

Estimation of Li-Ion Degradation Test Sample Sizes Required to Understand Cell-to-Cell Variability**

Philipp Dechent,^{+, *^[a, b]} Samuel Greenbank,⁺^[c] Felix Hildenbrand,^[a, b] Saad Jbabdi,^[d] Dirk Uwe Sauer,^{*^[a, e]} and David A. Howey^{*^[c]}

Ageing of lithium-ion batteries results in irreversible reduction in performance. Intrinsic variability between cells, caused by manufacturing differences, occurs throughout life and increases with age. Researchers need to know the minimum number of cells they should test to give an accurate representation of population variability, since testing many cells is expensive. In this paper, empirical capacity versus time ageing models were fitted to various degradation datasets for commercially available

cells assuming the model parameters could be drawn from a larger population distribution. Using a hierarchical Bayesian approach, we estimated the number of cells required to be tested. Depending on the complexity, ageing models with 1, 2 or 3 parameters respectively required data from at least 9, 11 or 13 cells for a consistent fit. This implies researchers will need to test at least these numbers of cells at each test point in their experiment to capture manufacturing variability.

1. Introduction

Lithium-ion batteries have grown in importance over the past decade and are now the key technology underlying applications from electric vehicles to grid energy storage.^[1] High specific energy, low internal resistance and long lifetime have already led Li-ion cells to dominate the market for consumer electronics applications. A crucial issue that strongly impacts overall system performance is the intrinsic variability in capacity, resistance and degradation rate between cells, caused by small variations in manufacturing processes.^[2–4] Quantifying

the typical variability in mature commercially available Li-ion cells is crucial for understanding the trade-offs involved in designing battery packs and estimating pack performance and lifetime on the basis of individual cell performance.

Battery state of health (SOH) is usually defined as the capacity or impedance/resistance of a cell^[5] under standard test conditions, and it constrains the useful and safe operation of batteries. State of health changes with time and usage, and is influenced both by external factors (such as voltage, time and temperature) and by internal manufacturing and materials variations.^[6] Since it has a direct bearing on the economic value and operation of batteries, the estimation of current and future Li-ion SOH is a popular area of research.^[6]

Variability in cell capacity and resistance can create differing loads within a Li-ion battery pack, and depending on its extent, variability will inevitably impact performance, cost and safety.^[7–9] There are a variety of sources for this variability, for instance, thermal inhomogeneities influence SOH and increase cell-to-cell variability during usage.^[5,10,11] However, manufacturing variability, sometimes described as tolerance, is a significant contributor to cell-to-cell variability.^[2,3,12] New, nominally identical cells from the same batch exhibit a spread in capacity before they have been cycled.^[4,13,14] Simulations and experiments have demonstrated that this intrinsic manufacturing variability is a contributing factor to differing ageing rates between cells.^[2,3,13,15,16]

There are many possible sources of manufacturing variability,^[17] such as variability in electrode thickness and density,^[18] fraction of active material, liquid-to-solid ratio and coating gap.^[19] In this work we assume that all intrinsic variability may be expressed through a single lumped population distribution, for each parameter and dataset. A further contributor to cell-to-cell variability results from variance in experimental conditions, for example location in a given testing chamber.^[20]

[a] P. Dechent,⁺ F. Hildenbrand, D. U. Sauer
Electrochemical Energy Conversion and Storage Systems Group,
Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen
University, Germany
E-mail: philipp.dechent@isea.rwth-aachen.de

[b] P. Dechent,⁺ F. Hildenbrand
Jülich Aachen Research Alliance,
JARA-Energy,
Germany

[c] S. Greenbank,⁺ D. A. Howey
Battery Intelligence Lab,
Department of Engineering,
University of Oxford,
UK
E-mail: david.howey@eng.ox.ac.uk

[d] S. Jbabdi
Wellcome Centre for Integrative Neuroimaging,
FMRI, University of Oxford, UK

[e] D. U. Sauer
Helmholtz Institute Münster (HI MS),
IEK-12, Forschungszentrum
Jülich, Münster, Germany
E-mail: DirkUwe.Sauer@isea.rwth-aachen.de

[+] Authors contributed equally

[**] A previous version of this manuscript has been deposited on a preprint server (<https://arxiv.org/abs/2107.07881>)."

 © 2021 The Authors. Batteries & Supercaps published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Ageing reduces the performance of a given cell or pack and increases variability between cells.^[10,13,14,21–23] To compound this, no significant correlation between initial health and ageing rates and subsequent ageing rates in later life has been found experimentally,^[2,13,23,24] although features derived from early cycle life can be used to accurately predict lifetime.^[25,26] Some authors have found that initial battery health is sometimes bimodal.^[27,28] It has also been found that Weibull distributions, common in failure analysis, may be used to quantify battery end of life statistics.^[2,21,29,30]

In summary, cell-to-cell variability is a significant factor influencing the performance and value of batteries. As a result, modelling of battery ageing data is complicated by the question of how many cells should be tested at each experimental condition so as to adequately capture the intrinsic variability. Battery testing involves cost, in time and number of test channels, and therefore optimising the amount of information obtained from a test is a key consideration. The literature seems to have largely ignored the issue of how many cells should be tested to capture intrinsic variability, and for practical reasons, most ageing studies use only a small number of cells (e.g. 1–3) at each test condition. In this paper we address this question directly by fitting models to ageing data. There are many options for modelling battery capacity through life, from empirical curve fits^[31] through to physical models^[32] and machine learning approaches.^[6,33] Since our aim was to investigate intrinsic rather than extrinsic variability, we chose as our modelling approach to use simple empirical curve fits of health versus time. We examined the consistency of the resulting model parameters as we added data drawn from increasing numbers of cells within each dataset, using five different battery ageing datasets. An estimate of required sample size was drawn for each parameter in every model for every available dataset.

2. Datasets and models

2.1. Data

To study cell-to-cell variability, ideally we need data from a very large number of cells, perhaps thousands. The costs of such large scale testing would be prohibitive, requiring many battery test channels for multiple years, and no such datasets are openly available. As a compromise, however, some ageing datasets are available with order 10–100 cells cycled identically (or very similarly). We selected five datasets for analysis based on the requirement of wanting as many cells as possible to have been tested within each dataset. Two of these are open source, and three are from our own experiments. Each individual dataset used identical commercially available Li-ion cells, albeit having different manufacturers, chemistries and cell sizes from dataset to dataset. All datasets used 18650 cylindrical cells, although the methods discussed below can equally be applied to other form factors. Some of the datasets featured identical experimental conditions, i.e. each cell was tested in exactly the same way, whereas others varied the

testing conditions slightly beyond the expected uncontrollable experimental variability. The datasets are as follows:

1. **Baumhöfer-2014**^[3] consists of 48 Sanyo/Panasonic UR18650E NMC/graphite 1.85 Ah cells in a cycle ageing test each under the same operating conditions. Data available at [34].
2. **Dechent-2020** consists of 22 Samsung INR18650-35E NCA/graphite cells each with a nominal capacity of 3.4 Ah. The cells were cycled with C/2 constant current and a 20% cycle depth around an average SOC of 50%. Data available at [35].
3. **Dechent-2017** consists of 21 Samsung NR18650-15 L1 1.5 Ah NMC/graphite cells. Six of the cells were cycled with 1 C charge and 6 C discharge current between 3.3 V and 4.1 V (90% cycle depth), and 15 additional cells were cycled with the same voltage range but current rates varied by up to 15%. Data available at [36].
4. **Severson-2019** consists of 124 cells made by A123 APR18650 M1 A with LFP/graphite chemistry, each with nominal capacity 1.1 Ah. For this work a subset 67 of these cells with similar load profiles but slightly varying charging currents was chosen. Data available via.^[25] The cells in this dataset are from three different experimental batches so will have been subjected to higher variance in testing conditions than the other datasets.
5. **Attia-2020** replicates Severson-2019, but with a fixed charging window of 10 minutes. There are 45 cells in this set. Data available via [37].

Three of these datasets, namely Baumhöfer-2014, Severson-2019 and Attia-2020, exhibit an onset of rapid degradation in later life, sometimes called the ‘knee-point’.^[3,26] The other two datasets show only linear degradation over usage. The data prior to the knee-point in the Attia-2020 and Severson-2019 sets was separately extracted to produce two additional linear ageing datasets.

2.2. Models

The models and the corresponding datasets that they were fitted to are shown in Table 1. Here, Linear-1 and Linear-2 refer to the two linear models, having one and two parameters respectively. Alternatively, LinExp is a combined linear and exponential model that was used to capture the knee-point and later life health decay, where this was evident in the data.

Table 1. The dataset and empirical model combinations used here. Linear-1 is a single parameter linear model, Linear-2 a two parameter linear model, and LinExp a 3 parameter linear plus exponential model.

Dataset	Linear-1	Linear-2	LinExp	Ref.
Dechent-2017	X	X		[36]
Dechent-2020	X	X		[35]
Baumhöfer-2014			X	[3]
Severson-2019	X	X	X	[25]
Attia-2020	X	X	X	[37]

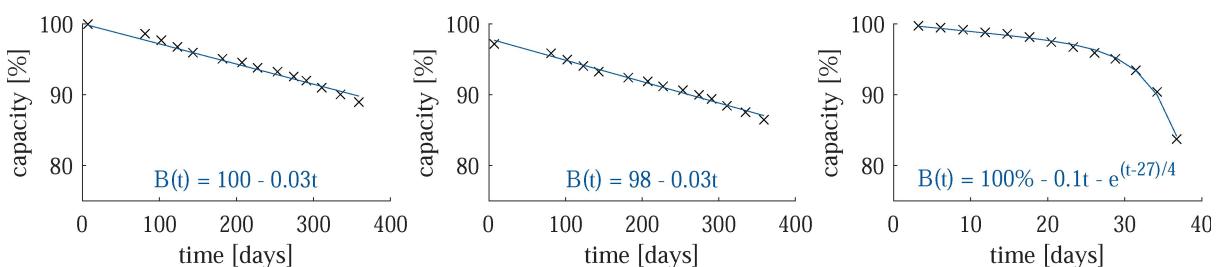


Figure 1. Examples of data and fitted curves. Left to right: Linear-1 model on a cell from Dechent-2020; Linear-2 model on a cell from Dechent-2020; LinExp model on a cell from Severson-2019.

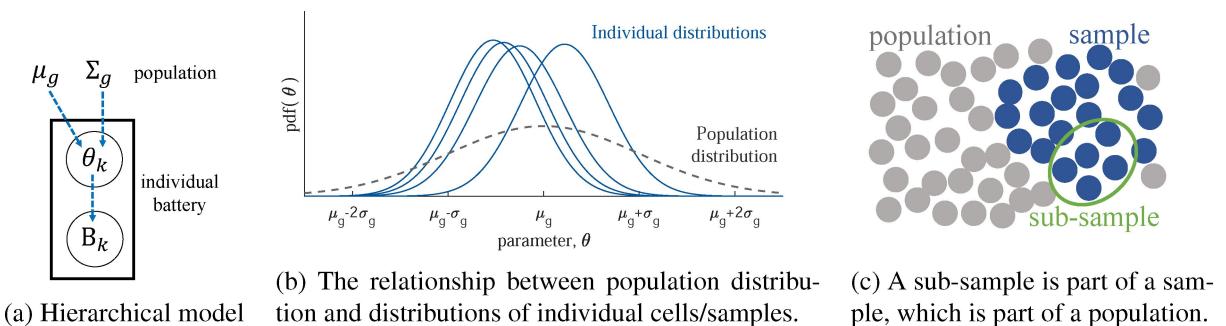


Figure 2. A hierarchical approach was used to infer population statistics.

The three models are given by the following expressions, where t is time, B is capacity, and all other parameters ($c_1, c_2, c_3, B_0, t_f, \tau$) were fitted to the data:

$$\text{Linear-1: } B(t) = 100\% + c_1 \times t \quad (1)$$

$$\text{Linear-2: } B(t) = B_0 + c_2 \times t \quad (2)$$

$$\text{LinExp: } B(t) = 100\% + c_3 \times t - \exp[(t - t_f)/\tau] \quad (3)$$

Linear-1 and Linear-2 differ only by the addition of the initial capacity B_0 as a fitted parameter in the latter. The cell capacities were normalised according to which model was in use. For Linear-1 and LinExp, the capacities were normalised relative to the initial capacity of each cell. Linear-2 used capacity curves normalised relative to the nominal capacity. In the LinExp model, the initial linear capacity decrease is followed by a faster exponential decrease with onset time t_f and time constant τ , as shown in Figure 1.

3. Methodology

To quantify cell-to-cell variability, we used an approach called multi-level Bayes (MLB), also known as hierarchical Bayes, where the parameters of an individual cell model are assumed to be drawn from a population distribution, as depicted in Figure 2a. In this framework, the first level of inference is on the parameters of an individual battery cell model, and the second level of inference is on the parameters of the underlying population distribution.^[38–40] Given some data sub-sampled

from the datasets described above, this approach provides an estimate of the individual (θ_k) and the population (μ_g, Σ_g) parameter values as well as their associated uncertainties, as depicted in Figure 2(a) and (b). Therefore one can explore the trade off between the number of cells' data used for fitting the models versus the stability and variance (or standard deviation) of the resulting population parameter estimates. As additional data from more cells is included in the estimation, the variance of the population mean and variance decreases (i.e. we become more certain of the population model). As illustrated in Figure 3, we considered the population estimates to be stable when the standard deviation of the population standard deviation estimate began to steadily decrease as a function of sub-sample size ($\sim \frac{1}{N}$). We set the condition of an acceptable variability as being within a threshold, α , of the stable

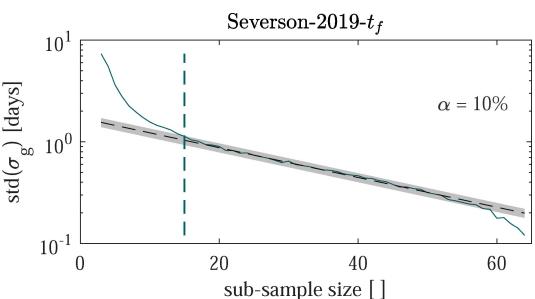


Figure 3. Example showing the decreasing standard deviation of the population standard deviation estimate as the number of sub-samples used for model fitting is increased. A threshold for an acceptable estimate is shown (vertical dashed line). Example taken from the t_f parameter for Severson-2019.

decreasing region. The value of α was set at 10% as shown by the grey shaded region in Figure 3.

The following conventions are used throughout the remainder of this work. Figure 2(c) shows the definitions of population, sample and subsample used. The ‘population’ means the very large (but unavailable) group of all possible similar batteries produced in the same manufacturing batch, from which a subset were tested in a lab. (Therefore, we expect the population statistics to be different for each dataset that was introduced in the previous section.) A ‘sample’ refers to all the available full data in a specific dataset. Therefore, a sample is drawn from a population. Conversely, any time a smaller subset was drawn from a full test dataset, it is referred to here as a ‘sub-sample’. Summary sample statistics are denoted with the Latin alphabet, while population estimates are written using the Greek alphabet. For example, mean and variance are (m, s^2) and (μ, σ^2) respectively. The letter k is used to denote value(s) for a specific cell, and K denotes the total number of cells. Probabilities and distributions are written in capitals: P, N. Battery capacities are represented by the letter B .

3.1. Multi-Level Bayes

The parameters of an individual cell degradation model are assumed to be drawn from a population distribution that is unique for each dataset, with unknown population mean and variance. Let B_k denote the capacity of cell k over a number of measurements (i.e. B_k is a vector) and θ_k denote the model parameters determining the time evolution of capacity. For example, for the LinExp model, $\theta_k = (c_k, t_{f,k}, \tau_k)^T$. Assuming additive and Gaussian measurement noise,

$$B_k = f(\theta_k) + \varepsilon_k, \quad (4)$$

where $\varepsilon_k \sim N(0, \sigma_{n,k}^2)$ and f is one of the three models introduced above.

The parameters of an individual cell model k are themselves drawn from a population distribution, assumed here to be Gaussian,

$$\theta_k \sim N(\mu_g, \sigma_g^2), \quad (5)$$

where μ_g is the group (i.e. population) mean and σ_g^2 the group variance. These population parameters are vectors of one, two, or three elements depending on whether the generative model is Linear-1, Linear-2, or LinExp, respectively.

To complete the specification of this generative model in the Bayesian framework requires prior distributions to be assumed. We used wide Gaussian priors (zero mean, 10^4 variance) on the population means μ_g , and uniform distributions on the population variances. The noise parameters $\sigma_{n,k}$

were assumed to follow a Jeffrey’s prior, $P(\sigma_{n,k}^2) \sim 1/\sigma_{n,k}^2$, and were integrated out analytically.

To infer the posterior distributions of the population parameters μ_g and σ_g , we used a two step process. In the first step (first level inference), we fitted individual cell parameters θ_k using Markov chain Monte Carlo (MCMC) sampling to obtain samples from $P(B_k|\theta_k) = \int P(B_k|\theta_k, \sigma_{n,k})P(\sigma_{n,k})d\sigma_{n,k}$. We then approximated the distributions for each cell with a Gaussian using their summary statistics (means μ_k and variances σ_k^2), which were then used in a second step (second level inference). This allowed the full posterior distribution of the population parameters to be written as follows:

$$P(\mu_g, \sigma_g | \{B_k\}) = \int \dots \int P(B_k|\theta_k)P(\theta_k|\mu_g, \sigma_g)P(\mu_g, \sigma_g) d\theta_1 \dots d\theta_K \quad (6)$$

The above multivariate integral can be evaluated analytically, owing to the Gaussian approximation to the first level inference. This yields:

$$P(\mu_g, \sigma_g | \{B_k\}) \propto \prod_{k=1}^K N(\mu_k, \sigma_k^2 + \sigma_g^2) P(\mu_g, \sigma_g) \quad (7)$$

Intuitively, this last expression shows that the posterior population mean is a Gaussian centred around the weighted average of the individual cell parameters, where the weights combine the first level variances (uncertainty on the fitting of each cell) and the group variance (mixed effect model). After this, we again used MCMC to draw samples from this posterior distribution to calculate its summary statistics.

The MLB approach was used to fit the parameters of all the model/dataset combinations shown in Table 1. That fitting was performed at all sub-sample sizes from minimum 3 cells, up to 3 less than the full number of cells in each dataset, with 1,000 repeats performed at each sub-sample size using random selection with replacement. A population distribution was deemed to have a stable fit if the standard deviation of the estimate of σ_g settled to follow a function $\log y = ax + b$ where x is sub-sample size, y is the standard deviation of σ_g and a, b are arbitrary slope and offset parameters. As a comparison, the equivalent result was also plotted when using a simple sub-sample distribution (SSD) by taking the mean, m_g , and variance, s_g^2 , of a given sub-sample:

$$m_g = \frac{1}{K} \sum_{k=1}^K \theta_k, \quad s_g = \sqrt{\frac{\sum_{k=1}^K (\theta_k - m_g)^2}{K-1}} \quad (8)$$

4. Results

As a reminder, the objective is to quantify the number of battery cells that are required for a stable fit of a population model, when cells are selected at random from a population. In particular, we wish to infer both the parameters of the capacity fade model for each cell, and the parameters of the underlying population, including their uncertainties. We now examine

both aspects in succession across the various datasets and models.

In most cases, the parameter values estimated by MLB for the population level distributions match that of the sample distributions. Figure 4 shows two examples of well matched population and sample distributions, namely from Severson-2019- c_1 and Attia-2020- τ . The parameter distribution in the sample is approximately Gaussian in both cases, matching an important assumption in the MLB derivation.

There was very little correlation between parameter distributions for any dataset/model combinations. The only significant case was for the Baumhöfer-2014-LinExp model, where Pearson's rank coefficients of 0.94 were found between all of c_1 , t_f and τ . Severson-2019 had correlation coefficients of ~0.6, suggestive of a weak relationship between the parameters of the LinExp model.

The MLB results were subject to significant variability when fitting a single incidence at each sub-sample size. Figure 5 demonstrates a typical set of parameter estimates where it was hard to interpret the values at small sub-sample sizes. The mean estimates appeared to be well fit at small values, but the population distributions only settled when the majority of the sample was used.

The summary results from 1,000 repeats, with replacement, were much smoother. The estimated standard deviation of σ_g for the Linear-1 models rapidly dropped with increasing numbers of cells in a sub-sample for all datasets as shown in Figure 6. The SSD approach produced a lower variance at all sub-sample sizes, but appears insensitive to small sub-samples.

The results for Linear-2 and LinExp were very similar as shown in Figures 7 and 8, although there were distinctly less stable fit for Baumhöfer-2014- τ . All three models shared a reduced standard deviation of σ_g when using SSD.

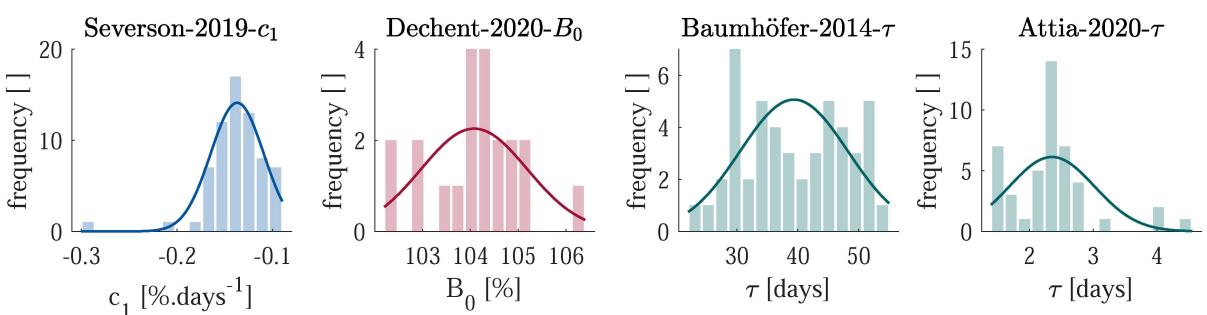


Figure 4. Generally, the inferred population distributions (lines) and sample distributions (bars) are similar, but there are some exceptions.

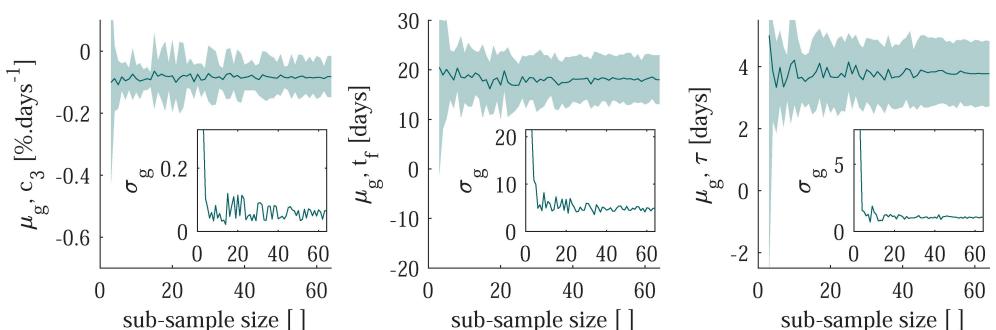


Figure 5. MLB results when fitting a single sub-sample draw have significant variability, shown here using Severson-2019 with the LinExp model. Lines show mean parameter estimates, shaded region is $1-\sigma$ about that mean, inset is σ_g

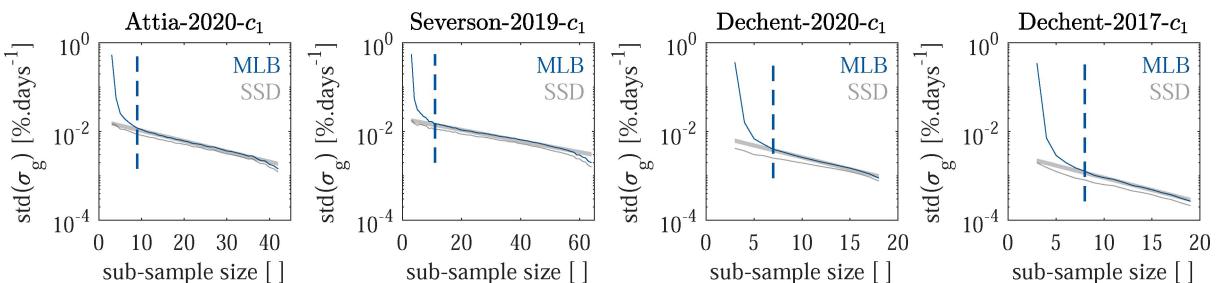


Figure 6. Smooth standard deviations are estimated by taking 1,000 random subsamples (with replacement), shown here with the Linear-1 model. MLB is Multi-Level-Bayes, SSD is sub-sample distribution, over the number of cells in the sub-sample.

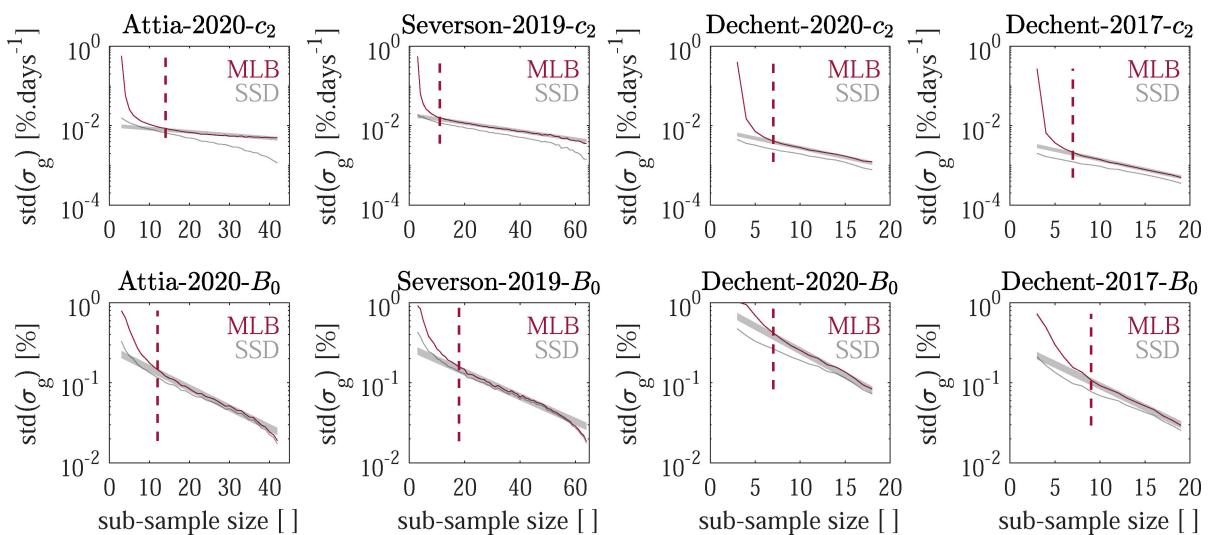


Figure 7. The standard deviations of the population-level standard deviation estimates for the Linear-2 model. MLB shows the Multi-Level-Bayes and SSD the sub-sample distribution results over the number of cells in the sub-sample.

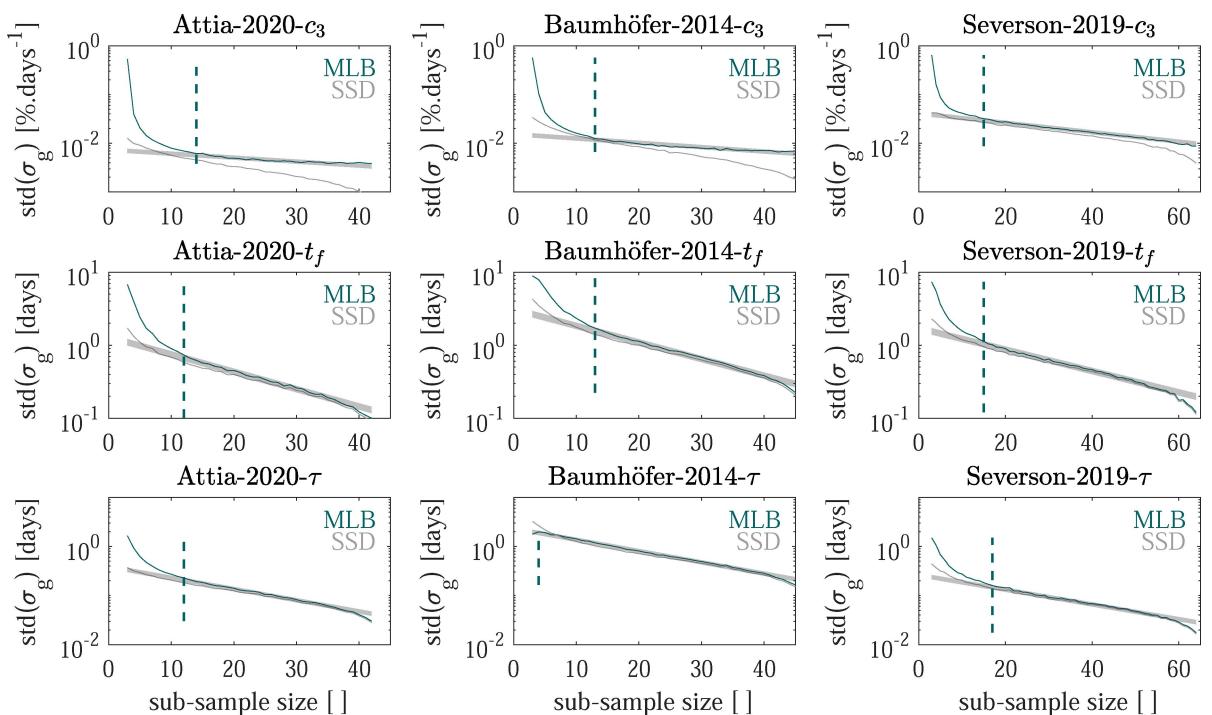


Figure 8. The standard deviations of the population-level standard deviation estimates for the LinExp model. MLB shows the Multi-Level-Bayes and SSD the sub-sample distribution results over the number of cells in the sub-sample.

The linear relationship between sub-sample size and the log of the standard deviations was deemed to represent a consistent fit. It was subsequently used to determine when an ‘effective’ sub-sample size had been reached. A model was considered well fit when the standard deviation of σ_g was within $\alpha = 10\%$ of this stable section, found using a linear extrapolation (as plotted in the figures).

Figure 9 shows the relationship between the number of cells required to achieve ‘stable’ population estimates, vs. the number of model parameters. The number is shown for all

model, dataset and parameter combinations. The mean required sub-sample sizes for a consistent fit were 9, 11 and 13 for the 1, 2 and 3 parameter models respectively.

5. Discussion

The number of cells required to fit the various models presented here and capture a stable estimate of the population variability is of order 10. For the simplest model, Linear-1, the

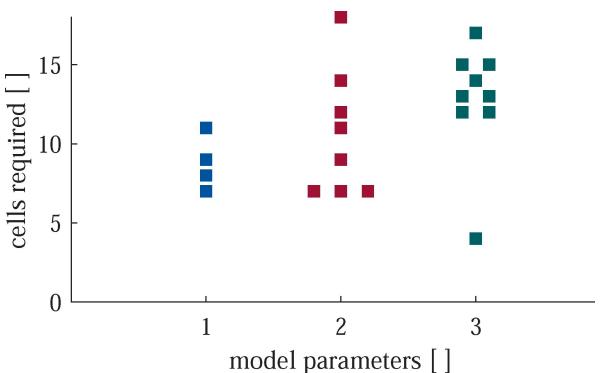


Figure 9. Testing with order 10 cells is required to achieve stable population estimates, with the number increasing as the model complexity increases. (Offset x-axis values were used to show identical results.)

number was 9 cells, and for the most complex LinExp model, the number increases to 13. The results understandably suggest that increased model complexity leads to an increase in the number of cells required to be tested to achieve a stable estimate of the population variability.

The multi-level Bayesian approach produced consistent parameter estimates from sub-samples. Given that cell-to-cell variability is an important phenomenon impacting battery performance, the estimated distributions are an invaluable tool to use in empirical modelling. Simple sample distribution techniques are limited to estimates of spread within the domain of the sample and hence showed less sensitivity to sub-sample size here when using random selection. The number of cells required to estimate population variability was fairly consistent across the datasets investigated here and was a stronger function of the model complexity than of the dataset. However, future work could test the robustness of this conclusion across a wider range of datasets.

The standard deviation of the σ_g estimates reduced as sub-sample sizes were increased. In most cases, the SSD and MLB results also approached the same values as sub-sample sizes increased because the two techniques will return similar results at high sub-sample sizes. At low sub-sample sizes SSD was limited to the variability of the sub-sample, whereas MLB was less certain, resulting in higher values for both σ_g and its standard deviation. In this case, SSD appears to have been artificially confident as an estimate of the population distribution.

The chosen threshold condition for a well fit σ_g parameter resulted in consistent results. The same consistency was also found when using other threshold values of α . The hypothesis that sub-sample size increases with model complexity appears to be supported. However, it would be useful to explore this in more depth using larger datasets.

In the derivation of the MLB approach, we assumed there to be no correlations between parameters in the prior probability distributions. That assumption was found to be questionable in two cases here. Future work should explore the impact of this on population modelling.

MLB was found to be a robust fitting mechanism when researching this work. For example, outliers were successfully ignored in Figure 4 where the sample distribution was not easily described as Gaussian. Our current approach assumes a Gaussian distribution at the population level. In the case of a bi-modal (or multi-modal) population distribution, we expect that the MLB method would respond by estimating a wide standard deviation, but this has not been tested. Extending the method proposed in this paper to other population distributions would be an interesting subject for future study.

Various ageing mechanisms are likely responsible for the degradation datasets considered in this paper. In the case of the Severson dataset, it is likely that degradation was largely caused by lithium plating,^[25] while on the Dechent-2020 dataset, covering layer formation and jellyroll deformation are key to degradation.^[41]

Between the datasets small differences were found. For the Linear-2 model the datasets Attia-2020 and Severson-2019 show a higher cell number threshold required for population estimation. The reasons for the trends we see are varied. The underlying mechanisms may be caused by increased variation of cycling conditions within the dataset. In order to capture higher usage variation in addition to intrinsic variability the number of cells will likely need to be increased. Future research could look into quantifying both use variability and manufacturing variability at the same time.

The fact that more complex models required more cells to be tested at each test point is challenging for battery lifetime experiments, since it could increase greatly the number of test channels and cells required in long term ageing experiments. We did not extrapolate to higher numbers of parameters or to other models, but it is reasonable to assume that the issues explored here will be present in other, more complex cases.

One challenge with the technique used here is that it relied on limited size samples from the population. Future work could explore whether larger sample sizes lead to similar results as found here.

Finally, further work is required to investigate the impacts of cell-to-cell variability using more complex physics-based models of battery ageing,^[32] which can have 5–10 or more parameters in addition to the 20+ parameters of the required underlying electrochemical model. Openly available ageing datasets are at present too small to enable meaningful calculation of population parameter variability for such complex models using the methods we outline, since the number of parameters is significantly more than 1–3. One approach in a future study could be using synthetic data to study more complex models with more parameters and test parameter identifiability.^[42]

6. Conclusions

Simple empirical battery capacity fade models were fitted to a variety of ageing datasets to quantify the number of cells required to estimate the variability of the underlying population. The number of cells required to give a stable population

variance estimate was found to vary according to the number of parameters in a given model. Respectively, 9, 11 and 13 cells are estimated to be required for models with 1, 2 and 3 parameters. Both sample statistics and population estimates were shown to stabilise with under 20 cells in most cases but this relied on there existing a Gaussian distribution of parameters within the sample, otherwise 20 cells were required.

For capacity curve fitting, perhaps the biggest challenge going forward is the selection of appropriate ageing model order and structure. This should be done not just by looking to what functions fit the capacity profiles best, but which functions produce the most reliable parameter distributions when looking at a dataset as a whole.

There was insufficient data here to test these results and conclusions as a function of variability caused by differences in usage, but this would be an interesting future exploration topic. Also, model selection across larger datasets is a challenging problem. For example, some of the battery capacity fade trajectories in this study fitted well to a linear degradation stage followed by an exponential decay starting from some knee point. However some of the resultant sample distributions cannot be confidently used to calculate basic summary statistics.

Understanding and quantifying battery cell-to-cell manufacturing variability is an open research topic, and this work represents an initial step. These results form a useful order of magnitude guide, for those undertaking long term battery ageing experiments, of what is needed to capture manufacturing variability.

Author contributions

P.D. and S.G.: conceptualisation, software, data curation, writing – original draft, visualisation. F.H.: software, writing – review and editing. S.J.: methodology, software, writing – review and editing. D.U.S.: supervision, funding acquisition. D.A.H.: conceptualisation, supervision, funding acquisition, writing – review and editing.

Acknowledgment

The authors would like to thank P.M. Attia, K. Severson and coworkers [25] and [37] S.G. is funded by EPSRC, UK, and Siemens Ltd. F.H. gratefully acknowledges the financial support of Engie. P.D., F.H. and D.U.S. gratefully acknowledge the financial support by Bundesministerium für Bildung und Forschung (BMBF 03XP0302C, 03XP0320A) and Engie. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: ageing · Bayesian approach · degradation · lithium-ion batteries · modeling

- [1] Nobel Committee. Press release: The Nobelprize in chemistry 2019. <https://www.nobelprize.org/prizes/chemistry/2019/press-release/>, 2019.
- [2] S. J. Harris, D. J. Harris, C. Li, *J. Power Sources* **2017**, *342*, 589–597.
- [3] T. Baumhöfer, M. Brühl, S. Rothgang, D. U. Sauer, *J. Power Sources* **2014**, *247*, 332–338.
- [4] Markus Schindler, Johannes Sturm, Sebastian Ludwig, Julius Schmitt, Andreas Jossen, *eTransportation* **2021**, *100102*.
- [5] K. Rumpf, M. Naumann, A. Jossen, *J. Energy Storage* **2017**, *14*, 224–243.
- [6] Y. Li, Kailong Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, H. E. Hoster, *Renewable Sustainable Energy Rev.* **2019**, *113*, 109254.
- [7] S. Santhanagopalan, R. E. White, *International Journal of Electrochemistry* **2012**, *1*–10.
- [8] M. Dahmardeh, Z. Xi, *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg* **2019**, *5*, (020909).
- [9] S. Orcioni, A. Ricci, L. Buccolini, C. Scavongelli, M. Conti, In *2017 13th Workshop on Intelligent Solutions in Embedded Systems (WISES)*, pages 15–21, June **2017**.
- [10] X. Liu, W. Ai, M. N. Marlow, Y. Patel, B. Wu, *Appl. Energy* **2019**, *248*, 489–499.
- [11] S. Paul, C. Diegelmann, H. Kabza, W. Tillmetz, *J. Power Sources* **2013**, *239*, 642–650.
- [12] D. Shin, M. Poncino, E. Macii, N. Chang, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **2015**, *34*, 252–265.
- [13] S. Rothgang, T. Baumhofer, D. U. Sauer, *2014 IEEE Vehicle Power and Propulsion Conference (VPPC)* **2014**, *1*–6.
- [14] M. Baumann, L. Wildfeuer, S. Rohr, M. Lienkamp, *J. Energy Storage* **2018**, *18*, 295–307.
- [15] B. Kenney, K. Darovich, D. D. MacNeil, I. J. Davidson, *J. Power Sources* **2012**, *213*, 391–401.
- [16] M. Dubarry, C. Pastor-Fernández, G. Baure, T. F. Yu, W. D. Widanage, J. Marco, *J. Energy Storage* **2019**, *23*, 19–28.
- [17] D. Beck, P. Dechent, M. Junker, D. U. Sauer, M. Dubarry, *Energies* **2021**, *14*, 3276.
- [18] G. Lenze, H. Bockholt, C. Schilcher, L. Froböse, D. Jansen, U. Kreuer, A. Kwade, *J. Electrochem. Soc.* **2018**, *165*, A314–A322, 2018.
- [19] M. Duquesnoy, I. Boyano, L. Ganborena, P. Cereijo, E. Ayerbe, A. A. Franco, *Energy and AI* **2021**, *5*, 100090.
- [20] M. Elliott, L. G. Swan, M. Dubarry, G. Baure, *J. Energy Storage* **2020**, *32*, 101873, 12.
- [21] S. F. Schuster, M. J. Brand, P. Berg, M. Gleissenberger, A. Jossen, *J. Power Sources* **2015**, *297*, 242–251.
- [22] I. Zilberman, S. Ludwig, A. Jossen, *J. Energy Storage* **2019**, *26*, 100900.
- [23] A. Devie, G. Baure, M. Dubarry, *Energies* **2018**, *11*, 1031.
- [24] M. Lucu, E. Martínez-Laserna, I. Gandiaga, K. Liu, H. Camblong, W. D. Widanage, J. Marco, *J. Energy Storage* **2020**, *30*, 101410.
- [25] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, R. D. Braatz, *Nat. Energy* **2019**.
- [26] P. Fermín-Cueto, E. McTurk, M. Allerhand, E. Medina-Lopez, M. F. Anjos, J. Sylvester, G. dos Reis, *Energy and AI* **2020**, *1*, 100006.
- [27] S. Lehner, *Reliability Assessment of Lithium-Ion Battery Systems with Special Emphasis on Cell Performance Distribution*. Shaker, PhD thesis, RWTH Aachen University, 1 edition, March **2017**.
- [28] F. An, L. Chen, J. Huang, J. Zhang, P. Li, *Sci. Rep.* **2016**, *6*, 35051.
- [29] C. I. Ossai, N. Nagarajan Raghavan, *Batteries* **2017**, *3*, 32.
- [30] Y. Jiang, J. Jiang, C. Zhang, W. Zhang, Y. Gao, Q. Guo, *J. Power Sources* **2017**, *360*, 180–188.
- [31] M. Ecker, J. B. Gerschler, J. Vogel, S. Käbitz, F. Hust, P. Dechent, D. U. Sauer, *J. Power Sources* **2012**, *215*, 248–257.
- [32] J. M. Reniers, G. Mulder, D. A. Howey, *Journal of The Electrochemical Society* **2019**, *166*, A3189–A3200.
- [33] X. Hu, L. Xu, X. Lin, M. Pecht, *Joule* **2020**, S2542435119305859.
- [34] D. U. Sauer, Time-series cyclic aging data on 48 commercial nmc/graphite Aanyo/Panasonic UR18650E cylindrical cells, **2021**, doi.org/10.18154/RWTH-2021-04545.
- [35] P. Dechent, L. Willenberg, M. Eckert, D. U. Sauer. Time series data on cycle tests on Samsung NR18650-35E, **2021**, doi.org/10.18154/RWTH-2021-07965.

- [36] P. Dechent, S. Lehner, D. U.Sauer, Time series data on cycle tests on Samsung NR18650-15L1, **2021**, doi.org/10.18154/RWTH-2021-08740.
- [37] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, W. C. Chueh, *Nature* **2020**, 578, 397–402.
- [38] M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, M. Jenkinson, S. M. Smith, *NeuroImage* **2004**, 21, 1732–1747.
- [39] K. P. Murphy, *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. MIT Press, Cambridge, MA, **2012**.
- [40] S. Theodoridis, Chapter 13 - Bayesian Learning: Approximate Inference and Nonparametric Models. In Sergios Theodoridis, editor, *Machine Learning (Second Edition)*, pages 647–730. Academic Press, January **2020**.
- [41] L. Willenberg, P. Dechent, G. Fuchs, M. Teuber, M. Eckert, M. Graff, N. Kürten, D. U. Sauer, E. Figgemeier, *J. Electrochem. Soc.* **2020**, 167, 120502.
- [42] A. Aitio, S. G. Marquis, P. Ascencio, D. Howey, *IFAC-PapersOnLine* **2020**, 53, 12497–12504.

Manuscript received: July 5, 2021

Revised manuscript received: August 4, 2021

Version of record online: September 24, 2021