Antti Aitio*
Department of Engineering Science,
University of Oxford,
Parks Road, Oxford, OX1 3PJ, U.K.,
email: antti.aitio@gmail.com

Dominik Jöst*
Chair for Electrochemical Energy
Conversion & Storage Systems,
Institute for Power Electronics & Electrical
Drives (ISEA),
& Center for Ageing, Reliability and Lifetime
Prediction for Electrochemical and Power
Electronic Systems (CARL),
RWTH Aachen University,
& Jülich Aachen Research Alliance,
JARA-Energy,
Campus-Boulevard 89, Aachen 52074,
Germany
email: dominik.joest@isea.rwth-aachen.de

Dirk U. Sauer
Chair for Electrochemical Energy
Conversion & Storage Systems,
Institute for Power Electronics & Electrical
Drives (ISEA),
& Center for Ageing, Reliability and Lifetime
Prediction for Electrochemical and Power
Electronic Systems (CARL),
RWTH Aachen University,
& Jülich Aachen Research Alliance,
JARA-Energy,
Campus-Boulevard 89, Aachen 52074,
Germany,
& Helmholtz Institute Münster (HI MS),
IEK 12, Forschungszentrum Jülich,
Wilhelm-Johnen-Straße, 52428 Jülich,
Germany
email: dirkuwe.sauer@isea.rwth-aachen.de

David A. Howey[1]
Department of Engineering Science,
University of Oxford,
Parks Road, Oxford, OX1 3PJ, U.K.,
& the Faraday Institution,
Quad One, Becquerel Avenue, Harwell
Campus, Didcot, OX11 0RA, U.K.,
email: david.howey@eng.ox.ac.uk

Learning battery model parameter dynamics from data with recursive Gaussian process regression

Estimating the state of health is a critical function of a battery management system, but remains challenging due to variability of operating conditions and usage requirements in real applications. As a result, existing techniques based on fitting equivalent circuit models may exhibit inaccuracy at extremes of performance and over long-term ageing, or instability of parameter estimates. Pure data-driven techniques, on the other hand, suffer from a lack of generality beyond their training dataset. Here, we propose a novel hybrid approach combining data- and model-driven techniques for battery health estimation, estimating both capacity loss and resistance increase. Specifically, we use a Bayesian method, Gaussian process regression, to estimate model parameters as functions of states, operating conditions, and lifetime. Computational efficiency is ensured by a recursive implementation, yielding a joint state-parameter estimator that learns parameter dynamics from data and is robust to gaps and varying operating conditions. Results show the efficacy of the method, on both simulated and measured drive cycle data, including accurate estimates and forecasts of battery capacity and internal resistance. This opens up new opportunities to understand battery ageing from field data.

Keywords: battery, health, estimation, observer, machine learning, Gaussian process

1 Introduction

Demand for battery systems is increasing rapidly through efforts to decarbonise electricity grids and electrify mobility [1]. Li-ion cells have become the workhorse here due to their long lifetime and high energy density [2]. Although cost has dramatically decreased [3], the economics of storage needs to improve further, notably in applications where battery systems are not yet competitive in terms of levelized cost [4]. Given the risks of Li-ion cell demand outpacing the supply of required raw materials [5], it is crucial that the performance and lifetime of existing systems is maximised. The overall cost-effectiveness of Li-ion batteries can be greatly enhanced by accurate estimation and prediction of battery state-of-health (SOH), which can improve lifetime, warranty and insurance costs, system safety, and timing of maintenance. Accurate SOH estimation and prediction, especially using field data, opens up additional possibilities for second-life applications and helps to 'close the loop' in terms of understanding the impact of design choices on real-world performance [6]. While these issues have been extensively studied at cell level in laboratory environments, little work has been done using realistic usage data [6]. Challenges in SOH estimation from real-world data arise due to the lack of controlled operating conditions, less accurate sensors, data gaps, cell-to-cell variability, and a lack of granularity of individual cell measurements when dealing with modules or packs.

Battery SOH estimation methods are usually categorised into model-driven (grey-box) and data-driven (black-box) approaches

[1]Corresponding Author.
*A. Aitio, D. Jöst are joint first authors.

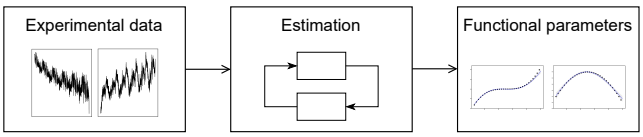


Fig. 1 We estimate (Algorithm 1) equivalent circuit model states and parameters from data; parameters are functions, represented by Gaussian processes, rather than constants.

[7], where the former consists of repeatedly fitting model parameters to input-output data and using the parameter estimates to track SOH, and the latter is concerned with mapping directly from measurements (e.g., voltage versus time) to calibrated capacity or resistance data. Note that SOH is defined in many different ways in literature, with common definitions based on the remaining capacity, internal resistance increase, combinations of these, or other metrics [7–9]. We remain agnostic about a specific definition of SOH in this paper, and instead focus on direct estimates of equivalent circuit model parameter values. Depending on the application, it may be more useful to track SOH via resistance, or capacity, or a combination of these parameters.

Battery SOH estimation has commonly been achieved with observers, including recursive state-parameter estimation techniques such as nonlinear approximations of Kalman filters [10–13], and methods from robust control that guarantee convergence of estimates via stability criteria [14, 15]. Prognosis (i.e., future prediction of health) in this framework is achieved using a separate model for the evolution of parameters over lifetime, and this can range from a random walk [11–13] to semi-empirical curve fits of trajectories that may be re-parameterised over lifetime using adaptive methods such as particle filtering [16, 17], a Bayesian approach that also provides uncertainty estimates.

Model-driven approaches tend to use simple equivalent-circuit models because they have relatively few parameters, however, assuming parameters to be constant leads to noisy estimates because they vary as functions of internal states and operating conditions [18–20]. Physics-based models, such as those within the Doyle-Fuller-Newman framework [21, 22], may suffer less from this issue, but they are plagued by poor identifiability [23] due to their large number of parameters and a lack of reference electrodes in commercial cells [24]. In contrast, as mentioned, data-driven methods for SOH diagnosis or prognosis attempt to map from operating data to SOH (usually defined in some consistent way, e.g. as the constant-current discharge capacity) either using the measurements directly [25–28] or via pre-defined features extracted from measurements [29–34]. To obtain these mappings, supervised machine-learning techniques such as neural networks [25, 26], Gaussian process regression [29, 35, 36] and relevance vector machines [37] have been used. Depending on the training data, these methods may be used to estimate either present or future SOH.

In this study, we introduce a 'hybrid' approach to SOH diagnosis and prognosis, merging model-driven and data-driven methods. Our strategy represents equivalent circuit model parameters as uncertain continuous smooth functions of time and operational conditions, and learns these functions from data using Gaussian process regression. A very high-level overview is shown in Fig. 1. Here, the 'estimation' step includes estimating the state (e.g., state of charge) and the parameters (e.g., capacity), which are functions. This approach is in contrast to existing studies employing hybrid approaches [38–40] that assume constant model parameters, or use ad hoc parameter update methods. Our approach enhances accuracy by allowing circuit parameters to reflect the underlying electrochemical processes more realistically. For instance, the reaction kinetics, similar to a nonlinear resistor, can be modeled as a function that is influenced by factors like current, temperature, and state of charge (SOC), as detailed in Brucker et al. [41]. To date, estimating battery equivalent circuit parameters as functions has received very limited attention, although there has been exploration into SOC dependencies using linear parameter-varying models [42], and SOC estimation using a direct GPR based observation model [43].

The novelty of our work is that we demonstrate simultaneous state and functional parameter estimation of a battery equivalent circuit model within an observer framework, including a computationally efficient implementation of Gaussian process regression [44], resulting in the computational effort for model fitting scaling linearly with the number of rows in the input-output data. The method is both battery-chemistry and form-factor agnostic, and the only prerequisite is a lab measurement of the full cell open-circuit voltage as a function of SOC. The approach gives both a current SOH estimate and a near-term future prediction of SOH with minimal extra computational effort.

2 Gaussian process regression

In this work we used Gaussian process (GP) regression to learn battery equivalent circuit model parameters as functions of time and usage conditions (SOC, current) from data. This section gives a short overview of Gaussian processes and discusses the recursive technique that was used for implementing GPs computationally efficiently in an observer-type framework. For readers unfamiliar with GPs, further information may be found in Rasmussen and Williams [45], Bishop [46], and Görtler et al. [47].

2.1 Overview. A Gaussian process (GP) is a collection of random variables where any subset is jointly Gaussian-distributed [45]. Consequently, a GP defines a distribution over functions over an input vector $\mathbf{x}$, characterised by a mean and a covariance,

$$\begin{aligned} f(\mathbf{x}) &\sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right), \\ m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))\left(f(\mathbf{x}') - m(\mathbf{x}')\right)], \end{aligned} \tag{1}$$

where $k(\mathbf{x}, \mathbf{x}')$ is the kernel function describing the covariance of the GP. Without loss of generality, we set the mean function $m(\mathbf{x}) = 0$. Given the definition of $f(\mathbf{x})$, the aim is to map inputs $\mathbf{x}$ to outputs $\mathbf{y}$ such that

$$\mathbf{y} = f(\mathbf{x}) + \epsilon\ ,\ \epsilon \sim \mathcal{N}(0, \sigma_n^2). \tag{2}$$

The fitting process consists of determining the posterior-predictive distribution, which can be used to predict values of $\mathbf{y}$ at any value of $\mathbf{x}$. Assuming zero-mean i.i.d. Gaussian measurement error ϵ in the observations $\mathbf{y}$, the posterior-predictive distribution is Gaussian, with mean and covariance for a test point $\mathbf{x}_* \in \mathbb{R}^{\mathrm{d}}$ given by

$$\mu(f_*) = \mathbf{k}_{*,\mathbf{X}} \left[\mathbf{K}_{\mathbf{X}} + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{y} \tag{3a}$$

$$\mathrm{cov}(f_*) = \mathbf{k}_{**} - \mathbf{k}_{*,\mathbf{X}} \left[\mathbf{K}_{\mathbf{X}} + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{k}_{*,\mathbf{X}}^{\mathrm{T}}, \tag{3b}$$

where $\mathbf{I}$ is the identity matrix, n the number of rows in the training data, and d the dimensions of the input. For notational brevity we use $\mathbf{k}_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$, $\mathbf{K}_{\mathbf{X}} = k(\mathbf{X}, \mathbf{X})$ and $\mathbf{k}_{*,\mathbf{X}} = k(\mathbf{x}_*, \mathbf{X})$, and $\mathbf{X} \in \mathbb{R}^{\mathrm{n} \times \mathrm{d}}$ denotes the training input data matrix.

A common assumption is to use a zero-mean prior; this is not particularly constraining because the *posterior* mean will be non-zero. Explicit mean functions can be included if desired, and later (section 3), for numerical stability in our particular application, we introduce a simple transformation that can be interpreted as a non-zero prior mean. For the sake of clarity in this introduction section, however, we assume a zero-mean prior. In this case, the model is defined by the training data and the parameters of the chosen kernel function $k(\mathbf{x}, \mathbf{x}')$—known as 'hyperparameters'—that describe

smoothness, magnitude, periodicity and so on [45]. Using GP regression to fit data requires estimation of the hyperparameter vector θ given the training data. In the Bayesian framework, the posterior distribution of the hyperparameters θ given the training data is

$$p(\theta|\mathbf{X},\mathbf{y}) = \frac{p(\theta)\int p(\mathbf{y}|\theta,f,\mathbf{X})p(f|\mathbf{X},\theta)\,\mathrm{d}f}{p(\mathbf{y}|\mathbf{X})}, \tag{4}$$

where $p(\theta)$ is the hyperparameter prior. Because the hyperparameter posterior is not tractable, it may be approximated using the Laplace method [48], Markov chain Monte-Carlo techniques, or variational inference [49]. If the full posterior can be approximated, then the hyperparameters may be marginalized (i.e. the whole distribution may be used) when making a prediction [50]. However, for computational efficiency, we used maximum likelihood estimates of θ, and these were obtained by assuming a uniform prior $p(\theta)$ and then maximising the logarithm of the numerator of Eq. (4) (i.e. the log marginal likelihood), given by [45]

$$\log\int p(\mathbf{y}|\theta,f,\mathbf{X})p(f|\mathbf{X},\theta)\,\mathrm{d}f = \\ -\frac{1}{2}\mathbf{y}^{\mathrm{T}}\left[\mathbf{K}_{\mathbf{X}}-\sigma_n^2\mathbf{I}\right]^{-1}\mathbf{y}-\frac{1}{2}\log|\mathbf{K}_{\mathbf{X}}|-\frac{n}{2}\log 2\pi, \tag{5}$$

where $|\mathbf{K}_{\mathbf{X}}|$ denotes the determinant of $\mathbf{K}_{\mathbf{X}}$. Note that the noise term, σ_n, is also fitted during hyperparameter optimisation.

Unfortunately, the standard approach for estimating hyperparameters or making model predictions with GP regression suffers from the so-called 'big-n' problem, i.e. poor computational scaling, because both Eq. (3) and (5) require the inversion of an $n\times n$ matrix. This usually scales computationally as $\mathcal{O}(n^3)$ and can become numerically unstable for larger matrices. Several solutions have been proposed, such as sparse GP regression [51], structured approaches [52], reduced-rank methods [53], or GPU parallelisation [54]. Here, we implemented a recursive method [44] that enabled a unified framework for the simultaneous estimation of states and parameters. This scales as $\mathcal{O}(n)$ with the number of data rows, making it an efficient option for time series data fitting.

2.2 Recursive GP regression. Previously, we considered the GP input to be a general vector $\mathbf{x}$. We now turn to the specific case of fitting a multidimensional function with inputs 'time', t, and 'space', $\mathbf{x}$. This separation becomes useful later for tracking parameter variations in time separately from those in operating conditions. Recursive estimation of the GP posterior-predictive distribution (3) and log marginal likelihood (5) may be achieved by interpreting a Gaussian process $f(\mathbf{x},t)$ as the solution to a linear time-invariant stochastic (partial) differential equation [44, 55],

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = \mathcal{F}f(\mathbf{x},t)+\mathbf{L}\omega(\mathbf{x},t) \tag{6a}$$

$$\mathbf{y}_{\mathrm{t}} = \mathcal{H}_{\mathrm{t}}f(\mathbf{x},t)+\epsilon_{\mathrm{t}}\ ,\ \epsilon\sim\mathcal{N}(0,\sigma_n^2), \tag{6b}$$

where $\mathcal{F}$ and $\mathcal{H}_{\mathrm{t}}$ are linear operators, $\mathbf{L}$ is a dispersion matrix, and $\omega(\mathbf{x},t)$ is spatially resolved white noise. The observation noise term ϵ_{t} is the same as in the standard ('batch') GP regression formulation (2). If the kernel function is assumed multiplicatively separable, i.e., a product of two kernels each only depending on time or space, respectively,

$$k\left((\mathbf{x},t),(\mathbf{x}',t')\right) = k(\mathbf{x},\mathbf{x}')k(t,t'), \tag{7}$$

then the linear operator $\mathcal{F}$ becomes a constant matrix (denoted $\mathbf{F}$). In this study, all kernel formulations over the input space $(\mathbf{x},t)$ are assumed separable because this simplifies the calculations significantly but still enables complex behaviours to be modelled [56]. The link between the dynamic system and the kernel function

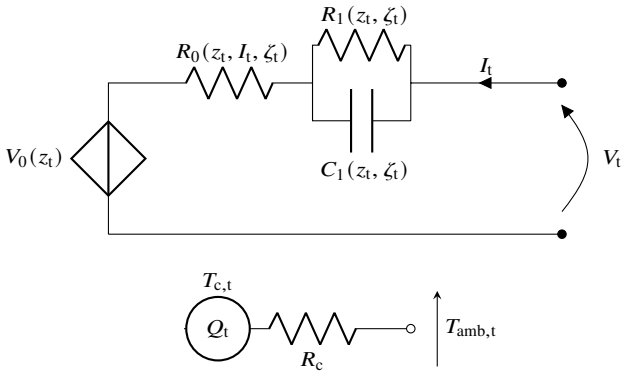


Fig. 2 Li-ion model with first-order RC electrical circuit and lumped thermal model. All electrical parameters are modelled as Gaussian processes.

describing the GP means that the kernel function may be directly mapped to $\mathbf{F}$, $\mathcal{H}_{\mathrm{t}}$ and the spectral density of the white noise process $\omega(\mathbf{x},t)$. The details of this may be found in Särkkä et al. [56, 57] and Solin [44].

Using this formulation, estimation of the posterior-predictive distribution and hyperparameter log marginal likelihood can be achieved recursively with a Kalman filter and Rauch-Tung-Striebel (RTS) smoother [58]. In the following sections, we describe a method for implementing this in the context of battery modelling, alongside state estimation, allowing us to estimate functional parameters in a computationally efficient way.

3 Combining circuit models and GP regression

The Li-ion cell model used here is shown in Fig. 2 and consists of a first-order resistor-capacitor (RC) electrical circuit with a coupled lumped thermal model. All electrical model parameters are considered to be GPs that are functions of SOC and/or applied current, as well as lifetime, denoted ζ_{t}. The thermal model consists of heat generation due to the total overpotential, and convection to the ambient environment, making the assumption that heat conduction through the cell is fast (i.e., the Biot number is small), such that the cell internal temperature is relatively uniform. Entropic heating was ignored. The (nonlinear) continuous-time dynamics of the 3-state electro-thermal model of Fig. 2 in state-space are

$$\begin{aligned}\frac{\mathrm{d}\mathbf{x}_{\mathrm{t}}}{\mathrm{d}t} &= f_{\mathrm{B}}(\mathbf{x}_{\mathrm{t}},I_{\mathrm{t}},\theta)\\ \mathbf{y}_{\mathrm{t}} &= h(\mathbf{x}_{\mathrm{t}},I_{\mathrm{t}},\theta),\end{aligned} \tag{8}$$

with the dynamics f_{B} for the battery state vector $\mathbf{x}_{\mathrm{t}} = [z_{\mathrm{t}}\ V_{1,\mathrm{t}}\ T_{\mathrm{t}}]^{\mathrm{T}}$ given by

$$\begin{aligned}\frac{\mathrm{d}z_{\mathrm{t}}}{\mathrm{d}t} &= I_{\mathrm{t}}Q^{-1}(\zeta_{\mathrm{t}})\\ \frac{\mathrm{d}V_{1,\mathrm{t}}}{\mathrm{d}t} &= -\alpha(z_{\mathrm{t}},\zeta_{\mathrm{t}})V_{1,\mathrm{t}}+\beta(z_{\mathrm{t}},\zeta_{\mathrm{t}})I_{\mathrm{t}}\\ \frac{\mathrm{d}T_{\mathrm{c,t}}}{\mathrm{d}t}C_{\mathrm{c}} &= -\frac{T_{\mathrm{c,t}}-T_{\mathrm{amb,t}}}{R_{\mathrm{c}}}+I_t\left[V_t-V_0(z_t)\right],\end{aligned} \tag{9}$$

where z_{t} is the SOC, I_{t} the applied current (positive for charging) and $Q^{-1}(\zeta_{\mathrm{t}})$ the inverse of battery capacity as a function of lifetime ζ_{t}, which here denotes normalised charge throughput. (In general, lifetime ζ_{t} could be denoted by calendar age, or by charge throughput, or by an arbitrary metric defined by the user.) The voltage $V_{1,\mathrm{t}}$ is across the RC pair, and its time dynamics are controlled by the functions $\alpha(z_{\mathrm{t}},\zeta_{\mathrm{t}})$ and $\beta(z_{\mathrm{t}},\zeta_{\mathrm{t}})$. These are related to the circuit

parameters via $\alpha = 1/R_1C_1$ and $\beta = 1/C_1$. The thermal model is parameterised by heat capacity C_c and thermal resistance R_c, which are considered known. Given these dynamics, the outputs $\mathbf{y}_t = [V_t\ T_t]^T$ are cell terminal voltage and temperature,

$$\begin{aligned} V_t &= V_0(z_t) + V_{1,t} + R_0(z_t, I_t, \zeta_t)I_t \\ T_t &= T_{c,t}\,. \end{aligned} \tag{10}$$

The model could be extended to include temperature dependencies for the circuit elements, but as the experimental data used here (see Section 6) only has a small temperature range of approximately 5 °C, this was not necessary.

To improve numerical performance, the functional parameters Q^{-1}, α, β and R_0 were expressed as affine transformations of zero-mean GPs, rather than directly as zero-mean GPs. Mathematically,

$$t_f(f) = c_f(1 + f), \tag{11}$$

where, for each respective parameter, c_f is a constant, f, defined in (1), represents a zero-mean GP, and t_f is a *non*-zero-mean GP representing the parameter in the model, i.e., either $Q^{-1}(\zeta_t)$, $\alpha(z_t, \zeta_t)$, $\beta(z_t, \zeta_t)$, or $R_0(z_t, I_t, \zeta_t)$. Including this transformation effectively sets a non-zero prior mean, c_f, for each circuit parameter. The exact value of this is not important, since this transformation is included purely to improve numerical scaling, but its value should be the correct order of magnitude and therefore was chosen to reflect an expert guess of the parameter value before fitting any data. This transformation scales the parameters so that the underlying zero-mean GPs (i.e., f) are in the unit range, making hyperparameter initialisation simpler, and it improves the numerical stability of the system dynamics (9). The four functions t_f describing the circuit parameters are also Gaussian processes because a Gaussian distribution remains Gaussian under arbitrary affine transformation.

The kernel function k_f for each GP is constructed so that the time-based input (ζ_t) is treated differently from the 'spatial' inputs, i.e., the instantaneous operating conditions (z_t and I_t). In this case, a non-stationary kernel describes each of the Gaussian processes in the time dimension, allowing uncertainty to change with time, and this enables better extrapolation than a stationary kernel since the latter reverts to the (zero) mean on long-range extrapolation. The non-stationary kernel is the Wiener velocity (WV) kernel [44] with magnitude parameter σ_ζ^2,

$$k_{WV}(\zeta, \zeta') = \sigma_\zeta^2 \left(\frac{\min^3(\zeta, \zeta')}{3} + |\zeta - \zeta'| \frac{\min^2(\zeta, \zeta')}{2} \right). \tag{12}$$

The kernel describing dependence of parameters on SOC (z_t) and applied current (I_t) is the squared exponential (SE) kernel [45],

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_{SE}^2 \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{x}') \right) \tag{13}$$

where σ_{SE}^2 is a magnitude hyperparameter, and $\boldsymbol{\Sigma}^{-1}$ is a diagonal two-dimensional matrix with elements $\gamma_{SE,z}$ and $\gamma_{SE,I}$ representing the inverse length scales of the SOC-dependence and current-dependence, respectively. The kernels (12) and (13) are combined by multiplying the two functions. In addition, a kernel describing process noise, i.e., short-term parameter fluctuations over time, is added—this adds equally to all points in $\mathbf{x}$, so that the overall kernel is therefore

$$k_f = \underbrace{k_{WV}(\zeta, \zeta') k_{SE}(\mathbf{x}, \mathbf{x}')}_{\text{spatially resolved, smooth}} + \underbrace{k_E(\zeta, \zeta')}_{\text{process noise}}, \tag{14}$$

where k_E is the exponential kernel [45],

$$k_E(\zeta, \zeta') = \sigma_{\zeta,r}^2 \exp\left(-\gamma_{\zeta,r} |\zeta - \zeta'| \right), \tag{15}$$

where $\sigma_{\zeta,r}$, $\gamma_{\zeta,r}$ are the magnitude and inverse length scale of the process noise. Note that the latter is separate to the *observation* noise, σ_n. Hence the evolution of parameters is decomposed into a longer-term smooth component and shorter-term fluctuations. When extrapolated, the latter decay quickly, giving a smooth trajectory. The hyperparameter $\gamma_{\zeta,r}$ estimates the autocorrelation of the process noise. If $\gamma_{\zeta,r}$ is high, then k_E effectively describes white noise.

4 Joint estimation of battery states and GPs

4.1 Discretisation and joint state vector. To construct a finite-dimensional state-space representation of Eq. (6), the input variable $\mathbf{x}$ for each GP, representing R_0, α, β respectively, has to be discretised. We chose n_z evenly spaced discretisation points over z (SOC) as inputs for the GPs for α, β, which are only functions of SOC, and n_{zI} points for the GP for R_0, which is a function of both SOC and applied current. In other words, α, β and R_0 are each represented by a state vector where each element corresponds to a point at a specific SOC and/or I. For R_0 the vector is effectively a stacked set of values at sampling points representing a grid of discrete SOC and current values. Therefore the state vectors for the three GPs can be written as

$$\begin{aligned} \mathbf{x}_\alpha &= \begin{bmatrix} \alpha_{z_1} & \alpha_{z_2} & \dots & \alpha_{z_{nz}} \end{bmatrix}, \\ \mathbf{x}_\beta &= \begin{bmatrix} \beta_{z_1} & \beta_{z_2} & \dots & \beta_{z_{nz}} \end{bmatrix}, \\ \mathbf{x}_{R_0} &= \begin{bmatrix} R_{0,(z_1,I_1)} & R_{0,(z_2,I_2)} & \dots & R_{0,(z_{nzI},I_{nzI})} \end{bmatrix}. \end{aligned} \tag{16}$$

A joint state vector may then be constructed to estimate the states of the GP and the battery model simultaneously given the n_z and n_{zI} points in the input space for the respective GPs. Let the vector $\mathbf{x}_{Batt,t}$ denote the means of the battery states at time t,

$$\mathbf{x}_{Batt,t} = \begin{bmatrix} z_t & V_{1,t} & T_{c,t} \end{bmatrix}^T, \tag{17}$$

and the vector $\mathbf{x}_{GP,s,t}$ denote the posterior predictive means of the GPs describing the parameters at time t,

$$\mathbf{x}_{GP,s,t} = \begin{bmatrix} \mathbf{x}_{Q^{-1},t} & \mathbf{x}_{\alpha,t} & \mathbf{x}_{\beta,t} & \mathbf{x}_{R_0,t} \end{bmatrix}^T, \tag{18}$$

where $\mathbf{x}_{Q^{-1},t} \in \mathbb{R}^2$, $\mathbf{x}_{\alpha,t} \in \mathbb{R}^{2n_z}$, $\mathbf{x}_{\beta,t} \in \mathbb{R}^{2n_z}$, $\mathbf{x}_{R_0,t} \in \mathbb{R}^{2n_{zI}}$. The dimensionality of the state vectors is due to the number of points used for the discretisation of $\mathbf{x}$ and the order of the Markov process using the Wiener velocity kernel $k_{WV}(\zeta, \zeta')$. The Wiener velocity kernel (12) has a time-domain representation [44] as follows,

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{d^2x}{dt^2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \frac{dx}{dt} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \omega(t), \tag{19}$$

where the spectral density of noise $\omega(t)$ is a function of $\sigma_{\zeta,s}$. This means that the kernel multiplication (14) results in each GP being represented by both the current state and its first-order time derivative at each spatial location in $\mathbf{x}$. The stationary exponential kernel (15), describing the short length scale noise process, has a single state representation,

$$\frac{dx}{dt} = -\gamma_{\zeta,r} x + \omega(t), \tag{20}$$

where the spectral density of the noise process $\omega(t)$ is a function of $\sigma_{\zeta,r}$. In this case, the GP state vector for k_E is

$$\mathbf{x}_{GP,r,t} = [x_{Q^{-1},1,t}\ \ x_{\alpha,1,t}\ \ x_{\beta,1,t}\ \ x_{R_0,1,t}]. \tag{21}$$

The overall joint state-parameter 'state' representation is given by the concatenation of the battery and parameter vectors,

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{x}_{Batt,t} \\ \mathbf{x}_{GP,s,t} \\ \mathbf{x}_{GP,r,t} \end{bmatrix}. \tag{22}$$

4.2 Initialisation and solution. The joint system is nonlinear and may be solved as a function of time using an appropriate Bayesian filter. For computational speed, the extended Kalman filter (EKF) was used; other variants, such as the unscented Kalman filter or particle filter, might provide more accurate results, but the EKF was adequate in initial tests using simulated data (see Section 5). There are two timescales involved—the first is given by the sampling frequency of current, voltage and temperature data during cycling, which here is 1 Hz, and the second is related to the 'extent of degradation', ζ, which covers the total lifetime of the battery and here is measured by cumulative normalised charge throughput, spanning months to years.

4.2.1 Initialisation. The initialisation of the two subsystems (i.e., the battery states and parameter-GPs, respectively) is consistent with the two timescales. A zero-mean GP is used to model the circuit parameters, via the affine transformation (11), which means that the initial mean estimates of the GPs are set so that $\mathbf{x}_{\mathrm{GP,t}} = \mathbf{0}$. The initial GP covariance matrix is block diagonal due to the assumption that all GPs are independent, so that

$$\mathbf{P}_{\mathrm{GP},0} = \oplus\Big(\oplus_{\mathrm{f}}\left(\mathbf{P}_{\mathrm{f,WV},0}\right), \oplus_{\mathrm{f}}\left(\mathbf{P}_{\mathrm{f,E},0}\right)\Big), \tag{23}$$

where, for compactness, the direct sum operator $\oplus$ denotes the construction of a block diagonal matrix of its arguments, and $\mathrm{f} \in \{Q^{-1}, \alpha, \beta, R_0\}$. The initial covariance for each WV-kernel-based GP is given by the Kronecker relation

$$\mathbf{P}_{\mathrm{f,WV},0} = k_{\mathrm{f},\mathbf{x}}(\mathbf{U}_{\mathrm{f}}, \mathbf{U}'_{\mathrm{f}}) \otimes \mathbf{P}_{\zeta_0,\mathrm{WV}}, \tag{24}$$

where $\mathbf{U}_{\mathrm{f}}$ are the coordinates of the discrete points chosen for each kernel function, i.e. the coordinates of (16). This is a discrete representation of the initial covariance of the spatially resolved white noise process.

The initial covariance for the Wiener velocity process, $\mathbf{P}_{\zeta_0,\mathrm{WV}}$, is determined by hyperparameters. In the standard formulation, the WV kernel (12) has *zero* covariance at time $\zeta = 0$. However, the parameters of the circuit model must be non-zero at the beginning of life, so a modification is required. To reconcile this, the WV kernel may be 'truncated' by replacing ζ with ζ_0, a non-zero value (see Appendix C). This gives the initial WV covariance as

$$\mathbf{P}_{\zeta_0,\mathrm{WV}} = \sigma^2_{\zeta,\mathrm{s}} \begin{bmatrix} \frac{1}{3}\zeta_0^3 & \frac{1}{2}\zeta_0^2 \\ \frac{1}{2}\zeta_0^2 & \zeta_0 \end{bmatrix}. \tag{25}$$

For the process noise,

$$\mathbf{P}_{\mathrm{E},0} = \sigma^2_{\zeta,\mathrm{r}}. \tag{26}$$

The GP mean and covariance describing the parameters only have to be initialised once for each battery model. The battery states, on the other hand, have to be re-initialised whenever there is a gap in telemetry data. For parameter estimation, not all measurement data are required because battery degradation is slow compared to the sampling frequency. Therefore, only a smaller number of specific charge/discharge cycles from within the larger dataset are needed. In this work, the simulated and experimental datasets each begin with a rest period, so we initialised the mean vector for battery states as

$$\mathbf{x}_{\mathrm{Batt},0} = \begin{bmatrix} V_0^{-1}(V_{\mathrm{t}}) \\ 0 \\ T_{\mathrm{amb,t}} \end{bmatrix} \tag{27}$$

at the start of each cycle. The state covariance was initialised with fixed values, so that

$$\mathbf{P}_{\mathrm{Batt},0} = \begin{bmatrix} P_{\mathrm{z},0} & 0 & 0 \\ 0 & P_{\mathrm{V_1},0} & 0 \\ 0 & 0 & P_{\mathrm{T},0} \end{bmatrix}, \tag{28}$$

and the overall system covariance $\mathbf{P}_0$ is formed by the block diagonal combination,

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{P}_{\mathrm{Batt},0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\mathrm{GP},0} \end{bmatrix}. \tag{29}$$

4.2.2 Solution in time. The mean and covariance of the joint system were solved in discrete time with an EKF [56],

$$\begin{aligned} \mathbf{x}_{\mathrm{t}}^{-} &= g(\mathbf{x}_{\mathrm{t-1}}, I_{\mathrm{t-1}}, T_{\mathrm{t-1}}) \\ \mathbf{P}_{\mathrm{t}}^{-} &= \mathbf{G}_{\mathrm{t-1}}\mathbf{P}_{\mathrm{t-1}}^{+}\mathbf{G}_{\mathrm{t-1}}^{\mathrm{T}} + \mathbf{Q}_{\mathrm{t-1}} + \lambda_{\mathrm{G,t-1}} \end{aligned} \tag{30}$$

where g describes both the system and parameter evolution dynamics, $\mathbf{G}_{\mathrm{t}}$ is the local Jacobian matrix of g at $\mathbf{x}_{\mathrm{t}}$, $\mathbf{Q}_{\mathrm{t}}$ is the joint discrete-time process covariance, and $\lambda_{\mathrm{G,t}}$ is an additional variance term arising from the posterior predictive variance of the GP. In addition, the predictive variances of α, β and R_0 have to be corrected for the uncertainty in the SOC, i.e., input variable z_{t}, as explained in Appendix B.

4.2.3 State and covariance propagation. Given estimates of $\alpha(z_{\mathrm{t}})$, $\beta(z_{\mathrm{t}})$, and $R_0(z_{\mathrm{t}}, I_{\mathrm{t}})$, discrete-time propagation of battery dynamics is approximated by using a zero-order hold on the applied current and linearising g with respect to the battery states at time t. The propagation of the GP states $\mathbf{x}_{\mathrm{GP,t}}$ is independent of the battery states and is linear, with the state transition given by

$$\mathbf{x}_{\mathrm{GP,t}} = \exp\Big(\oplus(\oplus_{\mathrm{f}}(\mathbf{I}_{\mathrm{f}} \otimes \mathbf{F}_{\mathrm{WV}}), \mathbf{F}_{\mathrm{E}})\Delta\zeta\Big)\mathbf{x}_{\mathrm{GP,t-1}}, \tag{31}$$

where $\mathbf{I}_{\mathrm{n}}$ is the identity matrix of size n, equal to the number of discrete 'spatial' points propagated through time for each GP, and

$$\mathbf{F}_{\mathrm{WV}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \ \mathbf{F}_{\mathrm{E}} = -\gamma_{\zeta,\mathrm{r}}\mathbf{I}_4. \tag{32}$$

The variable $\Delta\zeta_{\mathrm{t}}$ is the time step size in the GP 'degradation' timescale, which is much larger than the time step of the system dynamics. Within each discharge cycle this is assumed constant, therefore requiring only a single evaluation of the matrix exponential (31) for each discharge cycle. To improve computational efficiency in fitting the GPs representing α, β, and R_0 over lifetime, specific discharge cycles were down-selected from the raw data at an appropriate rate (in this case, one in thirty). Therefore, $\Delta\zeta$ is the cumulative charge throughput over a period of weeks rather than seconds. The discrete-time process noise covariance matrix $\mathbf{Q}_{\mathrm{t}}$ is block diagonal, with the values for the battery states fixed and the values for the GP determined by the kernel function hyperparameters. The WV kernel and exponential kernels have discrete-time process variances as a function of the step $\Delta\zeta$,

$$\mathbf{Q}_{\mathrm{t}} = \oplus(\mathbf{Q}_{\mathrm{Batt}}, \mathbf{Q}_{\mathrm{GP}}), \ \mathbf{Q}_{\mathrm{Batt}} = \oplus(q_{\mathrm{z}}, q_{\mathrm{V_1}}, q_{\mathrm{T_c}}), \tag{33}$$

and the block diagonal GP process covariance is

$$\begin{aligned} \mathbf{Q}_{\mathrm{GP}}(\Delta\zeta) = \oplus\Big(\oplus_{\mathrm{f}}\Big(k_{\mathrm{f},\mathbf{x}}(\mathbf{U}_{\mathrm{f}}, \mathbf{U}'_{\mathrm{f}}) \otimes \mathbf{Q}_{\mathrm{f,WV}}(\Delta\zeta)\Big), \\ \oplus_{\mathrm{f}}\left(Q_{\mathrm{f,E}}(\Delta\zeta)\right)\Big). \end{aligned} \tag{34}$$

where $\mathbf{U}_{\mathrm{f}}$ are the values of the coordinates of the discretisation points for each GP (16). The exponential process covariance is scalar for each of the functions f, and given by

$$Q_{\mathrm{E}}(\Delta\zeta) = \sigma^2_{\zeta,\mathrm{r}}\left(1 - \exp(-2\gamma_{\mathrm{r}}\Delta\zeta)\right). \tag{35}$$

Algorithm 1 Extended Kalman filter algorithm with NLML calculation. States are initialised at the beginning of every cycle and the GP is fitted over the battery lifetime. The NLML is calculated recursively for all available data.

Initialisation at $\zeta = \zeta_0$
  $\mathbf{x}^+_{\mathrm{GP}} = \mathbf{x}^+_{\mathrm{GP},0}$, $\mathbf{P}^+_{\mathrm{GP}} = \mathbf{P}^+_{\mathrm{GP},0}$
  $\phi_\mathrm{t} = 0$
for $\mathcal{D}_\mathrm{s} \in$ segments **do**
    Initialisation at start of segment
    $\mathbf{x}^+_{\mathrm{Batt}} = \mathbf{x}_{\mathrm{Batt},0}$, $\mathbf{P}^+_{\mathrm{Batt}} = \mathbf{P}_{\mathrm{Batt},0}$
    $\mathbf{x}^+_{\mathrm{GP}} = \exp(\mathbf{F}\Delta\zeta)\mathbf{x}^+_{\mathrm{GP}}$
    $\mathbf{P}^+_{\mathrm{GP}} = \exp(\mathbf{F}\Delta\zeta)\mathbf{P}^+_{\mathrm{GP}}\exp(\mathbf{F}\Delta\zeta)^\mathrm{T} + \mathbf{Q}_{\mathrm{GP}}(\Delta\zeta)$
    $\mathbf{x}^+ = \begin{bmatrix}\mathbf{x}^+_{\mathrm{Batt}} & \mathbf{x}^+_{\mathrm{GP}}\end{bmatrix}^\mathrm{T}$
    $\mathbf{P}^+ = \oplus\left(\mathbf{P}^+_{\mathrm{Batt}}, \mathbf{P}^+_{\mathrm{GP}}\right)$
    for $\mathrm{I_t}$,$\mathrm{V_t}$,$\mathrm{T_t} \in \mathcal{D}_\mathrm{s}$ **do**
      Propagation
      $\mathbf{x}^-_\mathrm{t} = g(\mathbf{x}_{\mathrm{t}-1}, I_{\mathrm{t}-1}, T_{\mathrm{t}-1})$
      $\mathbf{P}^-_\mathrm{t} = \mathbf{G}_{\mathrm{t}-1}\mathbf{P}^+_{\mathrm{t}-1}\mathbf{G}^\mathrm{T}_{\mathrm{t}-1} + \mathbf{Q}_{\mathrm{t}-1} + \lambda_{\mathrm{G},\mathrm{t}-1}$
      Observation and update
      $\mathbf{e}_\mathrm{t} = [V_\mathrm{t}\ T_\mathrm{t}]^\mathrm{T} - h(\mathbf{x}_\mathrm{t}, I_\mathrm{t})$
      $\mathbf{S}_\mathrm{t} = \mathbf{H}_\mathrm{t}\mathbf{P}^-_\mathrm{t}\mathbf{H}^\mathrm{T}_\mathrm{t} + \mathbf{R} + \lambda_{\mathrm{H},\mathrm{t}}$
      $\mathbf{K}_\mathrm{t} = \mathbf{P}^-_\mathrm{t}\mathbf{H}^\mathrm{T}_\mathrm{t}\mathbf{S}^{-1}_\mathrm{t}$
      $\mathbf{x}^+_\mathrm{t} = \mathbf{x}^-_\mathrm{t} + \mathbf{K}_\mathrm{t}\mathbf{e}_\mathrm{t}$
      $\mathbf{P}^+_\mathrm{t} = (\mathbf{I} - \mathbf{K}_\mathrm{t}\mathbf{H}_\mathrm{t})^\mathrm{T}\mathbf{P}^-_\mathrm{t}(\mathbf{I} - \mathbf{K}_\mathrm{t}\mathbf{H}_\mathrm{t}) + \mathbf{K}_\mathrm{t}\mathbf{R}\mathbf{K}^\mathrm{T}_\mathrm{t}$
      $\phi_\mathrm{t} = \phi_\mathrm{t} + \frac{1}{2}\mathbf{e}^\mathrm{T}_\mathrm{t}\mathbf{S}^{-1}_\mathrm{t}\mathbf{e}_\mathrm{t} + \frac{1}{2}\log|2\pi\mathbf{S}_\mathrm{t}|$
    end for
end for

4.3 Observation model. Following the forward solution of the system by (30), the predicted voltage and temperature are given by (10), which involves re-evaluating $R_0(z_\mathrm{t}, I_\mathrm{t})$ with the latest estimate of the state vector $\mathbf{x}_{\mathrm{R}_0,\mathrm{t}}$. The output equation (10) is nonlinear in SOC (z_t), due to the open-circuit voltage V_0 and $R_0(z_\mathrm{t}, I_\mathrm{t})$, but may be locally linearised to give the observation Jacobian $\mathbf{H}_\mathrm{t}$. The dependency of R_0 on z_t, which itself is a Gaussian random variable, can be calculated using the method discussed in Appendix A. The uncertainty in R_0 arising from GP extrapolation to the current operating point, $\lambda_{\mathrm{H},\mathrm{t}}$, is also incorporated into the predictive distribution for the output, which is

$$\begin{bmatrix}V_\mathrm{t}\\ T_\mathrm{t}\end{bmatrix} \sim \mathcal{N}(h(\mathbf{x}^-_\mathrm{t}, I_\mathrm{t}), \mathbf{S}_\mathrm{t}), \tag{36}$$

where the covariance $\mathbf{S}_\mathrm{t}$ of the output is given by

$$\mathbf{S}_\mathrm{t} = \mathbf{H}_\mathrm{t}\mathbf{P}^-_\mathrm{t}\mathbf{H}^\mathrm{T}_\mathrm{t} + \mathbf{R} + \lambda_{\mathrm{H},\mathrm{t}}, \tag{37}$$

and h is given by Eq. (10), $\mathbf{R}$ is the measurement noise covariance matrix that is estimated, and the calculation of $\lambda_{\mathrm{H},\mathrm{t}}$ may be found in the Appendix B.

4.4 Hyperparameter fitting and smoothed posterior. The battery model dynamics and outputs depend on the hyperparameters of the kernel functions k_ζ and $k_\mathbf{x}$, i.e., the magnitudes σ_ζ, $\sigma_{\zeta,\mathrm{r}}$, $\sigma_\mathbf{x}$ and (inverse) length scales $\gamma_\mathbf{x}$. In addition, we estimate from the data the observation noise parameters $\sigma_{\mathrm{n,V}}$ and $\sigma_{\mathrm{n,T}}$ for the output voltage and temperature. To obtain maximum likelihood estimates of the hyperparameters, the EKF algorithm may be augmented to include calculation of the negative log marginal likelihood (NLML) (5) over the observations [56]—the last step of Algorithm 1. Calculating this enables us to optimize the hyperparameters using a gradient-based optimiser. Analytical solutions for the gradient of the NLML may also be obtained [20, 59], but in this case we used automatic differentiation.

The final step after hyperparameter estimation is to calculate a posterior distribution of the latent states $\mathbf{x}_{\mathrm{f},\mathrm{t}}$ that is consistent with 'batch-mode' GP regression. The forward filtering distribution $p(\mathbf{x}_{\mathrm{f},\mathrm{t}}|\mathbf{y}_{1:\mathrm{t}})$, Algorithm 1, only includes observations up until point t. However, we require the so-called smoothing distribution $p(\mathbf{x}_{\mathrm{f},\mathrm{t}}|\mathbf{y}_{1:\mathrm{T}})$, i.e., the marginal distributions of the latent states conditioned on *all* available data. This is obtained using the RTS smoother [58], which operates by moving backward in time through the filtering distribution calculated earlier. This only needs to be calculated for the linear GP subsystem in the ζ timescale, therefore it only requires a number of steps equal to the number of discharge cycles in the data.

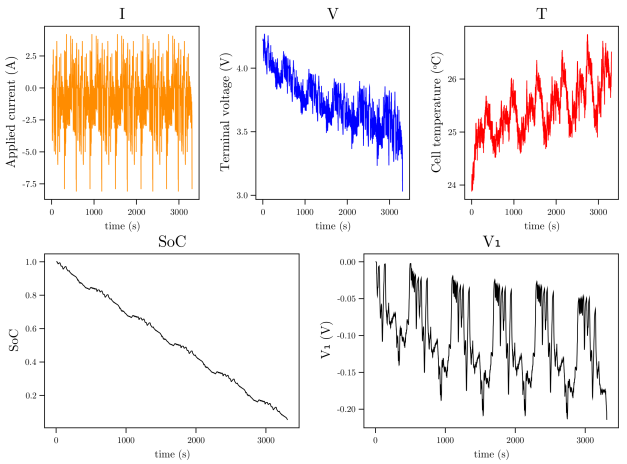


Fig. 3 Synthetic data for simulation

5 Simulation results

We now use synthetic data from simulations to obtain a known ground truth and demonstrate the effectiveness of the parameter estimation method. We simulated the voltage and temperature responses from the model (9) using a current profile from a US06 drive cycle [60]. The functions α, β and R_0 were chosen arbitrarily, and are shown in Table 1 alongside the other simulation parameters. To simulate observation noise, zero-mean Gaussian noise was added to the voltage and temperature measurements with standard deviations of 5 mV and 0.1 K respectively. The current, voltage, temperature, and internal model states (z and V_1) are shown in Fig. 3. The joint state estimator was applied to the simulated data, with 6 evenly spaced points over the range of SOC (z) used to describe the GPs for α and β, and an evenly spaced 4×15 grid over observed values of z and I for R_0. The constants in the transformation (11) for Q^{-1}, α, β and R_0 were set at $1.09\,\mathrm{Ah}^{-1}$, $0.01\,\mathrm{s}^{-1}$, $0.0007\,\mathrm{F}^{-1}$ and $0.04\,\Omega$ respectively. The GP hyperparameters, consisting of the length scales for α, β, R_0, the magnitudes for all four GPs, and the observation noise parameters $\sigma_{\mathrm{n,V}}$ and $\sigma_{\mathrm{n,T}}$, were estimated using a box-constrained Broyden-Fletcher-Goldfarb-Shanno (BFGS-B) optimisation algorithm implemented in the `Optim.jl` package in Julia [61], using forward-mode automatic differentiation to calculate the NLML gradients [62]. The constraints in the optimisation routine were imposed to guarantee numerical stability. We only considered in this synthetic data section the case where battery age ζ is constant, therefore requiring only a single discharge cycle in order to estimate the functions over operating conditions. In other words, the identifiability of the parameter functions can be considered independent of the input ζ in each case as long as the availability of data is similar throughout lifetime. If this is not the case, the GP framework will simply produce predictive posteriors with wider credible intervals where data are sparser.

The estimation results are shown in Fig. 4 and Table 1. From the projections of Fig. 4, it is clear that the Gaussian process estimator accurately retrieves the circuit parameter ground-truth values. However, a small loss of accuracy occurs for R_0 at low current I where the ground truth function has an inflection point. The maximum likelihood hyperparameter estimates are such that the GP

Parameter	Description	Value	Units	GPR RMSE (%)
$\alpha(z)$	Inverse RC pair time constant	$0.015 - 0.09\,(0.05 - z)^3$	s^{-1}	3.1
$\beta(z)$	Inverse RC pair capacitance	$0.002\left(1 - (z - 0.5)^2\right)$	F^{-1}	0.73
$R_0(z, I)$	Series resistor	$0.05\sinh^{-1}(\lvert I\rvert)/\lvert I\rvert + 0.04\,(z-1)^2$	Ω	0.97
Q^{-1}	Inverse cell capacity	1.2	Ah^{-1}	0.15
R_c	Thermal resistance	5.5	$K\,W^{-1}$	Given
C_c	Heat capacity	15.7	$J\,K^{-1}$	Given
V_0	Open-circuit potential	$3.64 + 0.55x - 0.72x^2 + 0.75x^3$	V	Given
$\sigma_{n,V}$	Voltage measurement noise st. dev.	0.005	V	0.4†
$\sigma_{n,T}$	Temperature measurement noise st. dev.	0.1	K	2.8†

Table 1 Simulation parameters and errors in estimates. The error in GPR estimates after a single discharge cycle is <4% for each circuit parameter, and is not strongly affect by adding more data. † these are a subset of the GP hyperparameters.

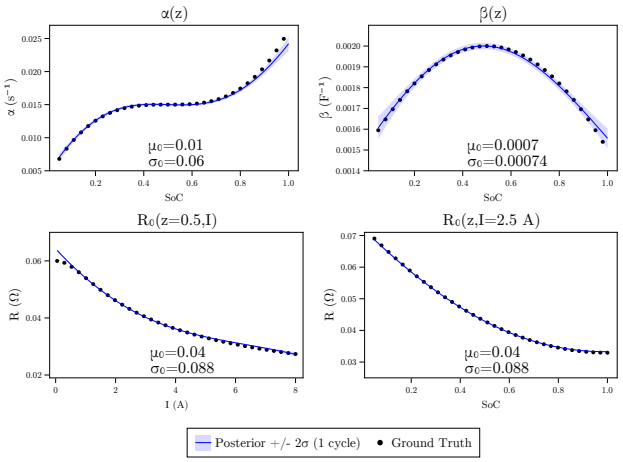


Fig. 4 Ground truth functions for $\alpha(z)$, $\beta(z)$, $R_0(z, I)$ together with their GP estimates using 1 cycle of input/output data. Estimation errors are reported in Table 1.

assumes a long length scale for $R_0(z, I)$ over both input dimensions, giving a predictive posterior that extrapolates to a higher value as $I \to 0$. This is partially due to poor numerical conditioning at very low applied current where a small error in voltage and/or current causes a large change in estimated R_0. The GP in this case relies on extrapolation from regions of higher current where the function is more identifiable.

6 GP Estimates from experimental data

To apply the framework to experimental data, we used time-series data of a cyclic ageing test including regular checkup tests with high energy 18650 Li-ion cells (Samsung SDI INR18650-35E, NCA vs. C+Si, 3450 mAh) with current, voltage and temperature measured. The data spans more than 800 full equivalent cycles (FEC), defined as half of the ratio of total Ampere-hour throughput to nominal capacity, and multiple cells. The latter were mounted in a holder, with active air cooling in a temperature chamber (Binder MK240), and cycled with a battery tester (Digatron MCFT 20-5-60 ME), which has a datasheet accuracy of ± 40 mA/10 mV and ± 20 mA/4 mV after calibration. The measured time series data is available online including metadata [63].

Cycle ageing was conducted at a temperature of 25 °C between 10 % and 90 % SOC. A constant-current constant-voltage (CCCV) charging protocol was used, with a current of 0.3C and a cutoff current of 0.02C, and discharging was achieved using a recorded and scaled drive-cycle profile with an average current of approximately 0.4C, shown in Fig. 5. We define C-rate in the usual manner, i.e., the reciprocal of the battery duration. The cycling profile was recorded from an electric delivery van, and is composed of four sections: inner-city driving (up to 1135 s), intercity driving (1135–2065 s), motorway driving (2065–2595 s) and rural hilly driving (2595 s to the end).

A checkup test procedure was conducted every 30 cycles and consisted of a 0.3C discharge capacity test after CCCV full charge followed by pulse tests at three SOC levels (80 %, 50 %, 20 %). The pseudo open-circuit voltage was determined at the beginning of life using a full discharge at 0.02C. To parameterise the thermal model in Eq. (9), the heat capacity of the cell ($43.5\,J\,K^{-1}$) was taken from literature [64]. The thermal resistance R_c was determined from the thermal relaxation following the first drive cycle at the beginning of life, using a least-squares fit, Fig. 5.

6.1 GP Estimation. Using the method of Section 4, we estimated the electrical parameters in the model (9) for two example Li-ion cells (cell numbers 009 and 015) in the dataset. We assumed that ageing was negligible within consecutive drive cycles between check-up sequences, and therefore only the cycling data from the final drive cycle in each set of 30 repeats was used for parameter estimation. Within each cycle, data were interpolated to a frequency of 1 Hz using the piece-wise cubic hermite method [65] to reduce the EKF computational load, avoiding the need to re-evaluate the matrix exponential in Eq. (31). For independent validation, we compared the estimated functional parameters against the checkup data, not used for model fitting, consisting of periodic 0.3C capacity test and internal resistance estimated from the average (over charge/discharge) $\Delta V/\Delta I$ from the first second of the pulse tests.

In addition, the dataset for each cell was split into two sections so as to investigate the method's ability to predict future degradation. Specifically, we left out the last 8 sets of cycles in each set, and these were used in an out-of-sample setting to assess the ability of the GP to forecast future SOH evolution. Overall, this resulted in an in-sample set of 19 and 20 discharge cycles for each of the two cells, respectively.

6.1.1 Hyperparameter estimation. The four GPs representing Q^{-1}, α, β and R_0 each have a multiplicative kernel function of the type (14), where in the case of Q^{-1} the function $k_x(\mathbf{x}, \mathbf{x}')$ is a constant because it has no dependency on z or I. In total, the hyperparameter vector $\boldsymbol{\Theta}_h$ controlling the properties of the 4 GPs over the inputs z, I and ζ contains 16 elements, so that

$$
\begin{aligned}
\boldsymbol{\Theta}_h &= [\boldsymbol{\Theta}_{SE,\mathbf{x}}\ \boldsymbol{\Theta}_{WV,\zeta}\ \boldsymbol{\Theta}_{E,\zeta}\ \sigma_{n,V}\ \sigma_{n,T}]\ , \text{ where} \\
\boldsymbol{\Theta}_{SE,\mathbf{x}} &= [\sigma_{Q^{-1}}\ \sigma_\alpha\ \sigma_\beta\ \sigma_{R_0}\ \gamma_{\gamma,z}\ \gamma_{\beta,z}\ \gamma_{R_0,z}\ \gamma_{R_0,I}], \\
\boldsymbol{\Theta}_{WV,\zeta} &= [\sigma_{Q^{-1},\zeta}\ \sigma_{\alpha,\zeta}\ \sigma_{\beta,\zeta}\ \sigma_{R_0,\zeta}], \\
\boldsymbol{\Theta}_{E,\zeta} &= [\sigma_{\zeta,r}\ \alpha_{\zeta,r}],
\end{aligned}
\tag{38}
$$

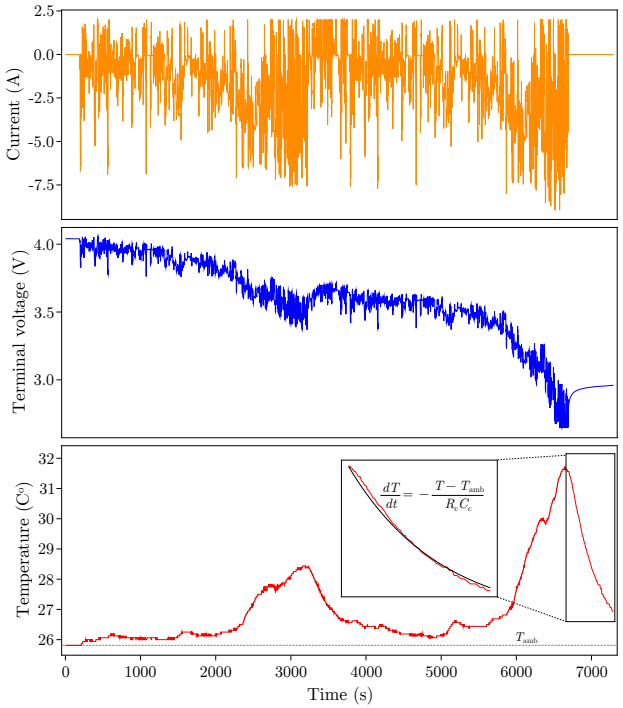


Fig. 5 Current, voltage and temperature profiles for the drive cycle used in experimental set-up. To parameterise the thermal model, we estimated R_c from the thermal relaxation at the end of the drive cycle using least-squares.

where $\boldsymbol{\Theta}_{\mathrm{SE},\mathbf{x}}$ are the hyperparameters for the SE kernel, and $\boldsymbol{\Theta}_{\mathrm{WV},\zeta}$, $\boldsymbol{\Theta}_{\mathrm{E},\zeta}$ are those for the WV and exponential kernels respectively, with measurement noise standard deviations given by $\sigma_{\mathrm{n,V}}$ and $\sigma_{\mathrm{n,T}}$ for the terminal voltage and cell temperature respectively. We assumed that both cells shared the same hyperparameters, and all estimation used the summed NLMLs across the both cells.

Several steps were taken to reduce the computational effort required for hyperparameter $\boldsymbol{\Theta}_{\mathrm{h}}$ estimation. Firstly, it was assumed that the length scale and magnitude parameters for the GPs for α and β could be shared, as they both relate to the behaviour of the RC pair over SOC, hence $\sigma_\alpha = \sigma_\beta$ and $\gamma_\alpha = \gamma_\beta$. Secondly, we grouped the WV magnitude parameters $\boldsymbol{\Theta}_{\mathrm{WV},\zeta}$ into two, so that $\boldsymbol{\Theta}_{\mathrm{WV},\zeta} = [\sigma_{0,\zeta}\ \sigma_{1,\zeta}]$, where $\sigma_{0,\zeta}$ was the WV kernel magnitude for Q^{-1} and $\sigma_{1,\zeta}$ the WV magnitude for α, β and R_0. The rationale for this grouping arises from assuming that the degradation process, *in relative terms* is similar for α, β and R_0, which are also dependent on operating conditions. This reduces the hyperparameter vector size to $\boldsymbol{\Theta}_{\mathrm{h}} \in \mathbb{R}^{12}$.

Finally, to use all available data (288,639 rows of I, V, T in total for the in-sample set) to estimate $\boldsymbol{\Theta}_{\mathrm{h}} \in \mathbb{R}^{12}$ would still require substantial computational effort due to the high dimensionality of the optimisation problem. The hyperparameter estimation problem was therefore split into two. As shown in Section 5, the dependency of each of the functions α, β and R_0 on battery states and operating conditions may be inferred from a single charge-discharge cycle.

Although this dependency may vary over the lifetime of the battery, and the multiplicative separable kernel formulation allows for this, we assumed that the hyperparameters controlling the GPs over z and I are constant throughout life. This reflects our assumption that the smoothness, lengthscale and periodicity of these functions does not change over life. With this assumption, we used the first available cycle from the beginning of life to estimate the subset of hyperparameters $\boldsymbol{\Theta}_{\mathrm{SE},\mathbf{x}}$. In addition to reducing the dimensionality of the estimation problem (as any hyperparameters related to ζ did not have to be estimated), the number of data rows required was reduced from 288,639 to 14,597.

A multi-start process was used in the optimiser. First, we estimated the NLML using 1000 randomly chosen points for $\boldsymbol{\Theta}_{\mathrm{SE},\mathbf{x}}$. We then chose the 25 lowest NLML points and applied the same gradient-based optimization algorithm as previously discussed, where the final $\boldsymbol{\Theta}_{\mathrm{SE},\mathbf{x}}$ values chosen were those with the lowest overall NLML value. Following this, $\boldsymbol{\Theta}_{\mathrm{WV},\zeta}$ and $\boldsymbol{\Theta}_{\mathrm{E},\zeta}$ were estimated using maximum likelihood together with fixed $\boldsymbol{\Theta}_{\mathrm{SE},\mathbf{x}}$ using the full in-sample dataset of 288,639 rows. In this case, the optimisation problem was only four-dimensional, so the number of iterations required was lower and only a single starting point was used, found by grid search. The box-constrained BFGS algorithm was again applied, with automatic differentiation. The box constraints additionally imposed a minimum on $\sigma_{\zeta,\mathrm{r}}$ to guarantee smoothness of the WV process. The battery model process noise covariance matrix was fixed,

$$\mathbf{Q}_{\mathrm{Batt}} = \begin{bmatrix} 10^{-12} & 0 & 0 \\ 0 & 10^{-6} & 0 \\ 0 & 0 & 10^{-4} \end{bmatrix}. \tag{39}$$

6.1.2 GP Posterior estimation and validation. Given maximum likelihood estimates for hyperparameters $\boldsymbol{\Theta}_{\mathrm{h}}$, the smoothed posterior of the ECM functional parameters was found using the RTS smoother over the battery lifetime, which consisted of 19-20 points, spaced through life, at the end of the down-selected discharge cycles. With the smoothed estimates of $\mathbf{x}_{\mathrm{GP,t}}$ and $\mathbf{P}_{\mathrm{GP,t}}$, we also extrapolated the GP estimates to the correct point in time to validate future life predictions. At these points in the time axis, we used the standard GP predictive equations (3) to retrieve the estimates of the functions over SOC and/or current, I.

6.2 Results and Discussion. Figure 6 shows 1-D projections of the GP estimates for capacity Q, resistance R_0, and circuit parameters α and β, over battery lifetime, for one of the experimental cells. As we used the summed NLML to determine hyperparameters using data from both cells, the results for the second cell are very similar (Appendix D). In Fig. 6, the GP posterior for the long-term smooth process is plotted, with the posterior of the noise process removed. The in-sample vs. out-of-sample (i.e., interpolation vs. extrapolation) division in each case is shown by the vertical black line. It is clear that the GP-estimated errors in capacity and R_0 are small and comparable to the values retrieved during the independent checkup tests. The errors are shown in Table 2, with an average of 0.016 Ah for capacity and 1.7 mΩ for internal resistance. The error between estimation and validation data for R_0 at SOCs of 20% and 80% appears larger than at 50% SOC. The reason is not clear but could relate to differences in experimental conditions between the fitting and validation steps. The trend for R_0 is approximately linear over time, and for capacity Q we observe a reduced rate of degradation as the battery ages, until the final checkup cycles where there is a possible increase in rate. Extrapolation of the GP with the WV kernel is linear along the last known trajectory, which in this case has similar accuracy to the interpolation for both Q and R_0. This extrapolation is thus suitable for short range predictions, but as can be seen, the uncertainty increases into the future. For longer range predictions a lifetime model would need to be built, but this is out of scope of the current work.

The entire R_0 function shifts upwards as the cell ages. The degradation in capacity (top left of Fig. 6) at early stages, measured during the first 20 checkup cycles, arguably has a slowing trend that could imply diffusion-limited SEI growth on the anode [66]. If this is the dominant ageing mechanism, the increase in R_0 could be due to the anode reaction current (i.e. the product of electrode surface area and exchange current density) decreasing. In the last few checkup cycles, a slightly increased rate of capacity degradation is observed, although this is not reflected as an acceleration in the increase in R_0. The upward shift in R_0 with age is not evenly

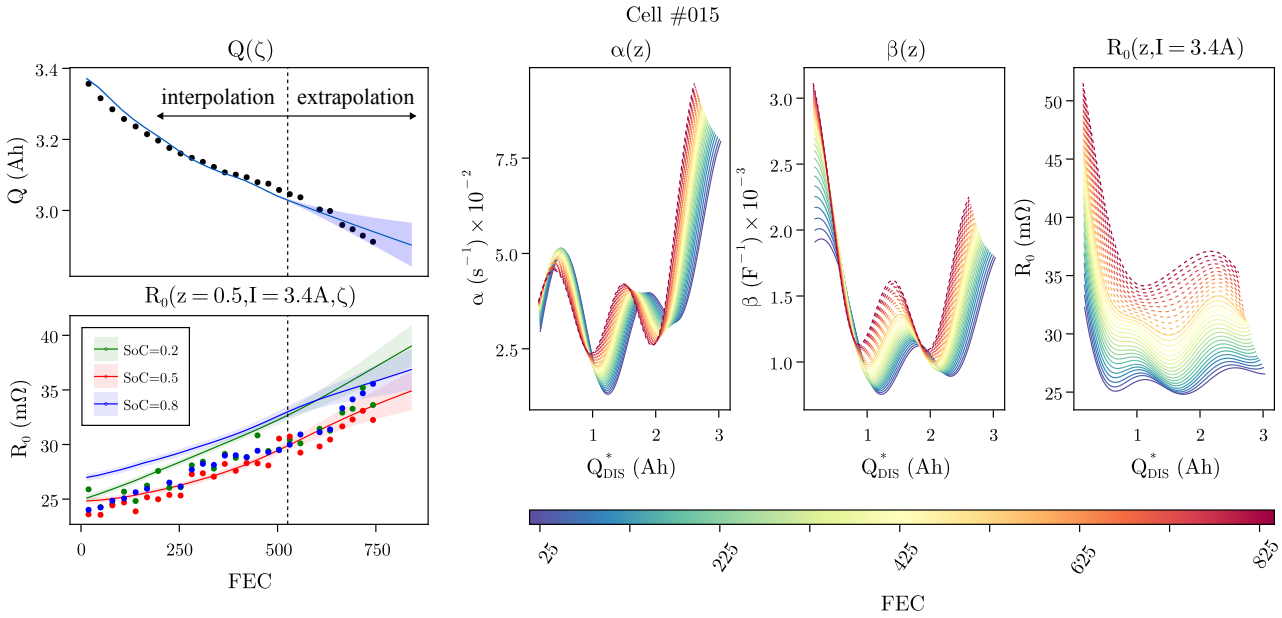


Fig. 6 Projections of GP posteriors for Q, R_0, α and β. Validation data for Q and R_0 (dots) show that the model accurately captures the evolution of SOH. The right three plots of α, β and R_0 as a function of discharge capacity show the strong dependency of ECM parameters on SOC (confidence bounds omitted for clarity). Colours indicate battery age as full equivalent cycles (FEC) and dashed lines denote extrapolated points in time equivalent to the Q and R_0 points to the right of the vertical dashed lines.

Cell	Case	Q (Ah)	R_0(z, I = 3.4 A) (mΩ)		
			z=0.2	z=0.5	z=0.8
9	Interpolated	0.018	2.0	0.92	2.7
	Extrapolated	0.014	2.2	0.98	1.8
15	Interpolated	0.017	2.0	0.83	2.8
	Extrapolated	0.016	1.9	1.5	1.7

Table 2 RMSE Values for capacity and internal resistance for GP mean vs. independent validation checkup tests, split by cell and GP interpolation or extrapolation cases.

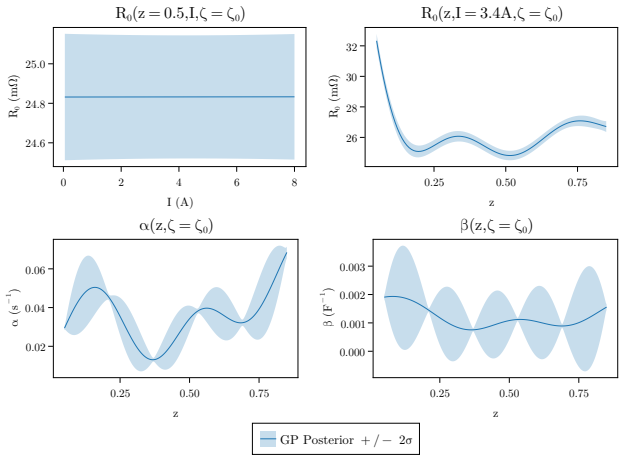


Fig. 7 ECM Parameter function estimates at the first discharge cycle for both cells. R_0 shows strong dependency on SOC but not current; α and β also show SOC-dependence, with a short GP length scale, implying high uncertainty.

spread as a function of SOC (right subplot of Fig. 6)—at high SOC (RHS of subplot), the curve shifts by approximately 8 mΩ, whereas low SOC (LHS of subplot), the shift is nearly 20 mΩ. The cells have a graphite anode with added silicon, and the latter often causes accelerated degradation of the electrode due to the large change in the volume of the silicon particles during charge and discharge, resulting in loss of active material [67]. As silicon participates most actively in the intercalation reactions at low SOC [68], it is consistent to see more substantial change in R_0 at low SOC due to mechanical degradation.

Each of the estimated functions for R_0, α, β are shown in more detail in Fig. 7, giving their dependencies on SOC and current, including beginning-of-life uncertainty bounds. We observe that R_0 is estimated to have no dependency on the applied current over the observed operating range, implying that linearisation of kinetics is probably reasonable, but the SOC-dependency of R_0 has an estimated range of 7 mΩ. Similarly α and β have substantial ranges over SOC, with an estimated GP length scale that is much shorter over SOC than for R_0. The short length scale is reflected in the rapid increase in predictive uncertainty in between the points at coordinates $\mathbf{U}_f$ that are propagated in state space. The evolution of the parallel RC-pair parameters over lifetime is not as clear as it is for the series resistance. Overall, we observe less change in α and β as a function of lifetime, but they both shift along the discharge capacity axis as the cell degrades.

The conventional RC parameterisation in terms of R_1 and C_1 may be retrieved from α and β and is shown in Fig. 8. Because a nonlinear transformation of Gaussian variables (inversion in this case) does not yield a Gaussian, we show the 50% percentile values for the parameters over battery lifetime, sampled from the posterior predictive ratios and inverses of α and β. The distributions are in fact heavily skewed because the GP for β has significant probability mass around zero. Overall, we observe an average 11% increase in R_1 and a 15% decrease in C_1 over the battery lifetime for both cells. A direct physical interpretation of the shape of the functions $\alpha(z,\zeta)$ and $\beta(z,\zeta)$, or the equivalent R_1 and C_1, is not straightforward.

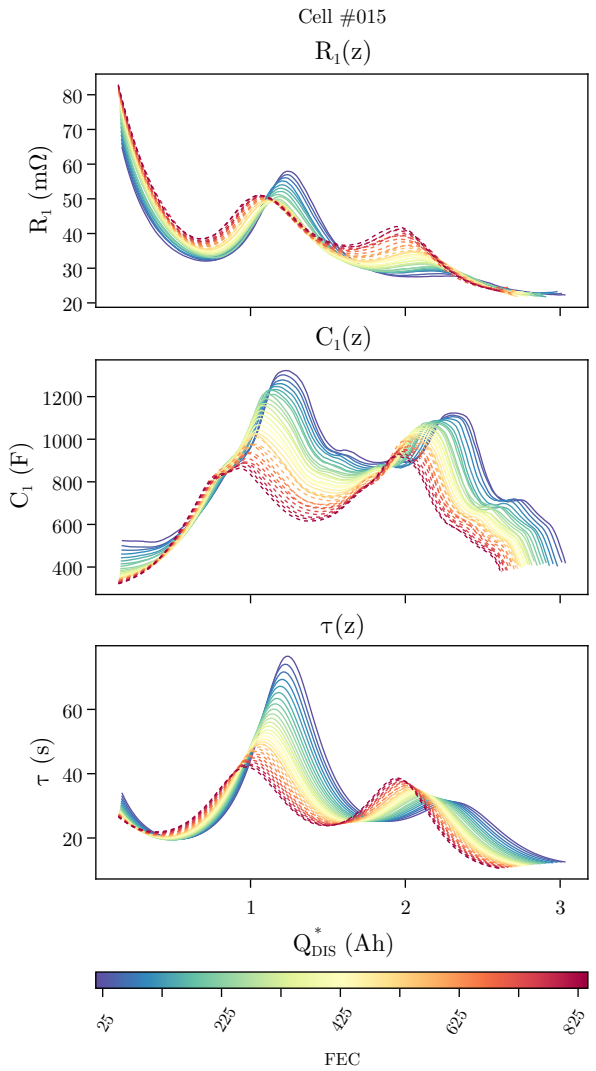


Fig. 8 Conventional RC circuit parameterisation calculated by sampling from the ratios of the estimated GPs β/α, $1/\beta$ and $1/\alpha$, over battery lifetime.

As the first order RC-circuit only includes a single time constant, the GP estimate of the time constant is likely to be a weighted mixture of different processes at both electrodes, such as charge transfer and diffusion. Another possible interpretation for α and β is the reaction relaxation time constant derived in Lin et al. [69], which describes the relaxation process of the inhomogeneity of SOC over the thickness of the electrode, originating from work by Newman and Tobias [70]. The time constant also shows little drift over lifetime, although at high SOC there is some increase matched by a decrease at low SOC, Fig. 8. This is the net effect of R_1 increasing by approximately 25% on average and C_1 decreasing equally. Ultimately, the time constant from a single RC pair ECM is difficult to connect with physics-based models as it likely reflects an average of several processes, as well as the local gradient of the open circuit voltage with respect to SOC.

7 Conclusions

This paper has demonstrated Gaussian process regression for estimating battery equivalent circuit parameters as functions of states and operating conditions using a computationally efficient recursive framework for the coupled estimation problem. Incorporating GP regression into the state-parameter estimation has multiple advantages. Firstly, using a GP kernel function to describe the evolution of parameters over battery lifetime gives a flexible method to extrapolate parameters into the future—existing literature has either used random walks or simple deterministic models for this. Secondly, by incorporating the operating-point dependency of each of the parameters, estimates are more stable across lifetime in real-world scenarios where conditions vary. Imposing a GP prior mitigates numerical ill-conditioning by acting as a regularisation mechanism in situations where parameters are not easily identifiable e.g. when estimating resistance with very low currents. Furthermore, the Bayesian framework provides estimates of parameter uncertainty, which is a function of model identifiability and the amount of data in the training set in the vicinity of the observed operating conditions and lifetime. In contrast to physics-based models, ECMs require minimal prior knowledge of battery parameters—only the open-circuit voltage curve is needed here.

In future, the applicability of this method could be explored where only partial charging or discharging is observed. In this case, the assumption of the independence of the different stochastic processes controlling battery evolution could be relaxed. For example, by introducing a non-zero prior covariance between parameters, parameters such as internal resistance could directly be mapped to remaining capacity, while taking into account the dependency of resistance on battery states and operating conditions. Additionally, the impact on model parameter estimates of changes in the open-circuit voltage curve over lifetime could be investigated.

Funding and conflict of interest

This work was supported by the Faraday Institution (ref. FIRG025) & Shell Foundation (ref. 22077). For the purpose of Open Access, the authors have applied a Creative Commons Attribution (CC BY) public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. D.A.H. is a co-founder of Brill Power Ltd.

Data and code availability

Code for this paper is may be requested from the authors. Experimental data are available at [63].

Appendix A: GP Prediction with input uncertainty

To evaluate the battery dynamics in (30), α, β and R_0 must be calculated. Ordinarily, the predictive means and variances are given by (3). However, because the SOC, z_t, is itself a Gaussian random variable, the predictive distributions must be marginalised (averaged) over the distribution of z_t by the integral

$$p\left(f\left(z_t\right)\right) = \int p\left(f(z_t)|z_t\right)p(z_t)\,\mathrm{d}z_t, \tag{A1}$$

evaluated at each time step, where $f(z_t)$ represents the predictive distribution of α, β and R_0. In other words, the average GP prediction across all possible values of the probability distribution of the present SOC, z_t, is calculated.

For an arbitrary kernel function k, this integral is intractable and can be approximated using, e.g., Taylor expansions [71]. However, for the SE kernel, (A1) has an analytical solution given by Quiñonero-Candela et al. [72], with the mean and variance of $p(f(z_t))$ for scalar input z being

$$\begin{aligned}\mathbb{E}[f(z)] &= \mathbf{l}_{\mathrm{f,z}}^{\mathrm{T}}\delta \\ \mathbb{V}[f(z)] &= \sigma_{\mathrm{f,GP}}^2(\mu_z) + \mathrm{Tr}\left((\mathbf{K}^{-1} - \mathbf{k}_{\mathrm{f,z,u}}\mathbf{k}_{\mathrm{f,z,u}}^{\mathrm{T}})\mathbf{L}\right) \\ &\quad + \mathrm{Tr}\left(\delta\delta^{\mathrm{T}}(\mathbf{L}_{\mathrm{f,z}} - \mathbf{l}_{\mathrm{f,z}}\mathbf{l}_{\mathrm{f,z}}^{\mathrm{T}})\right)\end{aligned} \tag{A2}$$

where Tr denotes the matrix trace and $\mathbf{K} = \mathbf{K}_{\mathbf{f,uu}} + \mathrm{diag}(\mathbf{P}_{\mathrm{f,t}})$ for each of the GPs that are functions of z_t. The parameter δ is given

by $\mathbf{K}^{-1}\mathbf{x}_{\mathrm{f,t}}$, where $\mathbf{x}_{\mathrm{f,t}}$ is the state vector for the GP at time t, and $\sigma^2_{\mathrm{GP}}(\mu_{\mathrm{z}})$ is the standard GP predictive variance evaluated at the mean of z_{t}. Finally, $\mathbf{l}_{\mathrm{z}}$, $\mathbf{L}_{\mathrm{z}}$ are given by

$$\begin{aligned}\mathbf{l}_{\mathrm{f,z}} &= \frac{\sigma^2_{\mathrm{f},\mathbf{x}}}{\sqrt{\gamma_{\mathrm{f,z}}\sigma^2_{\mathrm{z}}+1}}\exp\left(-\frac{1}{2(\gamma^{-1}_{\mathrm{f,z}}+\sigma^2_{\mathrm{z}})}(\mu_{\mathrm{z}}-\mathbf{u}_{\mathrm{f}})^2\right)\\ \mathbf{L}_{\mathrm{f,z}} &= \frac{\mathbf{k}_{\mathrm{f,z,u}}\mathbf{k}^{\mathrm{T}}_{\mathrm{f,z,u}}}{\sqrt{2\gamma_{\mathrm{f,z}}\sigma^2_{\mathrm{z}}+1}}\exp\left(\frac{2\gamma^2_{\mathrm{f,z}}}{2\gamma_{\mathrm{f,z}}+\sigma^2_{\mathrm{z}}}(\mu_{\mathrm{z}}-\overline{\mathbf{U}}_{\mathrm{f}})^2\right).\end{aligned} \tag{A3}$$

Here, the vector $\mathbf{u}_{\mathrm{f}}$ represents the coordinates of the discretisation points of z, and the element of matrix $\overline{\mathbf{U}}_{\mathrm{f,ij}} = \frac{1}{2}(\mathbf{u}_{\mathrm{f,i}} + \mathbf{u}_{\mathrm{f,j}})$. The variable $\gamma_{\mathrm{f,z}}$ is the inverse square of the length scale and $\sigma^2_{\mathrm{f,z}}$ is the magnitude of GP f over the SOC input z, and σ^2_{z} is the variance of z. The vector $\mathbf{k}_{\mathrm{f,z,u}}$ is the kernel function k_{SE} evaluated at μ_{z}, i.e. the same as $\mathbf{l}_{\mathrm{f,z}}$ where $\sigma^2_{\mathrm{z}} = 0$. In other words, uncertainty in z_{t} affects both the mean and variance of the GP output, changing the weighting of the linear combination of points x_{z} used to make a prediction and adding to the variance. As $\sigma^2_{\mathrm{z}} \to 0$, from equations (A2,A3) we recover the standard GP predictive equations.

Appendix B: Predictive variances in EKF recursion

The sparse matrices λ in (30) and (37) arise from accounting for extra uncertainties in the state dynamics and terminal voltage due to GP predictive variance. At each time-step, the GP predictive means and variances for α, β and R_0 are evaluated by Eqns. (A2, A3), which depend on z_{t} and/or applied current.

The 'extra variance' from GP interpolation or extrapolation in the λ terms results from the discretisation of the GPs over the inputs z_{t} and I_{t}, which affects the joint probability distributions relating to the state dynamics and output predictions [20, 73]. In the general case, the conditional distributions are such that

$$\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P}) \tag{B1a}$$

$$\theta|\mathbf{x} \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \mathbf{\Sigma}_{\mathrm{GP}}) \tag{B1b}$$

$$\mathbf{y}|\mathbf{x}, \theta \sim \mathcal{N}(g(\mathbf{x}, \theta), \Sigma_{\mathrm{n}}), \tag{B1c}$$

where the battery model parameter $\theta \in \{\alpha, \beta, R_0\}$ is a linear combination of the relevant GP states, which is then used in the state transition and observation function g to yield the predictive prior states and outputs. In the EKF, we locally linearise g, giving the joint distributions of $\mathbf{x}$, θ and $\mathbf{y}$,

$$\begin{bmatrix}\mathbf{x}\\ \theta\\ \mathbf{y}\end{bmatrix} \sim \mathcal{N}(\mathbf{m}^*, \mathbf{P}^*), \tag{B2}$$

where

$$\mathbf{m}^* = \begin{bmatrix}\mathbf{m}\\ \mathbf{H}\mathbf{m}\\ \mathbf{g}(\mathbf{m}, \mathbf{H}\mathbf{m})\end{bmatrix}, \tag{B3}$$

and

$$\mathbf{P}^* = \begin{bmatrix}\mathbf{P} & \mathbf{P}\mathbf{H}^{\mathrm{T}} & \mathbf{P}\mathbf{H}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\\ \mathbf{H}\mathbf{P} & \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}+\mathbf{\Sigma}_{\mathrm{GP}} & [\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}+\mathbf{\Sigma}_{\mathrm{GP}}]\mathbf{G}^{\mathrm{T}}\\ \mathbf{G}\mathbf{H}\mathbf{P}^{\mathrm{T}} & \mathbf{G}[\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}+\mathbf{\Sigma}_{\mathrm{GP}}] & \mathbf{G}[\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}+\mathbf{\Sigma}_{\mathrm{GP}}]\mathbf{G}^{\mathrm{T}}+\mathbf{\Sigma}_{\mathrm{n}}\end{bmatrix}, \tag{B4}$$

where $\mathbf{G}$ is the Jacobian of g. Because the model parameters that are discretised over z_{t} and I_{t} include α, β and R_0, $\mathbf{\Sigma}_{\mathrm{GP}}$ will be very sparse—it will only have non-zero terms for the state transitions of T and V_1 in (30) and the terminal voltage prediction in (37). To simplify the calculation, we refactored the bottom right-hand term in (B4) to obtain the λ terms in (30, 37), where the entries for each are given by

$$\lambda_{\mathrm{G,t}}[2,2] = \sigma^2_{\mathrm{GP},\beta}\left(\frac{I_{\mathrm{t}}(1-\exp(-\Delta t\mu_{\alpha,\mathrm{t}}))}{\mu_{\alpha,\mathrm{t}}}\right)^2 + \sigma^2_{\mathrm{GP},\alpha}V^2_{1,\mathrm{t}}, \tag{B5a}$$

$$\lambda_{\mathrm{G,t}}[3,3] = \sigma^2_{\mathrm{GP},\mathrm{R}_0}I^4_{\mathrm{t}}\left(1-\exp(-\frac{\Delta t}{R_{\mathrm{c}}C_{\mathrm{c}}})\right)^2, \tag{B5b}$$

$$\lambda_{\mathrm{H,t}}[1,1] = \sigma^2_{\mathrm{GP},\mathrm{R}_0}I^2_{\mathrm{t}} \tag{B5c}$$

Appendix C: Initialisation of WV process

The overall kernel describing each of the four GPs in the circuit model (9) is separable over the operating point and battery lifetime and has a form

$$\begin{aligned}k_{\mathrm{f}} &= k_{\zeta}(\zeta,\zeta')k_{\mathbf{x}}(\mathbf{x},\mathbf{x}')\\ &= \sigma^2_{\zeta}\left(\frac{\min^3(\zeta,\zeta')}{3}+|\zeta-\zeta'|\frac{\min^2(\zeta,\zeta')}{2}\right)\times\\ &\quad \sigma^2_{\mathbf{x}}\exp\left(-\frac{1}{2}\sum_{\mathbf{x}}\gamma_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')^2\right),\end{aligned} \tag{C1}$$

which zero covariance at beginning of life $\zeta = 0$. To correct for this and split the hyperparameter estimation problem in two (see Section 6.1.1), we reformulated the WV kernel so that its prior variance at beginning of life was equal to the $\sigma^2_{\mathbf{x}}$ for each of the GPs, where $\sigma^2_{\mathbf{x}}$ were estimated in the first phase of the estimation process. This is equivalent to shifting ζ to a non-zero starting point ζ_0. To find the correct initial conditions in the recursive formulation, it is sufficient to solve (25) for ζ_0 in the top-left hand element, so that

$$\frac{1}{3}\sigma^2_{\zeta}\zeta^3_0 = \sigma^2_{\mathbf{x}} \tag{C2}$$

for a given $\sigma^2_{\mathbf{x}}$ and σ^2_{ζ}. This then yields the consistent $\mathbf{P}_0(\zeta_0)$ for each of the GPs estimated.

Appendix D: Results from second experimental cell

The second cell tested (number 009) shows fitted functional parameter estimates (Figs. 9, 10) with very similar patterns to those of cell 015. This is due to the similarity in experimental conditions and the method of estimating GP hyperparameters by minimising the NLML using data from both cells.

References

[1] Figgener, J., Stenzel, P., Kairies, K. P., Linßen, J., Haberschusz, D., Wessels, O., Angenendt, G., Robinius, M., Stolten, D., and Sauer, D. U., 2020, "The development of stationary battery storage systems in Germany – A market review," Journal of Energy Storage, **29**(December 2019), p. 101153.

[2] Ioannis Tsiropoulos, Dalius Tarvydas, and Natalia Lebedeva, 2018, *Li-ion batteries for mobility and stationary storage applications - Scenarios for costs and market growth* .

[3] Ziegler, M. S., Song, J., and Trancik, J. E., 2021, "Determinants of lithium-ion battery technology cost decline," Energy & Environmental Science, **14**, pp. 6074–6098.

[4] DNV Renewables Advisory, 2021, "Techno-economic Analysis of Battery Energy Storage for Reducing Fossil Fuel Use in Sub-Saharan Africa," https://www.faraday.ac.uk/wp-content/uploads/2021/10/TEA_Report_Faraday_BESS_FINAL.pdf

[5] IEA, 2021, "The Role of Critical Minerals in Clean Energy Transitions," doi: 10.1787/f262b91c-en.

[6] Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y. T., Steinbacher, F., Khan, M. U., Lee, J. W., Siegel, J. B., Stefanopoulou, A. G., and Howey, D. A., 2021, "The challenge and opportunity of battery lifetime prediction from field data," Joule, **5**(8), pp. 1934–1955.

[7] Berecibar, M., Gandiaga, I., Villarreal, I., Omar, N., Van Mierlo, J., and Van Den Bossche, P., 2016, "Critical review of state of health estimation methods of Li-ion batteries for real applications," doi: 10.1016/j.rser.2015.11.042.

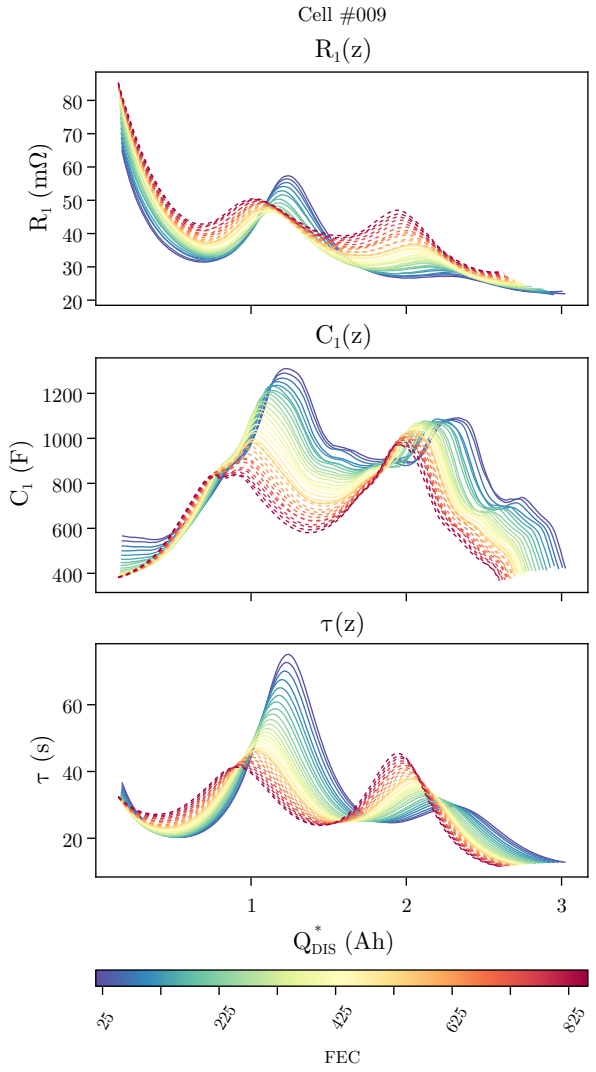


Fig. 9 Cell 009 RC parameterisation.

[8] von Bülow, F. and Meisen, T., 2023, "A review on methods for state of health forecasting of lithium-ion batteries applicable in real-world operational conditions," Journal of Energy Storage, **57**, p. 105978.

[9] Dubarry, M., Baure, G., and Anseán, D., 2020, "Perspective on state-of-health determination in lithium-ion batteries," Journal of Electrochemical Energy Conversion and Storage, **17**(4), p. 044701.

[10] Ling, L. and Wei, Y., 2021, "State-of-Charge and State-of-Health Estimation for Lithium-Ion Batteries Based on Dual Fractional-Order Extended Kalman Filter and Online Parameter Identification," IEEE Access, **9**, pp. 47588–47602.

[11] Baba, A. and Adachi, S., 2015, "Simultaneous state of charge and parameter estimation of lithium-ion battery using log-normalized unscented Kalman Filter," *2015 American Control Conference (ACC)*, IEEE, pp. 311–316, doi: 10.1109/ACC.2015.7170754.

[12] Plett, G. L., 2004, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation," Journal of Power Sources, **134**(2), pp. 277–292.

[13] Plett, G. L., 2006, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs Part 2: Simultaneous state and parameter estimation," Journal of Power Sources, **161**, pp. 1369–1384.

[14] Kim, I. S., 2010, "A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer," IEEE Transactions on Power Electronics, **25**(4), pp. 1013–1022.

[15] Ascencio, P., Smith, K., Howey, D., and Monroe, C. W., 2019, "Augmented State Observer for Simultaneous Estimation of Charge State and Crossover in Self-Discharging Disproportionation Redox Flow Batteries," CCTA 2019 - 3rd IEEE Conference on Control Technology and Applications, pp. 481–486.

[16] Saha, B., Goebel, K., Poll, S., and Christophersen, J., 2009, "Prognostics methods for battery health monitoring using a Bayesian framework," IEEE Transactions on Instrumentation and Measurement, **58**(2), pp. 291–296.

[17] Guha, A. and Patra, A., 2018, "State of Health Estimation of Lithium-Ion Batteries Using Capacity Fade and Internal Resistance Growth Models," IEEE Transactions on Transportation Electrification, **4**(1), pp. 135–146.

[18] Gomez, J., Nelson, R., Kalu, E. E., Weatherspoon, M. H., and Zheng, J. P., 2011, "Equivalent circuit model parameters of a high-power Li-ion battery: Thermal and state of charge effects," Journal of Power Sources, **196**(10), pp. 4826–4831.

[19] Remmlinger, J., Buchholz, M., Meiler, M., Bernreuter, P., and Dietmayer, K., 2011, "State-of-health monitoring of lithium-ion batteries in electric vehicles by on-board internal resistance estimation," Journal of Power Sources, **196**(12), pp. 5325–5331.

[20] Aitio, A. and Howey, D. A., 2021, "Predicting battery end of life from solar off-grid system field data using machine learning," Joule, **5**(12), pp. 3204–3220.

[21] Doyle, M., Fuller, T. F., and Newman, J., 1993, "Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymer/Insertion Cell," Journal of The Electrochemical Society, **140**(6), p. 1526.

[22] Fuller, T. F., Doyle, M., and Newman, J., 1994, "Simulation and Optimization of the Dual Lithium Ion Insertion Cell," Journal of The Electrochemical Society, **141**(1), p. 1.

[23] Aitio, A., Marquis, S. G., Ascencio, P., and Howey, D., 2020, "Bayesian parameter estimation applied to the Li-ion battery single particle model with electrolyte dynamics," IFAC-PapersOnLine, **53**(2), pp. 12497–12504.

[24] Bizeray, A. M., Kim, J., Duncan, S. R., and Howey, D. A., 2018, "Identifiability and parameter estimation of the single particle lithium-ion battery model," IEEE Transactions on Control Systems Technology, **27**(5), pp. 1862–1877.

[25] Chaoui, H. and Ibe-Ekeocha, C. C., 2017, "State of Charge and State of Health Estimation for Lithium Batteries Using Recurrent Neural Networks," IEEE Transactions on Vehicular Technology, **66**(10), pp. 8773–8783.

[26] Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D. U., 2021, "Online capacity estimation of lithium-ion batteries with deep long short-term memory networks," Journal of Power Sources, **482**, p. 228863.

[27] Qin, Y., Yuen, C., Yin, X., and Huang, B., 2023, "A Transferable Multistage Model With Cycling Discrepancy Learning for Lithium-Ion Battery State of Health Estimation," IEEE Transactions on Industrial Informatics, **19**(2), pp. 1933–1946.

[28] Wang, Q., Ye, M., Cai, X., Sauer, D. U., and Li, W., 2023, "Transferable data-driven capacity estimation for lithium-ion batteries with deep learning: A case study from laboratory to field applications," Applied Energy, **350**, p. 121747.

[29] Richardson, R. R., Birkl, C. R., Osborne, M. A., and Howey, D. A., 2019, "Gaussian Process Regression for in Situ Capacity Estimation of Lithium-Ion Batteries," IEEE Transactions on Industrial Informatics, **15**(1), pp. 127–138.

[30] You, G.-w., Park, S., and Oh, D., 2016, "Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach," Applied Energy, **176**, pp. 92–103.

[31] Greenbank, S. and Howey, D., 2021, "Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life," IEEE Transactions on Industrial Informatics, **3203**(c), pp. 1–9.

[32] Qin, Y., Arunan, A., and Yuen, C., 2023, "Digital Twin for Real-time Li-Ion Battery State of Health Estimation With Partially Discharged Cycling Data," IEEE Transactions on Industrial Informatics, **19**(5), pp. 7247–7257.

[33] Wang, Q., Wang, Z., Liu, P., Zhang, L., Sauer, D. U., and Li, W., 2023, "Large-scale field data-based battery aging prediction driven by statistical features and machine learning," Cell Reports Physical Science, **4**(12), p. 101720.

[34] Steininger, V., Rumpf, K., Hüsson, P., Li, W., and Sauer, D. U., 2023, "Automated feature extraction to integrate field and laboratory data for aging diagnosis of automotive lithium-ion batteries," Cell Reports Physical Science, **4**(10), p. 101596.

[35] Richardson, R. R., Osborne, M. A., and Howey, D. A., 2019, "Battery health prediction under generalized conditions using a Gaussian process transition model," Journal of Energy Storage, **23**(November 2018), pp. 320–328.

[36] Richardson, R. R., Osborne, M. A., and Howey, D. A., 2017, "Gaussian process regression for forecasting battery state of health," Journal of Power Sources, **357**, pp. 209–219.

[37] Zhang, C., He, Y., Yuan, L., and Xiang, S., 2017, "Capacity Prognostics of Lithium-Ion Batteries using EMD Denoising and Multiple Kernel RVM," IEEE Access, **5**, pp. 12061–12070.

[38] Dong, G., Yang, F., Wei, Z., Wei, J., and Tsui, K.-L., 2020, "Data-Driven Battery Health Prognosis Using Adaptive Brownian Motion Model," IEEE Transactions on Industrial Informatics, **16**(7), pp. 4736–4746.

[39] Xu, P., Hu, X., Liu, B., Ouyang, T., and Chen, N., 2022, "Hierarchical Estimation Model of State-of-Charge and State-of-Health for Power Batteries Considering Current Rate," IEEE Transactions on Industrial Informatics, **18**(9), pp. 6150–6159.

[40] Li, W., Chen, J., Quade, K., Luder, D., Gong, J., and Sauer, D. U., 2022, "Battery degradation diagnosis with field data, impedance-based modeling and artificial intelligence," Energy Storage Materials, **53**, pp. 391–403.

[41] Brucker, J., Bessler, W. G., and Gasper, R., 2021, "Grey-box modelling of lithium-ion batteries using neural ordinary differential equations," doi: 10.1186/s42162-021-00170-8.

[42] Fan, K., Wan, Y., and Jiang, B., 2022, "State-of-charge dependent equivalent circuit model identification for batteries using sparse Gaussian process regression," Journal of Process Control, **112**, pp. 1–11.

[43] Chen, X., Chen, X., and Chen, X., 2021, "A novel framework for lithium-ion battery state of charge estimation based on Kalman filter Gaussian process regression," International Journal of Energy Research, **45**(9), pp. 13238–13249, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.6649.

[44] Arno Solin, 2016, "Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression," Ph.D. thesis, Aalto

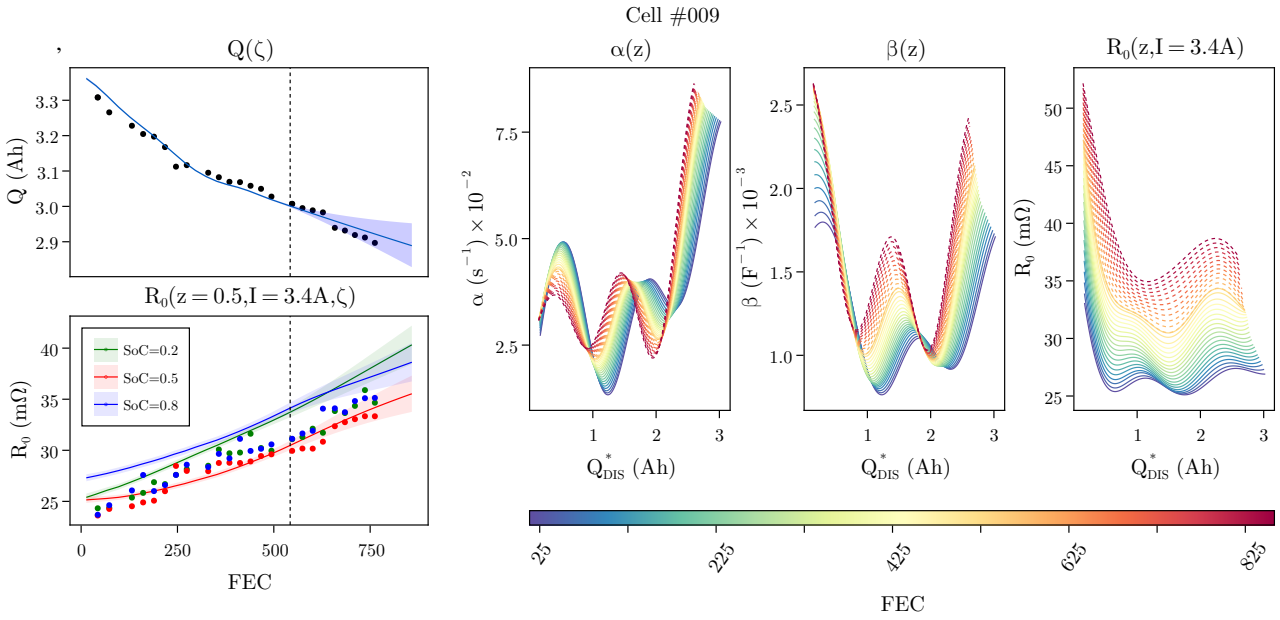


Fig. 10 GP posteriors for cell 009 in the experimental dataset, showing patterns consistent with cell 015.

University, https://aaltodoc.aalto.fi/bitstream/handle/123456789/19842/isbn9789526067117.pdf?sequence=1&isAllowed=y

[45] Rasmussen, C. E. and Williams, C. K. I., 2006, *Gaussian processes for machine learning*, Adaptive computation and machine learning, MIT, Cambridge, Mass. ; London.

[46] Bishop, C. M., 2006, "Pattern recognition and machine learning," Springer google schola, **2**, pp. 1122–1128.

[47] Görtler, J., Kehlbeck, R., and Deussen, O., 2019, "A Visual Exploration of Gaussian Processes," Distill, https://distill.pub/2019/visual-exploration-gaussian-processes.

[48] Pietiläinen, V., 2010, "Approximations for Integration over the Hyperparameters in Gaussian Processes," Ph.D. thesis, https://aaltodoc.aalto.fi/handle/123456789/3174

[49] Kucukelbir, A., Blei, D. M., Gelman, A., Ranganath, R., and Tran, D., 2017, "Automatic Differentiation Variational Inference," Journal of Machine Learning Research, **18**, pp. 1–45.

[50] Vanhatalo, J., Pietiläinen, V., and Vehtari, A., 2010, "Approximate inference for disease mapping with sparse Gaussian processes," Statistics in Medicine, **29**(15), pp. 1580–1607.

[51] Quiñonero-Candela, J. and Rasmussen, C. E., 2005, "A unifying view of sparse approximate Gaussian process regression," Journal of Machine Learning Research, **6**, pp. 1939–1959.

[52] Saatchi, Y., 2011, "Scalable Inference for Structured Gaussian Process Models," Dissertation.

[53] Solin, A. and Särkkä, S., 2019, "Hilbert space methods for reduced-rank Gaussian process regression," Statistics and Computing.

[54] Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., and Wilson, A. G., 2018, "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration," CoRR, **abs/1809.1**.

[55] Lindgren, F. and Rue, H., 2011, "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," Journal of the Royal Statistical Society. Series B, **73**(4), pp. 423–498.

[56] Särkkä, S., Solin, A., and Hartikainen, J., 2013, "Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering," IEEE Signal Processing Magazine, **30**(4), pp. 51–61.

[57] Sarkka, S. and Hartikainen, J., 2012, "Infinite-Dimensional Kalman Filtering Approach to Spatio-Temporal Gaussian Process Regression," *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, N. D. Lawrence and M. Girolami, eds., Proceedings of Machine Learning Research, Vol. 22, PMLR, La Palma, Canary Islands, pp. 993–1001, https://proceedings.mlr.press/v22/sarkka12.html

[58] Rauch, H. E., Tung, F., and Striebel, C. T., 1965, "Maximum likelihood estimates of linear dynamic systems," AIAA Journal, **3**(8), pp. 1445–1450.

[59] Mbalawata, I. S., Särkkä, S., and Haario, H., 2013, "Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering," Computational Statistics, **28**(3), pp. 1195–1223.

[60] Gregory L. Plett, 2015, *Battery Management Systems, Volume II: Equivalent-Circuit Methods*, Artech House Publishers.

[61] Mogensen, P. K. and Riseth, A. N., 2018, "Optim: A mathematical optimization package for {Julia}," Journal of Open Source Software, **3**(24), p. 615.

[62] Revels, J., Lubin, M., and Papamarkou, T., 2016, "Forward-Mode Automatic Differentiation in {J}ulia," arXiv:1607.07892 [cs.MS].

[63] Jöst, D., Ringbeck, F., Blömeke, A., and Sauer, D. U., 2021, "Timeseries data of a drive cycle aging test of 28 high energy NCA/C+Si round cells of type 18650," doi: 10.18154/RWTH-2021-02814.

[64] Steinhardt, M., Gillich, E. I., Rheinfeld, A., Kraft, L., Spielbauer, M., Bohlen, O., and Jossen, A., 2021, "Low-effort determination of heat capacity and thermal conductivity for cylindrical 18650 and 21700 lithium-ion cells," Journal of Energy Storage, **42**(August), p. 103065.

[65] Carlson, R. E. and Fritsch, F. N., 1980, "Monotone piecewise cubic interpolation," SIAM J Numer. Anal., **17**(2), pp. 238–246.

[66] Reniers, J. M., Mulder, G., and Howey, D. A., 2019, "Review and Performance Comparison of Mechanical-Chemical Degradation Models for Lithium-Ion Batteries," Journal of The Electrochemical Society, **166**(14), pp. A3189–A3200.

[67] Kirkaldy, N., Samieian, M. A., Offer, G. J., Marinescu, M., and Patel, Y., 2022, "Lithium-Ion Battery Degradation: Measuring Rapid Loss of Active Silicon in Silicon-Graphite Composite Electrodes," ChemRxiv.

[68] Ai, W., Kirkaldy, N., Jiang, Y., Offer, G., Wang, H., and Wu, B., 2022, "A composite electrode model for lithium-ion batteries with silicon/graphite negative electrodes," Journal of Power Sources, **527**(December 2021), p. 231142.

[69] Lin, J., Chu, H. N., Howey, D. A., and Monroe, C. W., 2022, "Multiscale coupling of surface temperature with solid diffusion in large lithium-ion pouch cells," Communications Engineering, **1**(1).

[70] Newman, J. S. and Tobias, C. W., 1962, "Theoretical Analysis of Current Distribution in Porous Electrodes," Journal of The Electrochemical Society, **109**(12), p. 1183.

[71] Girard, A., Rasmussen, C. E., Candela, J. Q., and Murray-Smith, R., 2003, "Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting," Advances in Neural Information Processing Systems, (December 2002), pp. 9–14.

[72] Quinonero-Candela, J., Girard, A., and Rasmussen, C., 2003, "Prediction at an Uncertain Input for Gaussian Processes and Relevance Vector Machines Application to Multiple-Step Ahead Time-Series Forecasting," Technical Report, (1).

[73] Särkkä, S., 2013, *Bayesian Filtering and Smoothing*, Cambridge University Press.