Proceedings of the ASME 2017 Dynamic Systems and Control Conference
DSCC2017
October 11-13, 2017, Tysons, Virginia, USA

DSCC2017-5365

BATTERY CAPACITY ESTIMATION FROM PARTIAL-CHARGING DATA USING GAUSSIAN PROCESS REGRESSION

Robert R. Richardson, Christoph R. Birkl, Michael A. Osborne and David A. Howey*
Department of Engineering Science
University of Oxford
Oxford, UK OX1 3PJ
Email: robert.richardson@eng.ox.ac.uk, christoph.birkl@eng.ox.ac.uk,
mosb@robots.ox.ac.uk and david.howey@eng.ox.ac.uk

ABSTRACT

Accurate on-board capacity estimation is of critical importance in lithium-ion battery applications. Battery charging/discharging often occurs under a constant current load, and hence voltage vs. time measurements under this condition may be accessible in practice. This paper presents a novel diagnostic technique, Gaussian Process regression for In-situ Capacity Estimation (GP-ICE), which is capable of estimating the battery capacity using voltage vs. time measurements over short periods of galvanostatic operation.

The approach uses Gaussian process regression to map from voltage values at a selection of uniformly distributed times, to cell capacity. Unlike previous works, GP-ICE does not rely on interpreting the voltage-time data through the lens of Incremental Capacity (IC) or Differential Voltage (DV) analysis. This overcomes both the need to differentiate the voltage-time data (a process which amplifies measurement noise), and the requirement that the range of voltage measurements encompasses the peaks in the IC/DV curves. Rather, GP-ICE gives insight into which portions of the voltage range are most informative about the capacity for a particular cell. We apply GP-ICE to a dataset of 8 cells, which were aged by repeated application of an ARTEMIS urban drive cycle. Within certain voltage ranges, as little as 10 seconds of charge data is sufficient to enable capacity estimates with $\sim 2\%$ RMSE.

*Address all correspondence to this author.

INTRODUCTION

Lithium ion (Li-ion) batteries experience capacity fade during use through a complex interplay of physical and chemical processes [1–3]. Knowledge of the present battery capacity is necessary to ensure reliable operation and facilitate corrective action when appropriate. To mitigate uncertainty in capacity, batteries are often over-sized and under-used, which results in increased system costs and reduced performance. Therefore, accurate online capacity estimation is a desirable function for battery management systems.

There are several different approaches to capacity estimation [4, 5]. The simplest approach is direct measurement: discharging the battery from a fully charged state until the lower cut-off voltage is reached. However, this technique is difficult or impossible to apply online in most practical applications. The most common approaches in practice involve parameter estimation of battery equivalent circuit models [6–9] or electrochemical models [10–13]. These approaches have been successfully applied in many studies; however, they all require the provision of an accurate battery model. Moreover, for high fidelity models, parameter identifiability can be a major challenge [14].

Incremental capacity (IC) and differential voltage (DV) analysis have also been used for capacity estimation. These techniques have conventionally been used for detailed cell analyses, such as understanding degradation mechanisms [15, 16], although recent studies have considered the use of portions of the IC/DV curve for online capacity estimation [17–20]. In particular, Berecibar et al. [20] demonstrated cell capacity estima-

tion using a selection of features of IC/DV curves as inputs. They demonstrated their approach using three different regression techniques: Linear Regression, Multilayer Perceptrons and Support Vector Machines, with the latter two methods showing best results. Although their approach showed good performance, the use of features derived from IC/DV curves as inputs to a regression problem has a number of drawbacks. Firstly, differentiating the voltage-time data amplifies the noise in the measurement, even when sophisticated ng algorithms are employed. In particular, the magnitude of the peaks were found to be especially sensitive to noise. Hence, this induces a loss of accuracy in the subsequent regression problem since the inputs are derived from the differentiated data. Secondly, since the inputs are the values and locations of the peaks, the voltage range must encompass the voltages at which these peaks occur. In some cases, one of these peaks may be located at a high SoC and another at a low SoC, and hence to identify all the inputs would require covering a large voltage range, and a long measurement duration. Lastly, the selection of the features is a cumbersome pre-processing step, since these are likely to vary between cells of different chemistries. Moreover, for a given chemistry, new peaks can develop and old ones disappear as the cell ages. Interpreting these peaks – determining how many there are, and the voltage ranges in which they occur – is a non-trivial task.

The present work introduces Gaussian Process regression for In-situ Capacity Estimation (GP-ICE) to address these issues. GP-ICE dispenses with the interpretation of the voltage data as IC or DV curves and instead operates directly on the voltage vs. time data itself. This is achieved by first smoothing the voltage curve using a Savitzky-Golay (SG) filter[1] [21], and then using the time values at equispaced voltages as the inputs to the regression problem. Furthermore, GP-ICE uses Gaussian processes (GPs) [22] rather than SVMs or neural networks for the regression step. GPs have previously been used in relatively few studies on battery diagnostics/prognostics [23–25], however, they possess a variety of desirable attributes. Firstly, they are non-parametric[2], and hence permit a model expressivity (e.g a number of parameters) that is naturally calibrated to the requirements of the data. Secondly, GPs are a Bayesian method, and hence handle uncertainty in a principled manner. An important aspect of diagnostics is not only estimating the capacity values but also expressing the uncertainty associated with these estimates. Bayesian methods provide a principled approach to dealing with uncertainty, giving rise to credible intervals comprising probabilistic upper and lower bounds, which are essential for making informed decisions. An advantage of this approach is that highly uncertain measurements (which would otherwise degrade the performance of our model) may effectively be neglected, as the results of this paper demonstrate.

The remainder of this article is organised as follows. In the following section we describe the GP-ICE approach, whilst the details of Gaussian process regression are provided in the appendix at the end of the paper. In the subsequent sections, we give details of the experiments conducted to obtain aging data, and then present and analyse the results of our method applied to these data. The final section discusses the practical applicability of the method, and elaborates on its advantages and disadvantages relative to other approaches.

METHOD

Overview

An overview of the general methodology is given below. The process is also depicted in Fig. 1. For simplicity, the following description assumes that charging (rather than discharging) data are used, although the procedure is equally applicable in either case.

Offline Assume we have a database of N_C cells, each of which has been cycled to varying states of health.

Inputs: At various stages throughout the life of each cell, a *full* constant-current charge cycle has been applied at a fixed pre-specified current and a fixed pre-specified ambient temperature, and the voltage vs. time data from this cycle are recorded. From here on we refer to these data as a Galvanostatic Voltage (GV) curve. The GV curves can be OCV or pseudo-OCV curves, although in general this is not necessary (although in practice smaller currents are likely to give better results). The GV curve is smoothed using a Savitzky-Golay (SG) filter (or any other simple, efficient smoothing algorithm), and the V-t data at 1 s intervals are acquired; a subset of these points will be used as the input data for a single sample as discussed in the next section.

Outputs: Since a full charge/discharge cycle is applied, the capacity of the cell at this C-rate is given by $y = \int_{t_0}^{t_{\text{end}}} I \, \mathrm{d}t$. We denote this known capacity as y since it will be the target value for this GV curve in the regression step.

Summary: Each cell can have a different number of GV curves, and the order of these curves is not important. Hence, the end result is just a labelled set of training data, consisting of a large set of smoothed GV curves (a subset of which will form the inputs), and an associated set of known cell capacities (the outputs). The total number of GV curves across all cells is the sample size, N_D, of the database.

Online The procedure for estimating the cell capacity using a short online diagnostic test is described next. Assume we have a cell with an unknown capacity and unknown SoC, and we wish to estimate the capacity.

[1] Savitztky-Golay filtering is often used when differentiating noisy data; differentiation is not our objective here, however we nonetheless use this filter since it reduces measurement noise, which is advantageous in any case.

[2] Support vector machines are, like GPs, non-parametric, but they do not provide confidence estimates in their predictions.

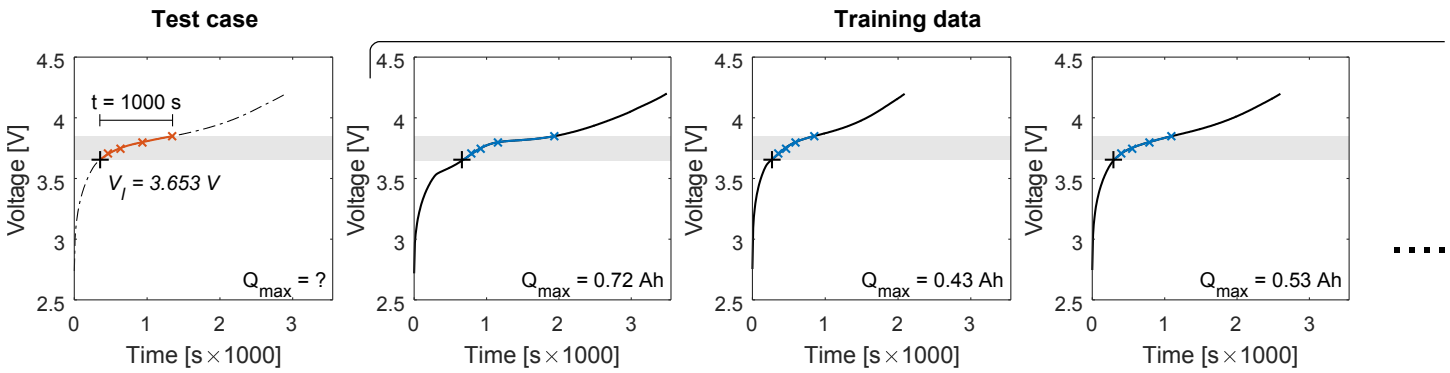

Figure 1. OVERVIEW OF THE GP-ICE METHOD. The time values at n equispaced voltage points between V_l and V_h are used as inputs to the regression model. The test inputs are shown as red crosses, and the training inputs are shown as blue crosses. The training inputs all have an associated known capacity. The figure shows just three GV curves for training, but in practice the model is trained on several hundred GV curves, obtained from multiple cells at different states of health (see Table 1).

1. Allow the cell to rest for a sufficient period to minimise electrical/thermal effects from the previous cycle. In practice the minimum length of this rest period will depend on the cell chemistry and the nature of the previous drive cycle.
2. Apply the pre-specified constant current for some duration, Δt, and measure the voltage throughout. In practice Δt would be dictated by the duration of time we can afford to take, or the duration of time a device happens to be charged for by the user. The voltage range of this test will span from some lower voltage, V_l, which is the cell voltage when the charge is first applied, to some higher voltage, V_h, which is the cell voltage at the instant the constant current is removed.
3. Smooth this voltage vs time data using an SG (or similar) smoothing filter, as before.
4. Identify the values of the time at n equispaced voltage points between V_l and V_h, and denote these values by the $n \times 1$ vector $\mathbf{x}^*$. For example, if we chose $n = 4$, and the voltage spanned from $V_l = 3.3$ V to $V_h = 3.5$ V, then $\mathbf{x}^*$ would consist of the time values at $\mathbf{V} = \{3.35, 3.40, 3.45, 3.50\}$V, i.e. $\mathbf{x}^* = \mathbf{t_V}$. We will later use $\mathbf{x}^*$ as the independent variable in the regression model[3], as shown in Fig. 1.
5. For each of the GV curves in the offline database, identify the corresponding input vectors, $\mathbf{x}$, given by the time taken to go from the lower voltage to each of the equispaced voltages, i.e. $\mathbf{x} = \mathbf{t_V} - t_{V_l}$. Since the cell capacities for each GV curve in the offline database are known, each time vector, $\mathbf{x}$, has an associated capacity, which we denote y.
6. Hence, for each GV curve in the training set, we have an input vector $\mathbf{x}$ and an output scalar, y. We use these as the inputs and outputs to a GP regression model for predicting the capacity, as described next.

[3]Intuitively, the inverse of this procedure (i.e. using the voltages sampled at uniformly spaced times as the inputs) might seem to be more logical. However, we chose the former approach because by using a fixed voltage range, we can avoid entering regions where there is no training data. For instance, if a large Δt is used in the test case, it might happen that this extends beyond the upper voltage region of the GV curve for a training case with smaller capacity.

Regression

The goal of a regression problem is to learn the mapping from inputs $\mathbf{x}$ to outputs y, given a labelled training set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_D}$, where N_D is the number of training examples. In the present case, the inputs $\mathbf{x}_i \in \mathbb{R}^{+n}$ are the time vectors for each GV curve, and the outputs $y_i \in \mathbb{R}^+$ are the corresponding measured capacities, as discussed in the previous section. We assume the underlying model takes the form $y = f(x) + \varepsilon$, where $f(x)$ represents a latent function and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is an independent and identically distributed noise contribution. The learned model can then be used to make predictions at a test index $\mathbf{x}^*$ (the vector of time values obtained online) for the unknown capacity, y^*.

In the present work, we employ Gaussian process regression with squared exponential kernel function to achieve this mapping. A full description of the mathematical machinery behind GPs is given in the appendix. The method was implemented in Matlab using the GPML toolbox [26].

A leave-one-out validation scheme was used, whereby each cell is used once as a test set while the data from the remaining cells form the training set. The performance was evaluated using the root-mean-squared error (RMSE) in the capacity estimation, defined as

$$\mathrm{RMSE}(\hat{y}_i{}^*, y_i^*) = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T}(\hat{y}_i{}^* - y_i^*)^2} \qquad (1)$$

where $\hat{y}^*$ is the estimated capacity, y^* is the true value, and N_T is the total number of test points.

Manufacturer	Kokam
Form factor	Pouch
# cells	8
# samples	519
Q range (Ah)	$0.74 \rightarrow 0.43$
Cycling	All cells cycled with same regime

Table 1. OVERVIEW OF THE AGING EXPERIMENTS. The '# samples' column represents the total number of voltage-time curves, across all cells. The 'Q range' column indicates the values of the maximum initial capacity and minimum final capacity respectively, across all cells.

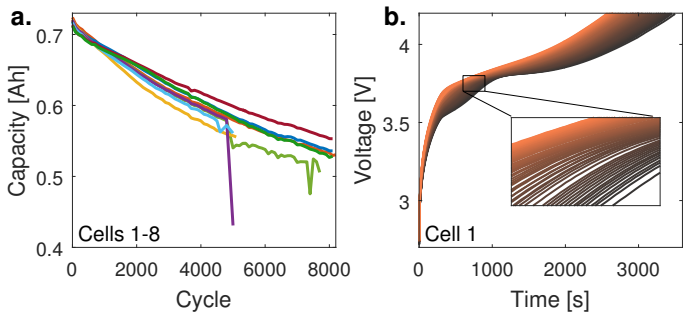

Figure 2. EXPERIMENTAL DATA. **a**, Capacity evolution of the tested cells. **b**, Evolution of the voltage curves for Cell 1 over the life of the cell. The colours range from light to dark orange as the cycle number increases.

EXPERIMENTS

We conducted aging experiments on 8 commercial Kokam pouch cells with 740 mAh nominal capacity. The cells negative electrode material is graphite and positive electrode material is a blend of lithium cobalt oxide (LCO) and lithium nickel cobalt oxide (NCO). The cells were cycled in a Binder MK53 thermal chamber at a constant ambient temperature of 40 °C throughout the lifetime of the experiments. Details of these experiments are given in Table 1.

All 8 cells were cycled by repeatedly discharging using the ARTEMIS urban drive cycle [27] and recharging at a constant current of 2C. After every 100 cycles, a characterisation test was carried out including a full charge-discharge cycle at 1C – these were the GV curves for this dataset. Fig. 2b shows the complete set of GV curves for Cell 1 over its entire lifecycle. Similar sets of curves were observed for the other cells. Each of these curves represents a single sample from which the inputs to the regression problem are sampled. A total of 519 charge curves were measured across all cells (i.e. $\sim$ 65 curves per cell).

The cell capacity was calculated by integrating the 1C charge curves. The calculated capacities for all 8 cells are plotted as a function of the cycle number in Fig. 2a. The end of life (EoL) was deemed to occur if the manufacturer-specified lower voltage limit (2.7 V) was exceeded for more than 10 s during the discharge cycle or if the cell reached 0 V for any measurable period of time. In practice, the EoL was triggered by the latter condition for all cells due to the presence of a large current pulse in the drive cycle. The EoL typically occurred at around 7,000-8,000 cycles (Fig. 2a) although one of the cells failed much earlier than this ($\sim$5,000 cycles). We also note that one of the cells (light green line in Fig. 2a) entered a change of regime around 5,000 cycles where a sudden drop in capacity occurred – this provides an interesting challenge for the capacity estimation algorithm as we discuss in the following section.

RESULTS

Fig. 3 shows results for selected cells for two selected combinations of online measurement duration, Δt, and lower voltage, V_l. For each plot, the model is tested on the cell shown and trained on all other cells. Note that for the test set, we do not actually carry out a separate online diagnostic test as described in the previous section; rather we simply select the relevant portion of the data from the full GV curve, as though it had come from a short diagnostic test. Fig. 3a shows that reasonably good performance can be achieved using a relatively short measurement duration of just 50 s. Where the predictions are less accurate, the error bars are quite honest and generally extend to encompass the true values. For instance, Cell 2 exhibits an unusual drop in capacity at $\sim$5000 cycles, a behaviour which is not manifested by any of the other cells (which were used for training in this case). Hence, the estimates made for Cell 2 after $\sim$5000 cycles are slightly erratic, but their uncertainty is accurately reflected by their correspondingly larger error-bars. On the other hand, Fig. 3b shows that consistently high performance can be achieved if a long measurement duration is utilized. The estimates for all cells in this case have an RMSE value below 1%. Interestingly, the method performs well for Cell 2 even in the regime beyond $\sim$5000 cycles, and expresses high confidence in these estimates. In practice the provision of such confidence estimates has significant implications. For instance, in an online setting, as capacity measurements are received sequentially from diagnostic tests of varying duration, a Kalman filter [28] (or other probabilistic filter) could effectively discount the uncertain measurements and retain the certain ones. This would result in an overall more robust diagnosis over multiple cycles.

Fig. 4 shows the overall results, where each cell is used once as the test set. Fig. 4a shows actual vs. predicted capacities across all cells for a selection of Δt and V_l values. It is apparent that larger Δt values (lower rows on the grid of plots in Fig. 4a) have higher accuracy, whereas differences in V_l (columns of the same

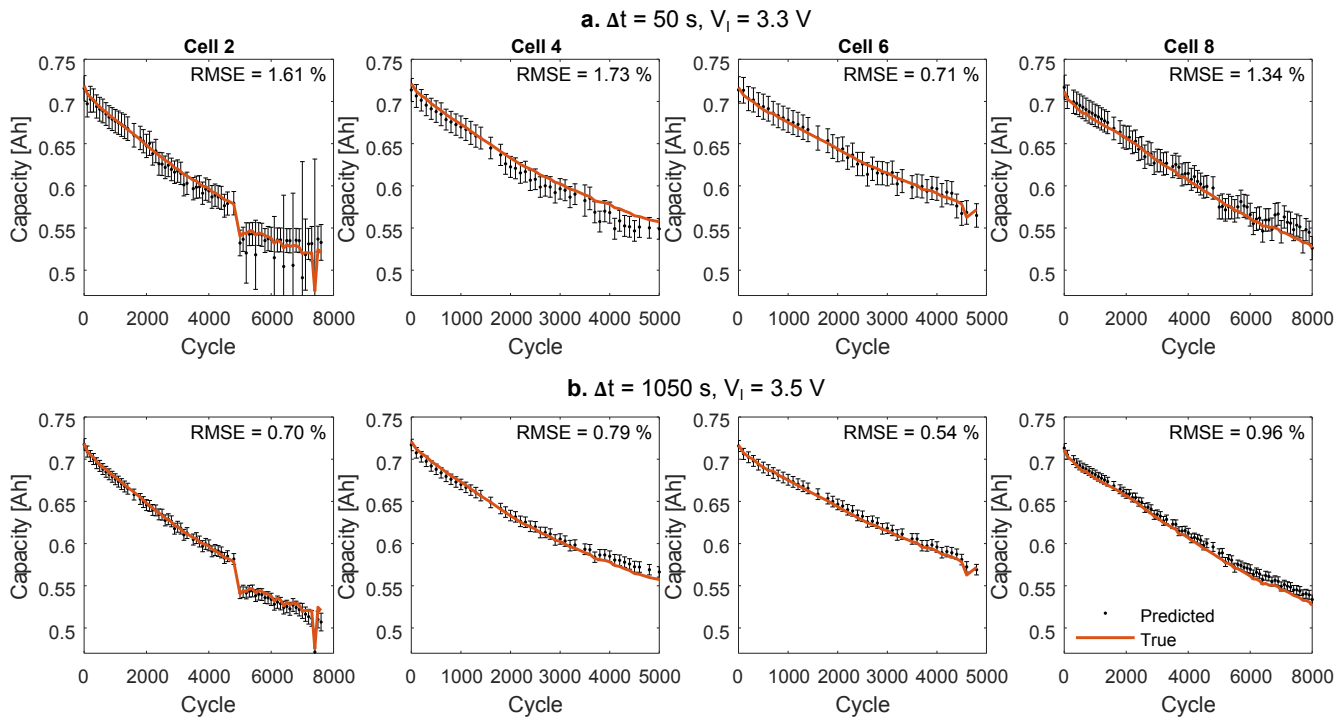

Figure 3. RESULTS FOR SELECTED CELLS/PARAMETERS. Red lines indicate measured capacity and black markers with error-bars indicate the GP-ICE estimates $\pm 2\sigma$. **a**, $\Delta t = 50$ s and $V_l = 3.3$V, **b**, $\Delta t = 1050$ s and $V_l = 3.5$V.

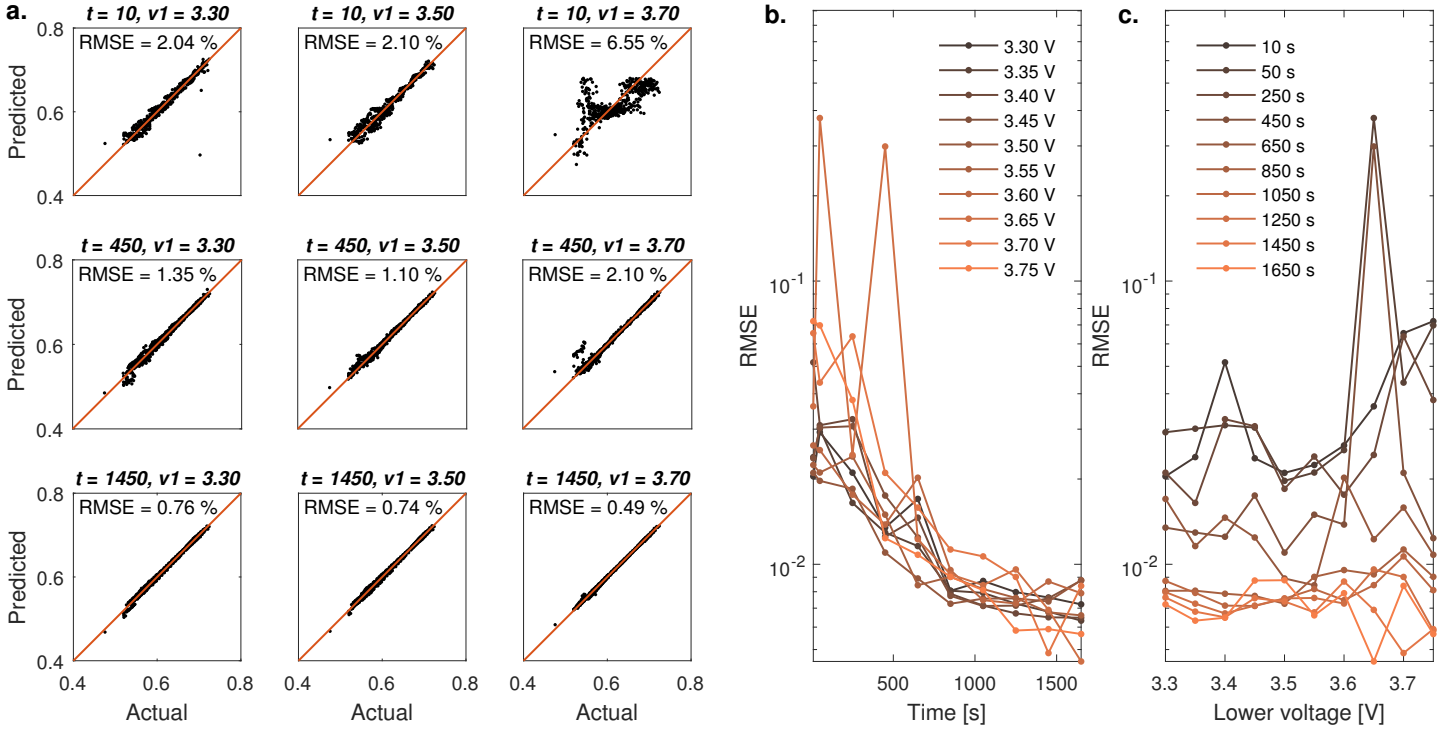

Figure 4. OVERALL RESULTS FOR ALL CELLS/PARAMETERS. RMSE values are based on the entire dataset with each cell used once as the test set. **a**, Actual vs. predicted capacities for different starting voltages and measurement durations. The red line represents $y^* = \hat{y}^*$. The closer the data points lie to this line, the smaller the difference between the actual and predicted value. **b**, RMSE vs. measurement duration for different starting voltages. **c**, RMSE vs. starting voltage for different measurement durations.

grid) have a less consistent effect on the RMSE. This is shown explicitly in Figs. 4b and c, which show the overall RMSE values plotted against Δt and V_l respectively. For all starting voltages there is a clear decreasing trend in RMSE as the measurement duration is increased. This is intuitively what we would expect, since larger measurement durations capture larger portions of the charge curve and hence have the potential to identify more variation between different inputs.

For the measurement duration of $\Delta t = 1450$ s (the bottom row of the grid of plots in Fig. 4a), the capacity is accurately estimated even at extreme values. For instance, the lone data point at just under 0.5 Ah lies very close to the red line despite not having other nearby training examples from which to learn. One of the advantages of Bayesian methods such as GPs over deterministic methods is that they can generalise better from relatively small datasets such as the one used here by properly expressing their uncertainty about the underlying model.

On the other hand, when smaller measurement durations are used (such as the middle and upper rows of plots) this outlier is over-estimated. However, we note that in most cases where the estimate was inaccurate, the error bars were found to be correspondingly larger, hence accurately conveying the model's uncertainty. For the sake of clarity the error bars are not shown in the plots of Fig. 4a since with such a large number of data points the plots would be excessively cluttered; however, a specific example of this was the estimates for Cell 2 in the region beyond $5,000$ cycles discussed previously. In contrast, non-probabilistic approaches (such as the previously used neural networks or SVMs [20]) implicitly assign equal credibility to the accurate and inaccurate measurements.

DISCUSSION

In the present study, the cells used for training were aged identically to the cells used for testing, whereas in practice, this is unlikely to be the case. In principle, the application of the method in the latter case would be identical – however, the performance would depend on how strongly the GV curves are correlated with battery capacity across different aging regimes. In practice, this may mean that larger datasets are required to achieve similar performance. Hence, future work will consider the application of this technique to cells that have been aged in differing ways.

There are many practical scenarios in which GP-ICE could be feasible. For instance, in EV applications, the vast majority of charging stations output a power of less than 22 kW [29], which would equate to < 0.5 C for a typical EV battery pack. One hurdle to practical implementation is the applicability of the diagnostic test under variable ambient temperature conditions. Although the present results apply to a single temperature, the method should support variable ambient temperature conditions, provided sufficient training data were available.

Since GP-ICE does not rely on cell specific knowledge such as the expected locations and numbers of peaks in IC/DV curves, it could be directly applied to other cell chemistries without modification. Of course, there is no guarantee of equivalent results to those obtained here – the performance of the method is dependent on how strongly the galvanostatic voltage-time data are correlated with the cell capacity, something which may vary from cell to cell and across voltage ranges, as the earlier results show. However, the important point is that we do not need to encode any cell-specific information in our model – the capacity estimation is achieved automatically in any case. This generality also opens up the possibility of applying the method to portions of constant-current data within otherwise dynamic drive cycles. This is likely to be non-trival due to dynamics in the cell; however, if long enough portions of constant current are available, then it may give satisfactory results.

Recently, You et al. [30] presented an approach which uses a Recurrent Neural Network trained on partial charge curves for estimating cell capacity. This is similar to our approach but with some key differences. Specifically, our approach: (i) employs a Gaussian process method for the regression step, which provides confidence in the capacity estimates, (ii) uses Savitzky-Golay filtering as a preprocessing step to improve signal to noise ratio, (iii) selects a subset (4 points) of the smoothed data in order to minimise computational overhead – this is a necessary requirement given the higher computational overhead of GPs compared to neural networks. Moreover, our method shows how the performance of the capacity estimates varies as a function of the starting voltage and measurement duration, something which has not been demonstrated in previous work. On the other hand, the method of [30] exploits the sequential nature of the charge curves, unlike our approach, which ignores any correlation between the inputs. An interesting area of future work could involve accounting for correlations between the inputs by encoding recurrent behaviour into the kernel of the GP function (such as in the method presented in [31]) in order to achieve the benefits of both of these approaches.

CONCLUSIONS

This paper has introduced GP-ICE, a technique for estimating battery capacity using small portions of voltage-time data under constant current (galvanostatic) operation. GP-ICE dispenses with the interpretation of galvanostatic voltage (GV) data as incremental capacity or differential voltage curves, and instead involves directly performing regression using the voltage/time data as inputs. Key to achieving accurate results with this approach was the use of a Bayesian non-parametric regression technique, Gaussian processes regression, which adapts to the complexity of the data and avoids over-fitting.

GP-ICE provides various advantages over IC/DV peak tracking such as greater flexibility, shorter diagnostic test re-

quirements, and the provision of estimates of uncertainty in its predictions. It also provides insight into which voltage ranges are most informative, and hence may inform a BMS as to when best to perform a diagnostic test.

The present work considered training data in which the cells were aged identically to the cells used for testing. Future work will consider the more challenging application to cells that have been aged in differing ways. Future work should also consider accounting for variable ambient temperatures during the diagnostic test – this should be feasible provided sufficient training data are available under the required range of temperatures.

ACKNOWLEDGMENT

This work was funded by an RCUK Engineering and Physical Sciences Research Council grant, ref. EP/K002252/1.

REFERENCES

[1] Broussely, M., Biensan, P., Bonhomme, F., Blanchard, P., Herreyre, S., Nechev, K., and Staniewicz, R., 2005. "Main aging mechanisms in Li ion batteries". *Journal of power sources,* ***146***(1), pp. 90–96.

[2] Vetter, J., Novák, P., Wagner, M., Veit, C., Möller, K.-C., Besenhard, J., Winter, M., Wohlfahrt-Mehrens, M., Vogler, C., and Hammouche, A., 2005. "Ageing mechanisms in lithium-ion batteries". *Journal of power sources,* ***147***(1), pp. 269–281.

[3] Jung, S.-K., Gwon, H., Hong, J., Park, K.-Y., Seo, D.-H., Kim, H., Hyun, J., Yang, W., and Kang, K., 2014. "Understanding the degradation mechanisms of LiNi0. 5Co0. 2Mn0. 3O2 cathode material in lithium ion batteries". *Advanced Energy Materials,* ***4***(1).

[4] Farmann, A., Waag, W., Marongiu, A., and Sauer, D. U., 2015. "Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles". *Journal of Power Sources,* ***281***, pp. 114–130.

[5] Zhang, J., and Lee, J., 2011. "A review on prognostics and health monitoring of li-ion battery". *Journal of Power Sources,* ***196***(15), pp. 6007–6014.

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

[7] Plett, G. L., 2006. "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1: Introduction and state estimation". *Journal of Power Sources,* ***161***(2), pp. 1356–1368.

[8] Plett, G. L., 2006. "Sigma-point kalman filtering for battery management systems of lipb-based HEV battery packs: Part 2: Simultaneous state and parameter estimation". *Journal of power sources,* ***161***(2), pp. 1369–1384.

[9] Waag, W., and Sauer, D. U., 2013. "Adaptive estimation of the electromotive force of the lithium-ion battery after current interruption for an accurate state-of-charge and capacity determination". *Applied Energy,* ***111***, pp. 416–427.

[10] Chaturvedi, N. A., Klein, R., Christensen, J., Ahmed, J., and Kojic, A., 2010. "Algorithms for advanced battery-management systems". *IEEE Control Systems,* ***30***(3), pp. 49–68.

[11] Moura, S. J., Chaturvedi, N. A., and Krstic, M., 2012. "PDE estimation techniques for advanced battery management systemsPart ii: SOH identification". In American Control Conference (ACC), 2012, IEEE, pp. 566–571.

[12] Prasad, G. K., and Rahn, C. D., 2013. "Model based identification of aging parameters in lithium ion batteries". *Journal of power sources,* ***232***, pp. 79–85.

[13] Bole, B., Kulkarni, C. S., and Daigle, M., 2014. "Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use". In Proceedings of Annual Conference of the Prognostics and Health Management Society, Fort Worth, TX, USA, Vol. 29, pp. 1–9.

[14] Bizeray, A. M., Kim, J.-H., Duncan, S. R., and Howey, D. A., 2017. "Identifiability and parameter estimation of the single particle lithium-ion battery model". *arXiv preprint arXiv:1702.02471*.

[15] Dubarry, M., Truchot, C., and Liaw, B. Y., 2012. "Synthesize battery degradation modes via a diagnostic and prognostic model". *Journal of power sources,* ***219***, pp. 204–216.

[16] Birkl, C. R., Roberts, M. R., McTurk, E., Bruce, P. G., and Howey, D. A., 2017. "Degradation diagnostics for lithium ion cells". *Journal of Power Sources,* ***341***, pp. 373–386.

[17] Weng, C., Cui, Y., Sun, J., and Peng, H., 2013. "On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression". *Journal of Power Sources,* ***235***, pp. 36–44.

[18] Weng, C., Feng, X., Sun, J., and Peng, H., 2016. "State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking". *Applied Energy,* ***180***, pp. 360–368.

[19] Berecibar, M., Garmendia, M., Gandiaga, I., Crego, J., and Villarreal, I., 2016. "State of health estimation algorithm of LiFePO 4 battery packs based on differential voltage curves for battery management system application". *Energy,* ***103***, pp. 784–796.

[20] Berecibar, M., Devriendt, F., Dubarry, M., Villarreal, I., Omar, N., Verbeke, W., and Van Mierlo, J., 2016. "Online state of health estimation on NMC cells based on predictive analytics". *Journal of Power Sources,* ***320***, pp. 239–250.

[21] Savitzky, A., and Golay, M. J., 1964. "Smoothing and

differentiation of data by simplified least squares procedures.". *Analytical chemistry,* ***36***(8), pp. 1627–1639.
[22] Rasmussen, C. E., and Williams, C. K. I., 2006. *Gaussian processes for machine learning*. Citeseer.
[23] Saha, B., Goebel, K., Poll, S., and Christophersen, J., 2009. "Prognostics methods for battery health monitoring using a bayesian framework". *IEEE Transactions on instrumentation and measurement,* ***58***(2), pp. 291–296.
[24] Liu, D., Pang, J., Zhou, J., Peng, Y., and Pecht, M., 2013. "Prognostics for state of health estimation of lithium-ion batteries based on combination gaussian process functional regression". *Microelectronics Reliability,* ***53***(6), pp. 832–839.
[25] Richardson, R. R., Osborne, M. A., and Howey, D. A., 2017. "Gaussian process regression for forecasting battery state of health". *arXiv preprint arXiv:1703.05687*.
[26] Rasmussen, C. E., and Nickisch, H., 2010. "Gaussian processes for machine learning (GPML) toolbox". *J. Mach. Learn. Res.,* ***11***, Dec., pp. 3011–3015.
[27] André, M., Keller, M., Sjödin, Å., Gadrat, M., Mc Crae, I., and Dilara, P., 2009. "The ARTEMIS European tools for estimating the transport pollutant emissions". In Proc. 18th International Emission Inventories Conference, pp. 1–10.
[28] Grewal, M. S., 2011. *Kalman filtering*. Springer.
[29] Sbordone, D., Bertini, I., Di Pietra, B., Falvo, M. C., Genovese, A., and Martirano, L., 2015. "EV fast charging stations and energy storage technologies: A real implementation in the smart micro grid paradigm". *Electric Power Systems Research,* ***120***, pp. 96–108.
[30] You, G.-w., Park, S., and Oh, D., 2017. "Diagnosis of electric vehicle batteries using recurrent neural networks". *IEEE Transactions on Industrial Electronics*.
[31] Al-Shedivat, M., Wilson, A. G., Saatchi, Y., Hu, Z., and Xing, E. P., 2016. "Learning scalable deep kernels with recurrent structure". *arXiv preprint arXiv:1610.08936*.
[32] Murphy, K. P., 2012. *Machine learning: a probabilistic perspective*. MIT press.
[33] Stein, M. L., 2012. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.
[34] Iman, R. L., 2008. "Latin hypercube sampling". *Encyclopedia of Quantitative Risk Analysis and Assessment*.

GAUSSIAN PROCESS REGRESSION

This section gives an overview of Gaussian process regression; a more detailed presentation of this topic is given in ref. [22] and Chapter 15 of ref. [32].

A Gaussian process (GP) defines a probability distribution over functions, and is denoted as:

$$f(\mathbf{x}) \sim \mathcal{GP}\big(m(\mathbf{x}), \kappa(\mathbf{x},\mathbf{x}')\big), \tag{2}$$

where $m(\mathbf{x})$ and $\kappa(\mathbf{x},\mathbf{x}')$ are the mean and covariance functions respectively, denoted by

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{3}$$

$$\kappa(\mathbf{x},\mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))\left(f(\mathbf{x}') - m(\mathbf{x}')\right)^T]. \tag{4}$$

For any finite collection of input points, say $X = \mathbf{x}_1, ..., \mathbf{x}_{N_D}$, this process defines a probability distribution $p(f(\mathbf{x}_1), ..., f(\mathbf{x}_{N_D}))$ that is jointly Gaussian, with some mean $\mathbf{m}(\mathbf{x})$ and covariance $\mathbf{K}(\mathbf{x})$ given by $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$.

Gaussian process regression is a way to achieve non-parametric regression with Gaussian processes. The key idea is that, rather than postulating a parametric form for the function $f(\mathbf{x},\phi)$ and estimating the parameters ϕ (as in parametric regression), we instead assume that the function $f(\mathbf{x})$ is a sample from a Gaussian process as defined above.

The most common choice of covariance function is the squared exponential (SE), defined by

$$\kappa_{\mathrm{SE}}(\mathbf{x},\mathbf{x}') = \theta_f^2 \exp\left(-\frac{1}{\theta_l^2}\|\mathbf{x}-\mathbf{x}'\|^2\right). \tag{5}$$

The covariance function parameters (i.e. the hyperparameters), θ_f and θ_l, control the y-scaling and x-scaling, respectively. The choice of the SE kernel makes the assumption that the function is very smooth (infinitely differentiable). This may be too strict a condition for many physical phenomena [33], and so a common alternative is the Matérn covariance class:

$$\kappa_{\mathrm{Ma}}(x-x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{(x-x')}{\rho}\right)^{\nu} \mathcal{R}_\nu\left(\sqrt{2\nu}\frac{(x-x')}{\rho}\right), \tag{6}$$

where ν is a smoothness hyperparameter (larger ν implies smoother functions) and $\mathcal{R}_\nu$ is the modified Bessel function. This equation simplifies considerably for half-integer ν. The most common examples are $\nu = 5/2$ and $\nu = 3/2$, which we denote as Ma5 and Ma3 in this work. All of the aforementioned kernels are *stationary* kernels, since the correlation between points is purely a function of the difference in their inputs, $x - x'$. In the present study, we used the Ma5 kernel, although similar performance was observed using the SE and Ma3 kernels.

The mean function is commonly defined as $m(\mathbf{x}) = 0$, and for convenience we follow this convention here.

Now, if we observe a labelled training set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_D}$, predictions can be made at test indices X^* by computing the conditional distribution $p(\mathbf{y}^*|X^*, X, \mathbf{y})$.

This can be obtained analytically by the standard rules for conditioning Gaussians [32], and (assuming a zero mean for notational simplicity) results in a Gaussian distribution given by:

$$p(\mathbf{y}^*|X^*,X,\mathbf{y}) = \mathcal{N}(\mathbf{y}^*|\mathbf{m}^*,\sigma^*) \tag{7}$$

where

$$\mathbf{m}^* = \mathbf{K}(X,X^*)^T\mathbf{K}(X,X)^{-1}\mathbf{y} \tag{8}$$

$$\sigma^* = \mathbf{K}(X^*,X^*) - \mathbf{K}(X,X^*)^T\mathbf{K}(X,X)^{-1}\mathbf{K}(X,X^*). \tag{9}$$

The values of the covariance hyperparameters θ may be optimised by minimising the negative log marginal likelihood defined as NLML $= -\log p(\mathbf{y}|X,\theta)$. Minimising the NLML automatically performs a trade-off between bias and variance, and hence ameliorates over-fitting to the data. Given an expression for the NLML and its derivative w.r.t θ (both of which can be obtained in closed form), we can estimate θ using any standard gradient-based optimizer. In our case, we used the GPML toolbox [26] implementation of conjugate gradients. Since the objective is not convex, local minima can be a problem. However, this was not an issue in the present study, as was verified by repeated diverse initialisations using Latin hypercube sampling [34] yielding identical results.