

Predicting battery end of life from solar off-grid system field data using machine learning

Antti Aitio, David Howey

October 15, 2021

Abstract

Hundreds of millions of people lack access to electricity. Decentralised solar-battery systems are key for addressing this whilst avoiding carbon emissions and air pollution, but are hindered by relatively high costs and rural locations that inhibit timely preventative maintenance. Accurate diagnosis of battery health and prediction of end of life from operational data improves user experience and reduces costs. But lack of controlled validation tests and variable data quality mean existing lab-based techniques fail to work. We apply a scaleable probabilistic machine learning approach to diagnose health in 1027 solar-connected lead-acid batteries, each running for 400-760 days, totalling 620 million data rows. We demonstrate 73% accurate prediction of end of life, eight weeks in advance, rising to 82% at the point of failure. This work highlights the opportunity to estimate health from existing measurements using ‘big data’ techniques, without additional equipment, extending lifetime and improving performance in real-world applications.

Keywords: battery, health, machine learning, rural electrification, Gaussian process

Context and scale

Off-grid solar-battery systems provide clean electricity to those who lack access, enabling education and enterprise. However, these systems are in remote areas, and when their batteries fail, it can be difficult to replace them quickly. To improve reliability, uptime and cost-effectiveness, a scalable and non-invasive method to diagnose and predict battery health is required. We demonstrate how real-world battery operating data may be used to infer health and detect end-of-life failure using only the measured voltage, current and temperature data from users in the field, with no additional sensors or requirements to take systems offline for specific tests. This work highlights the opportunity to complement laboratory battery data with large field datasets analyzed through machine learning to improve performance and understanding of aging.

1 Introduction

To achieve universal electricity access, the number of decentralised solar-battery systems and solar mini-grids in areas without grid access will need to increase ten-fold [1], but this is inhibited by the relatively high costs and uncertain lifetimes of batteries [2]. Diagnosis and prediction of battery state of health (SoH) in real-world operating environments is required for operational safety, warranties, and planning of maintenance, as well as for improving designs by understanding the impact of varying usage on battery life. Diagnosis in the field is challenging because direct measurement of SoH using standardised performance tests is usually not possible due to the costs of the required service interruption and testing equipment. Therefore health diagnosis should be performed directly from monitored operational data, for example battery terminal voltage, temperature and current. However, this means that the controlled operating conditions that would normally ensure consistent health estimation in laboratory tests are missing. Further complication arises because the most common battery health metrics—capacity and internal resistance—are influenced by operating conditions. Finally, the current, voltage and temperature sensors used in many real-world applications are not of the same accuracy as those in laboratories, and data recording is often incomplete, resulting in gaps in the time series.

Techniques for battery health estimation can broadly be classified into model-based and data-driven methods [3]–[6]. Model-based approaches typically use an electrical equivalent circuit model combined with techniques from feedback control to track internal states, such as state of charge (SoC), and parameters, such as resistance and capacity. Gradual evolution in the parameters enables state of health estimation using Bayesian filtering [7]–[9] or adaptive observers [10]–[12]. The choice of battery model is a trade-off between parsimony/computational resources, and accuracy/flexibility. Equivalent circuit models are ubiquitous and widely employed [7], [8], but may suffer from a lack of accuracy across the wide range of operating points experienced in real world usage [13]. Higher fidelity ‘physics-based’ models derived from porous electrode theory are also available [14], [15], but are generally considered too complex and computationally demanding for state of health tracking despite recent improvements [16]–[20].

In contrast, data-driven methods try to estimate state of health either directly from raw current, voltage and temperature measurements, or from features such as time spent in certain operating regions. Machine learning techniques used for this purpose include Gaussian process (GP) regression [21], [22], support vector machines [23] and neural networks [24], [25]. Data-driven health estimation has mainly been investigated using laboratory data under controlled conditions at relatively small scales. Many publicly available experimental laboratory datasets exist [26], but they are limited in size.

Supplementing laboratory aging data, field data from end users enables failure detection, gives insight into real-world performance, and improves understanding of manufacturing and usage variability [27]. There have been few studies of battery SoH estimation using field data, even though this presents a significant opportunity to broaden understanding of performance [28], [29]. Although existing demonstrators show promise, their ‘black-box’ nature reduces interpretability in comparison to model-driven approaches. Additionally, if charging and discharging patterns are highly dynamic, it might not be possible to calculate consistent features over the battery lifetime. So-called ‘hybrid’ models that combine data-driven and physics-based methods [30] offer promise by balancing robustness, flexibility and transparency.

In real-world applications, constraints are imposed by changing usage conditions, less accurate sensors, lack of controlled usage and lack of prior knowledge of model parameters. Here, we implement physics-informed probabilistic machine learning for SoH estimation that is robust to changing operating conditions and data gaps. Our approach, illustrated in Fig. 1, is demonstrated

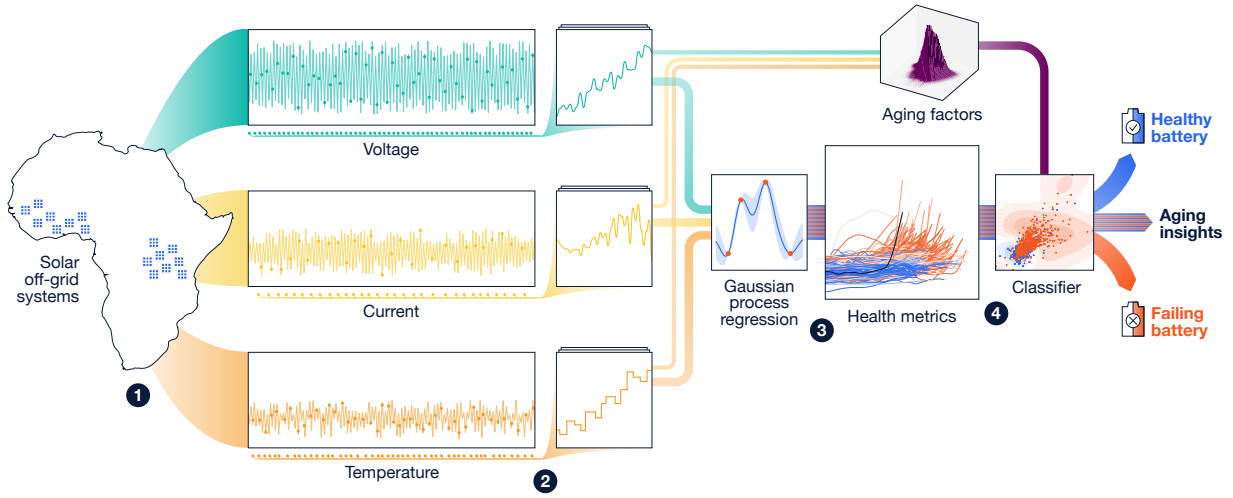


Figure 1: Workflow, left to right: (1) Data are retrieved from PV-connected batteries—the raw dataset contains 620 million rows of current, voltage and temperature measurements. (2) Data are sub-sampled for computational efficiency to yield 39 million rows from an average of 261 charging segments per battery. (3) GP regression is applied to obtain estimates for health metrics (calibrated internal resistance) over the lifetime of each battery. (4) Batteries are classified as healthy or failing up to 2 months prior to end of life using these metrics plus other aging factors calculated from the raw data; this also gives insight into the key factors driving aging.

using measured current, voltage and temperature data from lead-acid batteries, but is applicable to any chemistry that can be represented with a low order electrical model, including lithium-ion cells. We employ Gaussian process regression [31]–[33] to construct health trajectories of 1027 batteries connected to photovoltaic systems in sub-Saharan Africa, using internal resistance as a health metric, since it may be estimated from data more accurately than discharge capacity, because the observable impact of capacity changes on measured voltage is very small. Resistance is estimated as a function of instantaneous operating conditions, such as temperature and state of charge, enabling calibration of data, resulting in smooth estimates. Next, an end-of-life failure detection algorithm based on a Gaussian process classifier [31] predicts the probability of future failure by combining the estimated health trajectories with stress factors also extracted from the raw data. To train the classifier, batteries were labelled as failed or healthy using repair data from field workshops. We demonstrate, using 5-fold stratified cross-validation, that this technique gives 82% balanced accuracy of end-of-life failure prediction at the time of failure versus a 66% benchmark, and 73% accuracy 8 weeks in advance of failure versus a 50% benchmark.

2 Real-world dataset

Our dataset was generated by 1027 lead-acid batteries, each with nominal voltage 12 V (internally comprising 6 cells in series), nominal capacity 20 Ah, and attached to a 50 Wp photovoltaic panel. These systems are used for lighting, phone charging and small appliances, and are located across sub-Saharan Africa. Each battery was in use for 400–760 days, giving a total dataset size of 620

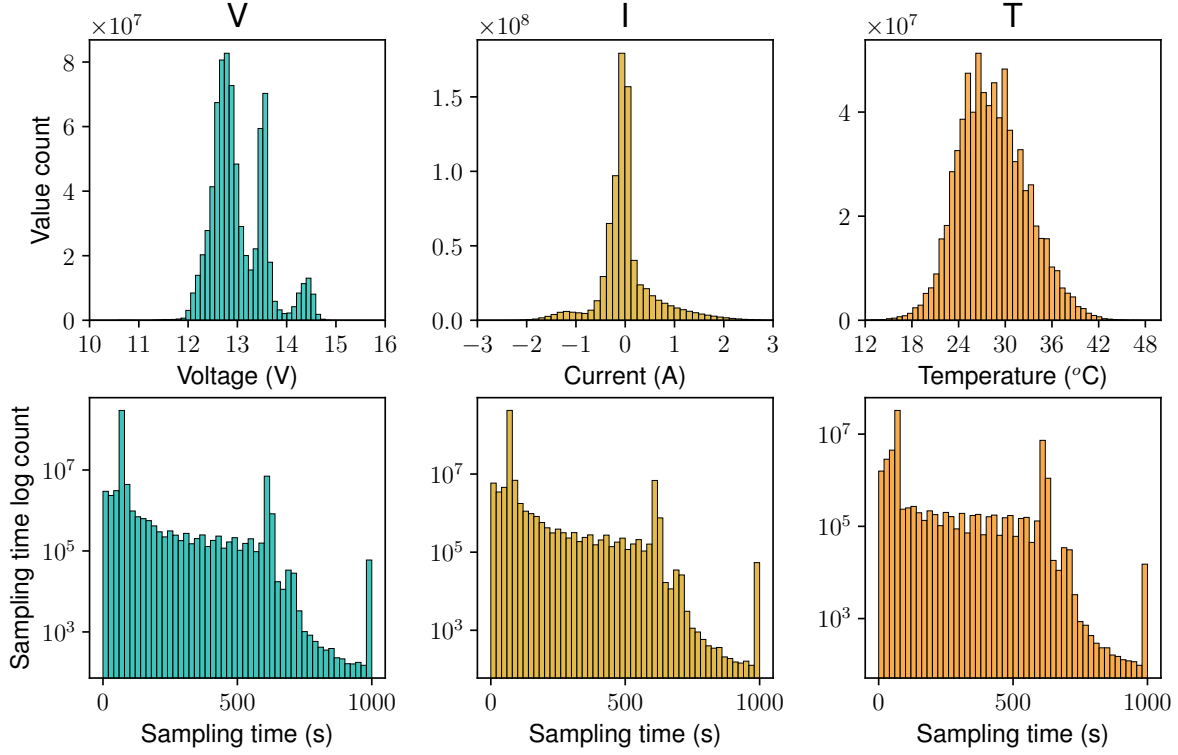


Figure 2: Solar off-grid systems experience varying states of charge, currents and temperatures, and the systems we used had non-uniform data sampling. Shown here (top) are distributions of voltage, current, temperature and (bottom) histograms of their sampling times. Sampling time has a mode at $t=60$ s for each measurement (counts are plotted on a logarithmic scale).

million rows (49 GB). This dataset is a small subset of the total number of systems deployed, and was selected to ensure that each time series was at least 400 days in length and also that the set contained approximately the same number of failed versus healthy batteries. A full explanation of the data selection process is given in Supplementary Material S1. The distributions of terminal voltage, current (where, by convention, charging is positive), temperature and sampling times are shown in Fig. 2. Sampling times are irregular with a mode of 60 s, with temperature recorded less frequently on average.

There are some particular challenges for battery health diagnosis from the measured field data in this application. Firstly, the depth of discharge is commonly quite small, for example the majority of usage in our dataset is between 46–100% SoC, making it difficult to observe changes in the discharge voltage curve caused by capacity changes as the cell ages. Secondly, since average currents are small (99th percentile approximately 0.1C, where C-rate is defined as the current divided by the nominal Ampere-hour capacity of the battery) and often constant, estimation of the internal resistance is numerically poorly conditioned, in other words small errors in measured voltage cause large errors in estimated resistance. These difficulties are compounded by unknown sensor accuracy.

Furthermore, the average currents, temperatures and depths of discharge vary over time and across the population of batteries. We therefore require a methodology for health diagnosis that is robust to changes in usage and that does not require controlled diagnostic tests. In most battery systems, either capacity or internal resistance (or both) may be used as a health metric (e.g.

[34], [35]). In lead-acid systems, internal resistance is a clear health indicator [36], although the techniques outlined here can also be used to estimate capacity if sufficiently deep discharges are available. In our dataset, charging segments typically provide a more varied and higher amplitude input signal for resistance estimation compared with discharging segments, where the current is on average small and relatively constant. We therefore chose to estimate the resistance during charging as the state of health metric.

3 Data-driven modelling

We now describe the modelling and inference process, outlined in Fig. 1, in detail. First, measured current, voltage and temperature data with non-uniform sampling were obtained from a central database, and suitable charging segments selected (see Methods)—on average, 261 such segments were identified per battery. Then, within each charging segment, data were interpolated onto a 1 minute uniform grid in time, and electrolyte acid concentration as a state of charge metric was estimated using Coulomb counting, with initial conditions calibrated by a known open circuit voltage versus state of charge relationship (see Methods). This process resulted in an average of 38000 rows of data for each battery, reduced 16-fold down from 603000 rows per battery.

The internal resistance of all batteries depends on age, current, temperature and state of charge. In lead-acid cells the reasons for this include non-linear kinetics [18], [19], nucleation and dissolution of lead sulfate [37], hydrolysis during charging [38], [39], and degradation mechanisms such as sulfation, loss of active material and electrode corrosion [40], [41]. Because of this, resistance estimates are not a reliable SoH metric unless they are first calibrated to remove the impacts of current, temperature and state of charge. Comprehensively modelling all the underlying physics would result in a model with a large number of parameters [38], and these are challenging to estimate from data [19]. However, machine learning techniques can be used to learn from data the dependency of internal resistance on other factors. To ensure principled treatment of uncertainty associated with field data, we used a Bayesian approach, expressing the battery internal resistance as a Gaussian process over applied current, temperature, estimated electrolyte concentration (i.e. state of charge), and time.

In contrast to parametric models [35], Gaussian processes offer a flexible probabilistic technique that makes fewer assumptions about the structure of the underlying data [31]. We assumed that internal resistance over the lifetime of each battery consisted of a sum of two independent Gaussian processes, the first capturing the dependency on the instantaneous operating point, and the second the degradation. This improves computational efficiency, although it assumes that the dependence of resistance on age can be decoupled from the dependence of resistance on temperature, state of charge and current. We modelled the operating point dependency using a standard squared exponential kernel and the degradation dependency using a Wiener velocity kernel [42], both shown in Fig. 3. The former reflects our assumption that the variability of resistance with instantaneous temperature, state of charge and current should be relatively smooth. The latter is a non-stationary kernel [43], enabling degradation at the beginning of life to be zero for each individual battery, and then to grow as the battery ages. This means that extrapolation of future battery health, beyond available data, follows the trajectory learned from the data more accurately.

Gaussian process models have ‘hyperparameters’ that describe the smoothness, magnitude and periodicity of the data being modelled. Fitting to data and then using the resulting model involves two steps. First, the hyperparameter values must be estimated, then the posterior distribution (i.e. mean and variance) of the internal resistance must be evaluated. Both of these steps can be computationally expensive, each by default scaling with $\mathcal{O}(n^3)$, where n is the number of

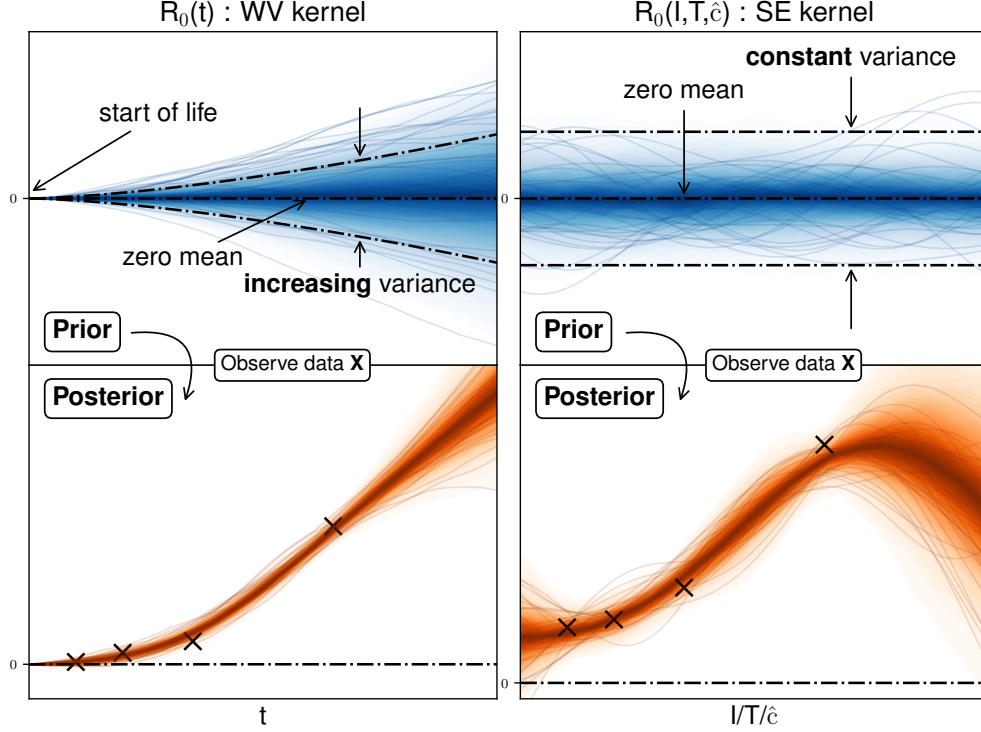


Figure 3: Illustration of the two kernels used to model battery internal resistance as a SoH metric. A Gaussian process defines a prior distribution over functions and is conditioned with observations to produce a posterior distribution. Here, random draws from priors (top row) and posteriors (bottom row) are shown for a non-stationary Wiener velocity (WV) kernel used to model the time-dependency of resistance, and a stationary squared exponential (SE) kernel with short scale smoothness used to model the dependence of resistance on operating point. Dotted lines show 1σ in the priors. The WV kernel anchors the aging process to zero at start of life and extrapolates its last known trajectory, whereas the SE kernel is non-zero everywhere and reverts to the mean when extrapolated.

data points being fitted [31], equal to approximately 38000 per battery. To overcome this challenge, we applied recursive techniques [32], [33] (see Methods). To benchmark our approach against an existing method [7], [8], we considered a separate case where R_0 follows a random walk through time and is assumed to be independent of operating conditions. This process is also controlled by a set of hyperparameters for which we found maximum-likelihood estimates using the same methodology.

4 Constructing a consistent battery health metric

To ensure that comparisons of health between batteries were like-for-like, we estimated internal resistance as a function of temperature, state of charge, current and time in the down-selected dataset, and then chose a *single* constant set of values of the operating conditions at which to evaluate SoH, for all batteries. This can be thought of as learning a function $R_0 = f(T, I, \hat{c}, t)$ from data, using the GP technique described in the previous section, and then evaluating or slicing through this function at fixed values of all independent variables apart from time.

Fig. 4a shows the variability of internal resistance as a function of temperature, current and state of charge estimated at the population level (see Methods). A wide range of operating points are observed in the down-selected dataset, as shown in the Fig. 4a insets, and consequently the estimated value of R_0 changes substantially with the instantaneous operating conditions. If this is not accounted for, then these effects would mask the variability caused solely by degradation.

The function describing R_0 that is learned from data is consistent with battery physics. First, the inverse relationship between internal resistance and temperature is due to the Arrhenius dependence of reaction rate on temperature, and results in a variation of up to $0.16\ \Omega$ in estimated resistance. Second, the shape of internal resistance with respect to acid concentration is due to competing effects. The exchange current density is an increasing function of electrolyte acid concentration [18], but the transport limitation caused by the reduced rate of dissolution of lead-sulfate during charging [37] becomes dominant at higher states of charge. The size of this SoC effect was estimated to be approximately $0.09\ \Omega$ across our dataset, with relatively high uncertainty at lower SoC values, which are visited less often, as shown by the wider credible intervals. Finally, internal resistance reduces as current increases because the relationship between reaction rate and overpotential is nonlinear, specifically, the activation overpotential increases more slowly at higher current densities, as described by Butler-Volmer kinetics [18]. Accounting for this dependency on applied current is key because it has the largest impact on R_0 in the observed operating range, resulting in changes of up to $1.58\ \Omega$, although there is high uncertainty associated with the estimate at higher applied currents. This is caused by the batteries spending relatively little time at higher current, resulting in fewer data points and therefore a wider variance of the posterior predictive distribution obtained through GP regression.

To calibrate estimates of battery internal resistance across all batteries we must choose a reference set of operating conditions (temperature, current, SoC). As indicated in Fig. 4b, there is a trade off between choosing a single fixed set of calibration conditions for the entire population of batteries versus using separate calibration conditions for each battery trajectory. To obtain a low variance estimate of the resistance of an individual battery, it is best to calibrate using the mean operating conditions for that individual battery, since conditions further from this are associated with higher uncertainty. However, a standardised population-wide calibration condition allows truly like-for-like comparisons to be made between batteries, and gives 37% lower overall standard deviation, so this is what we chose. The remaining population variance reflects cell-to-cell variability at beginning of life resulting from manufacturing variations, storage time and conditions prior to field deployment.

As a result of the calibration process we produce trajectories of R_0 that are only a function of time, all referred to a common operating point which was the population mean of temperature, current and acid concentration. The trajectories are shown in Fig. 5. In addition to the absolute value of resistance R_0 , the derivative with respect to time $\partial R_0 / \partial t$ is also a useful health indicator (calculation given in Supplementary Material). Together these two metrics capture whether the battery degradation is beyond the so-called 'knee point', i.e. the onset of accelerated degradation towards end of life.

5 Predicting and understanding battery end of life

Having constructed consistent battery health trajectories from raw measured data by estimating internal resistance and calibrating it for operating conditions, we can now (i) validate the health metrics by using them to predict end-of-life failure, and (ii) gain insight into the factors that drive battery aging. To predict battery end of life due to aging, we used repair data provided by BBOX

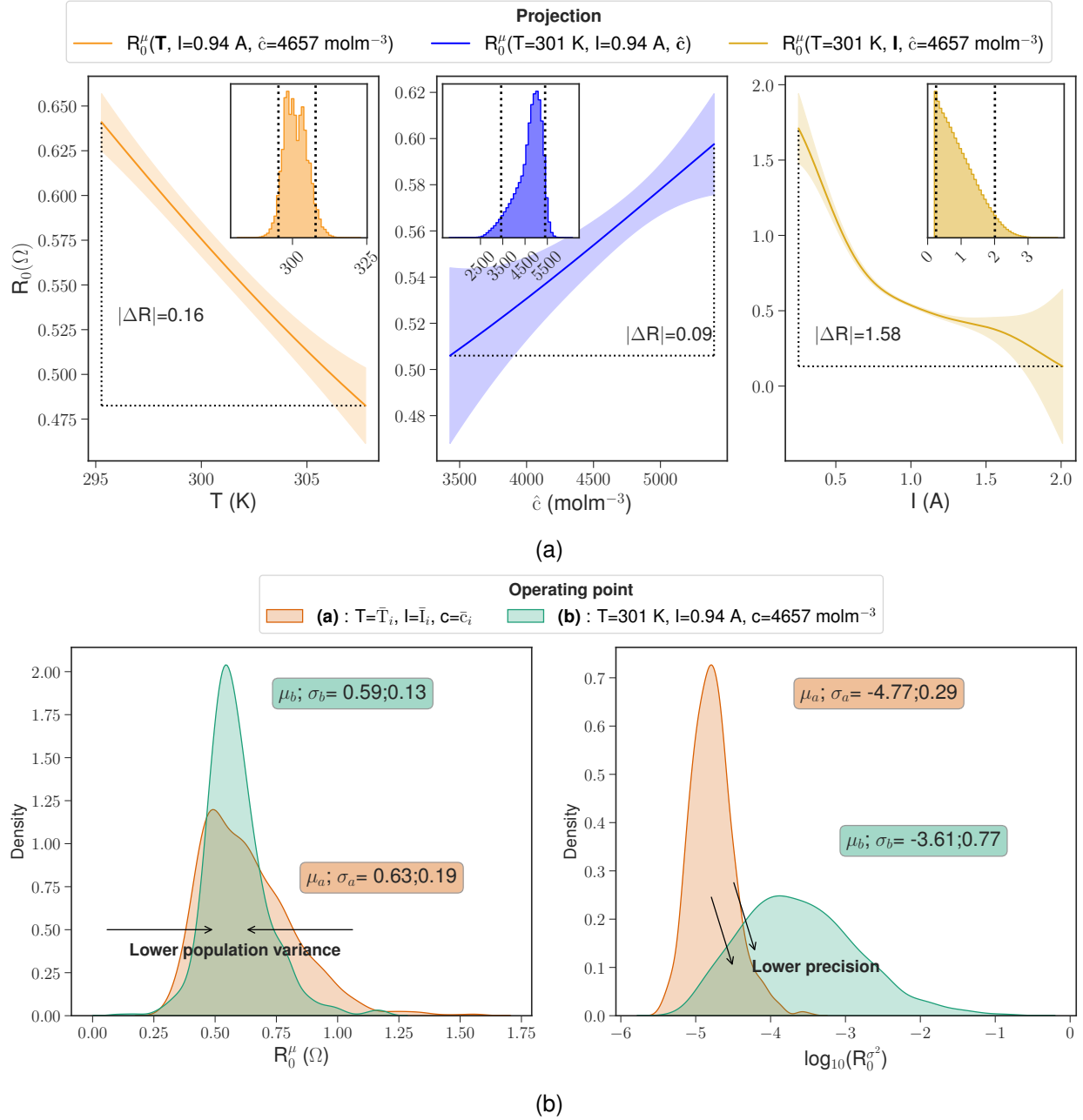


Figure 4: (a) Internal resistance varies significantly with operating conditions, as shown by projections of estimated R_0 as a function of temperature, applied current and SoC (acid concentration) between 5th and 95th percentiles of each. Shaded regions show $\pm 2\sigma$ credible intervals. Insets show histograms of the independent variables for the down-selected dataset. (b) There is a trade-off between using a single population-wide calibration point vs. a local calibration point per battery. Left-hand image shows lower variance in estimated mean of R_0 values with a single calibration point. However, in this case the uncertainty associated with each R_0 estimate increases, as right-hand image shows.

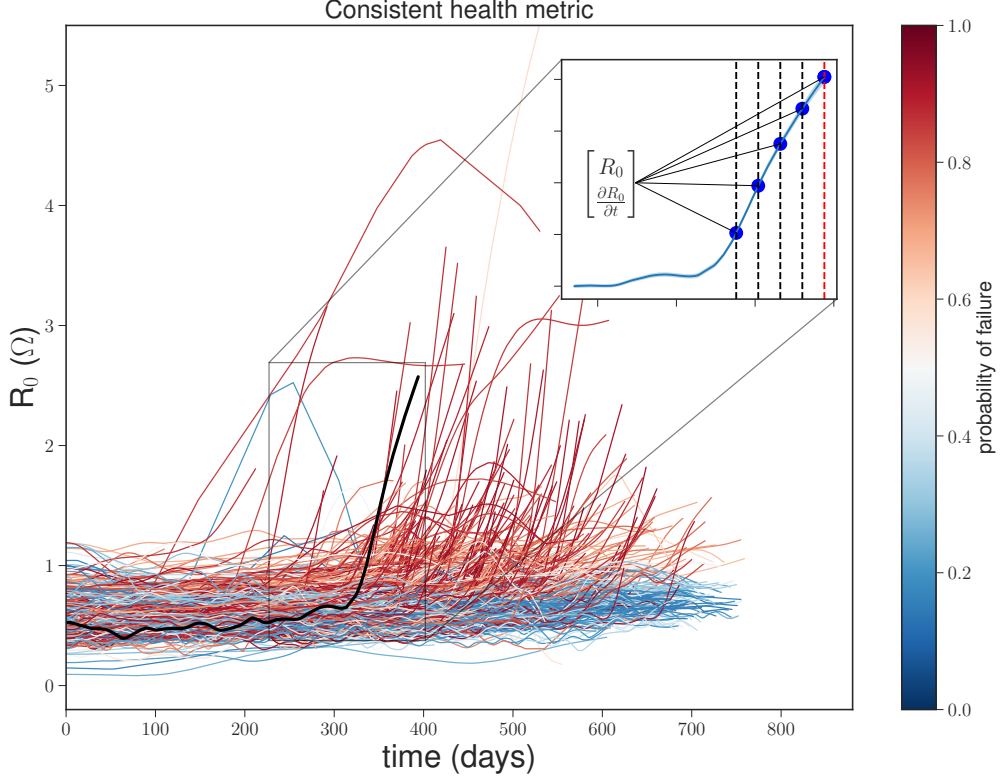


Figure 5: Our approach results in smooth SoH trajectories across the entire dataset of 1027 batteries, shown here as calibrated internal resistance profiles estimated from measured data. Color indicates the probability of a battery having failed at the end of each time line, as predicted by a classifier using the features calculated as inputs. The ‘knee point’ is a key indicator of end of life, and as the inset shows, after this the resistance R_0 and its gradient $\partial R_0 / \partial t$ increase.

Ltd. to label each battery as failed or healthy at the end its respective data time series, and trained a GP classifier [31] to predict this based on health and stress factors at various different prediction horizons. This provides an indirect method of validating the health estimates, although there is uncertainty associated with the labeling of healthy and failed batteries, i.e. there could be false positives and false negatives since repairs are driven by customer decisions. The available data were split into training and test sets, and the classification performance was measured by predicting whether a repair would occur in the test set.

We assessed the performance of this end-of-life classifier under multiple scenarios. First, we used only the health features R_0 and $\partial R_0 / \partial t$ described above as inputs. Second, to benchmark the performance, we used a health metric consisting solely of internal resistance fitted with a random walk, as is common in the literature [7], [8], without consideration of operating-condition dependency. Third, we considered the case where the health metrics R_0 and $\partial R_0 / \partial t$ were augmented with stress factors known to affect lead-acid battery health [40]. The latter are features indicative of usage that are also extracted from the raw data, and they consisted of calendar age, charge throughput, cycle count, cumulative time spent at float charge voltage, as well as average temperature and voltage (see Supplementary Material S5 for calculation details). Fourth, we tested the performance using solely the stress factors, omitting information relating to the current state of health as indicated by resistance. To demonstrate the ability to predict end of life in advance with different horizons from zero to eight weeks, and with varying ratios of failed versus healthy batteries,

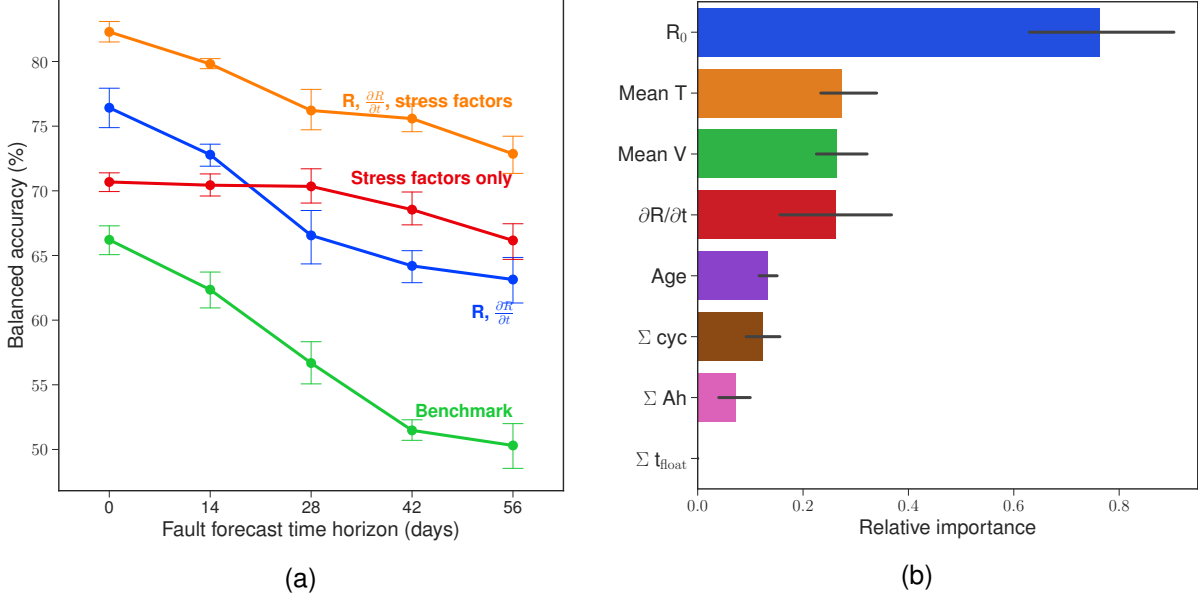


Figure 6: (a) Combining estimated SoH (resistance and gradient of resistance) with aging stress factors gives best overall end-of-life prediction. Classification performance is quantified using balanced accuracy, for prediction horizons from 0-56 days. Error bars show the standard deviation in the average test accuracy grouped by test set, including a varying proportion of failed batteries. (b) Internal resistance, mean temperature, rate of change of internal resistance with time and mean voltage are important predictors of end of life. The relative importance of various inputs was quantified using the inverse length scales of the hyperparameters averaged across training sets. The standard deviation of each inverse length scale is indicated by the error bar.

we structured the classification performance tests with nested cross validation (see Methods).

The classifier performance as a function of prediction horizon and according to the various scenarios is shown in Fig. 6(a). The best overall end-of-life diagnosis and prediction performance occurs when we combine both the current state of health, as indicated by estimated resistance and its gradient, with the stress factors. In this case 82% balanced accuracy (see equation 12 in Methods for definition) is achieved at end of life and 73% when the time horizon for prediction is extended to 8 weeks. The Supplementary Material gives the full confusion matrix showing classifier performance. In comparison, the benchmark approach and the approach using only estimated SoH—without stress factors—give a lower accuracy which rapidly drops with increasing prediction horizon. This is due to the absence of knee points in SoH as one moves further back in time from the end of life; two months before they fail, relatively few batteries have experienced the onset of accelerated degradation. At this point the benchmark model performance is near 50% balanced accuracy, which is equivalent to random classification. The classifier using only the stress factors—without resistance estimates—performs well in comparison to the benchmark, achieving an average accuracy of 69% over all horizons. This indicates that aging is being driven by usage conditions and calendar life, and is fairly consistent across the population, although introducing battery-specific SoH information near end of life significantly improves predictive accuracy.

The classification technique also gives insight into the factors driving aging. Relative importance was quantified using the inverses of the length-scale hyperparameters of the classifier. Resistance and its rate of change over time are important end-of-life indicators, but other factors, in particular mean temperature and mean voltage, are significant, as shown in Fig. 6(b). Temperature

is known to have a complex impact on lead-acid battery degradation, and the full dataset spans more than 20 °C range. Elevated temperatures, especially during charging, may improve lifetime due to improved solubility of lead sulfate [40], although they also increase electrode grid corrosion. Similarly, mean voltage is indicative of mechanisms such as sulfation, which occurs at low voltage [40]. Although some aging factors, such as cycle count, are considered unimportant, this is likely because they are correlated with other factors such as charge throughput and age (see Supplementary Material S5).

6 Conclusions

Real-world operating data from battery systems in the field may be used to detect end-of-life failure before it happens, improving maintenance, value, safety and customer experience and giving insights into degradation and performance. We have developed and demonstrated at large scale a data-driven approach for battery state of health estimation and end-of-life prediction using only measured current, voltage and temperature data whilst in use, without requiring controlled cycling or systems to be taken offline. Using data from 1027 solar off-grid lead-acid batteries, each running for 400-760 days, we obtain an end-of-life prediction performance of 73% balanced accuracy, 8 weeks prior to end of life, rising to 82% accuracy at end of life. This accuracy was achieved by combining estimates of state of health with aging stress factors also calculated from the measured data. We used probabilistic machine learning techniques to learn internal resistance as a function of current, temperature, state of charge and time, enabling us to calibrate state of health to consistent reference conditions across the entire population dataset. The success of the approach may be explained by the combination of a population-wide health model with a battery-specific indicator that becomes increasingly important towards end of life. These techniques are applicable to any battery that can be represented with a simple electrical circuit model assuming that sufficient data are available, that degradation manifests as a resistance change, and that C-rates are small. Like any data-driven approach, a limitation of the work is that a sufficiently large dataset is required to train the models and demonstrate effectiveness, although a Bayesian approach provides a principled framework for relating uncertainty to the amount of data available. Broadly, this work highlights the opportunity to complement laboratory battery data with large field datasets analyzed through machine learning to improve performance and understanding. An immediate practical benefit is cost savings in the supply chain for battery spare parts through reducing storage times and inventory requirements. Additionally, more accurate lifetime models can be built from the population data—this would be particularly powerful if a large number of solar-battery system organisations are willing to share operational data. Finally, these algorithms could be used for second life screening if sufficient first life battery data are available.

7 Methods

7.1 Data selection

To obtain a clean dataset and reduce the computational load required for model training, several pre-processing steps were necessary. First, segments of charging data were chosen over the lifetime of each battery. To be eligible, segments had to meet the conditions listed in Table 1.

Charging segment duration	> 6000 s
Starting voltage range	11.5 V-12.9 V
Starting current	< 0.1 A
max(Voltage in segment)	> 14 V
max(Time gap in recorded data)	< 610 s

Table 1: Qualifying charging segment conditions

These conditions ensured that charging segments covered a reasonable state of charge range. Additionally, each charging segment was truncated to include voltages only up to 14 V due to increased uncertainty in estimating state of charge at higher voltages when using the method described in section 7.2, because the magnitude of the side reactions increases exponentially with terminal voltage. After down-selecting appropriate charging segments, data were interpolated to a 1-minute time grid using piece-wise cubic hermite interpolation [44]. During charging, data points where $I_t < 0.2$ A, or the estimated open circuit voltage was above the measured terminal voltage (which may occur sometimes due to inaccuracy of the state of charge estimate), were removed to improve numerical conditioning and ensure resistance estimates were always positive.

7.2 Inferring acid concentration from current and voltage data

In lead-acid batteries, electrolyte acid concentration is a direct measure of state of charge—the latter is the normalised version of the former, but requires the maximum concentration to be known, and this changes with age. We therefore simply use acid concentration as a state of charge indicator throughout, and use the two terms interchangeably in this paper. To infer acid concentration from measured data, we first measured the battery open circuit voltage in a laboratory (see Supplementary Material S2), then used this to determine initial acid concentration from measured voltage at points where the charging current was at a minimum. The required open circuit voltage curve was measured experimentally using the galvanostatic intermittent titration technique (Biologic SP-150 potentiostat), placing the battery in a thermal chamber (Binder MK53) at 25 °C. From this data, the electrolyte volume was also inferred by a least squares fit using Bode’s well known result [45]. Finally, the trajectory of acid concentration for each charging segment was obtained by Coulomb counting from the initial concentration and using the measured current data, accounting for the known side reactions in lead-acid systems [38], [39] by using a lumped term for the gassing reactions [41] according to

$$\frac{d\hat{c}_t}{dt} = \frac{I_t - I_{\text{gas},0} e^{c_T(T_t - T_0) + c_V(V_t - V_0)}}{FV_{\text{elec}}}, \quad (1)$$

where I_t , T_t , V_t are the measured current, temperature and terminal voltage, F is Faraday constant, V_{elec} is the estimated electrolyte volume, and gassing current parameters $I_{\text{gas},0}$, c_T , T_0 , c_V , V_0 were from literature [41]. The substantial uncertainty due to the variation in these parameters over the lifetime of the battery was taken into consideration by projecting the input uncertainty in the acid concentration $\hat{c}$ to be measurement noise variance in GP regression (equations 6) .

7.3 Input data normalisation

To ensure comparability of hyperparameters, the current, temperature and estimated acid concentration I_t , T_t , $\hat{c}_t$ were normalised using the population-level moments according to

$$X = \frac{x - \bar{x}}{\sigma_x}, \quad x \in \{T_t, I_t, \hat{c}_t\}, \quad (2)$$

where $\bar{x}$ and σ_x represent the population mean and variance of the down-selected dataset. A normalised time-scale was obtained by dividing the time since beginning of life in seconds by 34 560 000 s, i.e. a nominal 400 days life expressed in seconds, bringing it to a similar range as the other inputs. This method of normalisation (2) was also used for the inputs for the GP classifier.

7.4 Modelling internal resistance as a Gaussian Process

7.4.1 Battery-specific estimates

Given that the C-rates during charging are low ($<0.2C$), we may ignore the effect of concentration overpotentials [18], [19]. Therefore the terminal voltage is given by the sum of the open circuit voltage and a lumped linearised internal resistance term,

$$V_t = V_0(\hat{c}_t) + R_0(t, I_t, T_t, \hat{c}_t)I_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{n,t}^2) \quad (3)$$

where V_t is the terminal voltage and t , I_t , T_t , $\hat{c}_t$ are the (normalised) time since beginning of life, applied current (positive for charging by convention), measured temperature and estimated bulk sulfuric acid concentration at time t respectively. The experimentally obtained open circuit voltage as a function of acid concentration is V_0 and the dependency $R_0(I_t, T_t, \hat{c}_t, t)$ is modelled by a zero mean Gaussian process,

$$R_0(x) \sim \mathcal{GP}(0, k(x)), \quad x \in \{t, I, T, \hat{c}\}, \quad (4)$$

where k is the covariance function. We chose to model R_0 as a sum of two GPs, where the degradation process as a function of time is described by a Wiener velocity (WV) kernel [43] and the dependency of R_0 on operating point $(I_t, T_t, \hat{c}_t)$ is described by a squared exponential (SE) kernel, such that

$$k(I, T, \hat{c}, t) = \sigma_{t,0}^2 \left(\frac{\min^3(t, t')}{3} + |t - t'| \frac{\min^2(t, t')}{2} \right) + \sigma_{t,1}^2 \exp \left(\sum_{x \in \{I, T, \hat{c}\}} -\frac{(x - x')^2}{2l_x^2} \right), \quad (5)$$

where $|\cdot|$ denotes the absolute distance between two points. The choice of the WV process, used in target tracking applications [46], allows for better extrapolation outside the observed data because the kernel is non-stationary. In comparison, using the zero-mean SE kernel would result in extrapolation over longer time horizons tending to the prior distribution. Examples of random draws from the prior distributions for each kernel are shown in Fig. 3.

Extrapolation over time is necessary in our case because the down-selected charging segments are not evenly distributed over time across the population. Extrapolation is thus required to give estimates of R_0 at points 0-56 days preceding the data series end for each battery. Additionally, expressing R_0 as a sum of kernels, equation 5, makes the assumption that the degradation is purely additive, and independent of the operating point chosen. This significant simplification reduces the degrees of freedom of the system in comparison to a product of kernels over all inputs and makes inference computationally lower cost.

To fit the model defined by equations 3-5, the hyperparameters to be estimated are the two process variances, $\sigma_{f,0}^2, \sigma_{f,1}^2$, and length scales l_x across the inputs, $(l_t, T_t, \hat{c}_t)$. The model is fitted by adjusting these hyperparameters according to a cost function (defined in equation 15 of the Supplementary Material) which includes the error between the modelled and measured voltage. The fitted function is given by the posterior distribution of R_0 for each battery. Special treatment was given to the noise ϵ_t in equation 3. There may be considerable uncertainty in the open circuit voltage function $V_0(\hat{c})$, for example due to hysteresis, as well as in the estimate of acid concentration $\hat{c}$, due to uncertainty in the parameters in equation 1. Additionally, our experimental estimation of the OCV curve and electrolyte volume (see Supplementary Material S2) was done using a single battery, so variations in these across the population are another source of uncertainty. To account for this, we assumed a 10% uncertainty in $d\hat{c}/dt$, together with a 150 mV standard deviation caused by voltage measurement and open circuit voltage uncertainty, giving a total variance per charge segment as

$$\sigma_{n,t}^2 = 0.0225 + \text{Var}(\hat{c}_t) \left(\frac{dV_0}{d\hat{c}} \bigg|_{\hat{c}} \right)^2 \quad (6a)$$

$$\text{Var}(\hat{c}_t) = \sum_t 0.01 \Delta \hat{c}_t^2, \quad (6b)$$

where $\Delta \hat{c}_t = (dc_t/dt) \Delta t$, assuming dc_t/dt is constant in each time segment Δt . In summary, the GP regression includes heteroskedastic noise $\sigma_{n,t}^2$ which we pre-calculate rather than estimate as a hyperparameter in the fitting process.

To impose smoothness on the function R_0 over all input dimensions, we assumed a prior distribution over hyperparameters,

$$p(\sigma_{f,0}, \sigma_{f,1}, l_t, l_T, l_{\hat{c}}) = \prod_{m \in \{0,1\}} \chi(\sigma_{f,m}, k=1, s=0.2) \prod_{x \in \{l, T, \hat{c}\}} \Gamma^{-1}(l_x, \alpha=1, \beta=2), \quad (7)$$

where χ represents the chi distribution and Γ^{-1} the inverse gamma distribution. The chi distribution with the degree of freedom $k=1$ is equal to a half-normal distribution with standard deviation $s=0.2$. We assumed an inverse gamma prior for the length scales, with shape and scale parameters α and β chosen to give a mode of 1.

To estimate the posterior distribution of R_0 for each battery and to recover maximum-a-posteriori (MAP) estimates of the hyperparameters, we employed a recursive estimation framework for GPs [33], [47], [48], where they are interpreted to be the solution of a stochastic partial differential equation of the form

$$\frac{\partial R_0(t, l, T, \hat{c})}{\partial t} = F R_0(t, l, T, \hat{c}) + L w(t, l, T, \hat{c}), \quad (8)$$

where the transition matrix F , dispersion matrix L and the properties of the white noise process $w(t, l, T, \hat{c})$ are determined by the kernel function, equation 5. This framework allows the use of standard Kalman filtering and smoothing techniques [43], [49] to estimate both the posterior distribution of R_0 and the so-called energy function, that is the negative unnormalised logarithm of the posterior probability of the hyperparameter vector. Crucially, this method scales as $\mathcal{O}(n)$ over the number of data rows, which in our case is order 10^4 for each battery.

To obtain a finite-dimensional representation of the system in equation 8, we used a similar approach to that of Särkkä et al. [48]. First, we applied a k-means algorithm to choose 20 representative points across $(l_t, T_t, \hat{c}_t)$, at which to estimate R_0 through the lifetime of the battery. In order to obtain estimates for R_0 for the observed $(l_t, T_t, \hat{c}_t)$ through all charging segments, the

predictive distribution of the GP over the operating points was calculated and added to the value predicted by the degradation GP.

Using these recursive techniques, we fitted regression hyperparameters across the entire population of batteries, consisting of 39 million rows of data, in approximately 80 minutes using an Apache SparkTM cluster running 30 cores.

7.4.2 Population level model over operating points

Population level estimates of R_0 as a function of all the operating points were calculated by fitting a Gaussian process over the individual battery R_0 estimates (i.e. the solutions of equation 8) at each battery's mean operating point at the beginning of life. This was done using the standard batch GP approach [31], with the SE kernel and hyperprior for the magnitude and lengthscales equivalent to the recursive case. The hyperprior including the measurement noise was then

$$p(\sigma_f, \sigma_n, l_T, l_I, l_C) = \chi(\sigma_n, k = 1, s = 0.1) \chi(\sigma_f, k = 1, s = 0.2) \prod_{x \in \{I, T, C\}} \Gamma^{-1}(l_x, \alpha = 1, \beta = 2). \quad (9)$$

In addition, we added the variance estimates for each R_0 value to the measurement noise. Given MAP estimates retrieved by the L-BFGS-B algorithm [50], the posterior means and variances for the population level R_0 function were given by

$$\mu_p = k(x^*, X) \left[k(X, X) + \sigma_n^2 \mathbf{I} + \sigma_R^2(X) \right]^{-1} R_0(X) \quad (10a)$$

$$\sigma_p^2 = k(x^*, x^*) - k(x^*, X) \left[k(X, X) + \sigma_n^2 \mathbf{I} + \sigma_R^2(X) \right]^{-1} k(X, x^*), \quad (10b)$$

where k is the SE kernel function and $R_0(X)$ and $\sigma_R^2(X)$ denote the mean and variance of the battery-wise R_0 estimates at operating point X — in this case X was the mean operating for the battery. The predictive distribution was calculated on a grid x^* such that the range for each input variable (temperature, applied current and estimated acid concentration) was varied between its 5th and 95th percentile in turn while keeping the other two constant at the population mean.

7.4.3 Benchmark model

To benchmark our approach, we used the same model (equation 3) but without the dependency of R_0 on the operating point. Furthermore, we assumed that R_0 followed a random walk through time, an approach commonly taken in the literature to adapt parameters to data [51]. Using this recursive approach, the tuning parameters are the process and noise covariances as well as the initial variance estimate for R_0 , which are often set manually. For consistency with the main approach, we estimated the process noise covariance using maximum likelihood and set the initial variance to be a constant multiple of the process covariance and the noise covariance was calculated in the same way (equations 6) as for the main method.

7.5 End-of-life prediction using Gaussian process classification

To classify the batteries as failed or healthy at various time horizons up to the end date of each battery data series, we used a standard GP classification framework [31]. The inputs $R_0, \frac{\partial R_0}{\partial t}$ were extrapolated as required from the last observed charging segment to the appropriate point in time

preceding the end of the time series. In all cases, inputs were fed into a GP classifier using a SE covariance function with automatic relevance detection,

$$k_{\text{class}}(x, x') = \sigma_f^2 \exp \left(\sum_x -\frac{(x - x')^2}{2l_x^2} \right) \quad (11)$$

with an uniform hyperprior. Maximum likelihood estimates for the hyperparameters were then obtained by minimizing the negative log marginal likelihood of the data, using the GP classifier implementation in the Scikit-learn toolbox [52]. The performance metrics chosen for the classifier were such that the unevenness in labelling (536 healthy, 491 failed) was taken into account. The balanced accuracy metric was calculated as the average of classifier sensitivity and specificity,

$$\text{Balanced accuracy} = \frac{1}{2}(\text{Sensitivity} + \text{Specificity}) = \frac{1}{2} \left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right), \quad (12)$$

where TP, TN, FP, FN represent the true positive, true negative, false positive and false negative rates respectively.

To quantify classifier performance, first we used 5-fold stratified cross validation to split the dataset into training and test sets. The training set in each case had a 48/52% proportion of failed vs. healthy batteries using data up to the end of each time series, training 20 classifiers altogether (4 sets of inputs, and 5 training sets). Then, for each outer cross-validation stage, test subsets were selected, each containing varying proportions of failed batteries. These subsets were chosen by keeping the healthy battery set fixed and then randomly sampling 40%, 60% and 80% of the failed batteries in the test set, repeated ten times, plus the case when 100% of the failed batteries was used, giving a total of 3100 test cases. All performance metrics reported are the average for the test sets using 5-fold cross validation for each test case.

8 Acknowledgements

The authors thank BBOX Ltd. for access to data from PV-connected battery energy storage systems; D. Lovell for assistance with Fig. 1; V. Sulzer, R. Drummond, J. Reniers and S. Cooper for manuscript feedback. We acknowledge funding from the Faraday Institution (EP/S003053/1, grant number FIRG003) and Shell Foundation (agreement 22077).

9 Author contributions

Conceptualization, D.A.H.; data curation, software, visualization, A.A.; writing – original draft, review, editing, A.A. and D.A.H.; supervision, project administration, funding acquisition, D.A.H.

10 Competing interests

The authors have filed a patent related to this work: GB Application No. 2105995.1, dated 27 April 2021. D.A.H. is co-founder of Brill Power Ltd., and is a technical advisor at Habitat Energy Ltd. A.A. declares no competing interests.

References

- [1] ESMAP, "Energy Sector Management Assistance Program. 2019. Mini Grids for Half a Billion People : Market Outlook and Handbook for Decision Makers," World Bank, Tech. Rep., 2019. [Online]. Available: <https://openknowledge.worldbank.org/handle/10986/31926>.
- [2] J. T. Lee and D. S. Callaway, "The cost of reliability in decentralized solar power systems in sub-Saharan Africa," *Nature Energy*, vol. 3, no. 11, pp. 960–968, 2018, ISSN: 20587546. DOI: 10.1038/s41560-018-0240-y. [Online]. Available: <http://dx.doi.org/10.1038/s41560-018-0240-y>.
- [3] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer, "Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles," *Journal of Power Sources*, vol. 281, pp. 114–130, 2015, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2015.01.129. [Online]. Available: <http://dx.doi.org/10.1016/j.jpowsour.2015.01.129>.
- [4] M. Bercibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van Den Bossche, *Critical review of state of health estimation methods of Li-ion batteries for real applications*, 2016. DOI: 10.1016/j.rser.2015.11.042.
- [5] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *Journal of Power Sources*, vol. 405, no. 5, pp. 18–29, 2018, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2018.10.019. [Online]. Available: <https://doi.org/10.1016/j.jpowsour.2018.10.019>.
- [6] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Bercibar, E. Nanini-Maury, J. Van Mierlo, and H. E. Hoster, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renewable and Sustainable Energy Reviews*, vol. 113, p. 109254, Oct. 2019, ISSN: 1364-0321. DOI: 10.1016/J.RSER.2019.109254. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S136403211930454X>.
- [7] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation," *Journal of Power Sources*, vol. 134, no. 2, pp. 277–292, Aug. 2004, ISSN: 0378-7753. DOI: 10.1016/J.JPOWSOUR.2004.02.033. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378775304003611>.
- [8] —, "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*, vol. 161, pp. 1369–1384, 2006. DOI: 10.1016/j.jpowsour.2006.06.004. [Online]. Available: https://ac.els-cdn.com/S0378775306011438/1-s2.0-S0378775306011438-main.pdf?_tid=405e013f-8b51-4005-8034-e2f3fd93939b&acdnat=1551435649_2f4f89f933a745077c0cb012de28c
- [9] A. Baba and S. Adachi, "Simultaneous state of charge and parameter estimation of lithium-ion battery using log-normalized unscented Kalman Filter," in *2015 American Control Conference (ACC)*, IEEE, Jul. 2015, pp. 311–316, ISBN: 978-1-4799-8684-2. DOI: 10.1109/ACC.2015.7170754. [Online]. Available: <http://ieeexplore.ieee.org/document/7170754/>.
- [10] I. S. Kim, "A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer," *IEEE Transactions on Power Electronics*, vol. 25, no. 4, pp. 1013–1022, 2010, ISSN: 08858993. DOI: 10.1109/TPEL.2009.2034966.

- [11] J. Kim, S. Lee, and B. H. Cho, "Complementary cooperation algorithm based on DEKF combined with pattern recognition for SOC/capacity estimation and SOH prediction," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 436–451, 2012, ISSN: 08858993. DOI: 10.1109/TPEL.2011.2158554.
- [12] S. J. Moura, N. A. Chaturvedi, and M. Krstic, "Adaptive Partial Differential Equation Observer for Battery State-of-Charge/State-of-Health Estimation Via an Electrochemical Model," *Journal of Dynamic Systems, Measurement, and Control*, vol. 136, no. 1, p. 011 015, Oct. 2013, ISSN: 0022-0434. DOI: 10.1115/1.4024801. [Online]. Available: <http://dynamicsystems.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4024801>.
- [13] J. Gomez, R. Nelson, E. E. Kalu, M. H. Weatherspoon, and J. P. Zheng, "Equivalent circuit model parameters of a high-power Li-ion battery: Thermal and state of charge effects," *Journal of Power Sources*, vol. 196, no. 10, pp. 4826–4831, May 2011, ISSN: 0378-7753. DOI: 10.1016/J.JPOWSOUR.2010.12.107. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378775311000899>.
- [14] M. Doyle, T. F. Fuller, and J. Newman, "Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymer/Insertion Cell," *Journal of The Electrochemical Society*, vol. 140, no. 6, p. 1526, Jun. 1993, ISSN: 00134651. DOI: 10.1149/1.2221597. [Online]. Available: <http://jes.ecsdl.org/cgi/doi/10.1149/1.2221597>.
- [15] T. F. Fuller, M. Doyle, and J. Newman, "Simulation and Optimization of the Dual Lithium Ion Insertion Cell," *Journal of The Electrochemical Society*, vol. 141, no. 1, p. 1, Jan. 1994, ISSN: 00134651. DOI: 10.1149/1.2054684. [Online]. Available: <http://jes.ecsdl.org/cgi/doi/10.1149/1.2054684>.
- [16] A. Jokar, B. Rajabloo, M. Désilets, and M. Lacroix, "Review of simplified Pseudo-two-Dimensional models of lithium-ion batteries," *Journal of Power Sources*, vol. 327, pp. 44–55, 2016, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2016.07.036. [Online]. Available: <http://dx.doi.org/10.1016/j.jpowsour.2016.07.036>.
- [17] S. G. Marquis, V. Sulzer, R. Timms, C. P. Please, and S. J. Chapman, "An Asymptotic Derivation of a Single Particle Model with Electrolyte," *Journal of The Electrochemical Society*, vol. 166, no. 15, A3693–A3706, 2019, ISSN: 0013-4651. DOI: 10.1149/2.0341915jes.
- [18] V. Sulzer, S. J. Chapman, C. P. Please, D. A. Howey, and C. W. Monroe, "Faster Lead-Acid Battery Simulations from Porous-Electrode Theory: Part I. Physical Model," *Journal of The Electrochemical Society*, vol. 166, no. 12, A2363–A2371, Jul. 2019, ISSN: 0013-4651. DOI: 10.1149/2.0301910jes. [Online]. Available: <http://jes.ecsdl.org/lookup/doi/10.1149/2.0301910jes>.
- [19] —, "Faster Lead-Acid Battery Simulations from Porous-Electrode Theory: Part II. Asymptotic Analysis," *Journal of The Electrochemical Society*, vol. 166, no. 12, A2372–A2382, Jul. 2019, ISSN: 0013-4651. DOI: 10.1149/2.0441908jes. [Online]. Available: <http://jes.ecsdl.org/lookup/doi/10.1149/2.0441908jes>.
- [20] Z. Chu, G. L. Plett, M. S. Trimboli, and M. Ouyang, "A control-oriented electrochemical model for lithium-ion battery, Part I: Lumped-parameter reduced-order model with constant phase element," *Journal of Energy Storage*, vol. 25, no. July, p. 100 828, 2019, ISSN: 2352152X. DOI: 10.1016/j.est.2019.100828. [Online]. Available: <https://doi.org/10.1016/j.est.2019.100828>.

- [21] R. R. Richardson, C. R. Birkel, M. A. Osborne, and D. A. Howey, "Gaussian Process Regression for in Situ Capacity Estimation of Lithium-Ion Batteries," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 127–138, 2019, ISSN: 15513203. DOI: 10.1109/TII.2018.2794997.
- [22] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *Journal of Power Sources*, vol. 384, no. March, pp. 387–395, 2018, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2018.03.015. [Online]. Available: <https://doi.org/10.1016/j.jpowsour.2018.03.015>.
- [23] V. Klass, M. Behm, and G. Lindbergh, "A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation," *Journal of Power Sources*, vol. 270, pp. 262–272, 2014, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2014.07.116.
- [24] H. Chaoui, C. C. Ibe-Ekeocha, and H. Gualous, "Aging prediction and state of charge estimation of a LiFePO₄ battery using input time-delayed neural networks," *Electric Power Systems Research*, vol. 146, pp. 189–197, 2017, ISSN: 03787796. DOI: 10.1016/j.epsr.2017.01.032. [Online]. Available: <http://dx.doi.org/10.1016/j.epsr.2017.01.032>.
- [25] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, "Online capacity estimation of lithium-ion batteries with deep long short-term memory networks," *Journal of Power Sources*, vol. 482, p. 228 863, 2021, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2020.228863. [Online]. Available: <https://doi.org/10.1016/j.jpowsour.2020.228863>.
- [26] G. dos Reis, C. Strange, M. Yadav, and S. Li, "Energy and AI Lithium-ion battery data and where to find it," *Energy and AI*, vol. 5, 2021. DOI: 10.1016/j.egyai.2021.100081.
- [27] V. Sulzer, P. Mohtat, A. Aitio, S. Lee, Y. T. Yeh, M. U. Khan, J. W. Lee, J. B. Siegel, A. David, and A. G. Stefanopoulou, "The challenge of battery lifetime prediction from field data," *Joule*, pp. 1–20, 2021, ISSN: 2542-4351. DOI: 10.1016/j.joule.2021.06.005. [Online]. Available: <https://doi.org/10.1016/j.joule.2021.06.005>.
- [28] L. Song, K. Zhang, T. Liang, X. Han, and Y. Zhang, "Intelligent state of health estimation for lithium-ion battery pack based on big data analysis," *Journal of Energy Storage*, vol. 32, no. August, p. 101 836, 2020, ISSN: 2352152X. DOI: 10.1016/j.est.2020.101836. [Online]. Available: <https://doi.org/10.1016/j.est.2020.101836>.
- [29] Q. Wang, Z. Wang, L. Zhang, P. Liu, and Z. Zhang, "A novel consistency evaluation method for series-connected battery systems based on real-world operation data," *IEEE Transactions on Transportation Electrification*, vol. 7, no. 2, pp. 1–1, 2020, ISSN: 2332-7782. DOI: 10.1109/tte.2020.3018143.
- [30] M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. van Vlijmen, M. D. Berliner, M. Z. Bazant, R. D. Braatz, W. C. Chueh, and B. D. Storey, "Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime," *Journal of The Electrochemical Society*, vol. 168, no. 3, p. 030 525, 2021, ISSN: 0013-4651. DOI: 10.1149/1945-7111/abec55.
- [31] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, eng, ser. Adaptive computation and machine learning. Cambridge, Mass. ; London: MIT, 2006, ISBN: 9780262182539.
- [32] S. Särkkä and J. Hartikainen, "Infinite-dimensional kalman filtering approach to spatio-temporal Gaussian process regression," *Journal of Machine Learning Research*, vol. 22, pp. 993–1001, 2012, ISSN: 15337928.

- [33] S. Särkkä, A. Solin, and J. Hartikainen, "Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, 2013, ISSN: 10535888. DOI: 10.1109/MSP.2013.2246292.
- [34] G. L. Plett, "Recursive approximate weighted total least squares estimation of battery cell total capacity," *Journal of Power Sources*, vol. 196, no. 4, pp. 2319–2331, 2011, ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2010.09.048. [Online]. Available: <http://dx.doi.org/10.1016/j.jpowsour.2010.09.048>.
- [35] J. Remmlinger, M. Buchholz, M. Meiler, P. Bernreuter, and K. Dietmayer, "State-of-health monitoring of lithium-ion batteries in electric vehicles by on-board internal resistance estimation," *Journal of Power Sources*, vol. 196, no. 12, pp. 5325–5331, 2011, ISSN: 15213773. DOI: 10.1002/anie.201504913.
- [36] P. J. Kollmeyer and T. M. Jahns, "Aging and performance comparison of absorbed glass matte, enhanced flooded, PbC, NiZn, and LiFePO4 12V start stop vehicle batteries," *Journal of Power Sources*, vol. 441, no. September, p. 227 139, 2019, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2019.227139. [Online]. Available: <https://doi.org/10.1016/j.jpowsour.2019.227139>.
- [37] M. Huck, "Modeling transient processes in lead-acid batteries in the time domain," *Journal of Energy Storage*, vol. 29, no. April, p. 101 430, 2020, ISSN: 2352-152X. DOI: 10.1016/j.est.2020.101430. [Online]. Available: <https://doi.org/10.1016/j.est.2020.101430>.
- [38] J. Newman, "Simulation of Recombinant Lead-Acid Batteries," *Journal of The Electrochemical Society*, vol. 144, no. 9, p. 3081, 1997, ISSN: 00134651. DOI: 10.1149/1.1837963. [Online]. Available: <http://jes.ecsdl.org/cgi/doi/10.1149/1.1837963>.
- [39] Dawn M. Bernardi* and Michael K. Carpenter, "A Mathematical Model of the Oxygen-Recombination Lead-Acid Cell," *Journal of The Electrochemical Society*, vol. 142, no. 8, pp. 2631–2642, 1995, ISSN: 00219169. DOI: 10.1016/0021-9169(62)90162-9.
- [40] P. Ruetschi, "Aging mechanisms and service life of lead-acid batteries," *Journal of Power Sources*, vol. 127, no. 1-2, pp. 33–44, 2004, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2003.09.052.
- [41] J. Schiffer, D. U. Sauer, H. Bindner, T. Cronin, P. Lundsager, and R. Kaiser, "Model prediction for ranking lead-acid batteries according to expected lifetime in renewable energy systems and autonomous power-supply systems," *Journal of Power Sources*, vol. 168, no. 1 SPEC. ISS. Pp. 66–78, 2007, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2006.11.092.
- [42] A. Solin, "Applied Stochastic Differential Equations," 2019.
- [43] Arno Solin, "Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression," Ph.D. dissertation, Aalto University, 2016. [Online]. Available: <https://aaltodoc.aalto.fi/bitstream/handle/123456789/19842/isbn9789526067117.pdf?sequence=1&isAllowed=y>.
- [44] R. E. Carlson and F. N. Fritsch, "Monotone piecewise cubic interpolation," *SIAM J Numer. Anal.*, vol. 17, no. 2, pp. 238–246, 1980.
- [45] H. Bode, *Lead-acid batteries*. Wiley, 1977.
- [46] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. New York: Wiley, 2001, ISBN: 9780471221272.

- [47] J. Hartikainen and S. Särkkä, “Kalman filtering and smoothing solutions to temporal Gaussian process regression models,” *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2010*, no. Mlsp, pp. 379–384, 2010. DOI: 10.1109/MLSP.2010.5589113.
- [48] S. Särkkä and J. Hartikainen, “Infinite-Dimensional Kalman Filtering Approach to Spatio-Temporal Gaussian Process Regression,” Tech. Rep., 2012.
- [49] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [50] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A Limited Memory Algorithm for Bound Constrained Optimization,” *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995. DOI: 10.1137/0916069. [Online]. Available: <https://doi.org/10.1137/0916069>.
- [51] Gregory L. Plett, *Battery Management Systems, Volume II: Equivalent-Circuit Methods*. Artech House Publishers, 2015, ISBN: 9781630810276.
- [52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine Learning in {P}ython,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Supplementary Material for ‘Predicting battery end of life from solar off-grid system field data using machine learning’

Antti Aitio, David Howey

October 15, 2021

1 Selection of battery data

The database maintained by off-grid solar provider BBOXX Ltd. contains time series data for over 300,000 batteries. Each battery is equipped with telemetry to measure current, voltage and temperature. Current sensors are calibrated with a 0.2% accurate reference before leaving the factory. Voltage and temperature sensors have systematic uncertainties of 3% and 2 °C, respectively. The measured data include different chemistries, cell capacities and manufacturers. Ensuring comparability of batteries was crucial, so several steps were taken to obtain a consistent dataset. We obtained the final set of 1027 batteries through the following steps:

1. We chose lead-acid batteries from single manufacturer with nominal capacity 20 Ah.
2. We defined the lifetime of a battery as the time since the ‘activation date’. We chose batteries with a lifetime of over 400 days, with an activation date after 2018-09-01.
3. For those batteries that entered repair at some point after activation, end of life was defined as the date they entered repair. The dataset was filtered down so only batteries that either had no repairs or entered repair due to loss of capacity (as defined by BBOXX) were included. The (initial) failed set consisted of all the batteries that had been diagnosed with a loss of capacity when taken into a repair shop. To make up the total dataset with an approximate 52/48 split between healthy and failed cells, we used stratified sampling with respect to battery lifetime to generate a set of healthy batteries from the overall population.
4. We removed any batteries that had large gaps in their telemetry data (over 30 days).
5. We removed any batteries that had tampering alerts during operation.
6. We filtered out telemetry data for each battery where temperature readings were clearly incorrect.
7. We removed any batteries that did not have enough charging segments. We calculated the number of charging segments available for each battery as described in the Methods section and filtered out batteries with fewer than 50 qualifying segments. Additionally, batteries were omitted where the last qualifying segment was more than 30 days before the end of the dataset series.

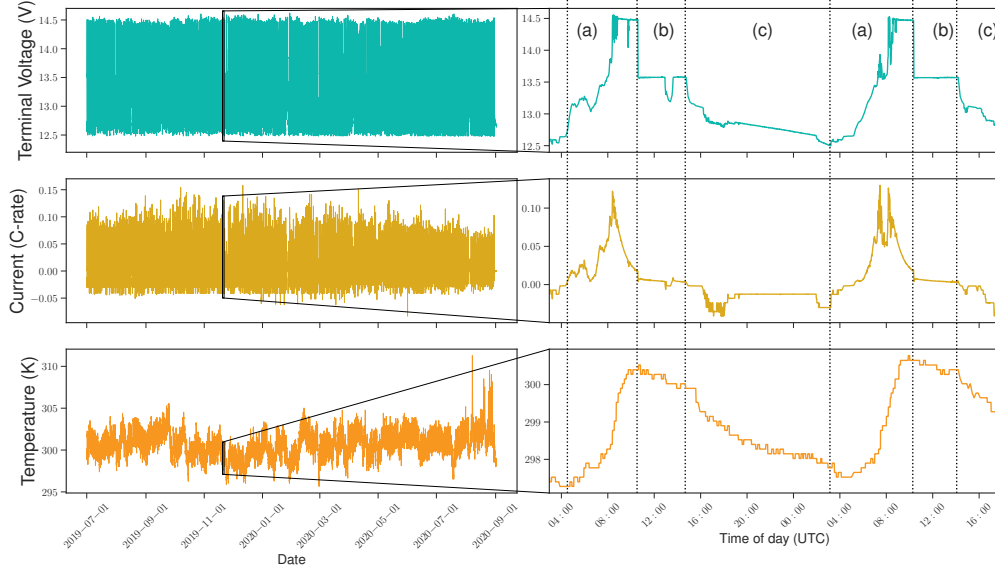


Figure 1: Solar off-grid systems exhibit variable temperatures and non-constant charging, with diurnal usage patterns including regions of (a) solar charge, (b) float charge, (c) discharge. Example data from a single battery system.

8. Batteries that entered repair were labelled ‘faulty’, and their timeseries were truncated to only include data up to the repair date.

Using the steps outlined above, we obtained a set of 1027 batteries with 491 faulty batteries and 536 healthy ones. The method above was designed to yield a roughly 50/50 split of failed/healthy batteries which provides good performance from the GP classifier. This ratio does not necessarily reflect real-world conditions, where battery failures in well-built systems may be rare. To illustrate the impact of this, we assessed the efficacy of our classifier in test cases where there are fewer failed batteries, as detailed in section 6 of the main text. The dataset consists of 620 million rows of telemetry data, each row including a timestamp, voltage, current and temperature and the arrival date at the repair shop for each battery where applicable. The total uncompressed size of the dataset (csv format) is 49 GB. The distributions of voltage, current and temperature in the raw data are shown in Fig. 2 in the main text. The distributions of current and voltage reflect the diurnal charge/discharge pattern we see in Fig. 1. During charging, voltage has an upper limit of approximately 14.5 V and once the current drops below a given threshold, float-charge starts at a voltage or roughly 13.5 V (this voltage has a temperature dependency). In general, it can be seen that the average currents are small (the nominal capacity is 20 Ah, making the 99th percentile C-rate approximately 0.09C). Due to the variability of incident sunlight on the solar panel attached to the battery, there is more variation in the charging current. Charging currents are also higher on average. Temperatures also have a diurnal cycle, with temperatures increasing during the charge phase and decreasing with discharge, partially explained by the ambient variation.

Fig. 2 in turn shows the distribution of battery ages in our dataset, where distribution of ages of healthy batteries reflects that of the overall population, as healthy batteries were sampled using stratification with respect to battery age. Also shown in Fig. 2 is the distribution of the number of qualifying charging segments for each battery in the population, obtained by applying the methods described in the main text section 7.

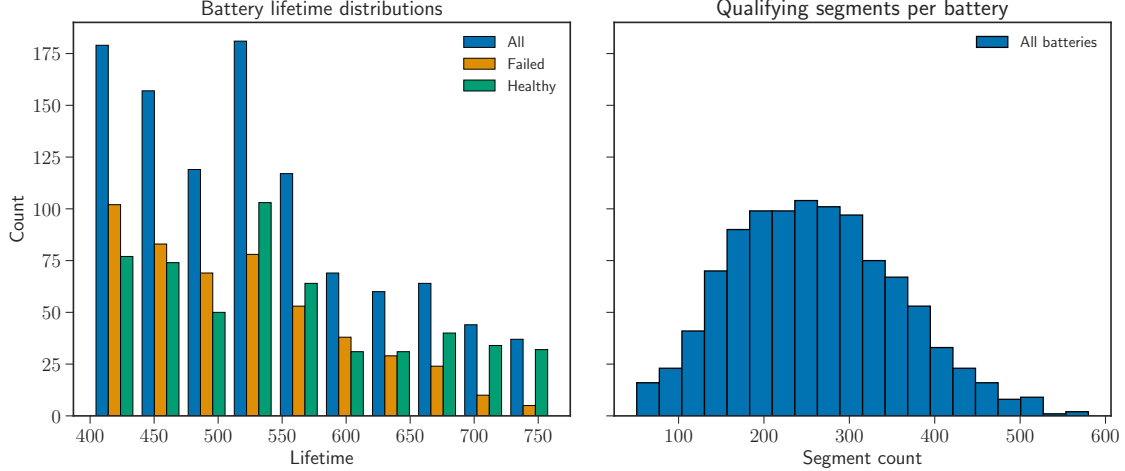


Figure 2: Histograms indicating the age distributions of the dataset and healthy and failed (repaired) subsets (left). Distribution of number of qualifying charging segments in the dataset (right).

2 Determination of open circuit voltage and electrolyte volume

For the Coulomb counting approach for state of charge estimation, described in the Methods section, we required an estimate for the electrolyte volume for the batteries in our dataset. This was obtained by a least squares fit of the OCV curve for lead acid cells from Bode [1] against experimentally measured lab OCV data. First, we carried out a GITT procedure with twenty 1 hour C/20 discharge segments, each with a 2 hour rest in-between them, at 25 °C, to retrieve pointwise estimates of the open circuit voltage as a function of capacity (see Fig. 3).

To determine the electrolyte volume, we note that the change in electrolyte concentration over discharge is given by

$$\Delta c_{HSO_4^-} = \frac{I \Delta t}{V_{elec} F},$$

where V_{elec} is the electrolyte volume and F is Faraday's constant. Bode expresses the OCV of a single cell as a function of acid molality m ,

$$V_{cell}(m) = 1.922 + 0.148 \log_{10}(m) + 0.064 \log_{10}^2(m) + 0.074 \log_{10}^3(m) + 0.034 \log_{10}^4(m), \quad (1)$$

so we made the simple conversion from molality to molarity [2],

$$m = \frac{V_w c_{HSO_4^-}}{(1 - c_{HSO_4^-} V_e) M_w},$$

where $c_{HSO_4^-}$ is acid molarity, V_w , V_e and M_w the partial molar volumes of water and acid, and the molar mass of water respectively. (Note that 12 V lead-acid batteries actually contain six cells internally in series, V_{cell} is the voltage of one of these, so $V_0 = 6 V_{cell}$.) We then were able to express our experimentally measured OCV curve (Fig. 3), which was originally a function of Coulombs, as a function of acid concentration instead. To find the best-fit value of V_{elec} we first inverted the Bode curve at a point in the middle on the range illustrated in Fig. 3, then used a least squares fit over the range shown to match the gradient, which is inversely proportional to V_{elec} . The optimal value for V_{elec} was found to be 143 cm³.

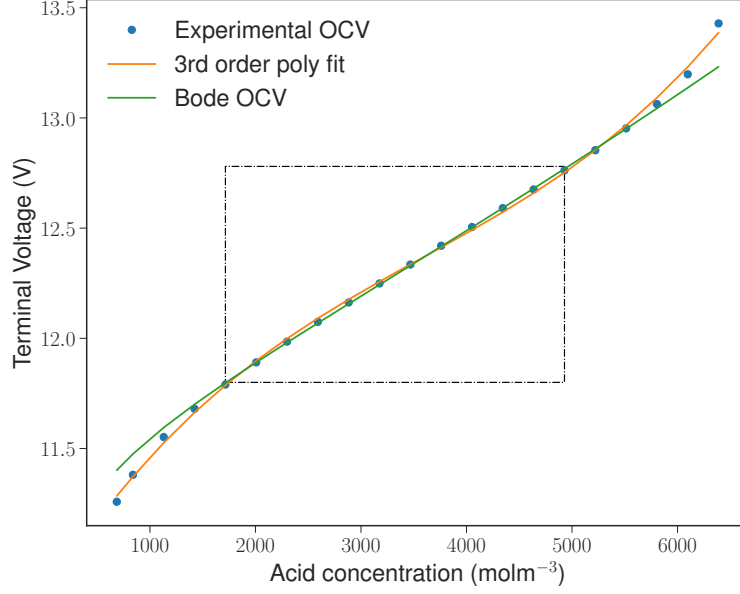


Figure 3: Experimentally determined point-wise estimates of OCV determined by GITT at 25 °C and a 3rd order polynomial fit vs. lead-acid OCV of Bode [1]. We matched the gradients by least squares over a range of acid concentrations in the range shown (central box).

For further use in modelling, we parameterized the experimental OCV obtained through GITT using a cubic polynomial as a function of estimated acid molarity. As the measured temperature range we observed in the dataset was approximately 10 K and the sensitivity of the OCV to temperature is on the order of 0.1 mV K⁻¹/cell [3], we ignored its temperature dependency.

3 Recursive estimation of the function $R_0(t, I_t, T_t, \hat{c}_t)$

To achieve linear computational scaling over n data rows when estimating the posterior distribution of R_0 for each battery, we employed the recursive framework pioneered by Särkkä et al. [4], [5]. This relies on the interpretation of a Gaussian Process as the solution of a linear stochastic differential equation, where the spectral density of the noise process is described by the covariance function of the GP in the frequency domain. As a result, the GP may be described by a linear dynamic system of the type

$$\frac{\partial R_0(t, I, T, \hat{c})}{\partial t} = F R_0(t, I, T, \hat{c}) + L w(t, I, T, \hat{c}), \quad (2)$$

where the transition matrix F , dispersion matrix L and the spectral density of the noise term $w(t, I, T, \hat{c})$ are parameterised by the hyperparameters of the GP kernel function. Here, the posterior of R_0 and the so-called energy function (the negative logarithm of the unnormalized posterior probability of the hyperparameters) may be obtained by employing standard Kalman filtering and smoothing techniques [6]. Fig. 4 illustrates the concept.

Recall that the kernel function of the GP is given by

$$k(I, T, \hat{c}, t) = \sigma_{f,0}^2 \left(\frac{\min^3(t, t')}{3} + |t - t'| \frac{\min^2(t, t')}{2} \right) + \sigma_{f,1}^2 \exp \left(\sum_{x \in \{I, T, \hat{c}\}} -\frac{(x - x')^2}{2l_x^2} \right). \quad (3)$$

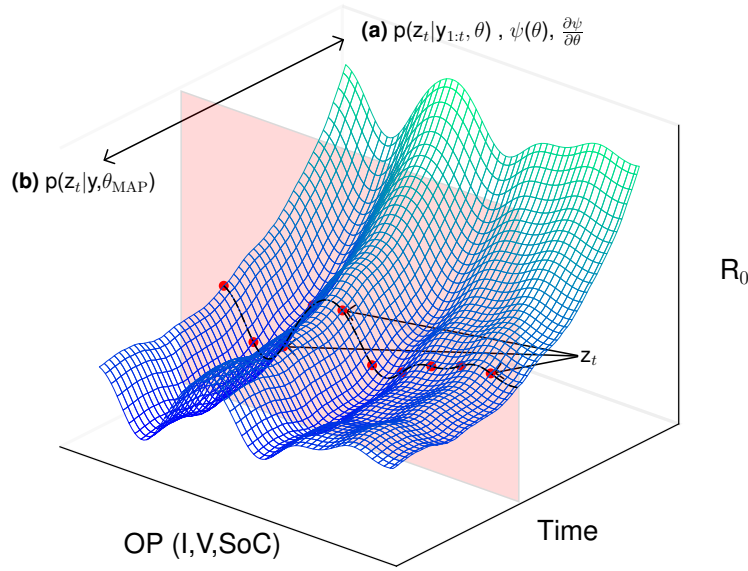


Figure 4: Fitting a GP recursively allows for linear scaling in computational effort in the time axis. State vector z_t represents the GP value at time t at 20 specific operating points, unique to each battery, which were chosen using the k-means algorithm. These 20 points were propagated and updated using a Kalman filter. A gradient based optimizer finds MAP estimates of the hyperparameters by repeatedly running the model ‘forwards’, (a), to evaluate the energy function and its gradient with respect to hyperparameters. The final posterior distribution is smoothed with a backward pass, (b).

As shown by Solin [7], the solution to the stochastic equation 2 (i.e. the mean and covariance of the solution trajectories) is equivalent to the posterior distribution of the Gaussian process. We find this using a (discrete time) Kalman filter, propagating the mean and covariance of the linear dynamic system using

$$z_{k+1} = A_{GP} z_k \quad (4a)$$

$$P_{k+1} = A_{GP} P_k A_{GP}^T + Q_{GP}, \quad (4b)$$

with state vector z being a concatenation of the two processes,

$$z = [z_{GP,0} \ z_{GP,1}]^T, \quad (5)$$

where $z_{GP,0}$ represents the degradation GP and consists of the vector $[R_t, \partial R_t / \partial t]^T$ and $z_{GP,1}$ represents the GP over operating points, which is constant over time. Equation (2) is infinite-dimensional, so we use a finite dimensional representation over the operating points by choosing, via k-means clustering, 20 inducing points over $(I, T, \hat{c})$ for each battery, making $z_{GP,1} \in \mathbb{R}^{20}$. Given the state vector, the discrete time transition matrix A_{GP} is block diagonal,

$$A_{GP} = \expm \left(\begin{bmatrix} F_{GP,0} & 0 \\ 0 & F_{GP,1} \end{bmatrix} \Delta t \right), \text{ where } F_{GP,0} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, F_{GP,1} = \mathbf{0}, \quad (6)$$

where $\mathbf{0}$ indicates a square matrix of zeros of size 20 and Δt denotes the (normalised) time-step between charging segments. The process covariance Q_{GP} is also block-diagonal, given by

$$Q_{GP} = \begin{bmatrix} Q_{GP,0} & \mathbf{0} \\ \mathbf{0} & Q_{GP,1} \end{bmatrix}, \text{ where } Q_{GP,0} = \begin{bmatrix} \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & \Delta t \end{bmatrix}, \quad Q_{GP,1} = \mathbf{0} \quad (7)$$

As we use a zero-mean GP, the state vector z_{GP} was initialized at zero. The covariance matrix for the dynamic system is likewise block-diagonal,

$$P_{GP} = \begin{bmatrix} P_{GP,0} & \mathbf{0} \\ \mathbf{0} & P_{GP,1} \end{bmatrix} \quad (8)$$

initialised separately for the two subsystems. The integrated Wiener process is defined to have zero initial covariance [7], so that $P_{GP,0} = \mathbf{0}$. For the second GP over operating points, which has no time dependency, the initial covariance is given by $P_{GP,1} = \sigma_{t,1}^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of size 20 due to the number of inducing points chosen.

We can, equivalently, express the propagation of the states and covariances in the form of a joint distribution of previous posterior and current prior states, z_{t-1}^+ and z_t^- respectively, given by [5]

$$p\left(\begin{bmatrix} z_{t-1}^+ \\ z_t^- \end{bmatrix}\right) \sim \mathcal{N}\left(\begin{bmatrix} m_{t-1}^+ \\ A_{GP} m_t^- \end{bmatrix}, \begin{bmatrix} P_{t-1}^+ & P_{t-1}^+ A_{GP}^T \\ A_{GP} P_{t-1}^+ & A_{GP} P_{t-1}^+ A_{GP}^T + Q_{GP} \end{bmatrix}\right), \quad (9)$$

where m_t^- and m_{t-1}^+ denote the current prior and previous posterior means. The observation model contains the GP describing the dependency on operating point. This means that the estimated internal resistance at any point in time may be expressed as a linear combination of the elements within state vector z_t . The resulting observation matrix $H_{GP,t}$ is determined by the Gaussian process over operating points. Summarising, observation model is given by the following distributions [8],

$$z_t \sim \mathcal{N}(m_t^-, P_t^-) \quad (10a)$$

$$R_{t,0}|z_t \sim \mathcal{N}(H_{GP,t} z_t^-, \sigma_{GP,t}^2) \quad (10b)$$

$$V_t|R_{0,t} \sim \mathcal{N}(V_0(\hat{\alpha}) + R_{0,t} l_t, \sigma_{n,t}^2). \quad (10c)$$

This gives the joint distribution

$$p\left(\begin{bmatrix} z_{t-1} \\ R_{t,0} \\ V_t \end{bmatrix}\right) \sim \mathcal{N}(m', P'), \text{ where} \quad (11a)$$

$$m' = \begin{bmatrix} m_t^- \\ H_{GP} m_t^- \\ V_0(\hat{\alpha}) + H_{GP} m_t^- l_t \end{bmatrix}, \quad (11b)$$

$$P' = \begin{bmatrix} P_t^- & P_t^- H_{GP,t}^T & P_t^- H_{GP,t} l_{d,t}^T \\ H_{GP,t} P_t^- & H_{GP,t} P_t^- H_{GP,t}^T + \sigma_{GP,t}^2 & [H_{GP,t} P_t^- H_{GP,t}^T + \sigma_{GP,t}^2] l_{d,t}^T \\ l_{d,t} H_{GP,t} P_t^- & l_{d,t} [H_{GP,t} P_t^- H_{GP,t}^T + \sigma_{GP,t}^2] & \mathbf{I}_{d,t} [H_{GP,t} P_t^- H_{GP,t}^T + \sigma_{GP,t}^2] \mathbf{I}_{d,t}^T + \sigma_{n,t}^2 \end{bmatrix}. \quad (11c)$$

The innovation covariance of the Kalman filter is the last block element on the leading diagonal (highlighted in bold). The matrix $\mathbf{I}_{d,t}$ is a diagonal matrix consisting of the current at each point of

the charging segment. The matrix $H_{\text{GP},t} \in \mathbb{R}^{n \times 22}$, n being the number of data points in the charging segment, is given by

$$H_{\text{GP},t} = [H_0 \quad H_1] = \begin{bmatrix} 1 & 0 & k(x_t, X_u) \left[k(X_u, X_u) + \text{diag}[P_{\text{GP},1,t}^-] \right]^{-1} \\ \vdots & \vdots & \vdots \end{bmatrix}, \quad (12)$$

where $k(\cdot)$ is the SE kernel covariance function in (3), $x_t = [I_t \ T_t \ \hat{c}_t]$, $X_u \in \mathbb{R}^{20 \times 3}$ is the location of the inducing points in terms of $(I_t, T_t, \hat{c}_t)$ and $\text{diag}[\cdot]$ denotes the operator returning the leading diagonal of a matrix. Note, we add the diagonal of the prior state covariance matrix estimate $P_{\text{GP},1,t}^-$ to $k(X_u, X_u)$ in (12) to reflect the uncertainty in the states as they are propagated through time. The uncertainty around the points can vary significantly due to the observed $(I_t, T_t, \hat{c}_t)$ changing from one segment to the next. The variance in the GP is given by

$$\sigma_{\text{GP},t}^2 = \text{diag} \left[k(x_t, x_t) - k(x_t, X_u) \left[k(X_u, X_u) + \text{diag} [P_{\text{GP},1,t}^-] \right]^{-1} k(X_u, x_t) \right], \quad (13)$$

and the calculation of $\sigma_{n,t}^2$, which takes into account the variance in $V_0(\hat{c}_t)$, is described in the Methods section. To calculate maximum-a-posteriori (MAP) estimates of the hyperparameters, we employ the standard approach via Bayes' theorem, noting that the denominator is independent of the hyperparameters,

$$p(\theta_h|D) = \frac{p(D|\theta_h)p(\theta_h)}{p(D)} \propto p(D|\theta_h)p(\theta_h). \quad (14)$$

where $\theta_h = [\sigma_{f,0} \ \sigma_{f,1} \ l_x]$ is a vector of hyperparameters and D represents data, consisting of the input-output pairs through each charging segment (current, state of charge, temperature being the inputs, terminal voltage the output).

For each pass through the data, pointwise estimates of the posterior probability $p(\theta_h|D)$ are given by the energy function, which can be included as part of the recursive Kalman filter approach. The energy function is [5]

$$\Phi(\theta_h) = -\log p(D|\theta) - \log p(\theta) = -\log p(\theta_h) + \frac{1}{2} \sum_t \log |2\pi S_t(\theta_h)| + \frac{1}{2} \sum_t e_t^T S_t(\theta_h)^{-1} e_t, \quad (15)$$

where $p(\theta_h)$ is the hyperprior specified in the Methods section, e_t the error vector for each charging segment, and $S_t(\theta_h)$ the innovation covariance for the segment, given by the lower right diagonal element in equation 11c. This was minimised to find the MAP estimate of θ_h using the L-BFGS-B algorithm [9], implemented in SciPy. To improve efficiency, the Jacobian of the energy function was determined analytically by extending the recursive method outlined by Mbalawata et al. [10] to include terms for the observation model, equations 10-13. Differentiating the energy function with respect to θ_h and dropping the dependency on θ_h in notation for convenience, we get

$$\begin{aligned} \frac{\partial \Phi}{\partial \theta_h} = & -\frac{\partial \log p(\theta_h)}{\partial \theta_h} + \frac{1}{2} \sum_n \text{Tr} \left(S_t^{-1} \frac{\partial S_t}{\partial \theta_h} \right) \\ & - \frac{1}{2} \sum_n \left(H_t \frac{\partial z_t^-}{\partial \theta_h} + \frac{\partial H_t}{\partial \theta_h} z_t^- \right)^T S_t^{-1} e_t \\ & - \frac{1}{2} \sum_n e_t^T S_t^{-1} \frac{\partial S_t}{\partial \theta_h} S_t^{-1} \\ & - \frac{1}{2} \sum_n e_t^T S_t^{-1} \left(H_t \frac{\partial z_t^-}{\partial \theta_h} + \frac{\partial H_t}{\partial \theta_h} z_t^- \right), \end{aligned} \quad (16)$$

where Tr denotes the matrix trace and

$$\frac{\partial S_t}{\partial \theta_h} = \frac{\partial H_t}{\partial \theta_h} P_t^- H_t^T + H_t \frac{\partial P_t^-}{\partial \theta_h} H_t^T + H_t P_t^- \left(\frac{\partial H_t}{\partial \theta_h} \right)^T + \frac{\partial \sigma_{r,t}^2}{\partial \theta_h}. \quad (17)$$

Given equations 16–17, the standard Kalman filter states can be augmented with their partial derivatives with respect to the hyperparameters, $\partial z_t / \partial \theta_h$, $\partial P_t^- / \partial \theta_h$, and propagated as described by Mbalawata et al. [10], using matrix fraction decomposition for the propagation of the augmented covariance. The downside of this approach is the repeated calculation over discharge segments of the large exponential of the matrix $M_t \in \mathbb{R}^{88 \times 88}$,

$$\text{expm}(M_t) = \text{expm} \left(\begin{bmatrix} F(\theta_h) & \Sigma(\theta_h) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -F^T(\theta_h) & \mathbf{0} & \mathbf{0} \\ G(\theta_h) & W(\theta_h) & F(\theta_h) & \Sigma(\theta_h) \\ \mathbf{0} & -G^T(\theta_h) & \mathbf{0} & -F^T(\theta_h) \end{bmatrix} \Delta t \right), \quad (18)$$

where $F(\theta_h)$ is the continuous time transition matrix of the state space system as defined by (6), $G(\theta_h) = \partial F(\theta_h) / \partial \theta_h$, and $W(\theta_h) = \partial \Sigma(\theta_h) / \partial \theta_h$, where Σ is the continuous time diffusion matrix of the GP. However, the additive kernel (3) has the convenient property that the matrix exponential (18) has an exact solution for each element in the hyperparameter vector θ_h , foregoing the need to evaluate the matrix exponential using standard numerical methods, therefore requiring a fraction of the computational cost. As a result, the dominant cost in the Kalman filter recursion becomes the inversion of the innovation covariance, S_t , which is of the order 100×100 for each discharge segment. This is already incorporated in the standard recursion, making the analytical method for evaluating the energy function gradient approximately 30% faster (and more stable) than the numerical equivalent.

Given the MAP estimates of the hyperparameters, the posterior of $R_0(x)$ was then obtained by using the RTS smoother [11]. This posterior could be evaluated to extrapolate the value of R_0 to any given operating point by using

$$R_0(x) \sim \mathcal{N}(\mu_p, \sigma_p^2) \quad (19a)$$

$$\mu = k(x, X_u) [k(X_u, X_u) + \text{diag}[P_{\text{GP},1}^s]]^{-1} z_{\text{GP},1}^s, \quad (19b)$$

$$\sigma_p^2 = k(x, x) - k(x, X_u) [k(X_u, X_u) + \text{diag}[P_{\text{GP},1}^s]]^{-1} k(X_u, x) \quad (19c)$$

where $z_{\text{GP},1}^s$ and $P_{\text{GP},1}^s$ denote the RTS smoothed mean and covariances, which are **constant** as a function of time for the GP over the operating point. For the degradation process, the value of $z_{\text{GP},0}$ and $P_{\text{GP},0}$ can be obtained by applying the forward propagating equations (9) using as a starting point the RTS smoothed values.

The method used for benchmarking our approach is similar to this, but because R_0 was considered to be independent of the operating point and modelled by a random walk through time, the state vector z_t , variance P_t and process variance Q become scalar, and transition matrix A_{GP} is simply I (in fact 1, since in this case the state is a scalar). The vector valued observations are the same and the measurement error was the fixed value $\mathbf{I}\sigma_{n,t}^2$, as in the main approach. The only hyperparameter Q was estimated by maximum likelihood using the same recursion technique as the main method. The initial variance of the state vector was set to $10Q$ in each case.

4 Estimated hyperparameters of R_0

We calculated the MAP estimates of the hyperparameters for the kernel function, equation 3, for each battery in our dataset using a parallel approach. Hyperparameters were estimated once using

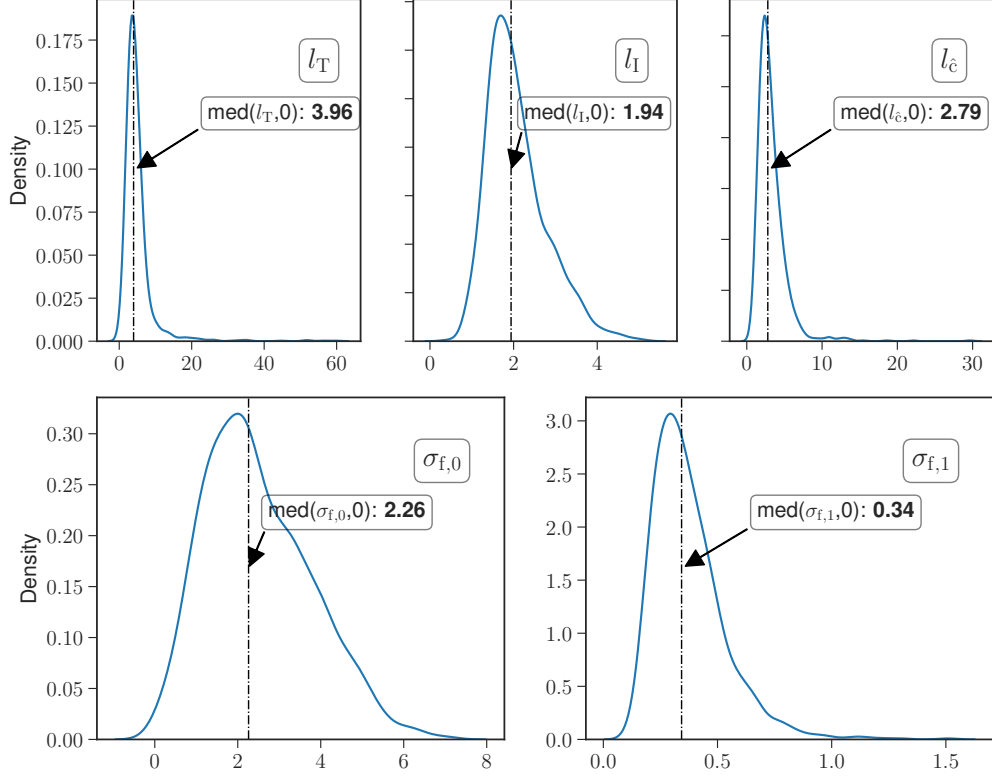


Figure 5: Distributions of each hyperparameter across the population. The median is shown for each hyperparameter.

the ‘zero’ time horizon dataset (i.e. the longest dataset). These hyperparameters were then used to estimate the posterior distributions in all cases. Fig. 5 shows the distribution of the hyperparameter MAP estimates and their median estimates in each case. The length scale l_I and magnitudes σ have relatively tight distributions, whereas the temperature and SoC length scales have some outliers. Also, the median length scales indicate that the applied current is the most significant input, which is consistent with the results showing the largest range in estimated R_0 across this input dimension.

5 Calculation of stress factors for GP classifier

The augmented input matrix for the GP classifier for fault diagnosis/prognosis consisted of R_0 , $\partial R_0 / \partial t$ and cumulative stress factors known to affect battery state of health. Firstly, battery age was defined as the time (in days) since the unit was marked as *activated* by BBOX. Cumulative time at float charge was calculated by simply summing time increments conditional on the battery being in the float charge region, so that

$$\Sigma_S = \sum_t S(t) \Delta t \quad \text{where } S = \begin{cases} 1 & \text{if in float charge} \\ 0 & \text{otherwise} \end{cases}, \quad (20)$$

where we define the float charge region as a period of over 600 seconds where the voltage is between 13.4 V and 13.6 V. We also calculated the total Ah throughput by taking the absolute

cumulative Coulomb count through battery lifetime, and the mean temperatures and voltages by using an expanding window from beginning of life. Finally, we calculated the discharge cycle count by adding up the discharge segments defined by any continuous discharge period of over 600 s where $I > 0.05$ A.

To investigate how many stress factors to include, we calculated the correlation between the health estimates R_0 and $\partial R_0 / \partial t$ and the stress factors introduced above. Fig. 6 shows the resulting Pearson correlation coefficients calculated at the end of the data series for each battery, for the entire population. Although the correlation coefficient does not capture non-linear interactions between features, it is clear that the health metrics R_0 and $\partial R_0 / \partial t$ are strongly related, which implies the existence of a degradation knee-point in many of the SoH trajectories. Other factors are generally not strongly correlated, indicating that a wide range of different usage conditions are explored in the data. This is an important advantage of learning battery degradation behaviour from large field datasets—creating a comparably wide variety of operating conditions in a laboratory would be prohibitively expensive and time-consuming. To assess which factors might be more important in causing end of life, we can compare their relative importance for fault prediction, as discussed in section 5 in the main text.

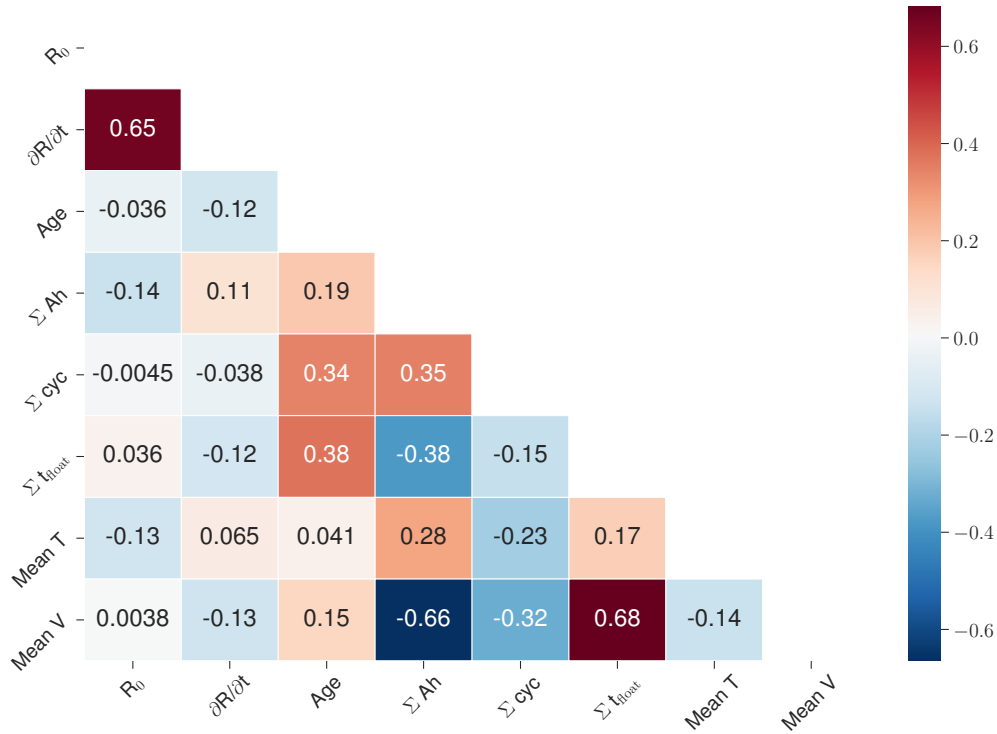


Figure 6: Battery SoH, measured by internal resistance, complements other aging factors, such as calendar age, as shown in this correlation matrix for health metrics R_0 , $\partial R_0 / \partial t$ and stress factors associated with battery degradation. Σ denotes the cumulative sum up to the point of diagnosis.

6 GP Classifier test set and performance matrix

As discussed in Methods, training and testing of the GP classifier was done by nested cross-validation. First, we used 5-fold stratified cross validation to split the ‘master’ dataset of 1027

Days before	Classifier input	TP	FP	FN	TN	Balanced Accuracy (%)
0	$R, \frac{\partial R}{\partial t}$	330	78	161	458	76
0	$R, \frac{\partial R}{\partial t}, \text{stress factors}$	389	78	102	458	82
0	Benchmark	244	95	247	441	66
0	Stress factors only	314	119	177	417	71
14	$R, \frac{\partial R}{\partial t}$	299	85	192	451	73
14	$R, \frac{\partial R}{\partial t}, \text{stress factors}$	378	95	113	441	80
14	Benchmark	203	91	288	445	62
14	Stress factors only	331	142	160	394	70
28	Stress factors only	349	161	142	375	71
28	Benchmark	175	123	316	413	56
28	$R, \frac{\partial R}{\partial t}, \text{stress factors}$	373	124	118	412	76
28	$R, \frac{\partial R}{\partial t}$	264	110	227	426	67
42	$R, \frac{\partial R}{\partial t}$	245	116	246	420	64
42	$R, \frac{\partial R}{\partial t}, \text{stress factors}$	373	132	118	404	76
42	Benchmark	123	119	368	417	51
42	Stress factors only	358	191	133	345	69
56	$R, \frac{\partial R}{\partial t}$	237	120	254	416	63
56	$R, \frac{\partial R}{\partial t}, \text{stress factors}$	363	151	128	385	73
56	Benchmark	107	117	384	419	50
56	Stress factors only	362	221	129	315	66

Table 1: Performance of GP classifier over all input sets and time horizons for the case where failed batteries constitute 48% of the population.

batteries with a 48/52% failed/healthy split in to training in test sets. In addition we created extra testing subsets within each test set, which had different proportions of failed vs. healthy batteries. These were created by taking 10 random samples of 40% , 60% and 80% of the existing failed batteries in the test set while keeping the healthy set fixed, making the sub-test sets smaller in size. Altogether we ended up with 3100 tests cases, where

$$3100 = (4 \text{ input sets}) * (5 \text{ CV sets}) * (5 \text{ time horizons}) * (10 * 3 \text{ random subsets} + 1 \text{ complete test set})$$

The main text summarises the performance by time horizon. The full performance data is shown in Table 1. In each case we show the number of true positives, false positives, false negatives and true negatives (TP, FP, FN, TN respectively) from the classifier predictions in the test set with 52/48% healthy/failed split using 5-fold cross-validation. The balanced accuracy is the average of sensitivity and specificity for each case (see Methods section in main paper). It is clear that the combined model including the inputs $R, \partial R / \partial t$ as well as the stress factors has the strongest performance in predicting faults.

7 Comparison to benchmark

Fig. 7 shows the comparison between R_0 timelines calculated using the method proposed in this paper vs. using the benchmark method (e.g. [12]), as discussed above and in the main text. Normalising for operating conditions and using a smooth GP kernel results in much smoother health trajectories over the lifetime of each battery, giving improved fault classification results.

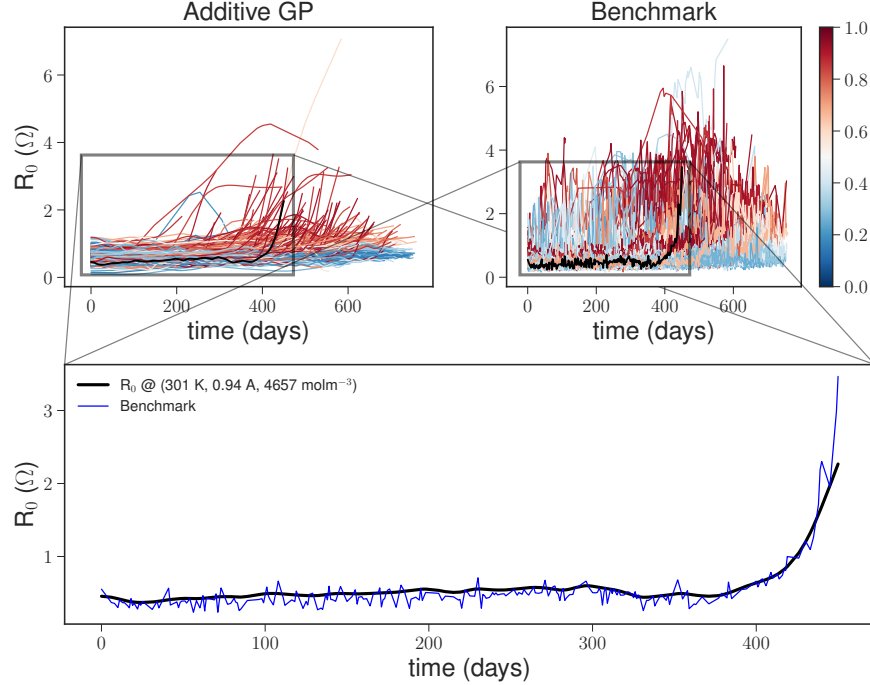


Figure 7: Estimated R_0 trajectories using our methods compared with the benchmark approach, illustrated for the population and for a single battery. The GP method results in much smoother estimates over the lifetime of each battery, by accounting for variations due to operating conditions and imposing smoothness by choice of kernel and prior distribution of hyperparameters.

8 Data quality sensitivity analysis

The distributions of sampling times and the number of qualifying segments per battery in the raw data are illustrated in Fig. 2 in the main text and Fig. 2. in the supplementary material respectively. In order to gauge the dependency of our results on data quality, we performed a sensitivity analysis by calculating the average classification performance while simulating deterioration in data availability. This was done in two stages. Firstly, we varied the interpolated sampling time within each charging segment in the downsampled set, comparing the baseline case of 1 minute interpolation to 2, 5 and 10 minutes. Secondly, we varied the number of segments available by comparing the case using all available segments obtained by filtering (see Methods) to choosing a maximum of 100 and 50 segments for each battery, with a 1-minute interpolated sampling time. Table 2 gives the average classification performance over all test cases. Varying the interpolated sampling time has very limited effect on performance up to a 10-minute window. Reducing the number of available segments (on average the dataset has 260 charging segments per battery) has a slight negative effect on the performance of the model—the performance of the strongest classifier, including both our health metric and stress factors, is reduced by 2% in terms of balanced accuracy when reducing the number of segments from an average of 260 to 50.

References

- [1] H. Bode, *Lead-acid batteries*. Wiley, 1977.
- [2] V. Sulzer, S. J. Chapman, C. P. Please, D. A. Howey, and C. W. Monroe, “Faster Lead-Acid Battery Simulations from Porous-Electrode Theory: Part I. Physical Model,” *Journal of The Electrochemical Society*, vol. 166, no. 12, A2363–A2371, Jul. 2019, ISSN: 0013-4651. DOI: 10.1149/2.0301910jes. [Online]. Available: <http://jes.ecsdl.org/lookup/doi/10.1149/2.0301910jes>.
- [3] T. Crompton, *Battery reference book*, Second edi. Butterworth-Heinemann, 1995. DOI: 10.1016/0140-6701(96)88678-4.
- [4] S. Särkkä and J. Hartikainen, “Infinite-dimensional kalman filtering approach to spatio-temporal Gaussian process regression,” *Journal of Machine Learning Research*, vol. 22, pp. 993–1001, 2012, ISSN: 15337928.
- [5] S. Särkkä, A. Solin, and J. Hartikainen, “Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering,” *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, 2013, ISSN: 10535888. DOI: 10.1109/MSP.2013.2246292.
- [6] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [7] Arno Solin, “Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression,” Ph.D. dissertation, Aalto University, 2016. [Online]. Available: <https://aaltodoc.aalto.fi/bitstream/handle/123456789/19842/isbn9789526067117.pdf?sequence=1&isAllowed=y>.
- [8] S. Särkkä, A. Solin, and J. Hartikainen, “Sparse Spatio-temporal Gaussian Processes with General Likelihoods,” in *Proceedings of the 21th international conference on Artificial neural networks - Volume Part I*, 2011, pp. 193–200.
- [9] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A Limited Memory Algorithm for Bound Constrained Optimization,” *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995. DOI: 10.1137/0916069. [Online]. Available: <https://doi.org/10.1137/0916069>.
- [10] I. S. Mbalawata, S. Särkkä, and H. Haario, “Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering,” *Computational Statistics*, vol. 28, no. 3, pp. 1195–1223, 2013, ISSN: 09434062. DOI: 10.1007/s00180-012-0352-y.
- [11] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, 1965, ISSN: 00011452. DOI: 10.2514/3.3166.
- [12] G. L. Plett, “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*, vol. 161, pp. 1369–1384, 2006. DOI: 10.1016/j.jpowsour.2006.06.004. [Online]. Available: https://ac.els-cdn.com/S0378775306011438/1-s2.0-S0378775306011438-main.pdf?_tid=405e013f-8b51-4005-8034-e2f3fd93939b&acdnat=1551435649_2f4f89f933a745077c0cb012de28c1c9.

Test case	Sampling (min)	Segments	Classifier mean balanced accuracy (%)			
			Benchmark	$R, \frac{\partial R}{\partial t}$	$R, \frac{\partial R}{\partial t}, \text{stress factors}$	Stress factors
baseline	1	~ 260	57	69	77	69
100 segments	1	100	58	67	76	69
50 segments	1	50	57	66	75	69
2 minute	2	~ 260	58	71	78	69
5 minute	5	~ 260	58	70	78	70
10 minute	10	~ 260	58	70	78	70

Table 2: Average classification accuracy across all test cases for varying interpolated sampling times and number of available charging segments.