eTransportation 26 (2025) 100486

Contents lists available at ScienceDirect

eTransportation

journal homepage: www.journals.elsevier.com/etransportation

Bayesian analysis of interpretable aging across thousands of lithium-ion battery cycles

Marc D. Berliner [a,1], Minsu Kim [a,1], Xiao Cui [b], Vivek N. Lam [b], Shakul Pathak [a], Yunhong Che [a], Patrick A. Asinger [a], Martin Z. Bazant [a,c], William C. Chueh [b], Richard D. Braatz [a,*]

[a] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America
[b] Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, United States of America
[c] Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America

ARTICLE INFO

Keywords:
Lithium-ion battery
Degradation
Parameter identification
Identifiability analysis
State-of-health
Uncertainty quantification

ABSTRACT

The Doyle–Fuller–Newman (DFN) model is a common mechanistic model for lithium-ion batteries. The reaction rate constant and diffusivity are key parameters that directly affect the movement of lithium ions, thereby offering explanations for cell aging. This work investigates the ability to uniquely estimate each electrode's diffusion coefficients and reaction rate constants of 95 T Model 3 cells with a nickel cobalt aluminum oxide (NCA) cathode and silicon oxide–graphite (LiC_6–SiO_x) anode. The four parameters are estimated using Markov chain Monte Carlo (MCMC) for a total of 7776 cycles at various discharge C-rates. While one or more anode parameters are uniquely identifiable over every cell's lifetime, cathode parameters become identifiable at mid- to end-of-life, indicating measurable resistive growth in the cathode. The contribution of key parameters to the state of health (SOH) is expressed as a power law. This model for SOH shows a high consistency with the MCMC results performed over the overall lifespan of each cell. Our approach suggests that effective diagnosis of aging can be achieved by predicting the trajectories of the aging parameters. As such, extending our analysis with more physically accurate models building on DFN may lead to more identifiable parameters and further improved aging predictions.

1. Introduction

Lithium-ion batteries are widely used in portable electronic devices as well as electric vehicles (EVs) due to their high energy density and long cycle life. To replace the expensive cycling experiments of batteries, various studies such as fast charging [1–3], aging analysis [4–7], lifetime prediction [8,9], and fault diagnosis [10] have been conducted using electrochemical-based battery models. The electrochemical models, represented by the Doyle–Fuller–Newman (DFN) model (also corresponding to the Porous Electrode Theory (PET) model or Pseudo-Two-Dimensional (P2D) model) [11–14], describe the cycling behavior of battery with considerable accuracy by expressing the physical/chemical phenomena inside the cell, such as the electrochemical reaction kinetics at the solid-electrolyte interface of the porous electrode and the lithium-ion transport through the electrolyte and solid particles, as governing equations. Building high-fidelity battery models is essential for enhancing the reliability of battery model-based analysis. To improve the fidelity of electrochemical models for real cycling behavior, fitting the parameters of the DFN model to battery cycling data is a common procedure.

An important consideration for selecting the optimal model parameters during the parameter identification stage of adjusting the DFN model to be suitable for cycling experiments is the *identifiability analysis*. In many practical applications, deriving the optimal point estimates and quantifying their uncertainties is of significant importance [15]. *Practical* identifiability pertains to whether the data contains enough information to specify the parameters uniquely. If parameters yield similar results over a broad range or exhibit multimodal distributions, such parameters are considered unidentifiable. Previous studies have used linearized identifiability analysis to show that various combinations of effective transport and kinetic coefficients can result in discharge voltage curves that are almost similar. However, the high nonlinearity of the DFN model suggests that linearized models may lead to inappropriate conclusions. Therefore, recent studies apply fully nonlinear quantitative identifiability analysis to provide accurate parameter estimations [16,17].

* Correspondence to: Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E19-551, Cambridge, MA 02139, United States of America.
E-mail address: braatz@mit.edu (R.D. Braatz).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.etran.2025.100486
Received 25 April 2025; Received in revised form 10 September 2025; Accepted 15 September 2025
Available online 22 September 2025

When conducting parameter estimation, both conventional optimization approaches and machine learning (ML) have been employed to analyze and enhance the identifiability of DFN models. Although ML is commonly used for predicting state of health (SOH) and state of charge (SOC), it has also shown promising performance in parameter estimation [18]. Chun et al. [19] developed a deep reinforcement learning (DRL) approach that generates optimal current profiles, thereby enhancing the identifiability of a stoichiometric range and demonstrating its effectiveness across various battery conditions. Laue et al. [20] propose a three-stage parameter estimation procedure, consisting of open-circuit potential (OCP), C-rate tests, and electrochemical impedance spectra, to address issues related to identifiability during the parameter estimation of the DFN model. These results suggest that widely used OCP and C-rate tests are insufficient for complete parameterization, thus necessitating electrochemical experiments like impedance spectroscopy. Deng et al. [21] developed a reduced-order physics-based model to address the issue of reduced efficiency in optimization-based parameter estimation due to the complexity of the model and the multitude of parameters. They estimated parameters and compared their approach with a commercial battery simulator. To differentiate identifiable parameters, they used criteria based on calculating the determinant and condition number of the Fisher Information Matrix (FIM). Kim et al. [22] selected a subset of 15 DFN model parameters through variance-based global sensitivity analysis and performed parameter estimation using discharge C-rate tests. In particular, they applied the Markov Chain Monte Carlo (MCMC) method to quantify parameter uncertainty and analyze identifiability. Bills et al. [23] performed an identifiability analysis on the degradation modes of 22 cells subject to an electric vertical takeoff and landing (EVTOL) aircraft profile. They track parameter trends across the cell lifetimes using a reduced-order model with median errors of 32.5 mV. Ramadesigan et al. [24] demonstrated that the behavior of aged cells can be described by decreased transport and kinetic rate coefficients. They used this approach to quantify the pathways and uncertainties for diffusion coefficients and rate constants.

Our previous works applied nonlinear identifiability analysis of electrochemical battery models to the single discharge voltage curve of a non-aged single cell (see Ref. [17] for PET model application and Ref. [16] for multiphase PET model (MPET) [25] application). Here we extend this work to a dataset with silicon oxide–graphite/nickel cobalt aluminum lithium-ion (NCA/LiC_6–SiO_x) cells from a disassembled Tesla Model 3. Furthermore, this work applies identifiability analysis to describe the cycling behavior of both fresh and aged cells. Degradation diagnosis is divided into two approaches [26–28]: developing and incorporating degradation-related perturbation terms into the battery model, and tracking aging parameters. While degradation-related perturbation terms have been directly integrated into battery models in past studies, this work also analyzes the evolutionary trajectories of aging parameters. The *a posteriori* parameter distributions are estimated at every diagnostic cycle using MCMC. First, MCMC methods are presented and shown to provide global nonlinear identifiability trends for a small subject of parameters. Second, unknown physical properties of the NCA/LiC_6–SiO_x cell, which are estimated experimentally by inverting the model, are discussed. Third, the model specifications are established for the standardized cycling conditions of the fleet of NCA cells. Fourth, results are presented for the identifiability analysis and Bayesian estimation of each cycle with simple parameter fittings as a function of the SOH.

This article is organized as follows: Section 2 provides theoretical background on the DFN model and Bayesian parameter identification. Section 3 describes the parameterization for explaining the cycling behavior of the NCA cell. Section 4 presents the overall scheme for parameter identification and provides details on the NCA cell data. Section 5 represents the main results, which consist of identifiability and degradation diagnosis, and Section 6 summarizes the article.

2. Methods

This section introduces the mathematical representation of the battery model and Bayesian inference used in the degradation diagnosis framework represented by Fig. 1. Section 2.1 provides detailed information on the DFN model and simulation specification. Section 2.2 provides a mathematical description of Bayesian inference for parameter identification.

2.1. DFN model

The DFN model describes the microscopic physicochemical behavior of a lithium-ion battery by considering it as a two-dimensional structure consisting of a negative/positive porous electrode and an electrolyte. The DFN has two dimensions: the x direction, which moves across the length of the cell starting from the negative electrode through to the positive electrode, and the r direction, the distance from the center of a porous electrode particle to its surface. The two porous electrodes are emersed in an electrolyte solution which conducts the flow of lithium ions (see Fig. 2).

Solid-phase concentrations in each electrode control volume follow Fickian diffusion,

$$\frac{\partial c_s(x,r,t)}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{s,i}\frac{\partial c_s(x,r,t)}{\partial r}\right), \tag{1}$$

with Neumann boundary conditions at the center and surface of the particle,

$$\left.\frac{\partial c_s(x,r,t)}{\partial r}\right|_{x,r=0} = 0, \qquad \left.\frac{\partial c_s(x,r,t)}{\partial r}\right|_{r=R_i^p} = -\frac{j(x,t)}{D_{s,i}^{\text{eff}}}, \tag{2}$$

where j is the ionic flux in the electrodes, $D_{s,i}$ is the solid-phase diffusivity, R_i^p is the particle radius, $D_{s,i}^{\text{eff}}$ is the effective solid diffusion coefficient, and the superscript i refers to a section of the cell: $\{n,p\}$ are the positive and negative electrodes respectively, s is the separator section, and $\{a,z\}$ are the current collectors attached to the positive and negative electrodes respectively. The diffusion equation governs the electrolyte concentration,

$$\epsilon_i\frac{\partial c_e(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D_i^{\text{eff}}\frac{\partial c_e(x,t)}{\partial x}\right) + a_i(1-t_+)j(x,t), \tag{3}$$

where ϵ is porosity, D^{eff} is the effective electrolyte diffusion coefficient, c_e is the distribution of electrolyte concentration, a is the surface area to volume ratio of the solid particles, and t_+ is the transference number. The boundary conditions at the electrode-current collector interfaces are

$$\left.\frac{\partial c_e(x,t)}{\partial x}\right|_{x=0} = 0, \qquad \left.\frac{\partial c_e(x,t)}{\partial x}\right|_{x=L} = 0, \tag{4}$$

and the interfacial flux terms are

$$\begin{aligned} -D_n^{\text{eff}}\left.\frac{\partial c_e(x,t)}{\partial x}\right|_{x=L_n^-} &= -D_s^{\text{eff}}\left.\frac{\partial c_e(x,t)}{\partial x}\right|_{x=L_n^+}, \\ -D_s^{\text{eff}}\left.\frac{\partial c_e(x,t)}{\partial x}\right|_{x=L_s^-} &= -D_p^{\text{eff}}\left.\frac{\partial c_e(x,t)}{\partial x}\right|_{x=L_s^+}, \end{aligned} \tag{5}$$

where L is the total length of the cell and L_i is the distance measured from $x=0$ to the end of section i.

Interfacial ionic fluxes couple the two phases in the electrode sections modeled by Butler–Volmer reaction kinetics,

$$j(x,t) = 2k_i^{\text{eff}}\sqrt{c_e(x,t)\left(c_{s,i}^{\max}-c_s^*(x,t)\right)c_s^*(x,t)}\,\sinh\left(\frac{0.5F}{RT(x,t)}\eta(x,t)\right), \tag{6}$$

where k^{eff} is the effective rate constant, $c_s^{\max}$ is the maximum solid concentration, c_s^* is the solid particle surface concentration, T is temperature, η is the overpotential, F is Faraday's constant, and R is the ideal gas constant.

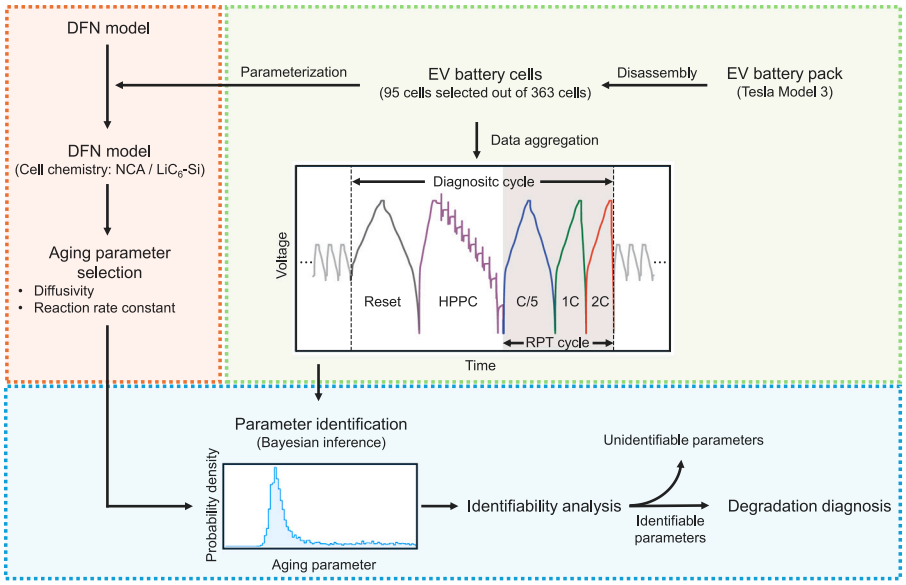

Fig. 1. Schematic diagram of this work on degradation diagnosis using Bayesian parameter identification and identifiability analysis.

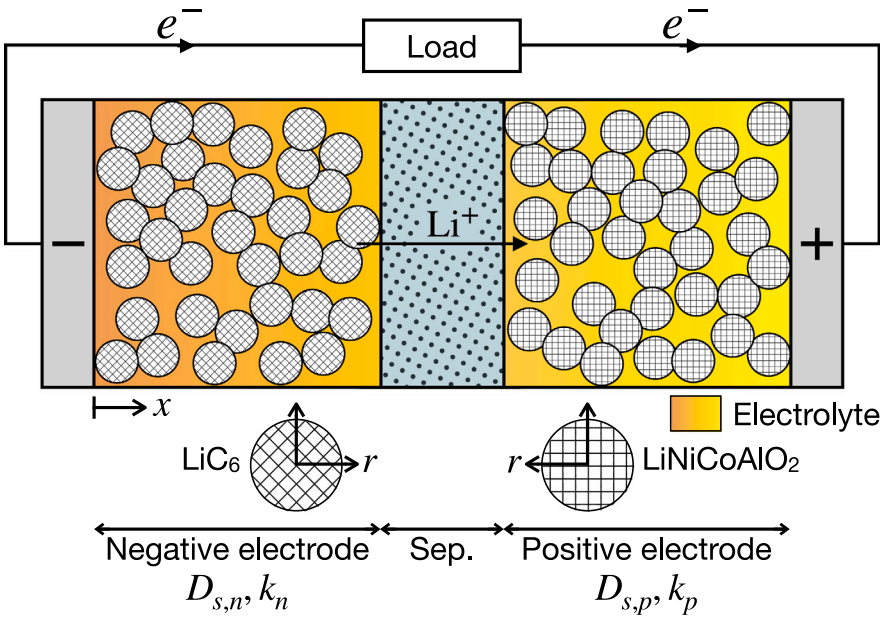

Fig. 2. Schematic of the DFN model for an NCA/LiC_6–SiO_x cell during discharge. The solid diffusion coefficients and reaction rate constants are listed under the sections whose physics they principally affect.

Partial differential equations for the solid and electrolyte potentials, Φ_s and Φ_e, respectively, are

$$\frac{\partial}{\partial x}\left(\sigma_i^{\text{eff}}\frac{\partial \Phi_s(x,t)}{\partial x}\right) = a_i F j(x,t), \tag{7}$$

$$\frac{\partial}{\partial x}\left(\kappa^{\text{eff}}(x,t)\frac{\partial \Phi_e(x,t)}{\partial x}\right) = -a_i F j(x,t) + \frac{\partial}{\partial x}\left(\frac{2\kappa^{\text{eff}}(x,t)RT(x,t)}{F}\left(1-t_+\right) \times \frac{\partial \ln c_e(x,t)}{\partial x}\right), \tag{8}$$

where σ^{eff} and κ^{eff} are the effective solid and electrolyte conductivities, respectively. Boundary conditions for Φ_s incorporate the applied current $I(t)$,

$$\left.\frac{\partial \Phi_s(x,t)}{\partial x}\right|_{x=0} = -I(t), \qquad \left.\frac{\partial \Phi_s(x,t)}{\partial x}\right|_{x=L} = -I(t), \tag{9}$$

$$\left.\frac{\partial \Phi_s(x,t)}{\partial x}\right|_{x=L_n} = 0, \qquad \left.\frac{\partial \Phi_s(x,t)}{\partial x}\right|_{x=L_s} = 0. \tag{10}$$

The boundary conditions for Φ_e mirror those for c_e,

$$\Phi_e(x,t)\big|_{x=0} = 0, \qquad \left.\frac{\partial \Phi_e(x,t)}{\partial x}\right|_{x=L} = 0, \tag{11}$$

$$\begin{aligned} -\kappa_n^{\text{eff}}\left.\frac{\partial \Phi_e(x,t)}{\partial x}\right|_{x=L_n^-} &= -\kappa_s^{\text{eff}}\left.\frac{\partial \Phi_e(x,t)}{\partial x}\right|_{x=L_n^+}, \\ -\kappa_s^{\text{eff}}\left.\frac{\partial \Phi_e(x,t)}{\partial x}\right|_{x=L_s^-} &= -\kappa_p^{\text{eff}}\left.\frac{\partial \Phi_e(x,t)}{\partial x}\right|_{x=L_s^+}, \end{aligned} \tag{12}$$

with the reference potential grounded to 0 V at $x = 0$. For further details of the governing equations, see Ref. [13]. We solve the equations using PETLION [29], which is an efficient open-source DFN simulation tool

in the Julia programming language.[2] PETLION discretizes the PDEs in space using the Finite Volume Method (FVM) and solves forward in time using the Method of Lines (MoL) [30]. The number of FVM control volumes (N) and solver tolerances (Δ_{abs}: absolute value, Δ_{rel}: relative value) affect the desired accuracy and computational complexity of the simulation. All solver settings are set to defaults ($N = 10$, $\Delta_{\text{abs}} = 10^{-6}$, $\Delta_{\text{rel}} = 10^{-3}$), as increasing the number of control volumes and decreasing the tolerance significantly increased the computational budget without meaningfully improving accuracy [1].

2.2. Bayesian parameter identification and identifiability

Parameter identification through Bayesian inference is a method for quantifying the uncertainty about estimated parameters [31,32]. Bayesian inference, based on Bayes' theorem, is explained by the relationship between the *posterior* distribution $\mathrm{P}(\theta|Y)$, *prior* distribution $\mathrm{P}(\theta)$, *likelihood* $\mathrm{P}(Y|\theta)$, and *marginal likelihood* $\mathrm{P}(Y)$,

$$\mathrm{P}(\theta|Y) = \frac{\mathrm{P}(Y|\theta)\,\mathrm{P}(\theta)}{\mathrm{P}(Y)}, \tag{13}$$

where Y is a vector of voltage measurements y_j. To estimate the most promising subset of parameters, the *Maximum A Posteriori* (MAP) is obtained by maximizing the posterior probability,

$$\theta^*_{\text{MAP}} = \underset{\theta}{\text{argmax}}\ \mathrm{P}(\theta|Y) = \underset{\theta}{\text{argmax}}\ \mathrm{P}(Y|\theta)\mathrm{P}(\theta), \tag{14}$$

where $\mathrm{P}(Y)$ is a normalization constant which may be ignored when solving the maximization. In practice, it is common to instead minimize a negative log transformation of Eq. (14),

$$\theta^*_{\text{MAP}} = \underset{\theta}{\text{argmin}}\ \ln \mathrm{P}(Y|\theta) + \ln \mathrm{P}(\theta), \tag{15}$$

which gives the same parameter value. If the prior distribution is a uniform or normal distribution with an infinitely large variance, the influence of the prior distribution can be neglected and the equation becomes the *Maximum Likelihood Estimate* (MLE),

$$\theta^*_{\text{MLE}} = \underset{\theta}{\text{argmin}}\ \ln \mathrm{P}(Y|\theta). \tag{16}$$

The MLE for a model with Gaussian error simplifies to the familiar sum of squared residuals,

$$\theta^*_{\text{MLE}} = \underset{\theta}{\text{argmin}} \sum_{j=1}^{N_d} \left(\frac{y_j - \hat{y}_j(\theta)}{\sigma_\epsilon} \right)^2. \tag{17}$$

The uncertainty associated with the estimated parameters are quantified through the confidence regions. The DFN model exhibits a highly nonlinear relationship between parameters and model predictions, which suggests that the confidence regions may not be a hyperellipsoid [33]. Linearization approaches may produce highly inaccurate estimates for highly nonlinear systems [16,17]. Nonlinear systems can be handled by relating the optimal estimates of parameters derived from the log-likelihood function to the chi-squared distribution [34]. With the definition

$$\chi^2(\theta) = \sum_{j=1}^{N_d} \left(\frac{y_j - \hat{y}_j(\theta)}{\sigma_\epsilon} \right)^2, \tag{18}$$

the nonlinear confidence region R_α can be expressed as all θ that satisfy the inequality

$$R_\alpha = \left\{ \theta : \chi^2(\theta) - \chi^2(\theta^*) \le \chi^2_{N_p}(1-\alpha) \right\}, \tag{19}$$

where $\chi^2_{N_p}$ is the chi-squared distribution with N_p degrees of freedom, N_p is the number of parameters, and α is the significance level (e.g., an α of 0.01 corresponds to a 99% confidence region).

[2] Due to the Just-in-Time (JIT) compilation of the Julia language, the first execution of a single discharge cycle in PETLION takes about 11 s, but subsequent executions only take about 3 ms [29].

MCMC is an efficient approach for uncertainty evaluation for highly nonlinear systems. MCMC employs the Metropolis–Hastings algorithm, which can sample from complex high-dimensional PDFs [35,36]. Parameters are initiated at an initial value, θ_0, and random perturbations are introduced to these parameters [37]. During each iteration t, an objective function $f(\theta)$, such as the sum of squared residuals, is calculated. The suggested parameter θ' is then accepted or rejected based on an acceptance ratio, serving as the criterion for determining whether to adopt the next parameter set.

3. Parameterization: Physical properties of NCA cell

The first step to performing an identifiability analysis for the NCA cells is accurately modeling the cell. The properties of the electrolyte are estimated as a function of electrolyte concentration and temperature. The cells have a small amount of silicon oxide in the anode, which differentiates them from conventional chemistries in the literature. To address this, the OCV functions for each electrode are regressed in cycling experiments of half-cells containing positive and negative electrodes with C/50 charge and C/60 discharge. Then, the half-cell OCVs are regressed against the full-cell OCV using nonlinear optimization to estimate the stoichiometry limits. The estimated functions for the properties of the electrolyte and electrode are presented in detail in Sections 3.1 and 3.2, respectively.

The DFN model parameters used for the pristine cell simulation are listed in Table 1. The reaction rate constants and diffusion coefficients at each electrode are estimated through identifiability analysis, and the other parameters are considered constants throughout the cell lifetime.

3.1. Electrolyte

The electrolyte conductivity κ and diffusivity D are tabulated as functions of electrolyte concentration c_e and temperature T using the Advanced Electrolyte Model [38] for an EC/EMC/DMC electrolyte mix. Empirical equations for $\kappa(c_e, T)$ and $D(c_e, T)$ are fit using response surface methodology [39] (see Tables 2 and 3),

$$\kappa(c_e, T) = \sum_{i=0}^{3}\sum_{j=0}^{1} a_{ij}(c_e^i T^j), \tag{20}$$

$$D(c_e, T) = \sum_{i=0}^{4}\sum_{j=0}^{2} a_{ij}(c_e^i T^j). \tag{21}$$

3.2. Electrode

The anode is graphite doped with a silicon oxide, producing a bimodal particle radius distribution with peaks in different regions for graphite and silicon oxide particles [40,41]. That is, although even small amounts of silicon oxide dopant in graphite lead to significant structural differences, for simplicity of calculations, we assume the anode is homogeneous, as a single particle with a radius of 16 μm.

The positive and negative OCVs are estimated with charge and discharge cycles at 0.1 and 0.2 mA, respectively (about C/60 and C/50). The OCVs are fit with empirical equations as a function of solid lithium concentration,

$$U_n(\Theta_n) = a_0 + a_1 \exp\left(\frac{\Theta_n - b_1}{c_1} \right) + \sum_{i=2}^{4} a_i \tanh\left(\frac{\Theta_n - b_i}{c_i} \right), \tag{22}$$

$$U_p(\Theta_p) = \sum_{i=1}^{8} a_i \exp\left(-\left(\frac{\Theta_p - b_i}{c_i} \right)^2 \right), \tag{23}$$

$$\Theta_i = c^*_{s,i}/c^{\max}_{s,i}, \tag{24}$$

where U_i is open-circuit voltage, Θ_i is the stoichiometry, and $c^*_{s,i}$ is solid-phase surface concentration. The coefficients for the empirical equations representing the open-circuit voltages of the negative electrode (U_n) and positive electrode (U_p) are detailed in Tables 4 and 5, respectively.

Table 1
Description of the parameter set of the NCA/LiC_6–Si cell.

Parameter	Unit	Description	Cathode current collector	NCA	Separator	LiC_6–Si	Anode current collector
$D_{s,i}$	m^2/s	Solid-phase diffusivity	–	8.716×10^{-14}	–	1.018×10^{-13}	–
k_i	$m^{5/2}/(mol^{1/2}s)$	Reaction rate constant	–	4.438×10^{-10}	–	6.837×10^{-12}	–
l_i	m	Thickness	1.0×10^{-5}	6.4×10^{-5}	1.0×10^{-5}	8.3×10^{-5}	1.0×10^{-5}
ϵ_i	–	Porosity	–	0.230	0.359	0.147	–
D_i	m^2/s	Electrolyte diffusivity	–	5.0×10^{-10}	5.0×10^{-10}	5.0×10^{-10}	–
R_i^p	m	Particle radius	–	1.1×10^{-5}	–	1.6×10^{-5}	–
$c_{e,i}^{init}$	mol/m^3	Initial concentration in the electrolyte	–	1200	1200	1200	–
$c_{s,i}^{max}$	mol/m^3	Maximum solid-phase concentration	–	54,422	–	28,967	–
ρ_i	kg/m^3	Density	2700	2500	1100	2500	8940
$C_{p,i}$	J/(kg K)	Specific heat	897	700	700	700	385
λ_i	W/(m K)	Thermal conductivity	237	2.1	0.16	1.7	401
σ_i	S/m	Solid-phase conductivity	3.55×10^7	100	–	100	5.96×10^7
$\epsilon_{s,i}$	–	Active material fraction	–	0.745	–	0.828	–
Brugg	–	Bruggeman coefficient	–	1.5	1.5	1.5	–
t_+	0.455	Transference number	–	–	–	–	–
$E_a^{D_i^s}$	J/mol	Solid-phase diffusion activation energy	–	5000	–	5000	–
$E_a^{k_i}$	J/mol	Reaction constant activation energy	–	5000	–	5000	–
Θ_i^{max}	–	Maximum stoichiometry limits	–	0.160	–	0.923	–
Θ_i^{min}	–	Minimum stoichiometry limits	–	0.859	–	0.014	–
T_{amb}	298.15 K	Ambient temperature	–	–	–	–	–
F	96485 C/mol	Faraday's constant	–	–	–	–	–
R	8.314472 J/(mol K)	Universal gas constant	–	–	–	–	–

Table 2
Coefficient (a_{ij}) for electrolyte conductivity ($\kappa(c_e, T)$) in Eq. (20).

	$j = 0$	$j = 1$
$i = 0$	-5.182×10^{-1}	1.696×10^3
$i = 1$	-6.518×10^{-3}	3.034×10^{-5}
$i = 2$	1.446×10^{-6}	-1.049×10^{-8}
$i = 3$	3.047×10^{-10}	0

Table 3
Coefficient (a_{ij}) for electrolyte diffusion coefficients ($D(c_e, T)$) in Eq. (21).

	$j = 0$	$j = 1$	$j = 2$
$i = 0$	1.864×10^{-8}	-1.392×10^{-10}	2.633×10^{-13}
$i = 1$	0	3.133×10^{-14}	-1.126×10^{-16}
$i = 2$	0	-7.301×10^{-17}	2.615×10^{-19}
$i = 3$	0	5.120×10^{-20}	-1.832×10^{-22}
$i = 4$	0	-1.151×10^{-23}	4.111×10^{-26}

Table 4
Coefficients (a_i, b_i, and c_i) for negative open-circuit voltage ($U_n(\Theta_n)$) in Eq. (22).

	a_i	b_i	c_i
$i = 0$	−48.99	–	–
$i = 1$	29.98	5.700×10^{-3}	-5.093×10^{-2}
$i = 2$	161.9	-1.057×10^{-1}	9.687×10^{-2}
$i = 3$	-2.833×10^{-1}	4.447×10^{-2}	4.235×10^{-2}
$i = 4$	−47.77	−18.95	7.041
$i = 5$	−65.06	2.268×10^{-3}	1.160×10^{-3}

Table 5
Coefficients (a_i, b_i, and c_i) for positive open-circuit voltage ($U_p(\Theta_p)$ in Eq. (23).

	a_i	b_i	c_i
$i = 1$	1.456×10^{-1}	7.961×10^{-1}	6.035×10^{-2}
$i = 2$	4.205×10^{-1}	9.489×10^{-1}	4.229×10^{-2}
$i = 3$	1.008	6.463×10^{-1}	1.034×10^{-1}
$i = 4$	1.350	7.378×10^{-1}	9.513×10^{-2}
$i = 5$	2.526	2.953×10^{-1}	2.019×10^{-1}
$i = 6$	2.636	5.372×10^{-1}	1.758×10^{-1}
$i = 7$	3.285	8.922	1.414×10^{-1}
$i = 8$	172.1	−1.344	7.371×10^{-1}

4. Problem description

Section 4.1 introduces detailed procedures for parameter identification and identifiability analysis. Section 4.2 provides a description of the selected aging parameters and a detailed description of the NCA dataset.

4.1. Bayesian estimation and identifiability procedure

The analysis of nonlinear identifiability has three steps: (1) estimating the posterior distribution of θ using the Metropolis–Hastings algorithm, (2) distinguishing between practically identifiable and unidentifiable parameters based on the probability densities, and (3) further categorizing the identifiable combinations through a gridded mesh showing the confidence regions. Once the parameter space is comprehensively mapped, a set of identifiable parameter groups is defined, which includes all identifiable and locally identifiable parameters.

The parameter space can be reduced by fixing unidentifiable parameters to an estimate or to physically meaningful upper and lower bounds. Likewise, equations considered insignificant because of the unidentifiable parameters can be excluded from the model to enhance computational efficiency. For instance, in the case of a very thin porous electrode that is not diffusion-limited, the effects of diffusion within the porous electrode could be disregarded in the model. Alternatively, a less restrictive approach would involve incorporating prior values from the literature.

Practical identifiability is confirmed through the results obtained from Bayesian estimation. A parameter is deemed practically identifiable if a sufficiently large number of chains indicate that it is bounded. Conversely, if the chains include parameter values that can be arbitrarily large or small, then the parameter is practically unidentifiable; further investigation is needed to determine whether there are any identifiable combinations. All identifiable combinations encompass every practically identifiable parameter.

Locally identifiable parameters that contribute to *identifiable combinations* are first evaluated using their probability densities. A parameter is likely part of the identifiable combinations if its probability density exhibits (1) a large peak and (2) a lower magnitude plateau at extreme values. The significant peak arises from identifiable combinations that include the parameter, while the plateau corresponds to the identifiable combinations that exclude it, indicating that the parameter is unidentifiable. Parameters with a uniform distribution do not belong

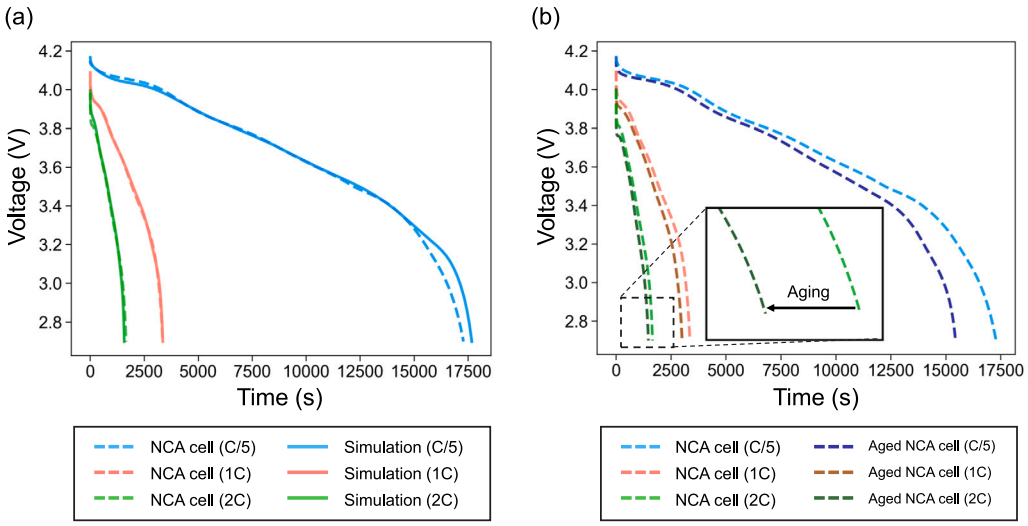

Fig. 3. (a) Generation of high-fidelity battery models through parameterization and comparison with real experiments and (b) comparison of cycling behavior of an aged cell and pristine cell.

to any identifiable combination sets. A detailed explanation of the methodology regarding identifiability, along with a simple example, is specified in previous work [16,17].

4.2. Model specifications

This article considers the key transport and kinetic parameters, $\theta = [D_{s,p}, D_{s,n}, k_p, k_n]^\top$, where $D_{s,p}$ and $D_{s,n}$ are the solid-phase diffusion coefficients of lithium in the cathode and anode respectively, and k_p and k_n are the electrochemical reaction rate constants for the cathode and anode respectively. In addition, transport and kinetic parameters were identified as key parameters in global sensitivity analysis using Sobol' indices (see Appendix B). These parameters lump the effects of multiple true material properties together [11]. The parameter identification is performed on a logarithmic basis of θ, which is a standard approach to improving numerical convergence for parameters that can vary by orders of magnitude.

The cycling data used for the identifiability analysis consisted of a representative sample of 95 cells, selected from a total of 363 cylindrical 21 700 NCA cells extracted from a Tesla Model 3 provided by van Vlijmen et al. [42]. Diagnostic cycles consisting of a reset cycle, hybrid pulse power characterization (HPPC), and reference performance test (RPT) cycles at three discharge C-rates (C/5, 1C, and 2C) are performed every 100 cycles for the whole lifetime of the 95 cells to measure the capacity fade (see Ref. [42] for a detailed description of the NCA cell dataset). Only the RPT cycles among the diagnostic cycles are used for the identifiability analysis. The discharge voltage curve used for parameter identification is represented with high fidelity in battery simulations by applying the DFN model parameters identified in Section 3 (Fig. 3a). As cycling continues, the capacity irreversibly decreases (Fig. 3b). Four parameters ($D_{s,p}$, $D_{s,n}$, k_p, k_n) are estimated through the discharge curves from all RPTs of each cell, and the trajectory of each parameter is tracked as aging progresses.

5. Results and discussion

5.1. Parameter identifiability

Visualizing the nonlinear confidence region can help interpret identifiability trends. A nonlinear confidence region depicts the error (e.g., the sum of squared residuals, chi-squared statistic, root-mean-squared error (RMSE)) as a function of the N_p-dimensional parameter space. Typically, the parameter space is gridded with sufficiently fine discretizations to show detailed resolution of the confidence region. Confidence regions of linear or linearized models depict hyperellipsoid confidence regions centered on θ^* [17,43], but nonlinear confidence regions are not restricted to hyperellipsoid shapes. Highly nonlinear models (such as PET-based battery models) have been shown to exhibit *banana-shaped* confidence regions in which an infinite number of parameter values give either the same or nearly the same quality of fit.

Fig. 4 describes an example confidence region based on the function of the cathode diffusion coefficient $D_{s,p}$ and the rate constant k_p. Both $D_{s,p}$ and k_p are *practically unidentifiable* because the darkly shaded extends towards $+\infty$. The parameters appear to be locally identifiable. The numerical optimization for the initial guesses converges to a point on the minimum curve that is a local minimum for k_p for fixed $D_{s,p}$, and also a point on the minimum curve that is a local minimum for $D_{s,p}$ for fixed k_p. As shown in Fig. 4, the confidence region as a function of the two parameters shows that the two parameters are not globally identifiable. The two extremes of the minimum curve, where $k_p \to \infty$ and $D_{s,p} \to \infty$ respectively, show very different sensitivities on k_p and $D_{s,p}$. The sensitivity to a parameter for the parameter identification objective can be very large or close to zero, depending on the values of other parameters. This observation has strong parameter identification implications and implies that relying only on local sensitivities can lead to misleading results.

As N_p increases, gridding the confidence region becomes prohibitively expensive as the number of model evaluations grows with the order of power N_p. The same identifiability trends can be interpreted through MCMC after a sufficient number of iterations, which exhibits better scaling. Fig. 5 shows the development of the posterior distribution from MCMC after various samples following a burn-in of 500, 2000, and 5000 samples. At 5000 samples (Fig. 5cf), the approximate posterior distributions closely resemble the true posterior distributions (where the true distribution was estimated by running MCMC for 10^6 iterations). The same local identifiability trends found with the gridded confidence region can also be interpreted with the posterior distributions from MCMC. Both parameters are practically unidentifiable from the posterior distribution because the distributions plateau towards $+\infty$. The parameters are also locally identifiable because of the prominent peak near their lower bounds.

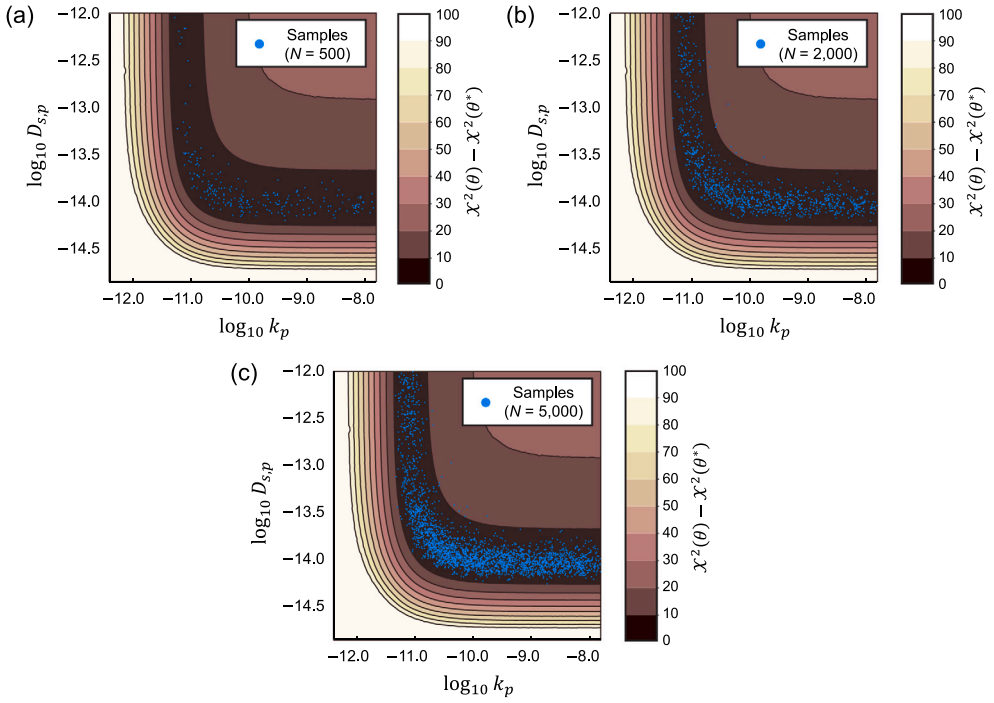

Fig. 4. MCMC sampling of the 2-dimensional confidence region: (a) 500 samples, (b) 2000 samples, (c) 5000 samples.

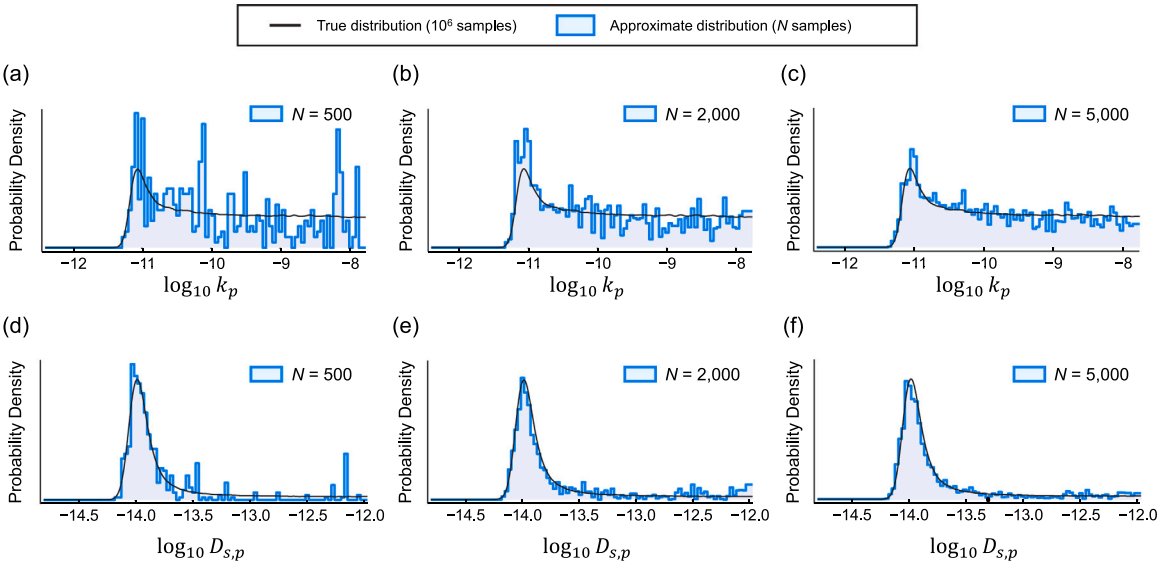

Fig. 5. Estimation of the posterior distribution by sampling the parameter space. After a few hundred iterations, the approximate posterior distributions resemble the true distributions with some noise. The true PDF was estimated by sampling the confidence region for 1,000,000 iterations: k_p through (a) 500 samples, (b) 2000 samples, (c) 5000 samples, and $D_{s,p}$ through (d) 500 samples, (e) 2000 samples, (f) 5000 samples.

A parameter identifiability analysis is performed for the diagnostic cycle of every cell using the C/5, 1C, and 2C discharge curves. These trends change as the battery degrades (Fig. 6). For a pristine cell, $D_{s,n}$ is the only identifiable parameter as its PDF is completely contained within an enclosed region, k_n is practically unidentifiable as its upper bound approaches infinity despite the large peak near the lower bound (i.e., locally identifiable), and both cathode parameters $D_{s,p}$ and k_p are practically unidentifiable. These identifiability trends are consistent with previous articles for LCO discharge curves [17,24]. After cycling and degrading the cell, the identifiable parameters are now $D_{s,n}$, $D_{s,p}$, and k_n while k_p remains practically unidentifiable. The mean of each identifiable parameter tends to decrease as a function of the cycle number. The confidence interval of each identifiable parameter tightens as capacity fade increases.

The diffusion and kinetic coefficients are directly related to the movement of Li^+ ions in the cell. Large values of $D_{s,i}$ and k_i correspond to faster movement of lithium inside the solid particles and in intercalation, respectively, and small values correspond to slow movement of lithium. In particular, it was demonstrated through an impedance model that both solid-phase and electrolyte diffusivity contribute significantly to degradation [44]. Therefore, the inverse of these coefficients, $1/D_{s,i}$ and $1/k_i$, can be interpreted as *resistances* to lithium flow in these cell sections. Unidentifiable parameters, whose upper bound approaches infinity, have zero resistance, and identifiable

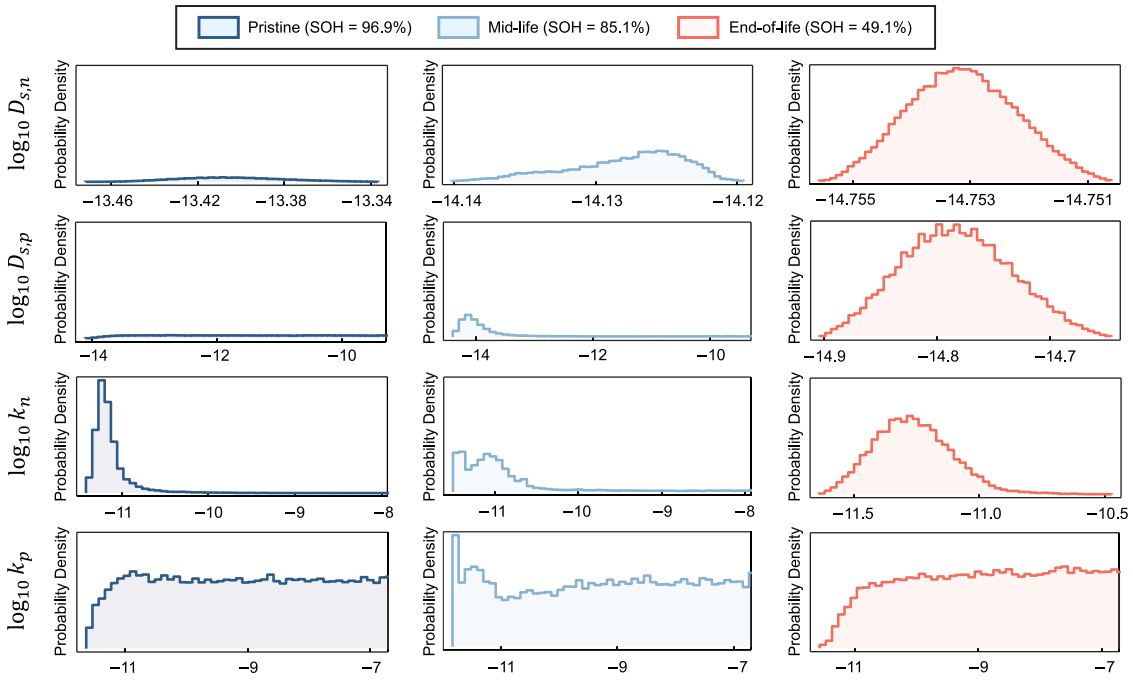

Fig. 6. Changing identifiability trends as the cell degrades. For a pristine cell, only $D_{s,n}$ is identifiable. At end-of-life, $D_{s,n}$, $D_{s,p}$, and k_n become identifiable while k_p remains unidentifiable.

Table 6
Best estimation (θ^*) and bounds of $\log_{10} D_{s,n}$, $\log_{10} D_{s,p}$, and $\log_{10} k_n$ for a single cell at selected cycles. Practically unidentifiable parameters have upper bounds of $+\infty$.

C-rate	SOH (%)	Cycle (RPT cycle)	$\log_{10} D_{s,n}$			$\log_{10} D_{s,p}$			$\log_{10} k_n$		
			θ^*	Lower bound	Upper bound	θ^*	Lower bound	Upper bound	θ^*	Lower bound	Upper bound
C/5	96.89	3 (1)	−13.4123	−13.4691	−13.3386	−11.7891	−14.1016	+∞	−11.2356	−11.4084	−10.7117
	92.03	353 (5)	−13.8416	−13.8578	−13.8201	−13.8258	−14.2561	+∞	−11.0996	−11.4488	−10.5916
	85.15	773 (9)	−14.1272	−14.1349	−14.1198	−14.0647	−14.3874	+∞	−11.0947	−11.4921	−10.4708
	73.11	1193 (13)	−14.3990	−14.4027	−14.3932	−14.4715	−14.6339	−14.1988	−10.9684	−11.4155	−10.2678
	60.44	1508 (16)	−14.5986	−14.6011	−14.5959	−14.6047	−14.7392	−14.4277	−11.1708	−11.5224	−10.7465
1C	94.00	4 (1)	−12.9504	−12.9833	−12.9032	−13.2378	−13.6134	+∞	−11.0926	−11.4292	−10.7501
	89.02	354 (5)	−13.2537	−13.2737	−13.2276	−9.5828	−13.7458	+∞	−11.6549	−11.7986	−10.9207
	80.55	774 (9)	−13.5330	−13.5443	−13.5213	−13.6734	−13.9970	+∞	−11.3667	−12.0971	−10.7626
	56.37	1194 (13)	−13.9357	−13.9441	−13.9196	−14.3648	−14.4676	−14.2102	−11.5877	−12.0410	−11.2227
	38.12	1509 (16)	−14.2170	−14.2235	−14.2116	−14.4165	−14.5365	−14.2574	−11.8804	−12.3274	−11.4914
2C	93.07	5 (1)	−12.5907	−12.6521	−12.4817	−12.7455	−13.2445	+∞	−11.5614	−11.8172	−11.2223
	86.38	355 (5)	−13.0125	−13.0387	−12.9654	−13.0800	−13.4779	+∞	−11.8226	−12.2882	−11.3668
	76.83	775 (9)	−13.2886	−13.3021	−13.2669	−13.5018	−13.7450	−13.0584	−11.8137	−12.3387	−11.3409
	46.50	1195 (13)	−13.7453	−13.7592	−13.7078	−14.0299	−14.1465	−13.9412	−13.1056	−13.3861	−12.8286
	24.40	1510 (16)	−14.0743	−14.1210	−13.9907	−14.5462	−14.6372	−14.4453	−13.5577	−13.8075	−13.3000

parameters have a unique, non-zero resistance. The series resistance of identifiable parameter groupings (e.g., $\{D_{s,p}, k_n\}$) have a meaningful and identifiable *total resistance* relationship, e.g.,

$$R_{\text{tot}} \propto \frac{\alpha}{D_{s,p}} + \frac{\beta}{k_n}, \tag{25}$$

where α, β are positive fitting constants [17]. The changing identifiability trends (Fig. 6) can be interpreted by changing resistances in the cell — anode resistances are dominant in early life but, as the cell degrades, cathode resistances become measurable (i.e., identifiable). Numerous studies have investigated the loss of active material (LAM) in each electrode and loss of lithium inventory (LLI) during cycling [45–48]. Studies have shown not only LLI but also anode LAM (LAM_{NE}) act as the dominant degradation mechanism, with the anode acting as the limiting electrode. As battery usage increases, the cathode LAM (LAM_{PE}) can become the primary degradation mechanism, causing a knee-point in the capacity fade curve (Table 6). See Appendix C for the relationship between LAM, LLI, and parameters.

5.2. Degradation diagnosis

A past study [24] found that, for the same set of four parameters considered in this article plus electrolyte diffusivity D, capacity fade could be predicted for future cycles while only regressing $D_{s,n}$ and k_n. Large uncertainties observed for $D_{s,p}$, D, and k_p were addressed by fixing their values to be constants, and reductions in the estimated $D_{s,n}$ and k_n with cycle number were observed to follow a power law. Although the approach in [24] accurately predicted future voltage discharge curves for one cell, the above nonlinear identifiability results indicate that empirical fits only to anode parameters may not apply to a broad range of cells that exhibit competing degradation mechanisms.

At all levels of degradation, k_p is unidentifiable (non-rate limiting) and may be replaced with a sufficiently large constant value ($k_p = 10^{-7}$ $\text{m}^{5/2}/\text{mol}^{1/2}\text{s}$). At low to moderate levels of degradation, $D_{s,p}$ and k_n are locally identifiable. By adding a weak Gaussian prior to $\log_{10} D_{s,p}$,

$$P(\log_{10} D_{s,p}) = \mathcal{N}(-15.2, 1^2), \tag{26}$$

Table 7
Fitted equations for the set of diffusion and kinetic parameters as a function of SOH in units of % (see Fig. 7).

Fitted parameter, $\hat{\theta}$(SOH)	Discharge C-rate	Fitted equation
$\log_{10}\hat{D}_{s,n}$(SOH)	C/5	$-14.75 + 0.6757\,\mathrm{atanh}(2.067(\mathrm{SOH}-50))$
	1C	$-14.04 + 0.6581\,\mathrm{atanh}(2.119(\mathrm{SOH}-50))$
	2C	$-13.71 + 0.6368\,\mathrm{atanh}(2.190(\mathrm{SOH}-50))$
$\log_{10}\hat{D}_{s,p}$(SOH)	C/5	$-16.11 + 2.684(\mathrm{SOH})$
	1C	$-15.33 + 2.504(\mathrm{SOH})$
	2C	$-15.03 + 2.561(\mathrm{SOH})$
$\log_{10}\hat{k}_{n}$(SOH)	C/5	$-12.18 + 1.135(\mathrm{SOH})$
	1C	$-14.36 + 3.409(\mathrm{SOH})$
	2C	$-15.68 + 4.502(\mathrm{SOH})$

we shift the median of the posterior distributions of $D_{s,p}$ and k_n towards their lower bounds without greatly affecting the error. Then, we regress the remaining identifiable set, $\{D_{s,n}, D_{s,p}, k_n\}$, for each cycle. The MCMC results for each cycle (Fig. 7) are fit to empirical equations using weighted least squares, where the weight is the reciprocal of the variance. $\log_{10} D_{s,n}$ is by far the most well-behaved across all C-rates and levels of degradation with low uncertainty, while $\log_{10} D_{s,p}$ and $\log_{10} k_n$ have more variation. This observation can also be explained in terms of the diffusion model. Solid particle diffusion, described by Fick's law (Eq. (1)), can lead to high computational demands. Therefore, reduced-order models such as the two-parameter polynomial approximation [49] expressed by

$$\frac{\partial c_s^{\mathrm{avg}}(x,t)}{\partial t} = -3\frac{j(x,t)}{R_i^p}, \tag{27}$$

$$c_s^*(x,t) - c_s^{\mathrm{avg}}(x,t) = -\frac{R_i^p}{\mathbf{D}_{s,i}^{\mathrm{eff}}}\frac{j(x,t)}{5} \tag{28}$$

can be applied at low or moderate C-rates (e.g., less than 2C). An increase in R_n^p is regarded as a parameter that results from the irreversible growth of the Solid Electrolyte Interphase (SEI) layer [50,51]. The notion that the growth of the SEI layer dominates other degradation mechanisms supports the high contribution of $D_{s,n}$ [52].

Starting at large SOH, $\log_{10} D_{s,n}$ quickly slopes downwards before entering a linear regime in SOH. The fit across SOH is well-approximated with an arctangent function (Table 7). The trends for $\log_{10} D_{s,p}$ and $\log_{10} k_n$ are noisier than the $\log_{10} D_{s,n}$ trend. The difference in uncertainty is likely $\log_{10} D_{s,n}$ encodes the dominant degradation mechanism, while the individual cell-to-cell variation appears in the optimized parameters for $\log_{10} D_{s,p}$ and $\log_{10} k_n$. At large discharge capacities, $\log_{10} D_{s,p}$ has a large level of uncertainty at its upper bound, which is consistent with the identifiability analysis for pristine and mid-life cells (Fig. 6) where $D_{s,p}$ is unidentifiable at low levels of degradation. The mean of $\log_{10} D_{s,p}$ is closer to its lower bound due to the weak prior in Eq. (26), which nudges the posterior distribution closer to the lower bound. As the cells reach about 60%–70% SOH, $D_{s,p}$ becomes identifiable. Broadly, $D_{s,p}$ and k_n decrease as degradation increases, representing greater internal resistances in the cell. The relationships for $\log_{10}\hat{D}_{s,p}$(SOH) and $\log_{10}\hat{k}_n$(SOH) are approximated with linear fits (Table 7).

The degrees of freedom of optimizing the parameters for every cycle scale with the number of cycles (about 2500) multiplied by the number of regressed parameters (3) – about 7500 degrees of freedom for each C-rate. In contrast, the degrees of freedom for the fitted parameters are significantly smaller — only 7 for the three equations for each C-rate (see Table 7). It is expected that the sets of optimized parameters, θ^*, will have uniformly smaller errors than those using the fitted parameter relationships, $\hat{\theta}$(SOH). Fig. 8 shows the error histograms for the three discharge C-rates with θ^* and $\hat{\theta}$(SOH). On average, the RMSEs increase by 35% when using $\hat{\theta}$(SOH) compared to θ^*, which is acceptable given the significantly smaller degrees of freedom with $\hat{\theta}$(SOH). Still, $\hat{\theta}$(SOH) is unable to capture significant variation in particular cells and cycles — the RMSE standard deviations for $\hat{\theta}$(SOH) increase by a factor of 2–3 compared to θ^*. The fitted parameters produce greater errors as the capacity fade increases, whereas the optimized parameters produce errors that do not greatly change with capacity fade. One possible explanation is that greater cell-to-cell variation appears as the cells become degraded, leading to greater deviation from the mean as capacity fade increases.

5.3. Beyond the DFN model

The DFN model provides a satisfactory coarse-grained description of real data because it includes more physics than the commonly used equivalent circuit models or single-particle models. However, we acknowledge that the DFN model still causes discrepancies with real data because it does not consider a number of physical aspects of the batteries for simplicity:

1. Staging phase separation in graphite, which leads to non-uniform lithium concentration distributions (in both MPET simulations and experimental imaging [25,53,54]) that bear little resemblance to the assumed shrinking core of the DFN model, except when diffusion dominates at high rates [55].
2. Lithium plating and SEI growth [54,56–58], the dominant side reactions in graphite anodes, and oxidation-induced cation disordering [59,60], an important degradation mechanism for layered-oxide cathodes, which are only indirectly modeled by re-fitting parameters with aging.
3. Coupled ion-electron transfer (CIET) kinetics of lithium intercalation [61,62], which can differ significantly from Butler–Volmer kinetics in the DFN model at extremes values of state of charge (perhaps explaining the larger activation overpotential in experiments compared to DFN simulation near the end of C/5 discharge in Fig. 3a).
4. Hybrid porous electrode theory [63], which accounts for significant electrochemical differences between silicon oxide and graphite in the composite anode, leading to nonuniform charging of the two components in both space and time during each cycle.
5. The Many-Particle DFN [64], which captures concentration heterogeneity and provides a high-fidelity description of the diffusion process, can represent voltage relaxation.

Future work could generalize our analysis to capture some of the missing physics using Hybrid MPET, which has recently been applied to similar electrode materials [63]. This could lead to different, more realistic values of the model parameters, as well as potentially improved aging predictions, albeit at the cost of greater computational complexity.

6. Conclusion

In this article, the trajectories of the diffusion coefficient and the reaction rate constant at each electrode over the lifetime are identified via Bayesian inference, and their functional relationship with the SOH is analyzed. A nonlinear identifiability analysis was performed using data across the lifetime of 95 NCA/LiC_6-SiO_x cells from a disassembled Tesla Model 3. 7776 diagnostic cycles were evaluated for C/5, 1C, and 2C discharges. Bayesian inference was performed with the DFN model and diffusion/kinetic parameters at each electrode using the MCMC algorithm. Histograms produced from MCMC are used to establish parameter confidence intervals. At low levels of degradation, only the anode solid diffusion coefficient could be uniquely identified from voltage discharge curves, indicating early anode-dominated degradation. At about 60%–70% SOH, the cathode diffusion coefficient and the anode reaction rate constant become identifiable, indicating that anode and cathode degradation pathways become significant and measurable later in life. Capacity fade is predicted by empirical models with two or

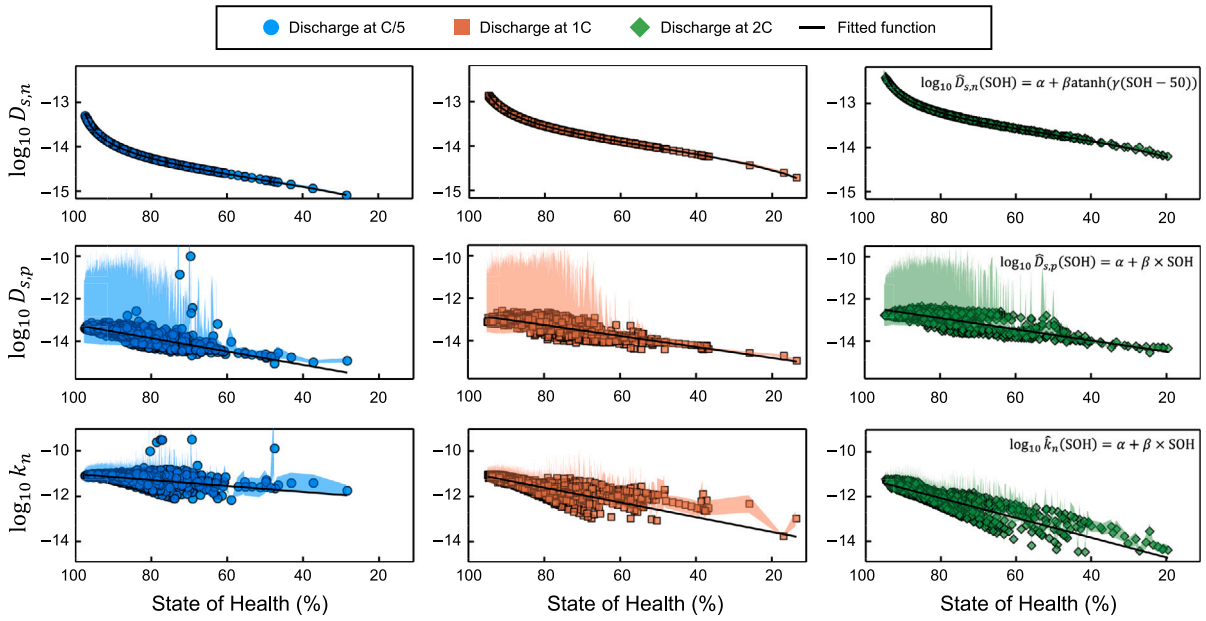

Fig. 7. MCMC results and simple fitting of every cycle for C/5, 1C, and 2C discharge rates of the diagnostic cycles with a weak Gaussian prior for $\log_{10} D_{s,p}$. The highlighted region is the uncertainty for each parameter, which is smoothed using an exponentially weighted moving average. Each dot is the logarithm of MAP at a particular SOH for all cells. Fitted parameter equations are generated to describe all MAPs well and are reported in Table 7.

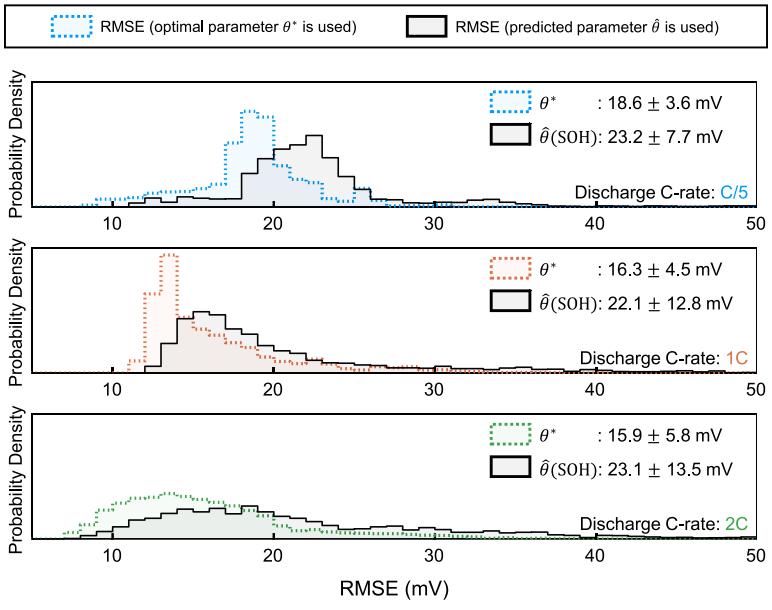

Fig. 8. Histogram of RMSEs from optimizing the parameters for each cycle, θ^*, and with a simple fitted relationship for each parameter, $\hat{\theta}$(SOH) (see Table 7). The mean and standard deviation are slightly larger using $\hat{\theta}$(SOH) compared to θ^*.

three parameters regressed on the optimal set of parameters, producing average errors of 23 mV.

In future works, we focus on identifying additional identifiable parameters that significantly contribute to aging and track their trajectories. This could provide deeper insights into lifetime prediction and the mechanisms of aging. Also, by thoroughly investigating various degradation modes such as LLI and LAM, and integrating the perturbation terms related to these modes into the parameter identification-based degradation diagnosis procedure, our approach will offer more comprehensive and quantitative information regarding degradation. The proposed framework for parameter identification-based aging mechanism analysis framework can be applied to various battery models and chemistries. The identifiability and trends of the diffusion and kinetic rate constants may vary, so it is essential to validate the methodology for new experimental conditions.

CRediT authorship contribution statement

Marc D. Berliner: Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Minsu Kim:** Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis. **Xiao Cui:** Writing – review & editing, Validation, Data curation. **Vivek N. Lam:** Writing – review & editing, Validation, Data curation. **Shakul Pathak:** Data curation, Validation, Visualization. **Yunhong Che:** Data curation, Validation, Visualization. **Patrick A. Asinger:** Investigation, Conceptualization. **Martin Z. Bazant:** Writing – review & editing, Funding acquisition. **William C. Chueh:** Writing – review & editing, Data curation. **Richard D. Braatz:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization.

Data and code availability

The data and code used in this work will be made available upon reasonable request to the corresponding author, Richard D. Braatz (braatz@mit.edubraatz@mit.edu)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Toyota Research Institute through the D3BATT Center on Data-Driven Design of Rechargeable Batteries.

Appendix A. Model validation: Degradation diagnosis

The accuracy of the DFN model in describing the cycling behavior of aged cell along the trajectory of $\log_{10} D_{s,n}$, $\log_{10} D_{s,p}$, and $\log_{10} k_n$ is depicted in Fig. A.1.

Appendix B. Global sensitivity analysis: Sobol' indices

The significance of transport and kinetic parameters as key indicators for diagnosing battery health is quantitatively assessed using Global Sensitivity Analysis (GSA). Sobol' indices are a widely used metric for analyzing parameter sensitivity in engineering applications where the system is not assumed to be linear or monotonic [65,66]. Sobol' indices quantitatively express the contribution of a specific parameter (i.e., partial variance) to the total variance of the system.

Here, Polynomial Chaos Expansion (PCE) was employed for variance calculation. PCE can compute variance more sample-efficiently than Monte Carlo simulation. Additionally, because of the orthogonality of the polynomials, the statistical information is simply calculated from polynomial coefficients, allowing the variance to be computed with simple numerical calculations. In this work, total Sobol' indices were used to calculate the overall contribution, including interactions with other parameters. See Ref. [1,66] for more information of PCE-based GSA.

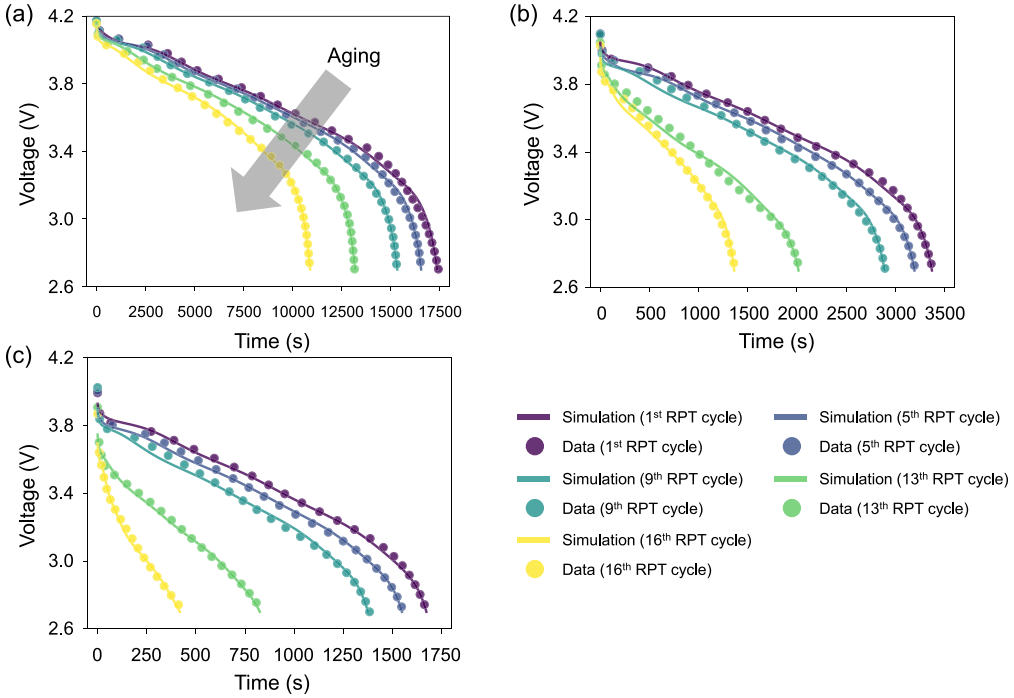

Fig. A.1. Validation of degradation diagnosis results in a single cell. Applied to five different RPT cycles, each corresponding to Table 6. For example, the discharge voltage curve at the 1st RPT of C/5 corresponds to 96.89% SOH, where the best estimation (θ^*) of $\log_{10} D_{s,n}$, $\log_{10} D_{s,p}$, and $\log_{10} k_n$ are −13.4123, −11.7891, and −11.2356, respectively.

Table B.1
Details of the model parameters for the global sensitivity analysis. Boundaries are determined based on the reference values of each parameter specified in Table 1.

Parameter	Unit	Description	Boundaries	Note
$D_{s,p}$	m^2/s	Cathode solid-phase diffusivity	[−60%, +20%]	[24]
$D_{s,n}$	m^2/s	Anode solid-phase diffusivity	[−60%, +20%]	[24]
D_e	m^2/s	Electrolyte solid-phase diffusivity	[−60%, +20%]	[24]
k_p	$m^{5/2}/(mol^{1/2}s)$	Cathode reaction rate constant	[−60%, +20%]	[24]
k_n	$m^{5/2}/(mol^{1/2}s)$	Anode reaction rate constant	[−60%, +20%]	[24]
l_a	m	Cathode current collector thickness	[−20%, +20%]	Assumed
l_p	m	Cathode thickness	[−20%, +20%]	Assumed
l_s	m	Separator thickness	[−20%, +20%]	Assumed
l_n	m	Anode thickness	[−20%, +20%]	Assumed
l_z	m	Anode current collector thickness	[−20%, +20%]	Assumed
ϵ_p	–	Cathode porosity	[−20%, +20%]	Assumed
ϵ_s	–	Separator porosity	[−20%, +20%]	Assumed
ϵ_n	–	Anode porosity	[−20%, +20%]	Assumed
R_p^p	m	Cathode particle radius	[−20%, +60%]	[24], Eq. (28)
R_n^p	m	Anode particle radius	[−20%, +60%]	[24], Eq. (28)
σ_p	S/m	Cathode solid-phase conductivity	[−20%, +20%]	Assumed
σ_n	S/m	Anode solid-phase conductivity	[−20%, +20%]	Assumed
Brugg_p	–	Cathode Bruggeman coefficient	[−20%, +20%]	Assumed
Brugg_s	–	Separator Bruggeman coefficient	[−20%, +20%]	Assumed
Brugg_n	–	Anode Bruggeman coefficient	[−20%, +20%]	Assumed
t_+	–	Transference number	[−20%, +20%]	Assumed
$\Theta_p^{\max}$	–	Cathode maximum stoichiometry limits	[−20%, +20%]	Assumed
$\Theta_p^{\min}$	–	Cathode minimum stoichiometry limits	[−20%, 1.0]	Assumed
$\Theta_n^{\max}$	–	Anode maximum stoichiometry limits	[−20%, 1.0]	Assumed
$\Theta_n^{\min}$	–	Anode minimum stoichiometry limits	[−20%, +20%]	Assumed

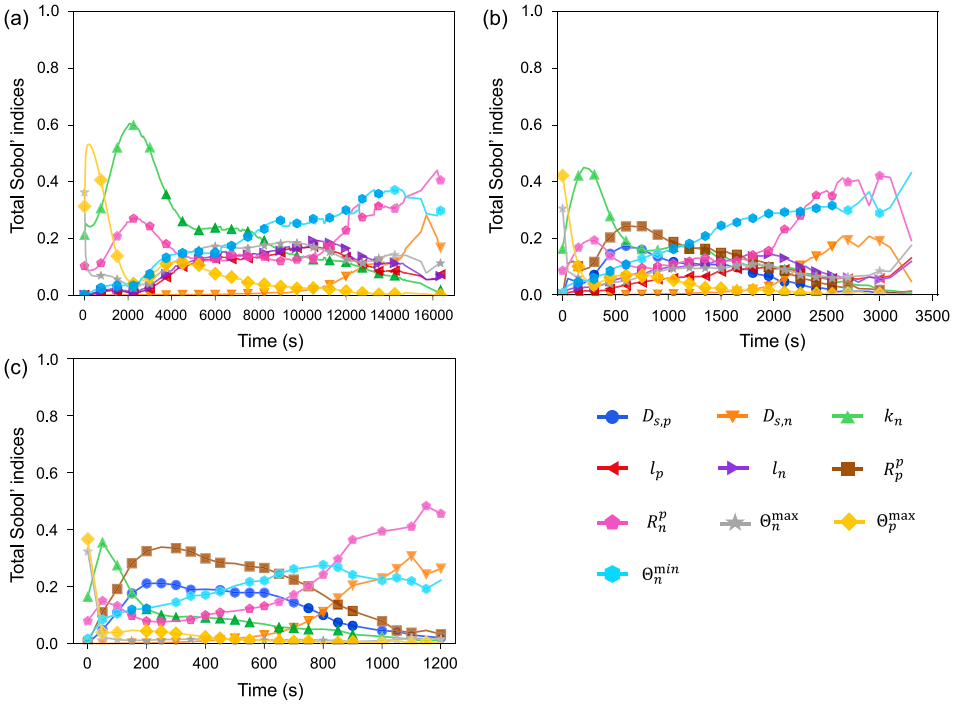

Fig. B.1. Global sensitivity analysis using Sobol' indices. See Table B.1 for the parameters used in the sensitivity analysis and their boundaries.

The boundaries for each parameter used to calculate the Sobol' indices are detailed in Table B.1. For non-aging scenarios, these boundaries are determined by accounting for the measurement uncertainty and variability of each parameter [1,67,68]. For aging analysis, which involves identifying parameter trajectories, this boundary determination is complicated. In this work, the boundaries for the diffusivity (D) and kinetic rate (k) were assumed to be identical to the uncertainty range of the parameters estimated in our previous work [24]. Similarly, the same criterion was applied to the particle radius (R^p), as it exhibits an inverse relationship with diffusivity in the diffusion model (Eq. (28)). For all other parameters, a 20% margin was assumed around their respective reference values.

Fig. B.1 shows the total Sobol' indices for three discharge C-rates. Parameters with a Sobol' index exceeding 0.1 were identified as key parameters. The key parameters for each discharge case are:

- C/5: $D_{s,n}$, k_n, l_p, l_n, R_n^p, $\theta_n^{\max}$, $\theta_p^{\max}$, $\theta_p^{\min}$
- 1C: $D_{s,p}$, $D_{s,n}$, k_n, l_p, l_n, R_p^p, R_n^p, $\theta_n^{\max}$, $\theta_p^{\max}$, $\theta_p^{\min}$
- 2C: $D_{s,p}$, $D_{s,n}$, k_n, R_p^p, R_n^p, $\theta_n^{\max}$, $\theta_p^{\max}$, $\theta_p^{\min}$

Excluding the k_p, which was not identifiable in this work, $D_{s,p}$, $D_{s,n}$, and k_n were identified as key parameters in at least two of the three discharge cases.

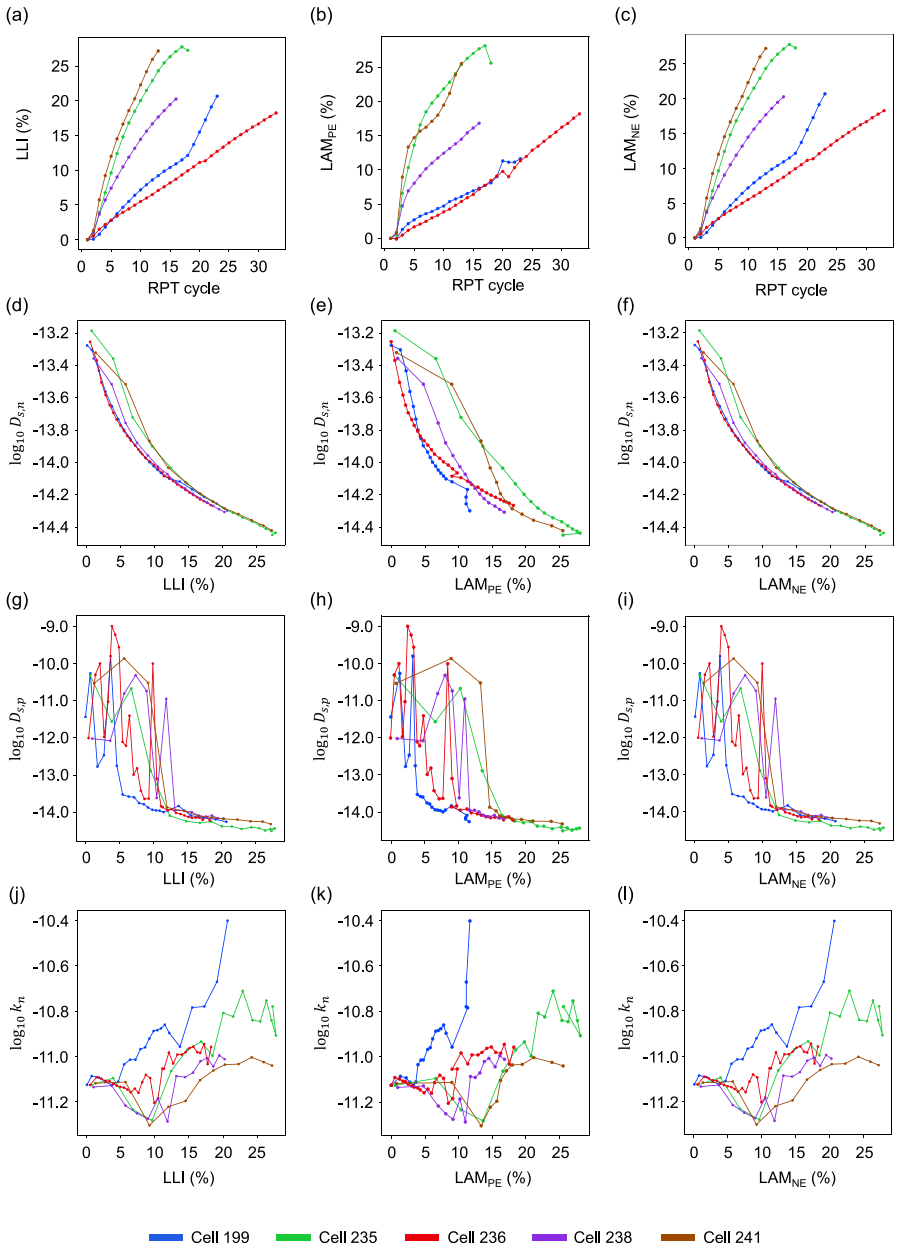

Fig. C.1. The trajectories of $\log_{10} D_{s,n}$, $\log_{10} D_{s,p}$, and $\log_{10} k_n$ at C/5 discharge for five randomly selected cells according to LLI, LAM_{PE}, and LAM_{NE} identified through DVA. (a) LLI, (e) LAM_{PE}, and LAM_{NE} according to RPT cycle. (a) LLI, (b) LAM_{PE}, and (c) LAM_{NE} according to the RPT cycle. $\log_{10} D_{s,n}$ trajectories according to (d) LLI, (e) LAM_{PE}, and (f) LAM_{NE}. $\log_{10} D_{s,p}$ trajectories according to (g) LLI, (h) LAM_{PE}, and (i) LAM_{NE}. $\log_{10} k_n$ trajectories according to (j) LLI, (k) LAM_{PE}, and (l) LAM_{NE}.

Appendix C. Aging mechanism: LAM and LLI

The aging modes of batteries are commonly classified into LAM and LLI [46,69]. In this section, the relationship between the parameter trajectories and LAM and LLI is discussed. First, LAM in each electrode is defined using electrode capacities:

$$\text{LAM}_{\text{PE}}(\%) = \left(1 - \frac{C_p^{\text{aged}}}{C_p^{\text{fresh}}}\right) \times 100, \tag{C.1}$$

$$\text{LAM}_{\text{NE}}(\%) = \left(1 - \frac{C_n^{\text{aged}}}{C_n^{\text{fresh}}}\right) \times 100, \tag{C.2}$$

where C_p is the electrode capacity of the cathode, and C_n is the electrode capacity of the anode. Second, LLI is defined as the amount of cyclable lithium ions:

$$\text{LLI}(\%) = \left(1 - \frac{n_{\text{Li}}^{\text{aged}}}{n_{\text{Li}}^{\text{fresh}}}\right) \times 100, \tag{C.3}$$

$$n_{\text{Li}} = \frac{3600}{\text{F}} \left(\Theta_p^{\max} C_p + \Theta_n^{\max} C_n\right). \tag{C.4}$$

For the same dataset, LAM and LLI were identified through differential voltage analysis (DVA), an approach that reconstructs OCV to identify aging modes [70]. Fig. C.1 shows LLI, LAM_{PE}, and LAM_{NE} according to the RPT cycles for five randomly selected cells that have different cycle lives. Furthermore, the parameter trajectories for each aging indicator are compared.

References

[1] Kim M, Schaeffer J, Berliner MD, Sagnier BP, Bazant MZ, Findeisen R, Braatz RD. J Electrochem Soc 2024;171(9):090517.
[2] Kim M, Kim J. ACS Sustain Chem Eng 2024;12(17):6786–96.
[3] Jiang B, Berliner MD, Lai K, Asinger PA, Zhao H, Herring PK, Bazant MZ, Braatz RD. Appl Energy 2022;307:118244.
[4] Lamorgese A, Mauri R, Tellini B. J Energy Storage 2018;20:289–97.
[5] Ashwin T, Chung YM, Wang J. J Power Sources 2016;328:586–98.
[6] Sordi G, Stecchini A, Evangelista R, Luder D, Li W, Sauer D, Casalegno A, Rabissi C. ETransportation 2025;24:100410.
[7] Liu Z, Ivanco A, Onori S. J Energy Storage 2019;21:519–27.
[8] Aykol M, Gopal CB, Anapolsky A, Herring PK, van Vlijmen B, Berliner MD, Bazant MZ, Braatz RD, Chueh WC, Storey BD. J Electrochem Soc 2021;168(3):030525.
[9] Sulzer V, Mohtat P, Aitio A, Lee S, Yeh YT, Steinbacher F, Khan MU, Lee JW, Siegel JB, Stefanopoulou AG, et al. Joule 2021;5(8):1934–55.
[10] Mallarapu A, Kim J, Carney K, DuBois P, Santhanagopalan S. ETransportation 2020;4:100065.
[11] Newman J, Tiedemann W. AIChE J 1975;21(1):25–41.
[12] Doyle M, Fuller TF, Newman J. J Electrochem Soc 1993;140(6):1526–33.
[13] Fuller TF, Doyle M, Newman J. J Electrochem Soc 1994;141(4):982–90.
[14] Fuller TF, Doyle M, Newman J. J Electrochem Soc 1994;141(1):1–10.
[15] Hines KE, Middendorf TR, Aldrich RW. J Gen Physiol 2014;143(3):401–16.
[16] Galuppini G, Berliner MD, Cogswell DA, Zhuang D, Bazant MZ, Braatz RD. J Power Sources 2023;573:233009.
[17] Berliner MD, Zhao H, Das S, Forsuelo M, Jiang B, Chueh WH, Bazant MZ, Braatz RD. J Electrochem Soc 2021;168(9):090546.
[18] Ren J, Ma J, Wang H, Yu T, Wang K. Prot Control Mod Power Syst 2024.
[19] Chun H, Yoon K, Kim J, Han S. IEEE Trans Transp Electrif 2022;9(1):995–1007.
[20] Laue V, Röder F, Krewer U. J Appl Electrochem 2021;51(9):1253–65.
[21] Deng Z, Deng H, Yang L, Cai Y, Zhao X. Energy 2017;138:509–19.
[22] Kim S, Kim S, Choi YY, Choi J-I. J Energy Storage 2023;71:108129.
[23] Bills A, Fredericks L, Sulzer V, Viswanathan V. ACS Energy Lett 2023;8(8):3578–85.
[24] Ramadesigan V, Chen K, Burns NA, Boovaragavan V, Braatz RD, Subramanian VR. J Electrochem Soc 2011;158(9):A1048–54.
[25] Smith RB, Bazant MZ. J Electrochem Soc 2017;164(11):E3291.
[26] Han X, Ouyang M, Lu L, Li J, Zheng Y, Li Z. J Power Sources 2014;251:38–54.
[27] Sadabadi KK, Jin X, Rizzoni G. J Power Sources 2021;481:228861.
[28] Prasad GK, Rahn CD. J Power Sources 2013;232:79–85.
[29] Berliner MD, Cogswell DA, Bazant MZ, Braatz RD. J Electrochem Soc 2021;168(9):090504.
[30] Schiesser WE. The numerical method of lines: Integration of partial differential equations. San Diego: Academic Press; 1991.
[31] Golberg DE. Genetic algorithms in search, optimization and machine learning. Boston: Addison-Wesley; 1989.
[32] Lee E, Kim M, Moon I, Kim J. Chem Eng J 2024;490:151484.
[33] Johnson RA, Wichern DW. Applied multivariate statistical analysis. 5th ed.. Upper Saddle River, NJ: Prentice Hall; 2002.
[34] Meeker WQ, Escobar LA. Amer Statist 1995;49(1):48–53.
[35] Haario H, Laine M, Mira A, Saksman E. Stat Comput 2006;16(4):339–54.
[36] Roberts GO, Rosenthal JS. J Comput Graph Statist 2009;18(2):349–67.
[37] Chib S, Greenberg E. Amer Statist 1995;49(4):327–35.
[38] Gering KL. Electrochim Acta 2017;225:175–89.
[39] Khuri AI, Mukhopadhyay S. Wiley Interdiscip Rev: Comput Stat 2010;2(2):128–49.
[40] Moyassari E, Streck L, Paul N, Trunk M, Neagu R, Chang C-C, Hou S-C, Märkisch B, Gilles R, Jossen A. J Electrochem Soc 2021;168(2):020519.
[41] Lee J-H, Kim W-J, Kim J-Y, Lim S-H, Lee S-M. J Power Sources 2008;176(1):353–8.
[42] van Vlijmen B, Lam V, Asinger PA, Cui X, Ganapathi D, Sun S, Herring PK, Gopal CB, Geise N, Deng HD, et al. ChemRxiv 2023.
[43] Beck JV, Arnold KJ. Parameter estimation in engineering and science. New York: Wiley; 1977.
[44] Zhou X, Huang J, Pan Z, Ouyang M. J Power Sources 2019;426:216–22.
[45] Devie A, Baure G, Dubarry M. Energies 2018;11(5):1031.
[46] Birkl CR, Roberts MR, McTurk E, Bruce PG, Howey DA. J Power Sources 2017;341:373–86.
[47] Edge JS, O'Kane S, Prosser R, Kirkaldy ND, Patel AN, Hales A, Ghosh A, Ai W, Chen J, Yang J, et al. Phys Chem Chem Phys 2021;23(14):8200–21.
[48] Koleti UR, Rajan A, Tan C, Moharana S, Dinh TQ, Marco J. Energies 2020;13(13):3458.
[49] Ramadesigan V, Boovaragavan V, Pirkle JC, Subramanian VR. J Electrochem Soc 2010;157(7):A854.
[50] Ploehn HJ, Ramadass P, White RE. J Electrochem Soc 2004;151(3):A456.
[51] Ashwin T, McGordon A, Jennings PA. Electrochim Acta 2017;232:203–14.
[52] Pinson MB, Bazant MZ. J Electrochem Soc 2012;160(2):A243.
[53] Ferguson TR, Bazant MZ. Electrochim Acta 2014;146:89–97.
[54] Thomas-Alyea KE, Jung C, Smith RB, Bazant MZ. J Electrochem Soc 2017;164(11):E3063.
[55] Fraggedakis D, Nadkarni N, Gao T, Zhou T, Zhang Y, Han Y, Stephens RM, Shao-Horn Y, Bazant MZ. Energy Environ Sci 2020;13(7):2142–52.
[56] Gao T, Han Y, Fraggedakis D, Das S, Zhou T, Yeh C-N, Xu S, Chueh WC, Li J, Bazant MZ. Joule 2021;5(2):393–414.
[57] Finegan DP, Quinn A, Wragg DS, Colclasure AM, Lu X, Tan C, Heenan TM, Jervis R, Brett DJ, Das S, et al. Energy Environ Sci 2020;13(8):2570–84.
[58] Lu X, Lagnoni M, Bertei A, Das S, Owen RE, Li Q, O'Regan K, Wade A, Finegan DP, Kendrick E, et al. Nat Commun 2023;14(1):5127.
[59] Zhuang D, Bazant MZ. J Electrochem Soc 2022;169(10):100536.
[60] Zhuang D, Bazant MZ. Phys Rev E 2023;107(4):044603.
[61] Bazant MZ. Faraday Discuss 2023;246:60–124.
[62] Fraggedakis D, McEldrew M, Smith RB, Krishnan Y, Zhang Y, Bai P, Chueh WC, Shao-Horn Y, Bazant MZ. Electrochim Acta 2021;367:137432.
[63] Liang Q, Bazant MZ. J Electrochem Soc 2023;170(9):093510.
[64] Kirk TL, Please CP, Chapman SJ. J Electrochem Soc 2021;168(6):060554.
[65] Homma T, Saltelli A. Reliab Eng Syst Saf 1996;52(1):1–17.
[66] Sudret B. Reliab Eng Syst Saf 2008;93(7):964–79.
[67] Hadigol M, Maute K, Doostan A. J Power Sources 2015;300:507–24.
[68] Chen J, Mattus SP, Cao W, Sauer DU, Li W. Adv Appl Energy 2025;100221.
[69] Mohtat P, Lee S, Siegel JB, Stefanopoulou AG. J Electrochem Soc 2021;168(10):100520.
[70] Che Y, Lam VN, Rhyu J, Schaeffer J, Kim M, Bazant MZ, Chueh WC, Braatz RD. Joule 2025.