shot battery degradation trajectory prediction with deep learning." *Journal of Power Sources*, **506**, 9 (2021).
35. L. Xu, Z. Deng, Y. Xie, X. Lin, and X. Hu, "A novel hybrid physics-based and data-driven approach for degradation trajectory prediction in Li-ion batteries." *IEEE Transactions on Transportation Electrification*, **9**, 2628 (2023).
36. N. Palm, F. Stroebl, and H. Palm, "Parameter individual optimal experimental design and calibration of parametric models." *IEEE Access*, **10**, 112515 (2022).
37. J. R. Koza and R. Poli, *Genetic Programming* (Springer, US)127 (2005).
38. M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data." *Science*, **324**, 81 (2009).
39. J. Wang and C. Zhang, "Software reliability prediction using a deep learning model based on the RNN encoder-decoder." *Reliability Engineering and System Safety*, **170**, 73 (2018).
40. Z. Wang, X. Su, and Z. Ding, "Long-term traffic prediction based on LSTM encoder-decoder architecture." *IEEE Transactions on Intelligent Transportation Systems*, **22**, 6561 (2021).
41. T. Ahmad and D. Zhang, "A data-driven deep sequence-to-sequence long-short memory method along with a gated recurrent neural network for wind power forecasting." *Energy*, **239**, 122109 (2022).
42. L. Xu, Z. Deng, Y. Xie, X. Lin, and X. Hu, "A novel hybrid physics-based and data-driven approach for degradation trajectory prediction in Li-ion batteries." *IEEE Transactions on Transportation Electrification*, **9**, 2628 (2022).
43. Y. Song, L. Li, Y. Peng, and D. Liu, "Lithium-ion battery remaining useful life prediction based on GRU-RNN." *12th International Conference on Reliability, Maintainability, and Safety (ICRMS)* (2018).
44. Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries." *IEEE Transactions on Vehicular Technology*, **67**, 5695 (2018).
45. P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in python." *Nature Methods*, **17**, 261 (2020).
46. M. A. Branch, T. F. Coleman, and Y. Li, "A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems." *SIAM Journal on Scientific Computing*, **21**, 1 (1999).
47. M. Cranmer, *Interpretable machine learning for science with PySR and SymbolicRegression.jl* (2023), arXiv:2305.01582.
48. Keras, https://keras.io/api/.
49. TensorFlow, https://www.tensorflow.org/.
50. E. Kalunga, S. Chevallier, Q. Barth, E. Kalunga, S. Chevallier, Q. Barth, and E. K. Kalunga, "Data augmentation in riemannian space for brain-computer interfaces." *STAMLINS - ICML Workshop on Statistics, Machine Learning and Neuroscience*, Lille, France (2015), https://hal.archives-ouvertes.fr/hal-01351990.
51. User Manual for GRUCell, https://pytorch.org/docs/stable/generated/torch.nn.GRUCell.html.
52. K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, *Learning Phrase representations using RNN encoder-decoder for statistical machine translation* (2014), arXiv:1406.1078.