Lawrence Berkeley National Laboratory
LBL Publications

Title

Machine-Learning-Based Efficient Parameterspace Exploration for Energy Storage Systems

Authors

Harris, Stephen J

Alghalayini, Maher

Noack, Marcus

et al.

Machine-Learning-Based Efficient Parameter Space Exploration for Energy Storage Systems

Maher B. Alghalayini[1,2*], Daniel Collins-Wildman[1], Kenneth Higa[1], Armina Guevara[1], Vincent Battaglia[1], Marcus M. Noack[2*], Stephen J. Harris[1]

[1]Energy Storage and Distributed Resources Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, 94720, California, USA.

[2]Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, 94720, California, USA.

*Corresponding author(s). E-mail(s): MAlghalayini@lbl.gov; MarcusNoack@lbl.gov;

Contributing authors: dcollinswildman@lbl.gov; khiga@lbl.gov; arminamguevara@berkeley.edu; vsbattaglia@lbl.gov; SJHarris@lbl.gov;

Abstract

Shifting towards sustainable energy sources requires developing new storage systems and estimating their remaining energy over their lifetime. The remaining energy of these systems depends on many operating parameters, resulting in a large high-dimensional parameter space to explore. Testing cells exhaustively on a dense grid in the parameter space is prohibitively expensive. This is especially true with considerable cell-to-cell variability in performance, even under the same cycling conditions. Here, we develop a framework based on Gaussian processes, equipped with domain knowledge, to implement Bayesian optimization to explore the parameter space efficiently and quantify remaining energy using failure distributions. Bayesian optimization identifies future experiments that maximize information gain and minimize uncertainty. Experimental results show accurate remaining energy predictions with significantly fewer experiments. However, laboratory cycling conditions, including those in literature, may not represent real-world cycling. We propose an approach based on laboratory results to predict remaining energy under real-world cycling conditions.

1

1 INTRODUCTION

In the fight against global warming, the demand for new energy storage technologies has increased dramatically. Although solar and wind have great potential to fight global warming, these intermittent sources hinder grid integration. Energy storage systems have risen as a popular solution [1]. Within the grid, energy storage systems store excess energy during peak generation periods and release energy when needed during low-energy generation periods. While commercially available, batteries for both electric vehicles and the grid require enhanced technology to improve energy density and safety. One essential requirement for developing new technology storage systems is the ability to predict their remaining energy over their lifetime under various operating conditions without testing every possible system to failure.

Published lifetime studies generally aim to predict a battery's expected life. We argue that even a perfect ability to predict the expected life has little commercial value to battery manufacturers or to utilities. For example, lifetime appears in a business plan primarily through the cost of providing a warranty, which depends only on the fraction of early outlier failures; and through inventory control, which depends on knowing the fraction of failures as a function of time. Both require knowing the failure probability distribution, which we will determine, rather than the expected life [2, 3].

Various parameters affect the remaining energy of storage systems throughout their lifetime [4], including operating conditions like temperature [5], charging rate [6], depth of discharge [7], rest periods and duty cycle [8]; as well as chemical and physical parameters of the cells [9]. Each parameter can take on multiple values, and any combination of these values may result in a different remaining energy. All possible combinations form an enormous multi-dimensional parameter space for researchers to explore when optimizing remaining energy. Such a space cannot be studied using a traditional design of experiments, even with a coarse grid. This daunting challenge is being addressed with both physics-based models (involving how batteries degrade) and machine learning-based models (focusing on when batteries degrade) [10, 11]. The most widely-used physics-based model, focusing on transport and electrochemistry, was developed by Newman and co-workers [12, 13, 14, 15, 16]. Multiple other physical/empirical models have been developed, for instance, empirical aging models [17, 18, 19], equivalent circuit models [20, 21, 22], and numerous electrochemical models [23, 24, 25].

Data-driven approaches based on machine learning models rely on experimental data to make predictions. These can be "black-box" models, where the input variables are external indicators extracted from the experimental data, and the output is some prediction for lifetime [26]. These approaches include Gaussian processes (GPs) models [27, 28, 29, 11], Deep Neural Networks (DNN) [30, 31, 32, 33], and Long Short-Term Memory Networks [34, 35, 36], support vector machine [28, 37], and relevance vector machines [38, 39]. In our recent work [40], we developed a technology-agnostic framework based on a domain-knowledge-informed Gaussian process (GP) model to estimate

the failure distribution. Still, that work implicitly assumed that lifetime depends only on the cycle number. Since degradation depends on the operating conditions of the energy storage system, the failure distribution should ideally be predictable at any point in the multidimensional parameter space. To achieve this aim, the parameter space must be explored with extreme efficiency. Here, we develop a framework based on Gaussian processes equipped with domain knowledge and a Bayesian optimization approach to explore the parameter space efficiently and quantify the remaining energy using failure probability distributions. Bayesian optimization identifies future experiments that maximize information gain and minimize uncertainty.

The battery community uses the term "state of health" (SOH), which, to be informative, should enable some estimate of the remaining useful life. However, consider two batteries that have lost 10% of their original capacity. The first battery was cycled at a 10C charging rate for ten cycles and the second at 0.1C for 1,000 cycles. Although a commonly used surrogate for SOH, remaining capacity, is 90% for both, and their total charge throughputs are identical, their remaining useful lives are almost certainly different and will depend on how they are used. Therefore, this work will not attempt to define a state of health. Instead, we focus on predicting the remaining energy (and the associated failure distribution) rather than the remaining capacity because energy loss incorporates both capacity and voltage fade. Furthermore, energy, unlike capacity, also translates directly to useful properties such as a vehicle's range or a grid battery's duration.

At the end of this paper, we briefly address that published research pays scant attention to the complex, dynamic cycling conditions experienced in the real world. That is, laboratory testing is usually done at a single point in the operating parameter space, where the operating conditions — charging rate (C-rate), temperature, maximum state of charge (SOC), depth of discharge, etc — are held constant throughout a given cycling test. Although the experiments described here are also carried out at a single point in the operating parameter space, we use our results to propose a method that allows us to predict, under certain conditions, the remaining energy under any arbitrary pathway through our parameter space.

For our experimental dataset, we use commercial $LiFePO_4$ cells. To generate data for our model within the project time frame (weeks rather than years), we used intuition and preliminary tests to identify the limits of the operating parameter space where energy loss occurs rapidly. This is accomplished for LFP cells by inducing Li-plating under fast charging conditions. [41, 42].

In what follows, we discuss our framework, represented in Figure 1, which efficiently explores a four-dimensional parameter space spanning the operating conditions of energy storage systems to estimate their remaining energy. In Section 2, we introduce GP modeling, explain our modifications to the model, and offer the reader an introduction to Bayesian optimization. In Section 3, we discuss experimental data collection methodology and efficient parameter space exploration. In Section 4, we report and discuss our framework results. Finally, we conclude this work in Section 5.

2 COMPUTATIONAL METHOD

One way to efficiently explore high-dimensional parameter spaces and accurately quantify the failure distribution is to use machine learning, specifically the Gaussian
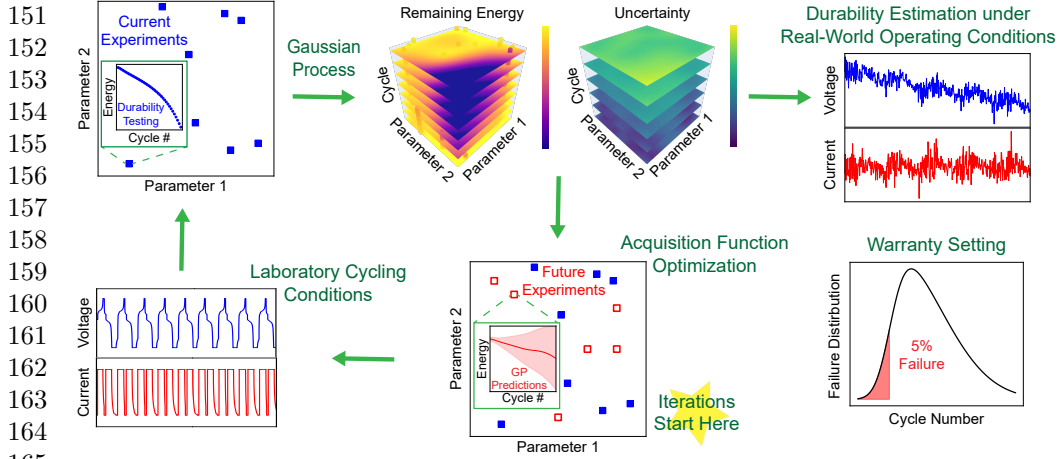
Fig. 1 Schematic representation of our framework that efficiently explores large parameter spaces and predicts remaining energy using failure probability distributions to set warranties. The framework starts by randomly choosing points in the parameter space and using laboratory cycling to extract cell performance for each point. Then, a domain-knowledge-informed Gaussian process is used to predict the remaining energy of cells and uncertainty at any point in the parameter space. Bayesian optimization is also implemented to identify future points to be tested that maximize information gain and minimize the overall uncertainty. Those experiments are then performed, and GP models are updated and cross-validated for prediction accuracy. The identification and testing of future experiments is repeated until the GP prediction accuracy becomes acceptable to the user. Then, the laboratory experimental results are proposed to be used to predict remaining energy under real-life cycling conditions.

process, enhanced with domain ("expert") knowledge [40, 43, 44, 45]. We choose a GP because it can stochastically approximate the (unknown) latent function relating the operating conditions to measurable properties, such as energy and capacity loss, using small datasets compared to the parameter space dimensionality. It can also quantify the uncertainty in its predictions [27, 28, 46, 2], allowing the estimation of failure distributions. Additionally, it differentiates between aleatoric uncertainty, which is linked to the inherent variability in the data, and epistemic uncertainty, which is linked to insufficient training data for accurate predictions. Moreover, a GP can capture complex, non-linear relationships without requiring explicit functional form specifications that may bias the predictions. Unlike other machine learning models, the parameters of a GP, the so-called hyperparameters, provide explicit insights into the characteristics of the estimated function, allowing for the interpretability of the GP. Another favorable characteristic of GP models is that the smoothness of the predicted function, among other properties, can be explicitly controlled using the kernel function. As we discuss below, we use the $\nu = 3/2$ Matèrn kernel since it allows for less smooth functions than the commonly used squared exponential kernel. The $\nu = 3/2$ Matèrn kernel provides flexibility in modeling smooth trends and local variations in the data, making it fit for our battery data. Overall, much work has been done to prove the approximation power of GP models of unknown functions [47, 48, 49]. A GP model can be entirely specified using three building blocks: the prior mean, noise, and kernel functions whose role we explore next.

The parameter space $\boldsymbol{X} \subset \mathbb{R}^4$ in this work is a bounded subdomain of the four-dimensional (4D) Euclidean space spanning the maximum SOC (x_1), C-rate (x_2), temperature (x_3), and cycle number (x_4) parameters. These parameters are used

4

following Konz et al. [23], who reported that the onset of lithium plating in their LFP batteries is a function of these operating parameters. After initial experiments are chosen by intuition, we use the proposed domain-knowledge-informed GP model to implement a Bayesian active learning approach that reliably identifies future experiments with the objective of efficiently exploring the parameter space.

Our GP framework aims to predict the remaining energy of a cell as a function of $\boldsymbol{x} = \{x_1, x_2, x_3, x_4\} \in \boldsymbol{X}$. The remaining energy is approximated by an unknown latent function $f(\boldsymbol{x})$ that can only be sampled via noisy measurements $y(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon(\boldsymbol{x})$ with $\epsilon(\boldsymbol{x})$ being the noise. Using $y(\boldsymbol{x})$, the GP provides a probabilistic representation of the latent function $f(\boldsymbol{x})$. Although we are interested in estimating the remaining energy using four parameters $\boldsymbol{x}$, our framework is not limited to this scenario. GP modeling can be easily modified to predict other quantities of interest using other input parameters.

GP models are defined by a prior multivariate normal probability distribution $\mathcal{N}(\boldsymbol{\mu}, \mathrm{Cov}[\boldsymbol{f}, \boldsymbol{f}])$ over the latent function values $\boldsymbol{f} = f(\boldsymbol{x}_i) \ \forall \ i \in \{1, 2, \ldots, n\}$ (remaining energy in this case), where n is the number of data points. Any finite subset of function values follows a Gaussian (normal) distribution. This multivariate distribution is characterized by a mean vector $\boldsymbol{\mu} = \mu(\boldsymbol{x}_i; \boldsymbol{\theta}) \ \forall \ i \in \{1, 2, \ldots, n\}$ and a covariance matrix $\mathrm{Cov}[\boldsymbol{f}, \boldsymbol{f}] = \boldsymbol{K} = k(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\theta}) \ \forall \ i, j \in \{1, 2, \ldots, n\}$, where k is the positive semi-definite kernel function acting on pairs of inputs $\boldsymbol{x}_i, \boldsymbol{x}_j$. A GP is a Bayesian method that requires a likelihood, in addition to the prior, to compute a posterior probability distribution. The likelihood is defined over noisy observations y as the normal distribution $\mathcal{N}(\boldsymbol{f}, \boldsymbol{V}(\boldsymbol{\theta}))$, where $\boldsymbol{V}_{ii} = \sigma_n^2(\boldsymbol{x}_i; \boldsymbol{\theta}) \ \forall \ i \in \{1, 2, \ldots, n\}$ defines the diagonal noise matrix. The hyperparameters $\boldsymbol{\theta}$ stemming from the prior mean, noise, and kernel must be learned. This is often done by maximizing the log marginal likelihood (maximum likelihood estimation or MLE) of the data. Marginalizing over the unknown $\boldsymbol{f}$ and conditioning on the data $\boldsymbol{y} = y(\boldsymbol{x}_i) \ \forall \ i \in \{1, 2, \ldots, n\}$ yields the posterior $p(\boldsymbol{f}^*|\boldsymbol{y}, \boldsymbol{\theta})$, where $\boldsymbol{f}^*$ are the predictions. Carefully choosing prior mean $\mu(\boldsymbol{x}; \boldsymbol{\theta})$, kernel $k(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$, and noise $\sigma_n^2(\boldsymbol{x}; \boldsymbol{\theta})$ functions improves the prediction capabilities of the GP model significantly, allowing the model to better extrapolate and quantify the uncertainty for the estimation of the failure distribution accurately [40].

To integrate domain knowledge, the framework builds on the standard GP model by customizing the prior mean, noise, and kernel functions. The standard GP model, which is usually used for predicting the remaining energy/capacity of batteries [46, 50, 51], involves a constant prior mean $\mu(\boldsymbol{x})$, a stationary kernel, here, the Matèrn kernel $k(\boldsymbol{x}, \boldsymbol{x}')_{\nu=3/2}$, and an independently identically distributed (i.i.d.) zero mean noise ϵ. This standard GP is characterized as

$$y(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon, \tag{1a}$$

$$f(\boldsymbol{x}) \sim \mathcal{GP}(\mu(\boldsymbol{x}) = c, k(\boldsymbol{x}, \boldsymbol{x}')), \tag{1b}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_s^2 \left(1 + \frac{\sqrt{3}d}{l}\right) \exp\left(-\frac{\sqrt{3}d}{l}\right), \tag{1c}$$

$$\epsilon \sim \mathcal{N}(0, \sigma_n^2), \tag{1d}$$

where d is the 4D-Euclidean-space distance between the data points. The hyperparameters $\boldsymbol{\theta} = \{c, \sigma^2, l, \sigma_n^2\}$ control the predictions of the GP. Within the kernel function, the signal variance (σ_s^2) controls the deviation of the predicted function from its mean,

5

and the length scale (l) controls the covariance decay rate based on the distance between the data points. The noise variance σ_n^2 quantifies the data variability. One of the main benefits of GP modeling is its adaptability to the application of interest by modifying the prior mean, noise, and kernel functions based on domain knowledge. For example, the literature shows that remaining energy of batteries degrades with cycling and that the degradation rate changes under different operating conditions [5, 6, 7]. We show that accounting for this decay using a proper prior mean function significantly enhances the extrapolation capabilities of the GP model. Moreover, the literature reports that cell-to-cell variability often increases with cycling [52, 53], and accounting for it enhances the aleatoric uncertainty quantification, which is linked to the inherent variability in the data, improving the failure distribution estimation. To improve the epistemic uncertainty quantification, we use a non-stationary kernel function. In what follows, we dive deeper into the three building blocks of GP modeling, demonstrating our integration of domain knowledge into GPs using different combinations of the mean, noise and kernel and quantifying their performance.

2.1 The Prior Mean Function

The prior mean function $\mu : \boldsymbol{X} \to \mathbb{R}$ encodes the expected general trend of the latent function and can be tailored to the domain knowledge about energy degradation. The prior mean function is usually set to a constant, resulting in poor extrapolation capabilities [40]. Domain knowledge indicates that the remaining energy generally degrades during cycling, depending on temperature, charging rate, and maximum SOC. Integrating this knowledge into the GP without biasing the results requires carefully choosing a flexible prior mean function that models this degradation. This flexibility allows the model to adjust assumptions when they are inaccurate, including switching to an uninformed prior if supported by the data.

A power law function in cycle number (x_4), typically with a negative slope, is one candidate for modeling energy degradation. However, any change in the degradation rate depends, in our experiments, on maximum SOC (x_1), charging rate (x_2), and temperature (x_3). Therefore, the power law parameters are considered functions of these operating condition parameters. For example,

$$
\begin{aligned}
\mu(\boldsymbol{x}) &= ax_4^p + b, \\
a(\tilde{\boldsymbol{x}}) &= c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4, \\
p(\tilde{\boldsymbol{x}}) &= c_5 x_1 + c_6 x_2 + c_7 x_3 + c_8, \\
b(\tilde{\boldsymbol{x}}) &= c_9 x_1 + c_{10} x_2 + c_{11} x_3 + c_{12}.
\end{aligned}
\tag{2}
$$

This prior mean represents a flexible power law model where the degradation rate (quantified by a), the non-linearity of the function (quantified by the power p), and the initial battery energy (quantified by the intercept b) all vary based on $\tilde{\boldsymbol{x}} = \{x_1, x_2, x_3\}$. Equation 2 depends on the hyperparameters $c_i \ \forall \ i \in \{1, 2, ..., 12\}$ estimated by MLE.

An alternative model composed of the sum of basis functions can also flexibly model degradation with varying rates. This model depends on identifying centers within the 4D parameter space where changes in degradation trends occur. This model may also be used to identify the location of the "knees" where new failure mechanisms, such as loss of active material and mechanical deformation [54, 55, 56], could have occurred while cycling. Here, we implement the sum of three linear basis functions (B_i) model

for each parameter in the 4D parameter space with $N = 3^4$ centers ($\boldsymbol{x}_0 \in \mathbb{R}^4$) and weights ($c \in \mathbb{R}$). The general equation of this model is as follows

$$\mu(\boldsymbol{x}) = \sum_{i=1}^{N} c_i B_i(\boldsymbol{x}, \boldsymbol{x}_{0i}). \tag{3}$$

The first and last centers in each dimension are chosen as the minimum and maximum of the corresponding dimension, whereas the ones in between are identified using MLE. This results in 85 prior mean hyperparameters optimized by MLE — 81 weights and 4 centers. This formulation can be easily extended to include more basis functions that allow for increased flexibility in the modeling of degradation rate. However, more hyperparameters would be required.

2.2 The Noise Model

The noise function $\sigma_n^2 : \boldsymbol{X} \to \mathbb{R}$ quantifies the aleatoric uncertainty in the GP predictions resulting primarily from the naturally occurring cell-to-cell variability, such as the onset of additional failure mechanisms during cycling [29]. This variability depends on the location of the data point in the parameter space (heteroscedastic). For example, studies have often shown that variability increases with cycling (x_4) [57, 3]. Although variability may also be affected by $\{x_1, x_2, x_3\}$, we ignore that to avoid overfitting. Since quantifying the aleatoric uncertainty should be performed with the least data possible, we integrate domain knowledge to reduce the required number of experiments. We showed in our previous work that using a noise model that models the increase in variability improves the quantification of failure distributions [40]. In addition to estimating the aleatoric uncertainty, the noise model is crucial for MLE and Bayesian inference, allowing the GP to better predict the function values at unobserved points in the parameter space.

A general, flexible function that models the increase in variability is the power law,

$$\sigma_n^2(\boldsymbol{x}) = m x_4^p + n. \tag{4}$$

The hyperparameters $\{m, n, p\}$ control the aleatoric uncertainty estimation as a function of x_4. Specifically, nonlinearity is controlled by p. In the simplest case, the rate can be assumed to be constant, setting $p = 1$. However, variability increases with cycling at an increasing rate [3] if the knee appears for different batteries at different locations. The variability would become more pronounced after some batteries suffered an additional failure mechanism and some did not. One drawback of this function is that it assumes that the variability continues increasing at an increasing rate with cycling. For cases where the variability increase slows down [57], we could use the sigmoid function,

$$\sigma_n^2(\boldsymbol{x}) = \frac{m}{1 + e^{-n(x_4 - x^*)}}. \tag{5}$$

In this case, m represents the upper asymptote of the variability, n quantifies the curve's steepness, and x^* denotes the inflection point signaling the transition of the variability's rate of increase from ascent to descent. A sum of linear basis functions model similar to Equation 3 could also allow for this pattern.

2.3 The Kernel Function

The kernel function $k : \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R}$ is the primary building block of a GP as it dictates the properties of the predictions and estimates the covariance among and between observed and unobserved function values. Within the context of efficient parameter space exploration, the kernel function identifies where data is needed for accurate predictions to help guide future experiments.

The kernel function serves as the covariance operator and estimates epistemic uncertainty. It uses the values of two input points, $\boldsymbol{x}$ and $\boldsymbol{x}'$, to estimate the covariance between the corresponding function values $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$, even if they were not observed yet. Additionally, the kernel function quantifies the epistemic uncertainty — corresponding to the lack of data — to identify where future experiments are needed for accurate predictions. Such an approach allows for efficient parameter space exploration, where only experiments that maximize information gain are chosen.

Most GP studies employ stationary kernels that compute the covariance based on the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ [58]. This dictates the epistemic uncertainty, which grows as the prediction points become more distant from the measured points. Although this is logical when quantifying uncertainty, distance is not the only aspect of data that controls it. One example of stationary kernel functions is the Matèrn kernel with $\nu = 3/2$, shown in Equation 1c. This isotropic kernel — all parameters in the 4D parameter space have an equal length scale — assumes identical decay in covariance in each direction as long as the distance is the same. This assumption is invalid in our work since the parameters have different ranges and, most likely, different covariance decay rates. To mitigate this, we make the kernel anisotropic with respect to the norm $||\cdot||^2$ by having an independent length scale for each dimension and determined using automatic relevance determination (ARD), resulting in the covariance decay rate varying from one dimension to another. We introduce the anisotropy by transforming the Euclidean metric using Equation 6(b-c), where $\{l_1, l_2, l_3, l_4\}$ are the length scales corresponding to each parameter and setting the length scale l in Equation 1c to 1.

Unlike stationary kernels that may result in poor predictions and inaccurate quantification of uncertainties, non-stationary kernels quantify the covariance based on the data points' explicit locations [59, 60]. These kernels have greater expressiveness and flexibility in their calculations. To give them their flexibility, non-stationary kernels require more hyperparameters. This increases their computational requirements but in a predictable manner. Additionally, for non-stationary kernels, it is harder to prove their positive semi-definite property, but past work has been done to develop them [61, 62, 63]. For instance, deep kernels use deep neural networks to transform the input parameter space to a latent space, possibly with a different number of dimensions, and use it to calculate the covariance using a given stationary kernel. Deep GPs are another example of how non-stationarity can be achieved; multiple GPs with stationary kernels are stacked, so the output of one serves as the input of the next to achieve non-stationarity [64]. Here, we implement a parametric non-stationary kernel where parametric equations replace the constant hyperparameters in stationary kernels [65, 66]. More specifically, we consider that the signal variance changes with maximum SOC, C-rate, and temperature but not with cycling since our previous work showed no significant non-stationarity in the cycling dimension [40]. Modifying the isotropic stationary kernel in Equation 1c with these modifications results in the

8

following anisotropic non-stationary kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = g(\tilde{\boldsymbol{x}})g(\tilde{\boldsymbol{x}}')\left(1 + \sqrt{3}d\right)\exp\left(-\sqrt{3}d\right), \tag{6a}$$

$$d = \sqrt{(\boldsymbol{x} - \boldsymbol{x}')^T M(\boldsymbol{x} - \boldsymbol{x}')}, \tag{6b}$$

$$M = \begin{bmatrix} 1/l_1^2 & 0 & 0 & 0 \\ 0 & 1/l_2^2 & 0 & 0 \\ 0 & 0 & 1/l_3^2 & 0 \\ 0 & 0 & 0 & 1/l_4^2 \end{bmatrix}, \tag{6c}$$

$$g(\tilde{\boldsymbol{x}}) = \sum_{i=1}^{27} w_i \exp\left(-\frac{\tilde{d}^2}{l_g}\right). \tag{6d}$$

The sum of exponential basis functions $g(\tilde{\boldsymbol{x}})$ captures the non-stationarity in the maximum SOC, C-rate, and temperature parameter space centered at each parameter's initial, middle, and final point, where $\tilde{\boldsymbol{x}} = \{x_1, x_2, x_3\}$ is the input data, and $\tilde{d}$ is the Euclidean distance in this 3D parameter space. This kernel function has 32 hyperparameters — four for the length scales in the distance (d), 27 for the weights (w_i), and one for the width (l_g).

2.4 Overview of Bayesian Optimization

Bayesian optimization (BO) is used to efficiently explore and predict the remaining energy at any point in parameter space $\boldsymbol{X} \subset \mathbb{R}^4$. BO has found its place in the energy storage field as an efficient approach to efficiently optimize black-box functions relating process parameters of energy storage systems to their key performance measures such as remaining energy [67, 68, 69, 70]. Like other active learning approaches, BO builds on the current experimental results to identify future points in $\boldsymbol{X}$ to be tested. Several studies have shown the superiority of BO over other approaches like random sampling, grid sampling, Genetic Algorithms, and Lagrange-Relaxation optimization by requiring fewer data points and providing faster convergence rates to the desired objective of the optimization [71, 72, 73, 74]. BO consists of two main components: a stochastic predictor model of the black-box function and an acquisition function. Generally, a GP model is used as the probabilistic predictor of the black-box function f as it provides the BO with an accurate prediction of the mean prediction and uncertainty using a small number of experiments. A typical BO approach starts with performing experiments at a few initial points in $\boldsymbol{X}$. The experimental results are then used to train a GP model. The BO builds on this GP to identify new points in $\boldsymbol{X}$ predicted to optimize the acquisition function. Experiments are then performed at the newly identified points, and their results are used to improve the black-box function estimation — by training a new GP model with all the available data — to identify the next new points. Fitting a GP model, identifying new points, and performing measurements at their locations constitute an optimization iteration.

Multiple acquisition functions have been developed throughout the years based on the optimization objective. One widely used acquisition function is the probability of improvement (PI), where the BO identifies the point(s) in parameter space with the maximum probability of being more optimal than the current optimal point. PI exploits the areas in the parameter space where it is likely to improve the objective

9

and may get stuck in local optima [75]. Another acquisition function is the expected improvement (EI) acquisition function, which balances the trade-off between exploitation and exploration of uncertain regions in the parameter space [69]. PI and EI are two examples of functions used to maximize an objective function. Here, we are interested in efficiently exploring the parameter space without exploiting a specific region. One example of such an acquisition function is the total correlation, which we use to identify multiple new points to test. Total correlation quantifies the expected information gained from future experiments. More specifically, it identifies new points in the parameter space whose function values are least correlated with each other — making the new points $\boldsymbol{x}^*$ mutually aware of each other — and with the function values at the previously tested points $\boldsymbol{x}$. Since the function values at the new points are least correlated, performing experiments at those points will lead to maximum information gain and uncertainty minimization throughout the parameter space. Mathematically, total correlation can be expressed as

$$\text{KL}\Big(p\big(\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x}^*)\big) \,||\, p\big(\boldsymbol{f}(\boldsymbol{x})\big).p\big(\boldsymbol{f}(\boldsymbol{x}^*)\big)\Big) = \text{KL}(A \,||\, B), \tag{7a}$$

$$A \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\boldsymbol{x}) \\ \boldsymbol{\mu}(\boldsymbol{x}^*) \end{bmatrix}, \begin{bmatrix} \boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}) & \boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}^*) \\ \boldsymbol{k}(\boldsymbol{x}^*, \boldsymbol{x}) & \boldsymbol{k}(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right), \tag{7b}$$

$$B \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\boldsymbol{x}) \\ \boldsymbol{\mu}(\boldsymbol{x}^*) \end{bmatrix}, \begin{bmatrix} \boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{k}(\boldsymbol{x}^*, \boldsymbol{x}^*)\mathbf{I} \end{bmatrix}\right), \tag{7c}$$

and is defined as the Kullback-Leibler (KL) divergence [76] between the joint prior distribution of the function values where tested points $\boldsymbol{x}$ are correlated with the untested points $\boldsymbol{x}^*$ and the joint prior where the tested points are uncorrelated with the untested points as shown in Eq. 7, where n is the number of points in $\boldsymbol{x}$ and $\boldsymbol{x}^*$. In this regard, the aim is to identify $\boldsymbol{x}^*$ that minimize the total correlation, and therefore the KL divergence between the joint distributions when $\boldsymbol{x}$ and $\boldsymbol{x}^*$ are correlated and not correlated. When the KL divergence is minimized, it shows that the new points are truly uncorrelated.

3 EXPERIMENTAL DATA COLLECTION

3.1 Data Collection

The bounds of the parameter space are identified by studying plating as a function of temperature and maximum SOC at a C-rate of 6C. The onset of lithium plating is detected by a drop in the coulombic efficiency of the cell after a fast charge cycle followed by a normal discharge cycle. As shown in Figure S1 in Supplemental Material, there does not appear to be an onset of plating at any maximum SOC charge at 303 K compared to 288 K and 293 K. For this reason, we consider the temperature range between 285 K and 291 K. Additionally, there did not appear to be sufficient plating at a maximum SOC of 10%, so we consider the maximum SOC to range from 20% to 80%. As plating is expected to grow rapidly with maximum SOC [23], we limited it to 80% to avoid excessive plating on any given cycle. Finally, since plating usually occurs at 6C in these temperature and maximum SOC ranges, we set our C-rate range to be 2C to 8C.

Table 1 The limits of the parameter space

Limits	Maximum SOC [%]	C-rate [C]	Temperature [K]	Loop Number
Minimum	20	2	285	0
Maximum	80	8	291	40

With the bounds of our parameter space identified, we designed our standard cycling protocol to monitor remaining energy loss with cycling. The protocol consists of repeated loops comprising fast and slow cycles to estimate the remaining energy while cycling accurately. The initial loop consists of five slow-charging cycles at 1C to assess the deliverable remaining energy from the cell. The remaining loops consist of two fast-charging cycles depending on the chosen C-rate and maximum SOC for the cycled cell and five slow-charging cycles at 1C. The remaining energy at the fifth slow cycle in each loop is the experiment's outcome to train the GP model at the corresponding input point in parameter space — the cell's maximum SOC, C-rate, temperature, and loop number. The damage caused by each fast cycle is reflected in the subsequent slow cycles as a loss in the remaining energy. Figure S2 in Supplemental Material shows an example dataset of the remaining energy as a function of cycle number, where the drops in remaining energy after each of the fast charging steps are apparent. Figure S2 also shows the relationship between loops and cycles. From hereafter, and due to this accelerated testing protocol, we consider loop number as our fourth parameter instead of cycle number. The full details of our experimental setup and cycling parameters can be found in the Experimental Procedure section. In our setup, some cells may not experience degradation in our testing time frame, especially those cycled at low SOC and C-rate. To limit our testing in the loop number dimension, all cells are stopped at loop number 40 if they did not fail earlier. Table 1 summarizes the bounds of the parameter space used in this study.

3.2 Parameter Space Exploration

The parameter space is explored using a BO approach over multiple iterations. First, two sets of $p = 16$ coordinates $\{x_1, x_2\}$ (maximum SOC, charging rate) are randomly selected to test $2p$ cells over two iterations, each iteration at a randomly chosen x_3 (temperature). In this regard, we define an iteration as the fitting of a GP model using the previous experiments (if any) and identification and simultaneous testing of p cells. A GP model is fitted using all collected data. Then, a third iteration is initiated where we extract a new set of p experiments $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_p^*\}$, with $\boldsymbol{x}_i^* = \{x_1, x_2, x_3, x_4\}$, that are expected to maximize the information gain and minimize the overall uncertainty. Although x_4 is inherently identified for all p new experiments, we do not consider it in our framework. Future cells are tested either until failure or until they reach the maximum identified loop number. All collected data are then added to the training dataset. Since the parameter space includes the loop number dimension, x_4 is implicitly determined by the BO. The loop number dimension cannot be excluded from the BO because it leverages data from different loop numbers to guide the selection of future experiments. The acquisition function identifying the future experiments is the total correlation. It quantifies the expected information gain from the future p experiments. Given our limitation of having only one temperature chamber, future experiments must have the same x_3. We employ a nested optimization approach, displayed in Algorithm 1, to achieve this: the upper level optimizes x_3 using a non-linear direct

11

search optimization algorithm, the lower level optimizes x_1, x_2, and x_4 (loop number) while x_3 held constant using a differential evolution global optimization algorithm. When performing the new experiments, the cells are tested from loop 0 to either failure or the upper limit of x_4, whatever occurs first, and the results of all loops are considered in the GP. The new data are added to the previous results to fit a new GP model. We conclude our exploration effort after these three iterations. A new iteration would have been initiated by asking for a new set of p experiments using the latest GP model. However, as discussed in Section 4.1, our results show good prediction performance; therefore, no additional iterations are considered.

Algorithm 1 Nested Optimization for $h(\boldsymbol{x}^*) = \mathrm{KL}(A \parallel B)$

1: **function** NESTEDOPTIMIZATION(h, $[x_{1\_init}, x_{2\_init}, x_{3\_init}, x_{4\_init}]$)
2: $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Upper level optm. (for x_3)
3: $\quad$ **function** UPPERLEVELOBJECTIVE(x_3)
4: $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Lower level optm. (for x_1, x_2, x_4)
5: $\qquad$ **function** LOWERLEVELOBJECTIVE(x_1, x_2, x_4)
6: $\qquad\quad$ **return** $h(x_1, x_2, x_3, x_4)$
7: $\qquad$ **end function**
8: $\qquad (x_{1\_opt}, x_{2\_opt}, x_{4\_opt}) \leftarrow \mathrm{Minimize}(\mathrm{LowerLevelObjective}, [x_{1\_init}, x_{2\_init}, x_{4\_init}])$
9: $\qquad$ **return** $h(x_{1\_opt}, x_{2\_opt}, x_3, x_{4\_opt})$
10: $\quad$ **end function**
11: $\quad x_{3\_opt} \leftarrow \mathrm{Minimize}(\mathrm{UpperLevelObjective}, x_{3\_init})$
12: $\quad$ **return** $(x_{1\_opt}, x_{2\_opt}, x_{3\_opt}, x_{4\_opt})$
13: **end function**

The GP model consists of power law mean and noise functions and the $\nu = 3/2$ Matèrn kernel. We used a power law function as the prior mean to account for that batteries may degrade with cycling at an increasing rate. To account for differences in the degradation rate at different operating conditions, we allowed the parameters of the power law function to vary with the operating conditions as discussed in Equation 2. We also used a power law as the noise function to account for the fact that variability may increase with cycling. We considered the noise function's parameters constant, assuming that the operating conditions do not affect the variability. Since our previous work [40] showed insignificant non-stationarity in the battery data in the loop dimension, we used the $\nu = 3/2$ Matèrn stationary kernel with ARD of the length scales of different dimensions in the parameter space.

4 RESULTS AND DISCUSSION

4.1 Efficient Exploration of the Parameter Space

The proposed framework efficiently and accurately explores the 4D parameter space. Here we show the prediction results of the GP model discussed in Section 3.2 when all the experimental data from the BO iterations are used. We also show how our GP model can quantify variability by repeating the testing of a randomly chosen cell with multiple other fresh cells. To showcase the efficient exploration and prediction accuracy of our framework, we also demonstrate the rapid decrease in prediction error with new data compared to the random sampling approach.
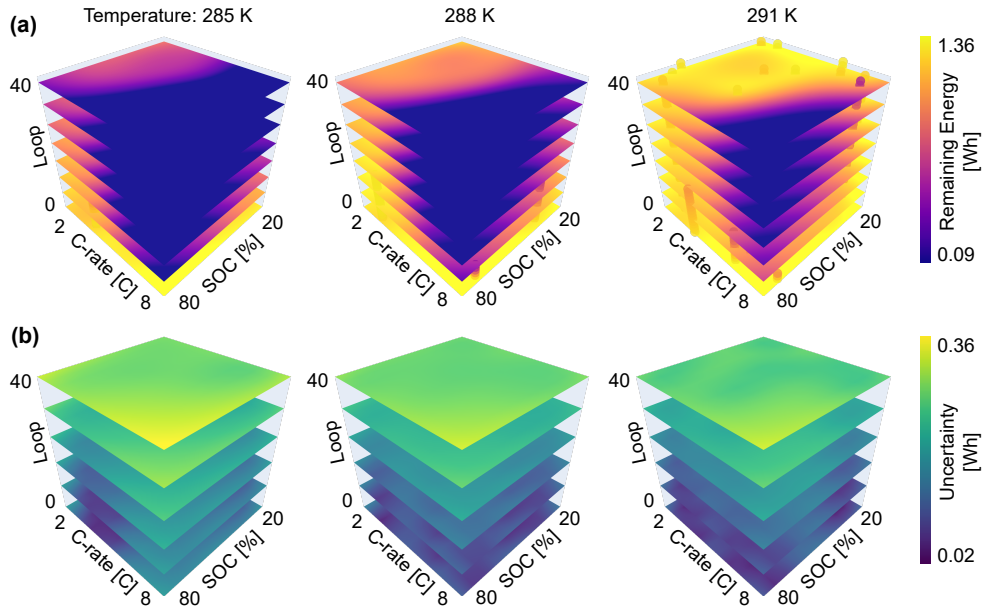
12

Fig. 2 GP prediction of the (a) posterior mean and (b)uncertainty of the remaining energy as a function of maximum SOC, C-rate, temperature, and loop number. Each cube represents the predictions at a given temperature, where the color of the slices within the represents the value of the posterior mean and uncertainty, respectively. These plots allow for the estimation of the failure distribution as a function of loop number at any point in the SOC, C-rate and temperature space using the method outlined in [40].

Figure 2 presents the GP posterior mean and uncertainty in the 4D parameter space using all the experimental data, where color represents the predictions of remaining energy (a) and uncertainty displayed by one standard deviation (b). Following this representation, the results of each tested cell are represented as vertical pillars within the cubes. The 291 K cube shows multiple pillars that stretch from the bottom to some loop number based on the extent to which the cells were tested. Some cells were tested until loop 40 and the pillar reaches that upper plane; the rest failed earlier. The 32 cells at 285 K and 288 K failed before loop 20. For a clearer display of experiments performed at each temperature, the tested points are displayed in Tables 1, 2, and 3 and as scatter plots in Figures S3, S4, and S5 in the Supplemental Material. Figure 2(a) clearly shows the effect of the process parameters that follow the expected decay in the remaining energy [5, 6, 7]. For instance, decreasing the temperature while keeping everything else fixed increases the decay rate (from plating). Similarly, increasing the maximum SOC, C-rate, or loop number individually while keeping everything else constant increases the degradation rate. Thus, our modeling approach of the prior mean, which allowed a change in the slope, power coefficient, and intercept of the prior mean based on maximum SOC, C-rate, and temperature, has resulted in accurate predictions.

Figure 2(b) shows the prediction uncertainty in the parameter space. The uncertainty starts at a minimum for all temperatures and increases with cycling due to the natural cell-to-cell variation (aleatoric uncertainty). To quantify this variability, we arbitrarily chose a cell ($x_1 = 30\%$, $x_2 = 7.46C$, and $x_3 = 291$ K) to replicate its testing with ten
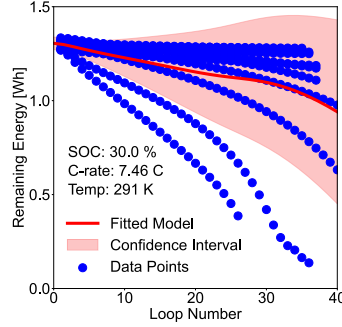
Fig. 3 Eleven identically cycled cells quantifying the cell-to-cell variability. The blue points represent the data, the red line represents the posterior mean, and the shaded region shows the 95% confidence interval. This plot displays the increase of uncertainty as a function of loop number.

other fresh cells. The results of the eleven cells, along with the GP predictions and 95% confidence interval of $p(y)$, are shown in Figure 3. Our results follow those in the literature that variability increases with cycling [57, 3]. Figure 3 shows two main clusters of cells, data points of nine cells within the confidence bounds and two outside. These bounds can be extended based on the user's confidence level preferences. They are set to cover 2 standard deviations around the mean to get the 95% confidence. Being outside the bounds suggests that the GP interprets the points as outliers at the corresponding loop number. The percentage of points included within the bounds changes with the loop number. All the data are within the bounds until loop number 5. For the loop numbers between 5 and 27, only 81% of data points are within the bounds. Testing the lowest cell stops at loop number 27, where afterward, 90% of the data becomes within the 95% confidence interval. However, these results may change upon the addition of more cells. GP uncertainty increases with variability because the GP becomes uncertain about the unseen true results. Additionally, Figure 2(b) shows that the uncertainty also seems to increase when the temperature decreases. This stems from the lack of data (epistemic uncertainty) at points beyond the 20 loops at temperatures 285 K and 288 K. This increase in epistemic uncertainty is represented by the lighter colors going from the right to the left cube while keeping everything else constant.

Cross-validation and root mean square error (RMSE) are used to test the GP prediction performance for unseen cells. Cross-validation is performed using 20 randomly chosen cells, where the results of a given cell are removed from the GP training dataset and used as the testing dataset for calculating the RMSE. Figure 4 shows three sets of representative cross-validation results. After training the GP model with the training dataset, the GP posterior mean (red line) and the posterior uncertainty (shaded region) are plotted on top of the testing data of the removed cell. Figure 4 shows accurate GP predictions for the three testing cells. The results of the remaining testing cells are shown in Figure S6 in the Supplemental Material. According to Figure S6, not all predictions are perfect. Some cells behave differently than what the GP predicts. However, these are still acceptable prediction errors since the testing data is within the uncertainty bounds. For instance, consider the cells in S6 (g), (n), and (p), which are roughly in the middle of the parameter space. According to Figure 2(a), for each temperature, the remaining energy degradation increases with increasing the C-rate and SOC for all loops. The GP model cannot accurately predict their degradation pathway for these cases because the cells lie between no degradation (low maximum
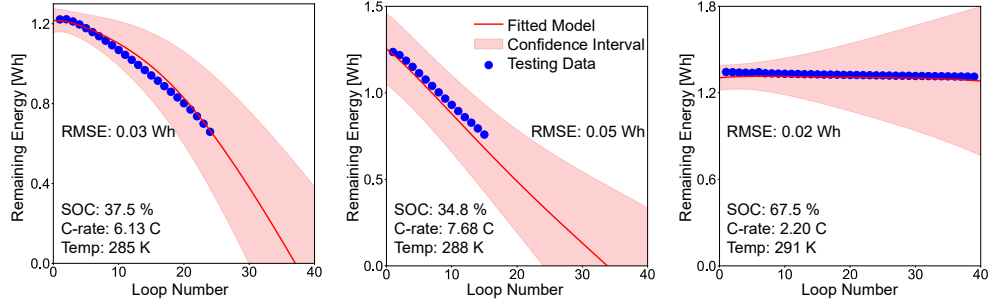
14

Fig. 4 Three representative cross-validation cases, where for each case, the data of a specific cell, represented by the blue points in each panel, are removed from the training dataset and used for testing the performance of the posterior predictions. The domain-knowledge-informed GP accurately predicts unseen testing experiments.
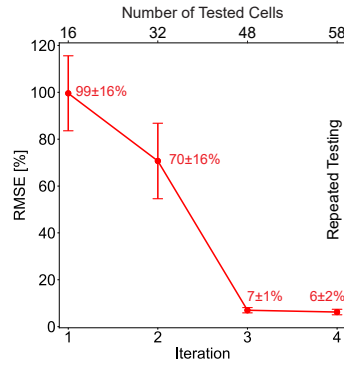

Fig. 5 Evolution of the GP prediction error as new experimental data are added. The markers represent the mean RMSE from the 20 cross-validation cells, and the error bars represent the standard error on the mean. Iteration 4 represents the 10 repeated cells. RMSE significantly decreases with more data points resulting from the efficient exploration of parameter space and the excellent prediction performance of the domain-knowledge-informed GP.

SOC and C-rate) and extreme degradation (high maximum SOC and C-rate) regions. Therefore, the GP followed its prior mean and predicted a decreasing trend in power law. However, as expected, the GP compensated for its prediction errors with increased uncertainty, showing the most likely trend based on domain knowledge and the other experiments. Even with these prediction errors, results show that the average RMSE for the 20 cells is 0.085 Wh, corresponding to around 6% error when normalized by the nominal cell energy (1.36 Wh). Figure 5 displays the decay in mean RMSE and standard error on the mean using the 20 cross-validation cells as a function of iteration number. Figure 5 clearly shows the significant improvement in the prediction performance with additional data, resulting from the efficient exploration capability of the parameter space and the use of domain-knowledge-informed GP. Therefore, these results show we acquired an accurate GP model with only $3p = 48$ cells and three iterations. More iterations would have been required if 6% were not an acceptable error to the user.

The proposed BO framework explores the parameter space more efficiently and accurately than the random sampling approach. The comparison starts by randomly identifying five points in the maximum SOC, C-rate, and temperature space and
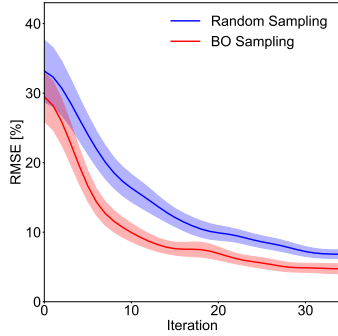
15

Fig. 6 Change in RMSE for the proposed BO framework and the random sampling approach with iterations, where the lines represent the mean of ten different exploration trials, and the shaded region represents the standard error on the mean. BO explores the parameter space more efficiently than random sampling.

then collecting data for cells cycled at these points. To ensure adequate comparison between the BO and random approaches, the fit in Figure 2 is considered a ground truth function to collect data instead of going through the physical cycling experiments. A separate GP model for each approach (BO and random) is fitted based on the collected data for the five points in the space. Using the two new GP models, cross-validation between these models and the ground truth is done using a grid of 4 points in each dimension in the parameter space (total of $4^4 = 256$ cross-validation points). Then, a new cell is identified using BO and random sampling, respective remaining energy data are collected, the GP models for the corresponding approach are updated, and cross-validation is performed. Identifying a new cell, collecting data, and updating the GP model constitute an interaction. Multiple iterations are performed, where Figure 6 shows the progression in RMSE as a function of iteration number. RMSE is displayed as a percentage of the ground truth maximum of remaining energy. Using a ground truth allows us to run many trials for each approach, where the lines in Figure 6 represent the mean of ten different exploration trials, and the shaded region displays the standard error on the mean. Figure 6 clearly shows the superiority of the BO approach where the RMSE drops relatively faster than the random sampling approach and achieves around 10% RMSE with less than ten iterations. These results suggest that the proposed BO approach more efficiently explores the parameter space than random sampling.

4.2 Varying GP Model Development

The prediction performance of the GP model depends on the choice of the prior mean, noise, and kernel functions, and here, we aim to identify the best combination of these functions based on the collected data. Multiple functions are tested for the prior mean, noise and kernel functions. We judge the performance based on maximum log marginal likelihood, lowest RMSE, and lowest continuous ranked probability score (CRPS) of cross-validation cases. RMSE and CRPS both measure the accuracy of predictions relative to the test data. However, CRPS also accounts for uncertainty in the predictions. This means that even if a model's predictions are inaccurate, a higher level of uncertainty can result in a better CRPS score [77].

16

The prior mean, noise, and kernel functions are varied to identify the combination with the best performance. Each function has multiple options, and to limit the number of tested combinations, we first identify the best prior mean function, then the best noise function, and finally, the best kernel function from the considered options. We start with the standard GP model with a constant prior mean, constant noise, and stationary $\nu = 3/2$ Matèrn kernel with ARD. Then, we fit GP models with different prior mean functions while keeping the noise and kernel functions fixed. Similarly, by fixing the other two functions, we repeat that for the noise and kernel functions, one at a time. Figure 7 shows the results of the different combinations based on the three prediction performance measures: (a) log marginal likelihood, (b) RMSE, and (c) CRPS. The bar graphs represent the mean of 20 cross-validation cases, and the error bars represent the standard error on the mean. Similar to Section 4.1, the cross-validation, which is repeated 20 times, is performed by removing the data of a random cell and using all the remaining to train a GP model. The predictions from this GP model are cross-validated with the removed cell. The results are divided within each panel based on the GP function being tested, and the best-performing combinations for the prior mean, noise, and kernel functions are highlighted. The power law prior mean function is superior to other alternatives as it has the maximum log marginal likelihood and the least RMSE and CRPS. This shows that the degradation rate increases with cycling. The power law performs better than the constant function since, in the latter case, the GP posterior mean approaches a constant value when extrapolating, and this results in increased prediction errors, especially for high loop numbers where the cells would have degraded. The power law captures the decreasing trend when extrapolating, keeping the prediction error small. The sum of linear basis functions (SLBF) provides a flexible general model that quantifies complex trends in the parameter space, especially when the location of the centers of the basis functions is optimized. Cross-validation results show that the log marginal likelihood measure is comparable to the power law, but the RMSE and CRPS are worse for SLBF. We believe that this poor performance is due to using only three centers for each parameter, causing sudden changes in the prediction trends of the remaining energy and resulting in higher RMSE and CRPS. The sum of cubic basis functions would be expected to show better results as it models a gradual change in the predictions, similar to a power law. Still, it would have required significantly more hyperparameters to optimize, increasing the computational requirements. The power law model has few hyperparameters and still results in favorable performance. The power law noise function also works best compared to the other alternatives. While the sigmoid noise function performs slightly better than the power law, it has higher RMSE and CRPS. These results suggest that the variability indeed increases non-linearly and then continues constant. Still, the constant variability is towards the upper edge of the loop dimension, and therefore, the power law model performed comparably well. As for the kernel function, using the non-stationary kernel did not improve the performance. This may indicate no non-stationarity in the maximum SOC, C-rate, and temperature operating parameter space or the number of data points ($3p = 48$) in that 3D parameter space is insufficient to capture the non-stationarity.

4.3 An Outlook Towards Real-World Cycling

Although we have shown an ability to predict the remaining energy in this 4D space, the experiments themselves, like almost all experiments in the battery literature, are
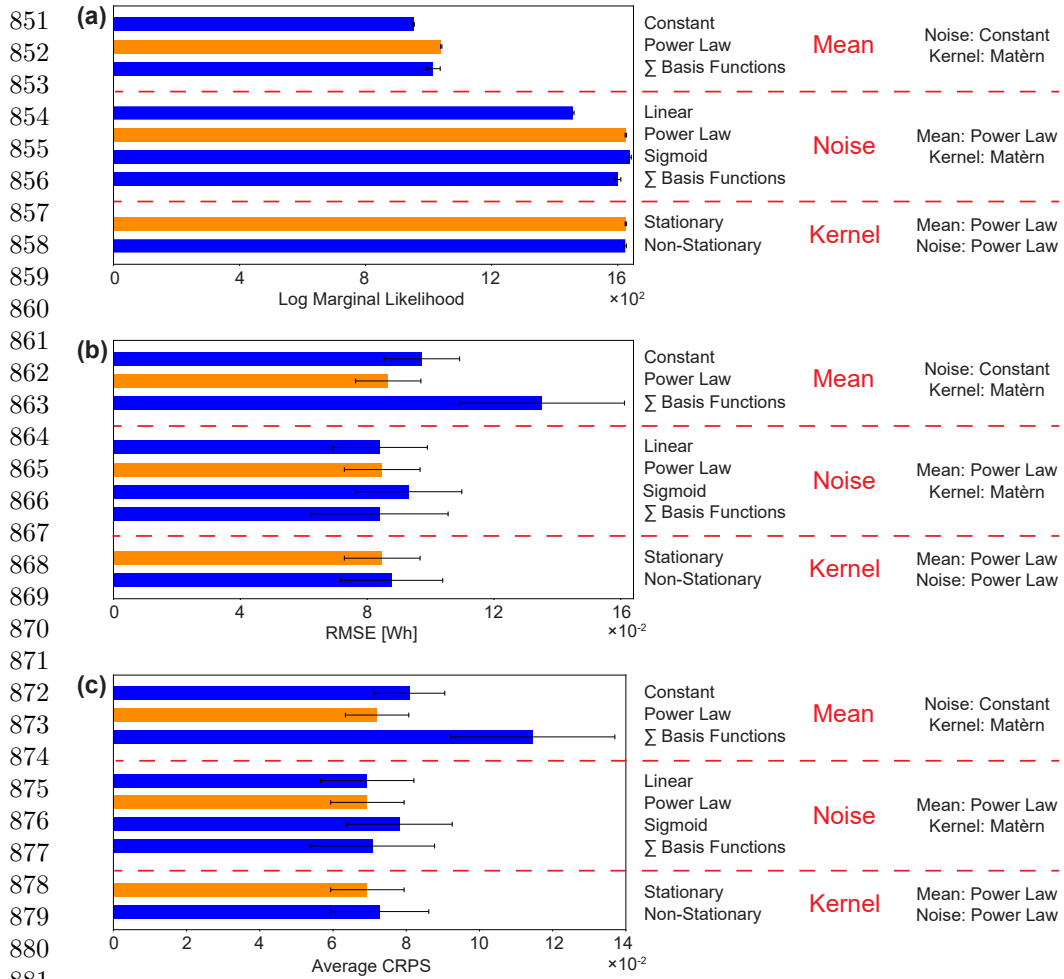
Fig. 7 Mean (bars) and standard error on the mean(error bar) of 20 cross-validation results for the (a) log marginal likelihood, (b) RMSE, and (c) CRPS. In each case, a randomly chosen cell is removed from the training dataset to be used as testing data. The orange bars represent the best-performing combinations when changing the prior mean, noise, and kernel functions. Among the compared combinations, a power law prior mean, a power law noise, and a stationary Matèrn kernel ($\nu = 3/2$) combination seems to be performing the best.

highly artificial because every cycle uses identical cycling parameters. In terms of our experiments, each cell retains the same maximum SOC, C-rate, and temperature as it increases the number of loops so that pathways through parameter space for a given cell show up as pillars in Figure 2. In real-world use, each cycle is generally different in duration, depth of discharge, C-rate, and nearly every other operating parameter [78]. This means that each cell may follow a complex pathway through parameter space. It is our goal to create a framework allowing the prediction of the remaining energy decay for any arbitrary pathway. The data in Figure 2 allow us to take the first step in this direction by differentiating each point with respect to loop number (holding other parameters constant) to produce Figure 8, where each point shows the energy loss for an incremental cycle there. The total degradation on any arbitrary pathway through parameter space is a path integral along that path, essentially adding up the

18

901
902
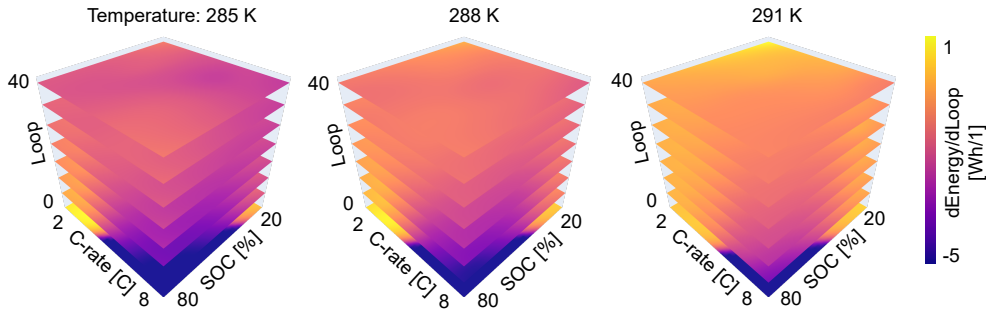903
904
905
906
907
908
909
910
911

Fig. 8 Energy loss at each point in the 4D parameter space, allowing for the estimation of the total degradation of cell operated under any real-world cycling history.

losses at each point visited. On the other hand, such a strategy does not account for previous history. The energy loss per cycle at a point (x,y,z,w) may depend on how the battery got to that point. These issues will be explored in future work.

5 CONCLUSION AND FUTURE WORK

Predicting the energy storage degradation rate under real-world cycling conditions requires efficiently exploring the parameter space. Results show that we can accurately predict the remaining energy using only 48 different points in the 3D parameter space (maximum SOC, C-rate, and temperature). This was achieved using the advanced GP model, which allows for accurate interpolating and extrapolating previously untested points in the space. This was also achieved using an active learning approach that identified future experiments expected to maximize the information gained and minimize the overall uncertainty in the predictions. We then quantified the performance of the GP model when considering different prior mean, noise, and kernel function combinations. We showed that using functions that follow the data trends improves the prediction performance of the GP model. We also proposed a framework to predict the decrease in the remaining energy under real-world cycling conditions. For this, the gradient of the GP predictions can be used to predict the incremental degradation level at each cycle and then sum or integrate the results to estimate the total degradation under any cycling history.

The work presented in this manuscript suffers from some limitations that will be addressed in future work. For instance, only one temperature chamber was available to do the experiments, limiting us to one temperature at a time when searching for future experiments. If it were not for this constraint, we expect that the parameter space would have been explored more efficiently — less total uncertainty — since our framework would have possibly identified a more optimal set of future experiments that are spread over the temperature domain as they were spread over the maximum SOC and C-rate domains. Additionally, we assumed that cell-to-cell variability only increases with cycling. Future work will allow the variability to change based on the other parameters. One way to accomplish this is to make the parameters of the noise model — in our power law model, these are the slope, power, and intercept — functions of maximum SOC, C-rate, and temperature similar to what we did for the power law prior mean function. Another limitation is that we only performed one active

912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950

learning step identifying future experiments at 291 K. With this one step, we got a GP model that accurately predicts unseen results — prediction error is $< 6\%$ — and this is considered a significant advantage to our work that we did not expect to achieve. However, future work should test this prediction capability by performing more active learning steps and quantifying the prediction performance as a function of the active learning steps. Finally, in our work, we proposed an approach to predict the degradation of batteries cycled under real-world conditions, ignoring the effects of cycling history. This may not be accurate, and future work should investigate the improvement in prediction performance when cycling history is considered.

In conclusion, the key outcomes from this work are as follows:

- Exploring the parameter space efficiently using domain-knowledge-informed GP machine learning and expected information gain
- Quantifying the prediction performance of GP modeling under different definitions
- Using laboratory cycling to predict cell degradation under real-world cycling

6 EXPERIMENTAL PROCEDURES

6.1 Resource availability

6.1.1 Lead contact

Further information and requests for resources should be directed to the lead contact, Maher Alghalayini (malghalayini@lbl.gov)

6.1.2 Materials availability

This study did not generate new materials.

6.1.3 Data and code availability

The data and the codes will be available upon the publication of the manuscript.

6.2 Cycling Protocol

LiFePO$_4$ cells (14430 format) with added tabs were purchased from BatterySpace.com / AA Portable Power Corp. These are listed as having a nominal potential of 3.2V and nominal capacity of 400 mAh, based on which C-rates were computed. Multiple cells were tested simultaneously, all within a single temperature-controlled chamber. Charging parameters, specific maximum SOC, C-rate, and temperature, were chosen to give rapid degradation and shorten the experimental time. Below are summaries of our cycling protocols to evaluate the energy loss of the batteries and our initial scan of temperatures to identify proper plating conditions. Leads from a Maccor battery tester were connected to the cell tabs, and the cells were placed in a vented, explosion-resistant ESPEC temperature chamber, which held each set of cells at a specified temperature for the duration of the tests.

6.2.1 Standard Cell Aging Protocol

1. Rest for 10 minutes
2. Repeat the following, 40 times or until the cell shows signs of shorting
   (a) Repeat the following, 5 times
      (i) Charge at a rate of 1.0C until cell potential increases to 3.75 V or higher
      (ii) Hold the cell potential at 3.75 V until the magnitude of the current decreases to 0.2C or lower
      (iii) Rest for 15 minutes
      (iv) Discharge at a rate of 1.0C until cell potential decreases to 2.5 V or lower
      (v) Hold the cell potential at 2.5 V until the magnitude of the current decreases to 0.02C or lower
      (vi) Rest for 15 minutes
   (b) Repeat the following, 2 times
      (i) Charge at a rate of RC until the charge passed has reached Q or higher or the elapsed time has reached T
      (ii) Rest for 15 minutes
      (iii) Discharge at a rate of 1.0C until cell potential decreases to 2.5 V or lower
      (iv) Hold the cell potential at 2.5 V until the magnitude of the current decreases to 0.02C or lower
      (v) Rest for 15 minutes

6.2.2 Procedure to Identify Plating Conditions

1. Rest for 10 minutes
2. Repeat the following, 5 times
   (a) Charge at 6.0C for until reaching 10% maximum SOC
   (b) Rest for 15 minutes
   (c) Discharge at a rate of 1.0C until cell potential decreases to 2.5 V or lower
   (d) Rest for 15 minutes
3. Repeat the following while increasing the maximum SOC with each charge cycle in intervals of 10% from 10-80% maximum SOC
   (a) Charge at 6.0C for until reaching the desired percent maximum SOC
   (b) Rest for 30 minutes
   (c) Discharge at a rate of 1.0C until cell potential decreases to 2.5 V or lower
   (d) Rest for 30 minutes

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

M.A. Gaussian process framework development and Bayesian optimization implementation, computational work execution, and manuscript writing. D.C.W., K.H., and A.G. experimental procedure development, data collection, and corresponding section writing. V.B. and S.J.H. idea creation and experimental procedure development supervision. M.N. idea creation and Gaussian process and implementation efforts supervision. All authors reviewed and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

[1] Matthieu Dubarry and Arnaud Devie. Battery durability and reliability under electric utility grid operations: Representative usage aging and calendar aging. *Journal of energy storage*, 18:185–195, 2018.

[2] Stephen J Harris and Marcus M Noack. Statistical and machine learning-based durability-testing strategies for energy storage. *Joule*, 7(5):920–934, 2023.

[3] Stephen J Harris, David J Harris, and Chen Li. Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells. *Journal of Power Sources*, 342:589–597, 2017.

[4] Sravan Pannala, Hamidreza Movahedi, Taylor R Garrick, Anna G Stefanopoulou, and Jason B Siegel. Consistently tuned battery lifetime predictive model of capacity loss, resistance increase, and irreversible thickness growth. *Journal of The Electrochemical Society*, 171(1):010532, 2024.

[5] Chico Hermanu Brillianto Apribowo, Sarjiya Sarjiya, Sasongko Pramono Hadi, and Fransisco Danang Wijaya. Optimal planning of battery energy storage systems by considering battery degradation due to ambient temperature: a review, challenges, and new perspective. *Batteries*, 8(12):290, 2022.

[6] JG Qu, ZY Jiang, and JF Zhang. Investigation on lithium-ion battery degradation induced by combined effect of current rate and operating temperature during fast charging. *Journal of Energy Storage*, 52:104811, 2022.

[7] Rutooj D Deshpande and Kotub Uddin. Physics inspired model for estimating 'cycles to failure'as a function of depth of discharge for lithium ion batteries. *Journal of Energy Storage*, 33:101932, 2021.

[8] Shiqi Liu, Junhua Wang, Haolu Liu, Qisheng Liu, Jia Tang, and Zhongxiang Li. Battery degradation model and multiple-indicators based lifetime estimator for energy storage system design and operation: Experimental analyses of cycling-induced aging. *Electrochimica Acta*, 384:138294, 2021.

[9] Laxman Timilsina, Payam R Badr, Phuong H Hoang, Gokhan Ozkan, Behnaz Papari, and Christopher S Edrington. Battery degradation in electric and hybrid electric vehicles: A survey study. *IEEE Access*, 11:42431–42462, 2023.

[10] Jiangtao He, Zhongbao Wei, Xiaolei Bian, and Fengjun Yan. State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage–capacity model. *IEEE Transactions on Transportation Electrification*, 6(2):417–426, 2020.

22

[11] Jiabo Li, Min Ye, Yan Wang, Qiao Wang, and Meng Wei. A hybrid framework for predicting the remaining useful life of battery using gaussian process regression. *Journal of Energy Storage*, 66:107513, 2023.

[12] John S Newman and Charles W Tobias. Theoretical analysis of current distribution in porous electrodes. *Journal of The Electrochemical Society*, 109(12):1183, 1962.

[13] John Newman and William Tiedemann. Porous-electrode theory with battery applications. *AIChE Journal*, 21(1):25–41, 1975.

[14] Marc Doyle, Thomas F Fuller, and John Newman. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *Journal of the Electrochemical society*, 140(6):1526, 1993.

[15] Marc Doyle and John Newman. The use of mathematical modeling in the design of lithium/polymer battery systems. *Electrochimica Acta*, 40(13-14):2191–2196, 1995.

[16] John Newman and Nitash P Balsara. *Electrochemical systems*. John Wiley & Sons, 2021.

[17] Gaizka Saldaña, José Ignacio San Martín, Inmaculada Zamora, Francisco Javier Asensio, Oier Oñederra, and Mikel González. Empirical electrical and degradation model for electric vehicle batteries. *IEEE Access*, 8:155576–155589, 2020.

[18] Shuoqi Wang, Dongxu Guo, Xuebing Han, Languang Lu, Kai Sun, Weihan Li, Dirk Uwe Sauer, and Minggao Ouyang. Impact of battery degradation models on energy management of a grid-connected dc microgrid. *Energy*, 207:118228, 2020.

[19] Matthew B Pinson and Martin Z Bazant. Theory of sei formation in rechargeable batteries: capacity fade, accelerated aging and lifetime prediction. *Journal of the Electrochemical Society*, 160(2):A243, 2012.

[20] Manh-Kien Tran, Manoj Mathew, Stefan Janhunen, Satyam Panchal, Kaamran Raahemifar, Roydon Fraser, and Michael Fowler. A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters. *Journal of Energy Storage*, 43:103252, 2021.

[21] Qi Zhang, Yunlong Shang, Yan Li, Naxin Cui, Bin Duan, and Chenghui Zhang. A novel fractional variable-order equivalent circuit model and parameter identification of electric vehicle li-ion batteries. *ISA transactions*, 97:448–457, 2020.

[22] Yang Li, Mahinda Vilathgamuwa, Troy W Farrell, Ngoc Tham Tran, Joseph Teague, et al. Development of a degradation-conscious physics-based lithium-ion battery model for use in power system planning studies. *Applied Energy*, 248:512–525, 2019.

[23] Zachary M Konz, Brendan M Wirtz, Ankit Verma, Tzu-Yang Huang, Helen K Bergstrom, Matthew J Crafton, David E Brown, Eric J McShane, Andrew M Colclasure, and Bryan D McCloskey. High-throughput li plating quantification for fast-charging battery design. *Nature Energy*, 8(5):450–461, 2023.

[24] Yizhao Gao, Kailong Liu, Chong Zhu, Xi Zhang, and Dong Zhang. Co-estimation of state-of-charge and state-of-health for lithium-ion batteries using an enhanced electrochemical model. *IEEE Transactions on Industrial Electronics*, 69(3):2684–2696, 2021.

[25] Xi Zhang, Yizhao Gao, Bangjun Guo, Chong Zhu, Xuan Zhou, Lin Wang, and Jianhua Cao. A novel quantitative electrochemical aging model considering side reactions for lithium-ion batteries. *Electrochimica Acta*, 343:136070, 2020.

[26] Xiaosong Hu, Yunhong Che, Xianke Lin, and Simona Onori. Battery health

prediction using fusion-based feature selection and machine learning. *IEEE Transactions on Transportation Electrification*, 7(2):382–398, 2020.

[27] Xin Xiong, Yujie Wang, Kaiquan Li, and Zonghai Chen. State of health estimation for lithium-ion batteries using gaussian process regression-based data reconstruction method during random charging process. *Journal of Energy Storage*, 72:108390, 2023.

[28] Xiaoyu Li, Changgui Yuan, Xiaohui Li, and Zhenpo Wang. State of health estimation for li-ion battery using incremental capacity analysis and gaussian process regression. *Energy*, 190:116467, 2020.

[29] Benjamin Larvaron, Marianne Clausel, Antoine Bertoncello, Sébastien Benjamin, and Georges Oppenheim. Chained gaussian processes to estimate battery health degradation with uncertainties. *Journal of Energy Storage*, 67:107443, 2023.

[30] Gae-Won You, Sangdo Park, and Dukjin Oh. Diagnosis of electric vehicle batteries using recurrent neural networks. *IEEE Transactions on Industrial Electronics*, 64(6):4885–4893, 2017.

[31] Noman Khan, Fath U Min Ullah, Amin Ullah, Mi Young Lee, Sung Wook Baik, et al. Batteries state of health estimation via efficient neural networks with multiple channel charging profiles. *Ieee Access*, 9:7797–7813, 2020.

[32] Fan Xu, Fangfang Yang, Zicheng Fei, Zhelin Huang, and Kwok-Leung Tsui. Life prediction of lithium-ion batteries based on stacked denoising autoencoders. *Reliability Engineering & System Safety*, 208:107396, 2021.

[33] Laisuo Su, Mengchen Wu, Zhe Li, and Jianbo Zhang. Cycle life prediction of lithium-ion batteries based on data-driven methods. *ETransportation*, 10:100137, 2021.

[34] Felix Heinrich and Marco Pruckner. Virtual experiments for battery state of health estimation based on neural networks and in-vehicle data. *Journal of Energy Storage*, 48:103856, 2022.

[35] Xing Shu, Jiangwei Shen, Guang Li, Yuanjian Zhang, Zheng Chen, and Yonggang Liu. A flexible state-of-health prediction scheme for lithium-ion battery packs with long short-term memory network and transfer learning. *IEEE Transactions on Transportation Electrification*, 7(4):2238–2248, 2021.

[36] Weihan Li, Neil Sengupta, Philipp Dechent, David Howey, Anuradha Annaswamy, and Dirk Uwe Sauer. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *Journal of power sources*, 482:228863, 2021.

[37] Xiaojun Tan, Di Zhan, Pengxiang Lyu, Jun Rao, and Yuqian Fan. Online state-of-health estimation of lithium-ion battery based on dynamic parameter identification at multi timescale and support vector regression. *Journal of Power Sources*, 484:229233, 2021.

[38] Xiaofei Sun, Kai Zhong, and Min Han. A hybrid prognostic strategy with unscented particle filter and optimized multiple kernel relevance vector machine for lithium-ion battery. *Measurement*, 170:108679, 2021.

[39] Akash Basia, Zineb Simeu-Abazi, Eric Gascard, and Peggy Zwolinski. Review on state of health estimation methodologies for lithium-ion batteries in the context of circular economy. *CIRP Journal of Manufacturing Science and Technology*, 32:517–528, 2021.

[40] Maher Alghalayini, Stephen J Harris, and Marcus Noack. Early prediction of the failure probability distribution for energy storage technologies driven by domain-knowledge-informed machine learning. 2024.

[41] Thomas Waldmann, Björn-Ingo Hogg, and Margret Wohlfahrt-Mehrens. Li plating as unwanted side reaction in commercial li-ion cells–a review. *Journal of Power Sources*, 384:107–124, 2018.

[42] Qianqian Liu, Chunyu Du, Bin Shen, Pengjian Zuo, Xinqun Cheng, Yulin Ma, Geping Yin, and Yunzhi Gao. Understanding undesirable anode lithium plating issues in lithium-ion batteries. *RSC advances*, 6(91):88683–88700, 2016.

[43] Marcus M Noack and James A Sethian. Advanced stationary and nonstationary kernel designs for domain-aware gaussian processes. *Communications in Applied Mathematics and Computational Science*, 17(1):131–156, 2022.

[44] Marcus M Noack, Gregory S Doerk, Ruipeng Li, Jason K Streit, Richard A Vaia, Kevin G Yager, and Masafumi Fukuto. Autonomous materials discovery driven by gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. *Scientific reports*, 10(1):17663, 2020.

[45] Marcus M Noack, Gregory S Doerk, Ruipeng Li, Masafumi Fukuto, and Kevin G Yager. Advances in kriging-based autonomous x-ray scattering experiments. *Scientific reports*, 10(1):1325, 2020.

[46] Xiaoyu Li, Changgui Yuan, and Zhenpo Wang. Multi-time-scale framework for prognostic health condition of lithium battery using modified gaussian process regression and nonlinear regression. *Journal of Power Sources*, 467:228358, 2020.

[47] Charles A Micchelli, Yuesheng Xu, and Haizhang Zhang. Universal kernels. *Journal of Machine Learning Research*, 7(12), 2006.

[48] Matthias W Seeger, Sham M Kakade, and Dean P Foster. Information consistency of nonparametric gaussian process methods. *IEEE Transactions on Information Theory*, 54(5):2376–2382, 2008.

[49] Aad W Van der Vaart and J Harry Van Zanten. Adaptive bayesian estimation using a gaussian random field with inverse gamma bandwidth. 2009.

[50] Xiaoqiong Pang, Xiaoyan Liu, Jianfang Jia, Jie Wen, Yuanhao Shi, Jianchao Zeng, and Zhen Zhao. A lithium-ion battery remaining useful life prediction method based on the incremental capacity analysis and gaussian process regression. *Microelectronics Reliability*, 127:114405, 2021.

[51] Jianwen Meng, Meiling Yue, and Demba Diallo. A degradation empirical-model-free battery end-of-life prediction framework based on gaussian process regression and kalman filter. *IEEE Transactions on Transportation Electrification*, 9(4):4898–4908, 2022.

[52] Zicheng Fei, Fangfang Yang, Kwok-Leung Tsui, Lishuai Li, and Zijun Zhang. Early prediction of battery lifetime via a machine learning based framework. *Energy*, 225:120205, 2021.

[53] Kristen A Severson, Peter M Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H Chen, Muratahan Aykol, Patrick K Herring, Dimitrios Fraggedakis, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 2019.

[54] Elisa Braco, Idoia San Martín, Alberto Berrueta, Pablo Sanchis, and Alfredo Ursúa. Experimental assessment of cycling ageing of lithium-ion second-life batteries from electric vehicles. *Journal of Energy Storage*, 32:101695, 2020.

[55] Weiping Diao, Jonghoon Kim, Michael H Azarian, and Michael Pecht. Degradation modes and mechanisms analysis of lithium-ion batteries with knee points. *Electrochimica Acta*, 431:141143, 2022.

[56] Peter M Attia, Alexander Bills, Ferran Brosa Planella, Philipp Dechent, Goncalo Dos Reis, Matthieu Dubarry, Paul Gasper, Richard Gilchrist, Samuel Greenbank, David Howey, et al. "knees" in lithium-ion battery aging trajectories. *Journal of*

1251   *The Electrochemical Society*, 169(6):060517, 2022.

1252 [57] Thorsten Baumhöfer, Manuel Brühl, Susanne Rothgang, and Dirk Uwe Sauer.
1253   Production caused variation in capacity aging trend and correlation to initial cell
1254   performance. *Journal of Power Sources*, 247:332–338, 2014.

1255 [58] Karl Ezra Pilario, Mahmood Shafiee, Yi Cao, Liyun Lao, and Shuang-Hua Yang.
1256   A review of kernel methods for feature extraction in nonlinear process monitoring.
1257   *Processes*, 8(1):24, 2019.

1258 [59] Marcus M Noack and Kristofer G Reyes. Mathematical nuances of gaussian
1259   process-driven autonomous experimentation. *MRS Bulletin*, 48(2):153–163, 2023.

1260 [60] Marcus M Noack, Hengrui Luo, and Mark D Risser. A unifying perspec-
1261   tive on non-stationary kernels for deeper gaussian processes. *arXiv preprint
1262   arXiv:2309.10068*, 2023.

1263 [61] Christopher Paciorek and Mark Schervish. Nonstationary covariance functions for
1264   gaussian process regression. *Advances in neural information processing systems*,
1265   16, 2003.

1266 [62] Paul D Sampson and Peter Guttorp. Nonparametric estimation of nonstationary
1267   spatial covariance structure. *Journal of the American Statistical Association*,
1268   87(417):108–119, 1992.

1269 [63] Christopher J Paciorek and Mark J Schervish. Spatial modelling using a new
1270   class of nonstationary covariance functions. *Environmetrics: The official journal
1271   of the International Environmetrics Society*, 17(5):483–506, 2006.

1272 [64] Piyush Tagade, Krishnan S Hariharan, Sanoop Ramachandran, Ashish Khandel-
1273   wal, Arunava Naha, Subramanya Mayya Kolake, and Seong Ho Han. Deep gaus-
1274   sian process regression for lithium-ion battery health prognosis and degradation
1275   mode diagnosis. *Journal of Power Sources*, 445:227281, 2020.

1276 [65] Yalong Yang, Siyuan Chen, Tao Chen, and Liansheng Huang. State of health
1277   assessment of lithium-ion batteries based on deep gaussian process regression
1278   considering heterogeneous features. *Journal of Energy Storage*, 61:106797, 2023.

1279 [66] Xizhe Wang, Xufeng Hong, Quanquan Pang, and Benben Jiang. Deep kernel
1280   learning-based bayesian optimization with adaptive kernel functions. *IFAC-
1281   PapersOnLine*, 56(2):5531–5535, 2023.

1282 [67] Hieu A Doan, Garvit Agarwal, Hai Qian, Michael J Counihan, Joaquín Rodríguez-
1283   López, Jeffrey S Moore, and Rajeev S Assary. Quantum chemistry-informed
1284   active learning to accelerate the design and discovery of sustainable energy storage
1285   materials. *Chemistry of Materials*, 32(15):6338–6346, 2020.

1286 [68] Nahid Sultana, SM Zakir Hossain, Salma Hamad Almuhaini, and Dilek
1287   Düştegör. Bayesian optimization algorithm-based statistical and machine learning
1288   approaches for forecasting short-term electricity demand. *Energies*, 15(9):3425,
1289   2022.

1290 [69] Adam Thelen, Murtaza Zohair, Jayanth Ramamurthy, Andrew Harkaway,
1291   Weimin Jiao, Mihir Ojha, Mahdi Ul Ishtiaque, Todd A Kingston, Cary L Pint,
1292   and Chao Hu. Sequential bayesian optimization for accelerating the design of
1293   sodium metal battery nucleation layers. *Journal of Power Sources*, 581:233508,
1294   2023.

1295 [70] Arpan Biswas, Anna N Morozovska, Maxim Ziatdinov, Eugene A Eliseev, and
1296   Sergei V Kalinin. Multi-objective bayesian optimization of ferroelectric materials
1297   with interfacial control for memory and energy storage applications. *Journal of
1298   Applied Physics*, 130(20), 2021.

1299 [71] Garvit Agarwal, Hieu A Doan, Lily A Robertson, Lu Zhang, and Rajeev S Assary.
1300   Discovery of energy storage molecular materials using quantum chemistry-guided

multiobjective bayesian optimization. *Chemistry of Materials*, 33(20):8133–8144, 2021.

[72] Pavlos Nikolaidis and Sotirios Chatzis. Gaussian process-based bayesian optimization for data-driven unit commitment. *International Journal of Electrical Power & Energy Systems*, 130:106930, 2021.

[73] Maho Harada, Hayami Takeda, Shinya Suzuki, Koki Nakano, Naoto Tanibata, Masanobu Nakayama, Masayuki Karasuyama, and Ichiro Takeuchi. Bayesian-optimization-guided experimental search of nasicon-type solid electrolytes for all-solid-state li-ion batteries. *Journal of Materials Chemistry A*, 8(30):15103–15109, 2020.

[74] Farshud Sorourifar, Naitik Choksi, and Joel A Paulson. Computationally efficient integrated design and predictive control of flexible energy systems using multi-fidelity simulation-based bayesian optimization. *Optimal Control Applications and Methods*, 44(2):549–576, 2023.

[75] Eric Brochu, Vlad M Cora, and Nando De Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

[76] Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1997.

[77] Marcus M Noack, Hengrui Luo, and Mark D Risser. A unifying perspective on non-stationary kernels for deeper gaussian processes. *APL Machine Learning*, 2(1), 2024.

[78] Robert R Richardson, Michael A Osborne, and David A Howey. Battery health prediction under generalized conditions using a gaussian process transition model. *Journal of Energy Storage*, 23:320–328, 2019.