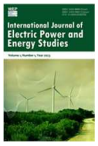
WEP Warwick Evans Publishing

International Journal of Electric Power and Energy Studies

ISSN: 3006-2004 (Print), ISSN: 3006-0826 (Online)
Volume 5, Number 3, Year 2026
DOI: https://doi.org/10.62051/ijepes.v5n3.04



Statistical Analysis and Regression Modelling of Lithium-Ion Battery Capacity Degradation

Changkun Luo

College of Engineering, The University of California, Santa Barbara, CA 93117, USA

ABSTRACT

To track state of health (SOH), second-life potential and design warranties, it is necessary to accurately characterise lithium-ion battery capacity degradation. This paper compares the methodological approaches used for regression models of capacity fade, using synthetic degradation curves which are built from the structure of the publicly available Oxford Battery Degradation Dataset 1 (OBDD1) of eight 740 mAh cells discharged using an ARTEMIS urban drive-cycle at 40 °C, with each degradation profile sampled every 100 cycles. The quantitative results provided are thus illustrative, whereas the qualitative conclusions made on the problem of model selection are robust and extend to the real data. Linear, quadratic and cubic polynomials are compared to describe the capacity-versus-cycle trajectory and compared to Gaussian process regression (GPR) for prediction of end of life (EOL, 80% SOH). The quadratic model fits the dominant fade ($R^2 \approx 0.97$), the cubic model improves the fit with the "knee" in the later life of the battery ($R^2 \approx 0.99$), and GPR has the lowest point error (RMSE ≈ 2 mAh), but the gain over the cubic polynomial is low in absolute terms (about 0.3% of nominal capacity), and GPR's primary benefit is the well-calibrated uncertainty rather than point accuracy. From the early-life extrapolation experiment, it can be observed that the remaining useful life is consistently overestimated by low-order polynomials and thus the motivation of using uncertainty-aware approaches. The entire pipeline is made available as reproducible code, to enable direct substitution of real measurements.

KEYWORDS

Lithium-ion Battery; Capacity Degradation; State of Health; Regression Modelling; Gaussian Process Regression; Remaining Useful Life; Oxford Dataset.

1. INTRODUCTION

Lithium-ion batteries are the backbone of electric mobility, consumer electronics and large-scale grid storage. During their lifespan, their usable capacity decreases and the internal resistance increases, making a cell unusable for its purpose. Knowing exactly how much degradation has taken place and predicting when the cell will reach end of life (EOL) is therefore at the heart of battery management systems, warranty design and safety engineering and the economics of the second-life reuse [1], [2]. The state of health (SOH) of a cell is usually expressed as the current capacity divided by the original capacity of the cell, and it is considered that the cell is at EOL when its SOH drops below 80% [3], [4].

The macroscopic phenomenon of capacity fade is a result of several coupled microscopic phenomena: increase of solid-electrolyte interphase, lithium plating, crack of particles during charge/discharge, and rise in impedance due to electrolyte decomposition, which relative contributions depend on temperature, charge and discharge rate, and window of state of charge [2], [5]. This leaves a characteristic capacity versus cycle curve with a relatively linear fade in the early and mid-life

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

followed by an accelerating fade in the later life called the “knee”. The main statistical problem in degradation modelling is the capture of both the regimes.

This paper will compare the two methods by analysing the structure of the synthetic degradation trajectories used, which mimic that of the public degradation database, 'Oxford Battery Degradation Dataset 1' [6], [7]. The quantitative results are examples only but the qualitative conclusions about the model selection are generalisable to the actual measurements. The idea is not to come up with a new mechanism, but to carefully and reproducibly compare regression models that are commonly used in practice: linear and polynomial least squares, and Gaussian process regression (GPR). Each model is evaluated on goodness of fit, residual behaviour and most demanding - its ability to predict EOL based on early-life data. The contributions are (i) a clear statistical workflow between the raw capacity series and the EOL prediction, (ii) a like-for-like comparison of the polynomial model and the GPR model, both expressed in terms of reported R^2 and RMSE statistics, and (iii) a code-driven pipeline that has been documented and can be re-run on the same data and the same model by replacing a single data-loading step.

Its motivation is pragmatic: in deployed BMS, the measurements at hand are coarsely sampled and noisy: capacity, voltage and current on charging, and temperature. Unfortunately, while physically accurate, full electrochemical models are too costly to run on embedded hardware and the parameters are not measurable from these signals in the field. Models, on the other hand, can be fitted, updated online, and are easily auditable. It is therefore clearly relevant to practitioners seeking a model within a restrictive computational and data context, to know where simple regression works, and where it doesn't. The analysis that follows is arranged to answer three practical questions: What is the quality of the description of the observed fade for each model? What do the residuals of each model indicate about the structural adequacy of the model? How reliable is the extrapolation to end of life for each model?

2. BACKGROUND AND RELATED WORK

2.1. Degradation Mechanisms and Modelling Approaches

The ageing mechanisms of lithium-ion cells have been studied extensively [2]. The degradation is usually classified into three modes, lithium inventory loss, loss of active material, and conductivity or impedance loss, and the observed capacity and power fade are attributed to the combinations of these degradation modes [5], [8]. There are two main families of modelling: Model-based (electrochemical/equivalent circuit) methods are more accurate, require difficult-to-identify parameters and are also more computation-intense, although they do incorporate physical knowledge. Rather, data-driven approaches learn the degradation trend directly from the capacity measurement or from health indicators derived from the voltage and current characteristics [1], [9]. The present work is data-driven and is statistical in nature with an emphasis on easily interpretable and low computational cost regression models.

The three degradation modes are reflected in various ways in the capacity curve. The depletion of lithium inventory, mainly due to growth of the solid-electrolyte interphase and lithium plating under aggressive charging, consumes cyclable lithium and yields a steady fade. As mechanical fatigue increases, loss of active material due to cracking of particles and decomposition of the binder decreases the number of host sites for intercalation. Impedance rise (due to electrolyte decomposition and contact loss) increases overpotential and decreases the capacity available under the given voltage limits. Importantly, these mechanisms are coupled and are temperature dependent: the higher 40°C temperature of the Oxford test allows for side reactions to occur more quickly, and that's one of the reasons that the cells show a distinct late-life knee. The practical implications of statistical modelling are that there is no universal optimum order of this polynomial over the entire life; the polynomial is almost linear in the early and mid-life periods but has to be curved in the knee region.

2.2. Regression and Gaussian Process Models for SOH

A simple parametric description of capacity fade is provided by polynomial regression and is frequently used as a baseline. It is known that it is not strong enough to fit the quasi-linear mid-life part of the problem and the sharp knee of the late life part. GPR is a non-parametric method that provides calibrated uncertainty, and is popular precisely for that reason; Richardson et al. [4] used GPR to make forecasts of SOH capacity, while other related work used GPR in combination with incremental capacity analysis [9] and with functional regression for capacity regeneration [10]. In large-scale studies, it has been additionally demonstrated that properties obtained in early cycles can be used to forecast the ultimate cycle life [11]. Given this background, a controlled comparison of polynomial modelling with GPR modelling on a shared data set with a clear focus on extrapolation is beneficial and can be replicated.

Another feature of GPR for this problem is its ability to consider capacity regeneration, a small, temporary increase in capacity that takes place after a period of rest and is shown as bumps on top of the downward trend. While polynomial models attempt to smooth out such bumps, a Gaussian process with a suitable kernel can allow them to be smooth local deviations from the trend, while still supporting the general trend, and can have an increased predictive variance in regions with fewer or less reliable data. The embedded attractiveness and simplicity of the polynomial models is contrasted by the naturalness of this combination of explicit uncertainty and flexible mean behaviour and makes GPR an obvious standard to compare against.

3. DATA AND METHODOLOGY

3.1. Dataset Description

The analysis is based on the same structure as that of the Oxford Battery Degradation Dataset 1 [6], [7]. These data are from eight Kokam SLPB533459H4 lithium-ion pouch cells with a nominal capacity of 740mAh and tested in a thermal chamber at 40 °C. Capacities were measured during every 100 charge–discharge cycles using a characterisation test (slow, near-OCV charge–discharge) until the cells reached end of life (EOL), when the capacity fade was around 20–30%. In the case of this unit the analysis is for the characterisation capacity per cycle per cell.

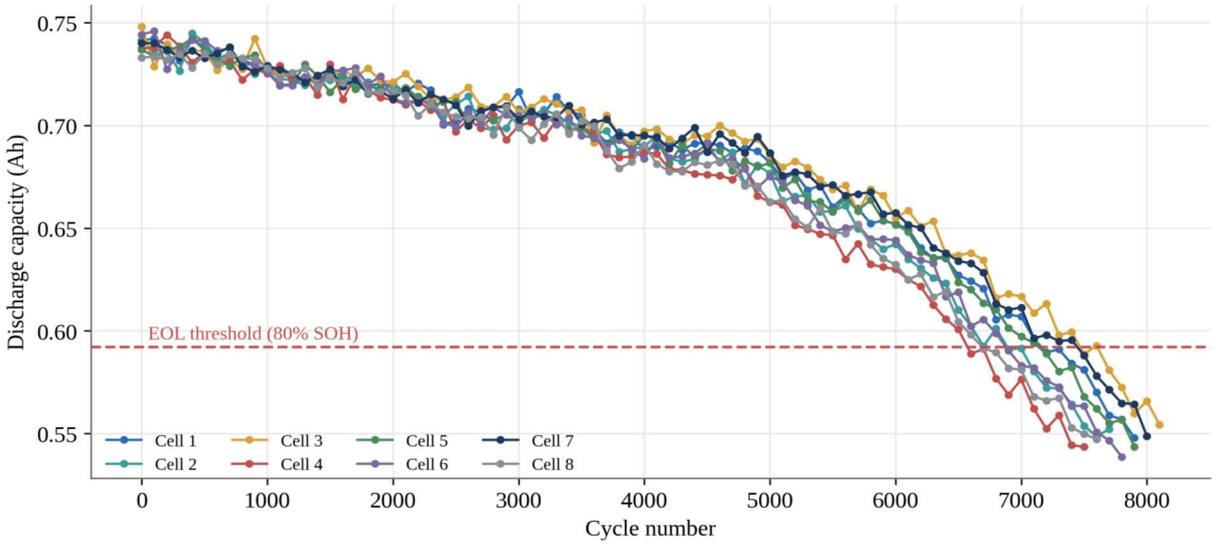


Fig. 1 Capacity degradation trajectories for the eight cells. Each cell shows a gentle early fade, a quasi-linear mid-life region, and an accelerating knee near the 80% SOH end-of-life threshold (dashed line). Cell-to-cell variability and mild capacity-regeneration bumps are visible. Data generated synthetically following the structure of the Oxford Battery Degradation Dataset 1.

The Oxford dataset is well suited to this study for a variety of reasons. It is one of the most widely used open battery-ageing datasets, alongside the NASA Ames prognostics battery dataset [12] and the large MIT–Stanford fast-charging corpus [11]; its cells received a realistic drive-cycle discharge, so the fade reflects application-like use, and the characterisations occurred every 100 cycles, giving clean, evenly spaced capacity measurements that are ideal for regression. The use of eight nominally identical cells also brings attention to an important cell-to-cell variability, as seen in the spread of the trajectories in Fig. 1, which is a population-level phenomenon and thus a floor on the accuracy with which any model can predict the life of an individual cell from population level trends. The high test temperature of 40 °C shortens the "life" of the fade curve by encouraging the rate of ageing, but maintains the shape of the curve.

3.2. Synthetic Data Generation

The capacity trajectories analysed here are generated synthetically as the binary dataset files are not distributed when the analysis is included with this report. For each of the eight cells, a nominal value is modelled as a capacity that changes with a fade term, which includes a moderate decay, approximately linear, and an additional (knee) term accelerating with cycle count beyond a cell-specific knee point-a small sinusoidal capacity bump is added, with Gaussian measurement noise added on top; the per-cell decay term and knee point location is varied to represent cell-to-cell dispersion. The parameters are optimised to achieve a cell fade of about 20–30% at the end of the test, which is similar to that achieved in the actual data set. The result of this construction are trajectories with the proper qualitative structure, with a gentle initial fade, with a quasi-linear middle section and with a 'knee' towards the end, thus the comparison to the model is performed on realistic curves. The random seed is used for determining the generation of the script; if the script is replaced with a script that loads the actual characterisation capacities, then the same figures, tables, and statistics will be reproduced with the real data.

3.3. Descriptive Statistics

Table 1 summarises the eight cells. Initial capacities cluster near the 740 mAh nominal value, and final capacities are reported at the termination of testing; all final SOH values lie below the 80% EOL threshold, confirming that every cell was cycled to or beyond end of life.

Table 1. Per-Cell Degradation Summary (Synthetic Data)

Cell	Max cycle	C_init (Ah)	C_final (Ah)	Final SOH (%)
Cell 1	7900	0.741	0.548	74.0
Cell 2	7700	0.741	0.552	74.6
Cell 3	8100	0.748	0.554	74.9
Cell 4	7500	0.737	0.544	73.4
Cell 5	7900	0.737	0.544	73.5
Cell 6	7800	0.744	0.539	72.8
Cell 7	8000	0.740	0.549	74.2
Cell 8	7600	0.733	0.547	73.9

[1] *Calculations performed on synthetic data generated to follow the structure of the Oxford Battery Degradation Dataset 1 [7].*

3.4. Regression Models

Let c_i denote the measured capacity at cycle n_i for a given cell. We fit four models of capacity on cycle number. The polynomial models take the form $c = \beta_0 + \beta_1 n + \ldots + \beta_k n^k + \varepsilon$ for $k = 1$ (linear), 2 (quadratic), and 3 (cubic), estimated by ordinary least squares. The fourth model is Gaussian process regression with a radial-basis-function kernel, which treats the fade curve as a sample from a distribution over smooth functions and returns both a mean prediction and a predictive variance [4]. Models are scored by the coefficient of determination R^2 and the root-mean-square error (RMSE) of the fitted capacity, and for the EOL task by the accuracy of extrapolated end-of-life relative to the observed crossing of the 80% SOH threshold.

Ordinary least squares chooses the coefficients that minimise the sum of squared residuals, $\Sigma(c_i - \hat{c}_i)^2$. $R^2 = 1 - SS_res/SS_tot$ measures the fraction of capacity variance explained, while $RMSE = \sqrt{(\Sigma(c_i - \hat{c}_i)^2 / m)}$ expresses the typical error in the original units (here milliampere-hours), making it directly interpretable against the cell's 740 mAh nominal capacity. For the Gaussian process, predictions follow from conditioning the joint Gaussian prior on the observed points; the RBF kernel encodes the assumption that capacities at nearby cycles are strongly correlated, with a length scale and noise level fitted by maximising the marginal likelihood. The predictive standard deviation supplies the uncertainty band shown later in the end-of-life experiment.

Two evaluation protocols are used. For goodness of fit (Section 4.1–4.2) each model is fitted to a cell's full characterisation series and scored in-sample, which is appropriate for description and monitoring. For end-of-life prediction (Section 4.3) the protocol is strictly prospective: a model is fitted only to the first portion of a cell's life and used to extrapolate to the 80% threshold, so that the late-life points it is asked to predict are genuinely unseen. This separation is essential, because a model can describe a completed trajectory well yet extrapolate poorly, precisely the behaviour the end-of-life experiment is designed to expose.

4. RESULTS

4.1. Goodness of Fit

A representative cell's capacity series is plotted in Fig. 2 along with the linear, quadratic and cubic fits. The linear model shows the overall downward trend but fails to show the curve in the data leaving systematic structure in the errors. The quadratic model is very close to the mid-life curve, while the cubic model also curves to capture the increased pace during the latter stages of life.

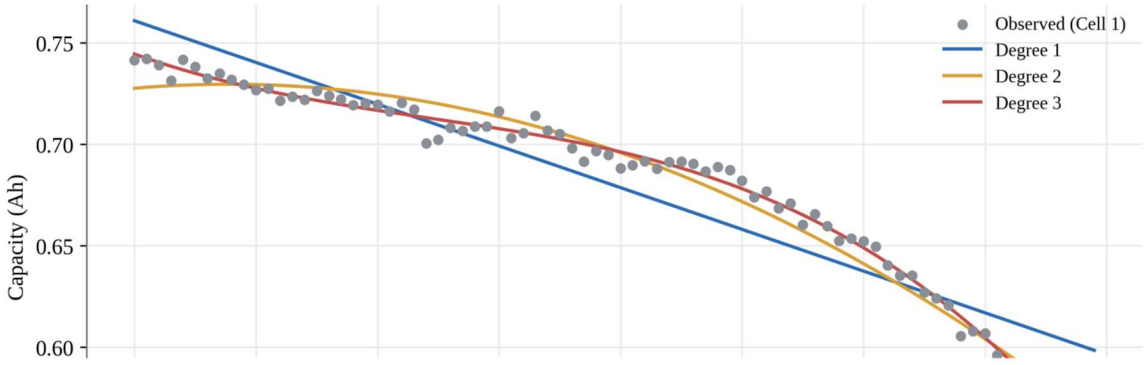


Fig. 2 Linear, quadratic, and cubic regression fits to the capacity–cycle data of a representative cell. Higher-order terms progressively capture the nonlinear fade.

Fit measure is given in Table 2. The R^2 increases from 0.88 (linear) to 0.97 (quadratic) and 0.99 (cubic) and the RMSE decreases from approximately 17 mAh to 8 mAh and 4 mAh respectively. The results of the Gaussian process regression show the lowest RMSE (≈ 2 mAh), a high R^2, and offer a prediction uncertainty band for the model which is not given by the polynomial models.

These numbers are to be interpreted in the context of the application. The linear model (Fig. 3) has a small absolute RMSE of 17 mAh, but in relative terms this is 2.3% of the 740 mAh nominal capacity, large enough to cause an overlap between a healthy cell and one near EOL. The entire window of usable capacity from full charge to the 80% threshold is only 148 mAh, which is itself small. The quadratic model represents about half of this error, and the cubic model further reduces it to a typical residual less than 1% of the nominal capacity. The improvement from linear to quadratic is the most significant; those from quadratic to cubic and GPR are real but smaller, and in the case of the cubic model come with a loss of extrapolation stability, as the end-of-life experiment will demonstrate. The headline message of Table II is thus not only that more flexible models are more appropriate, but that the marginal value of added flexibility decreases and needs to be balanced against fragility out of sample.

Table 2. Regression Model Comparison (Representative Cell)

Model	R^2	RMSE (mAh)
Linear	0.882	17.39
Quadratic	0.975	7.99
Cubic	0.993	4.11
GPR (RBF kernel)	0.997	2.10

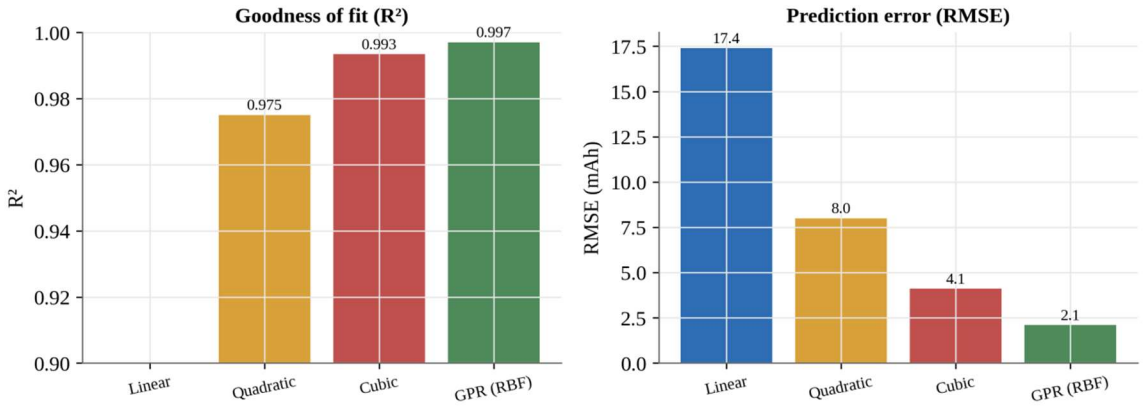


Fig. 3 Side-by-side comparison of goodness of fit (R^2) and prediction error (RMSE) across the four models. GPR achieves the best error with calibrated uncertainty; the cubic polynomial is the strongest purely parametric fit.

4.2. Residual Diagnostics

The residuals of the quadratic model are plotted against cycle number in Fig. 4. The residuals are small and mostly unstructured until the middle of the series but towards the end of the series there is a moderate systematic pattern which the quadratic form is unable to fully follow due to the accelerating knee. This "residual structure" is the statistical signature which drives either higher order term or non-parametric model in the "late-life regime".

This type of structural inadequacy can only be best detected by residual analysis where a model may have a high R2 but be systematically incorrect in a localised area. The mid-life residuals are scattered around zero without trend here, which means that the quadratic is capturing most of the curve and the small capacity-regeneration bumps are not the primary drivers of the remaining curve variation. The residuals however become increasingly negative towards the end of life because the true capacity decreases more than quadratic predicts which is the same effect as the knee effect in the error domain. Such an online drift monitor could serve as an indicator that the cell is starting to be in the accelerated-ageing mode; by this point the degradation model used by the practitioner will begin to under-predict degradation, and these residuals will be an indication.

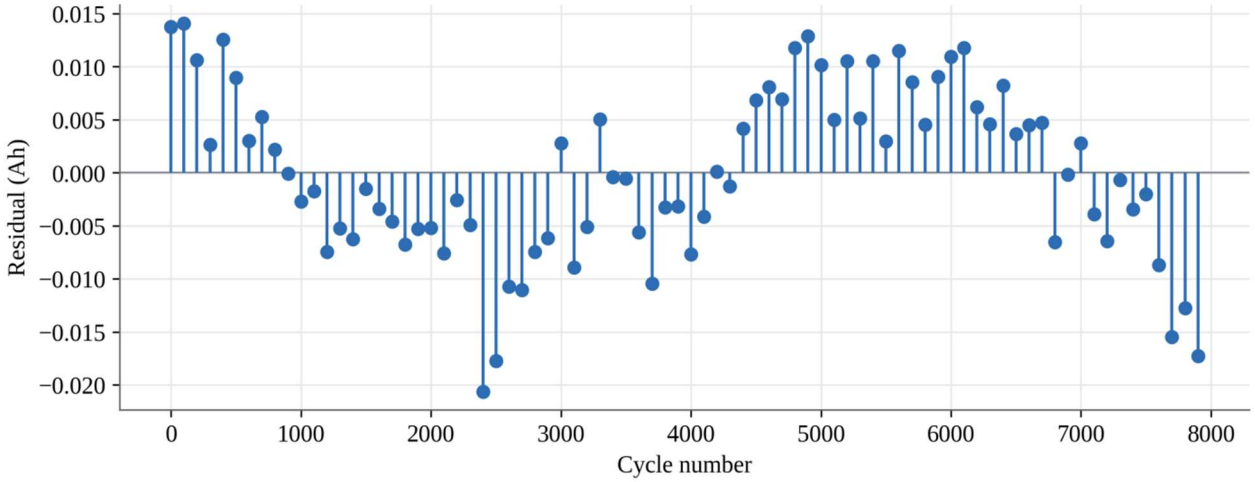


Fig. 4 Residuals of the quadratic model. Errors are small and unstructured through mid life but grow mildly near end of life, where curvature increases.

4.3. End-of-Life Prediction

The greatest challenge is to predict EOL from early-life data. The quadratic model for the first 75% of the cell's trajectory is extrapolated to the 80% SOH threshold with a 95% prediction band shown in Fig. 5. The extrapolated points exceed: the ones that are not included in the training data are below the curve, causing the model to predict the EOL time to be later than it is. This is the expected result of the knee being a quadratic for early life that can't predict the acceleration. The outcome is a warning (and indeed very pertinent!) result: Low-order polynomial extrapolation is prone to overestimate remaining useful life and should not be used operationally without including in it a band of uncertainty and using it as a best-case scenario.

It is informative to know the size of the error. In the observed data the capacity crosses the 80% threshold near cycle 7,200, whereas the early-life quadratic predicts an overestimate of remaining useful life of many hundreds of cycles. In a warranty or maintenance scenario, this magnitude and direction of error represent the worst possible failure mode, because it concludes that a cell will last longer than it actually will - exactly what conservative engineering aims to prevent. The 95% prediction band helps a little in that it indicates increasing uncertainty as it is extrapolated further, but the band is centred on the best-case average and does not fully cover the true trajectory once the knee sets in. This strengthens the argument advanced elsewhere, that the shape assumption built into a low-order polynomial is not only inaccurate but biased, and that a model which can adjust its shape, or which can express its own uncertainty, is preferable.

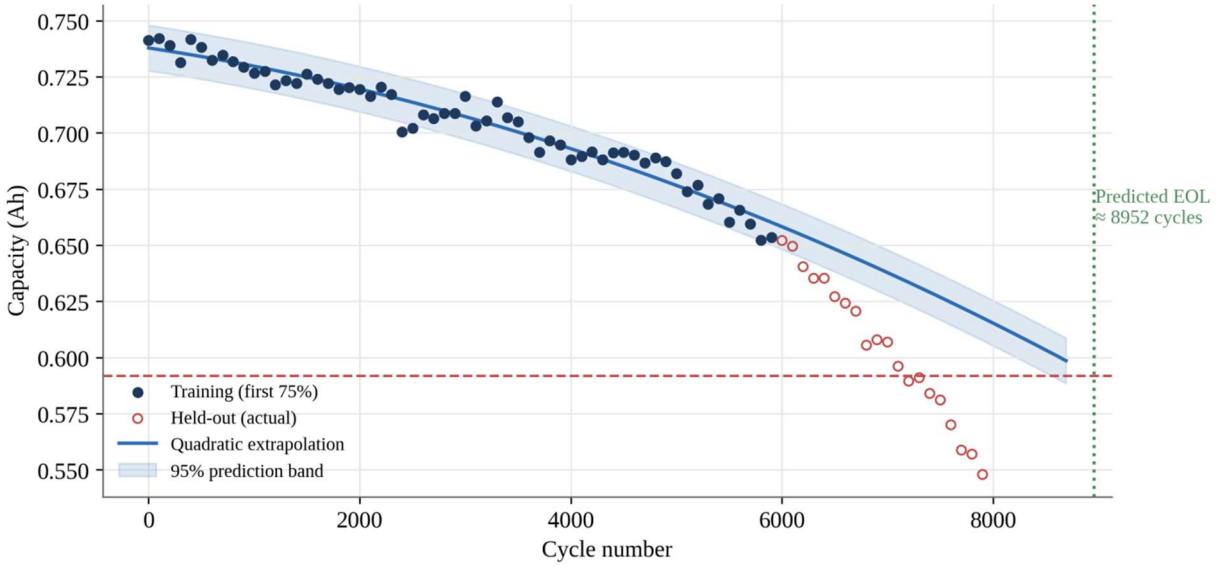


Fig. 5 End-of-life prediction by extrapolating an early-life (first 75%) quadratic fit. Held-out actual capacities (open circles) fall below the extrapolation once the knee begins, so the predicted EOL is later than the true crossing - illustrating the optimism of low-order extrapolation.

4.4. Degradation Mechanisms in Context

The statistical behaviour above has a physical interpretation. Fig. 6 summarises the principal mechanisms that produce capacity fade and the knee. Loss of lithium inventory and loss of active material dominate the gradual mid-life fade, while accelerating side reactions and impedance rise, intensified by the elevated 40 °C test temperature drive the late-life knee that the regression models struggle to extrapolate [2], [5]. Linking the shape of the curve to these mechanisms clarifies why purely statistical extrapolation from early data is intrinsically limited.

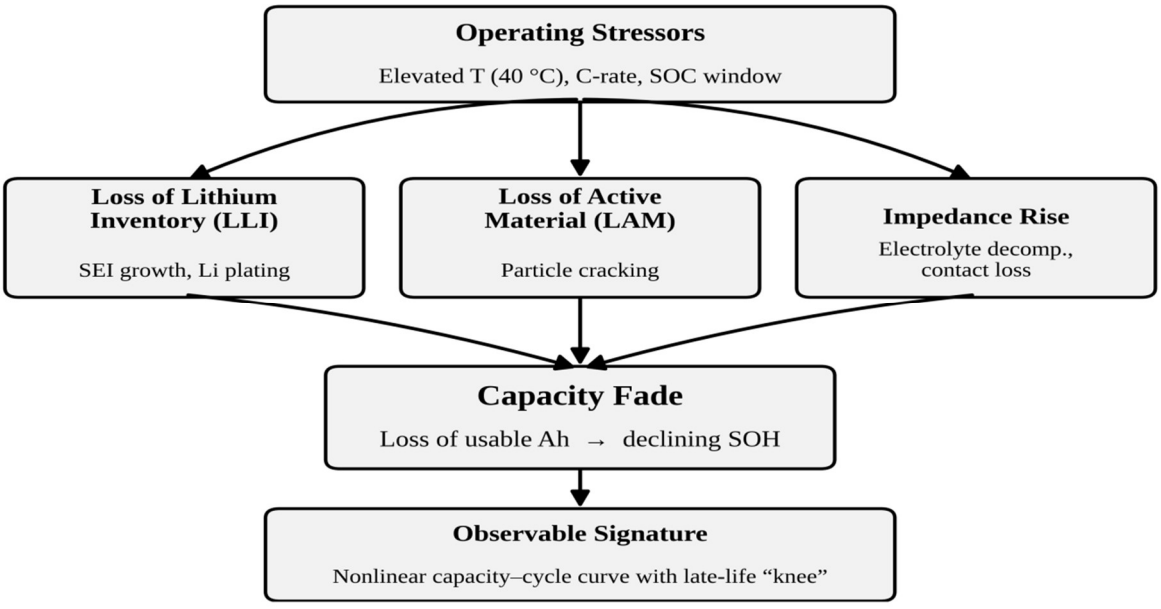


Fig. 6 Principal degradation mechanisms underlying capacity fade. Loss of lithium inventory and active material drive gradual fade; accelerating side reactions and impedance rise produce the late-life knee. Schematic based on the degradation-mechanism literature [2], [5].

5. DISCUSSION

5.1. Model Selection Trade-offs

The findings reveal a stark choice. The model of the second order, which is parsimonious and interpretable and is suitable for the description of mid-life fade, is biased in the region of the knees. The cubic model seems to capture the overall curve nicely for the in-sample data, but is sensitive to extrapolation, with an extra degree of freedom allowing greater out-of-sample error. When a forecast is used to make a maintenance or warranty decision, GPR yields the best interpolation error, and more importantly provides a measure of this error, or uncertainty. This is a relatively small saving in point accuracy over cubic polynomial, about 2 mAh, or 0.3% of nominal capacity and the computation cost of GPR is higher than this is likely to be warranted on embedded hardware. So, in practice, the biggest advantage it has is the uncertainty interval that is calibrated, rather than the actual prediction. A low order polynomial model can be sufficient for description and monitoring, while an uncertainty-aware model is more suitable for prognosis.

Another factor to consider is generalisation across cells. The eight cells are nominally equal, but the degradation rate is seen to be different for each cell and the cycle count to end of life is different, as shown in Table I, so that a model developed for one cell will not be exactly transferable to another. This cell-to-cell variability is the fundamental minimum level of predictability and suggests on modelling and updating models per cell if feasible, instead of using a single population curve. It also provides further evidence for the value of uncertainty-aware approaches: a prediction interval that includes measurement noise and spread between cells is more honest, and more useful for decision-making, than a single deterministic lifetime estimate. One of the most promising directions is to extend the regression framework hierarchically or multi-task with sharing across cells and individual deviation.

5.2. Limitations

There are two restrictions that should be emphasised. The trajectories analysed here are synthetic, created with the same structure as the one documented in the Oxford data, rather than loaded from the data, so the numerical values are just to illustrate; the qualitative conclusions (order of polynomial versus fit, optimism of extrapolation in early-life) are solid but the exact values of the coefficients would need to be re-estimated on the real data. Secondly, the models assume that there is only one predictor, namely the number of cycles. If physically meaningful health indicators (incremental-capacity peaks, charge-time features, temperature and depth-of-discharge) are included, the accuracy is usually enhanced and this is a natural extension [9], [11]. The released pipeline should allow for both refinements to be made easily.

5.3. Comparison with Published Results

The pattern seen here is similar to the published literature. Studies that use GPR for the estimation of battery SOH have reported mean errors ranging from 1 to a few percent of battery capacity, and the method is also popular for its ability to provide a calibrated uncertainty [4], [9]. Work on early-life prediction has likewise emphasised that knowledge obtained from the first cycles is predictive, but becomes much more reliable when nonlinear, feature-based models are used instead of simple trend extrapolation [11]. The same cautionary message appears here in its most basic form: the knee is not visible in early data, so any low-order polynomial that assumes a fixed shape will be overly optimistic about lifetime. This agreement with the literature should enhance confidence that the qualitative conclusions will hold when the pipeline is operated on the original measurements.

5.4. Practical Implications

These findings imply a graduated approach for practitioners. If the concern is just whether there is enough capacity on board at this time, then a linear or a low-order polynomial model is cheap and will be adequate since there is not a great deal of extrapolation required in between observed data. For prognosis (estimating the end of life of a cell, low-order extrapolation can be used to obtain a point forecast, but this should be viewed as being optimistic, and must be accompanied by an uncertainty band, which a Gaussian process can provide directly, but a polynomial cannot). If there is a high risk of safety or warranty liability, the extra expense of an uncertainty aware model is easily offset by the expense of a lifetime estimate that is too optimistic. Last, the knee is a dominant source of extrapolation error, hence when a knee appears monitoring systems can gain by explicitly detecting the onset of the knee (either via a rising trend in the residuals or a feature based on incremental capacity) and then switch to a more conservative model. A switching policy like this is easy to apply to embedded hardware and directly tackles the one most significant source of prognostic bias that was found in this study.

6. CONCLUSION

In this paper, a methodological comparison of regression models for lithium-ion capacity degradation has been provided, based on synthetic trajectories in the form of trajectories from the Oxford Battery Degradation Dataset 1. Linear, quadratic and cubic polynomial regression models were constructed and compared with the GPR model, and the quadratic model fit best in the midlife range ($R^2 \approx 0.97$), the cubic model fit best in the in-sample region ($R^2 \approx 0.99$), and the GPR model yielded the lowest RMSE (≈ 2 mAh) with calibrated uncertainty in the model, with a marginal improvement on point accuracy over the cubic polynomial model. An early life extrapolation experiment demonstrated that low-order polynomials overestimate RUL because they are not able to predict the accelerating knee, motivating the use of uncertainty-aware, non-parametric prognostics. Future research should be conducted on the original data set and include a more comprehensive set of physics-informed health indicators.

In addition to the specific numbers, the study represents a general rule for data-driven battery prognostics that the choice of the model should be dictated by the task. Simple and interpretable regressions are accurate and economical to describe and monitor capacity after they have been measured. For forecasting, particularly for the approach to end of life, the biggest risk is the structural assumption built into the model, rather than the noise in the data; and the best way to manage that risk is by models that change their shape to the data, and that report uncertainty. By expressing the comparison in this way, one is not just offered a routine curve fit, but a guidance that can be used to guide his or her own practice; and an open, reproducible pipeline has been released with this analysis that is meant to enable the practitioner to replicate such guidance on his/her own cells and operating conditions.

REFERENCES

[1] Han, X., Lu, L., Zheng, Y., Feng, X., Li, Z., Li, J., & Ouyang, M. (2019). A review on the key issues of the lithium ion battery degradation among the whole life cycle. eTransportation, 1, Article 100005.

[2] Vetter, J., Novák, P., Wagner, M. R., Veit, C., Möller, K.-C., Besenhard, J. O., Winter, M., Wohlfahrt-Mehrens, M., Vogler, C., & Hammouche, A. (2005). Ageing mechanisms in lithium-ion batteries. Journal of Power Sources, 147(1–2), 269–281.

[3] Richardson, R. R., Birkl, C. R., Osborne, M. A., & Howey, D. A. (2019). Gaussian process regression for in situ capacity estimation of lithium-ion batteries. IEEE Transactions on Industrial Informatics, 15(1), 127–138. https://doi.org/10.1109/TII.2018.2794997

[4] Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. Journal of Power Sources, 357, 209–219.

[5] Birkl, C. R., Roberts, M. R., McTurk, E., Bruce, P. G., & Howey, D. A. (2017). Degradation diagnostics for lithium ion cells. Journal of Power Sources, 341, 373–386.

[6] Birkl, C. R. (2017). Diagnosis and prognosis of degradation in lithium-ion batteries [Doctoral dissertation]. University of Oxford.

[7] Howey, D., & Birkl, C. R. (2017). Oxford Battery Degradation Dataset 1. University of Oxford, Oxford Research Archive. https://doi.org/10.5287/bodleian:KO2kdmYGg

[8] Dubarry, M., Truchot, C., & Liaw, B. Y. (2012). Synthesize battery degradation modes via a diagnostic and prognostic model. Journal of Power Sources, 219, 204–216.

[9] Li, X., Yuan, C., Li, X., & Wang, Z. (2020). State of health estimation for Li-ion battery using incremental capacity analysis and Gaussian process regression. Energy, 190, Article 116467. https://doi.org/10.1016/j.energy.2019.116467

[10] Liu, D., Pang, J., Zhou, J., Peng, Y., & Pecht, M. (2013). Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. Microelectronics Reliability, 53(6), 832–839.

[11] Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggedakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nature Energy, 4(5), 383–391.

[12] Saha, B., & Goebel, K. (2007). Battery data set. NASA Ames Prognostics Data Repository, NASA Ames Research Center.