Electrochemical Science Advances

WILEY-VCH

ELECTROCHEMICAL SCIENCE ADVANCES Open Access

RESEARCH ARTICLE OPEN ACCESS

Extending Equivalent Circuit Models for State of Charge and Lifetime Estimation

Limei Jin[1,2,3] | Franz Philipp Bereck[1,2] | Josef Granwehr[1,2] | Christoph Scheurer[2,3]

[1]Institute of Technical and Macromolecular Chemistry, RWTH Aachen University, Aachen, Germany | [2]Institute of Energy Technologies - Fundamental Electrochemistry (IET-1), Forschungszentrum Jülich GmbH, Jülich, Germany | [3]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany

Correspondence: Christoph Scheurer (scheurer@fhi.mpg.de)

Received: 7 August 2024 | **Revised:** 16 December 2024 | **Accepted:** 17 December 2024

Funding: This study was supported by the Helmholtz AI Cooperation Unit (HAICU), project 'Intelligent, individual battery management using spectroscopy and machine learning' (i2Batman).

Keywords: Chebyshev polynomials | electrochemical impedance spectroscopy | equivalent circuit modelling | state of charge | state of health

ABSTRACT

Equivalent circuit modelling (ECM) of electrochemical impedance spectroscopy (EIS) data is a common technique to describe the state-dependent response of electrochemical systems such as batteries or fuel cells. To use EIS for predictive assessments of the future behaviour of such a system or its state of health (SOH), a more elaborate digital twin model is needed. Developing a robust and continuous SOH estimation poses a formidable challenge. In this study, a framework is presented where ECM parameters are expanded in a high-dimensional Chebyshev space. It facilitates not only a mapping of the state of charge dependence with robust boundary conditions but also an extension towards a more abstract SOH description is possible. Such methods can bridge the gap between the experiment and purely data-driven techniques that do not rely on fitting of experimental data using a priori defined models. In the absence of long-time impedance measurements of a battery, quasi-Monte Carlo sampling can be employed to generate differently aged synthetic battery models with limited experimental impedance data. As additional data becomes available, the space spanning the possible states of a battery can be gradually refined. The developed framework, therefore, allows for the training of big data models starting with very little experimental information and assuming random fluctuations of the model parameters consistent with available data.

1 | Introduction

In recent years, the rapid adoption of electric vehicles and grid energy storage has underscored the importance of high-performance and reliable battery systems [1]. Their efficiency and longevity are crucial for achieving sustainable energy solutions. In an ideal battery, charging and discharging should be infinitely reproducible under a wide range of reasonable operating conditions. However, in reality, batteries degrade as they undergo multiple charging and discharging cycles [2]. To understand and characterize their distribution of relaxation time behaviour over time, the concept of separating timescales is useful. Over a long timescale, state of health (SOH) designates a cell's overall condition and performance compared to its initial, pristine state [3]. SOH is a fairly complex and nuanced concept that is not solely determined by the slow ageing of a battery. For example, SOH can be affected by sudden and unexpected events, such as overcharging, overheating or other adverse conditions, which can have a rapid and negative impact on the battery's health [4]. A battery can also experience partial recovery from specific types of damage or degradation, particularly when promptly addressing the adverse conditions [5]. Therefore, it is oversimplifying to express SOH as, for example, the number of charging/discharging cycles or the capacity ratio. Conversely, at a short timescale,

state of charge (SOC) is relatively straightforward to define and measure. It refers to the behaviour and changes that occur relatively quickly and often within one charging/discharging cycle and can be expressed as the ratio of the charge currently available to the total charge capacity of the battery [6].

The characterization of a battery's behaviour as a function of SOC can be obtained, for example from electrochemical impedance spectroscopy (EIS) measurements, which probe its voltage response to a sinusoidal current input (or load) signal (or vice versa) across a wide range of frequencies at different charge states. These experimental impedance spectra are then parameterized by equivalent circuit models (ECM) to create a robust battery model that reflects the dynamic changes in the battery's state and enables the reproduction of the battery's behaviour by simulating its response to a given current. The identification of relevant sets of ECM parameters, while preserving their physical significance, is achieved using distribution of relaxation time (DRT) methods [7] and curve fitting across multiple EIS spectra. By integrating EIS with ECM, a detailed characterization of the battery's behaviour at various SOC levels is enabled and can be stored in a look-up table format. The resulting look-up tables for ECM parameters can be further represented by interpolating polynomials, compressing the data while preserving significant trends that have been observed in these parameters as a function of SOC. Several conventional polynomial interpolations come with known inherent limitations when dealing with high-order approximations, making them susceptible to issues like over-fitting. To address these challenges and enhance the robustness of the parameter quantification process, orthogonal basis fitting methods, such as Chebyshev polynomials with their min-max properties, emerge as a potential approach as discussed in detail in Subsection 2.1.

The ECM parameters, or rather their explicit dependence on SOC, also vary with SOH and performing individual EIS fitting at each SOH could potentially yield a more accurate degradation model. However, conducting sets of EIS measurements for a manifold of SOH scenarios is time-consuming and challenging due to the complexity of determining every possible ageing state, possibly requiring accelerated ageing experiments. Additionally, battery degradation differs significantly depending on usage patterns, making it difficult to generalize findings across different operating conditions. We, therefore, propose to assess the variation of ECM parameters with ageing by generating populations of these parameters within a short range of anchor points in an abstract SOH space, using a baseline established from experimental data. As a showcase, we choose an experiment that provides data for only three representative ageing states: a fresh battery cell, a significantly aged cell and, finally, a damaged battery. These three systems are measured and parameterized at various SOC levels by EIS and ECM, resulting in three distinct sets of polynomial coefficient vectors that represent their ageing states. Based on the vectors derived from the experiment, small variations can be applied to generate sufficiently sized populations of synthetic systems that represent different ageing states. It allows for the recognition of the ageing progression from fresh to aged and ultimately to damaged conditions. The generation of polynomial vectors is conducted using the quasi-Monte Carlo (QMC) sampling method, which is characterized by its low discrepancy properties. It ensures that the generated samples are evenly distributed across the parameter space, enabling efficient exploration of the high-dimensional landscape of battery parameters while minimizing gaps and clustering typically associated with traditional random sampling methods. Details will be elaborated in Subsection 2.2.

2 | Methodology

As illustrated in Figure 1, this study demonstrates the application of capturing the nuances of battery degradation without the need for exhaustive EIS measurements at every SOH level. The generated numerical models are then used to simulate responses to small excitation signals, recording the corresponding voltage signals. The resulting pairwise current and voltage data can be expressed as impedance, which is then compared with measured impedance to validate the accuracy of the methods. The short-range population generation allows for rapid adjustments to the model as new data becomes available, ensuring that it remains relevant and reliable.

2.1 | Parameterization of ECM Elements as a Function of SOC

EIS can be used to measure a battery's response at multiple SOCs throughout charging/discharging cycles. The resulting impedance data are then commonly fitted to an assumed ECM for each individual cycle. However, this point-wise approach along the SOC dimension often yields noisy parameter values, mainly due to inherent measurement noise and the non-linear ECM fitting procedure. This can cause significant aliasing between the model parameters. Thus the process has to be repeated carefully without ageing the battery noticeably, to support the assumption of vanishing drift of the battery's condition on the slower SOH timescale.

For this work, we assume the simplified ECM for a coin cell that includes three different element types: a serial resistance, ZARC elements, and a Warburg element, to represent different aspects of its behaviour [8]:

$$Z(\omega) = R_s + \sum_{k=1}^{3} \frac{R_k}{1 + (j\omega\tau_k)^{\phi_k}} + \frac{\sigma}{\sqrt{\omega}}(1 - j)\coth(C_{\text{lim}}\sigma), \quad (1)$$

where R_s is the serial resistance. A ZARC element is functionally a parallel circuit of a constant phase element, representing a lossy capacitor, and a resistor. Here τ_k is the mean time constant, ϕ_k is the depression factor, and R_k is the charge transfer resistance of one of in total three ZARC elements. If $\phi_k = 1$, the ZARC behaves as a parallel RC circuit element and if $\phi_k = 0$, it behaves as an ideal resistor. For a finite space Warburg element, σ is the Warburg coefficient and C_{lim} is the limit of capacitance.

As recently shown, the initial and boundary conditions of ECM fitting can be extracted by a transform-based DRT analysis which allows one to reduce the ambiguity in the construction of ECM and thus over-fitting [7]. The fitted ECM parameters, such as R_s, R_k, τ_k, ϕ_k, σ and C_{lim}, vary as a function of SOC, thus Equation (1)

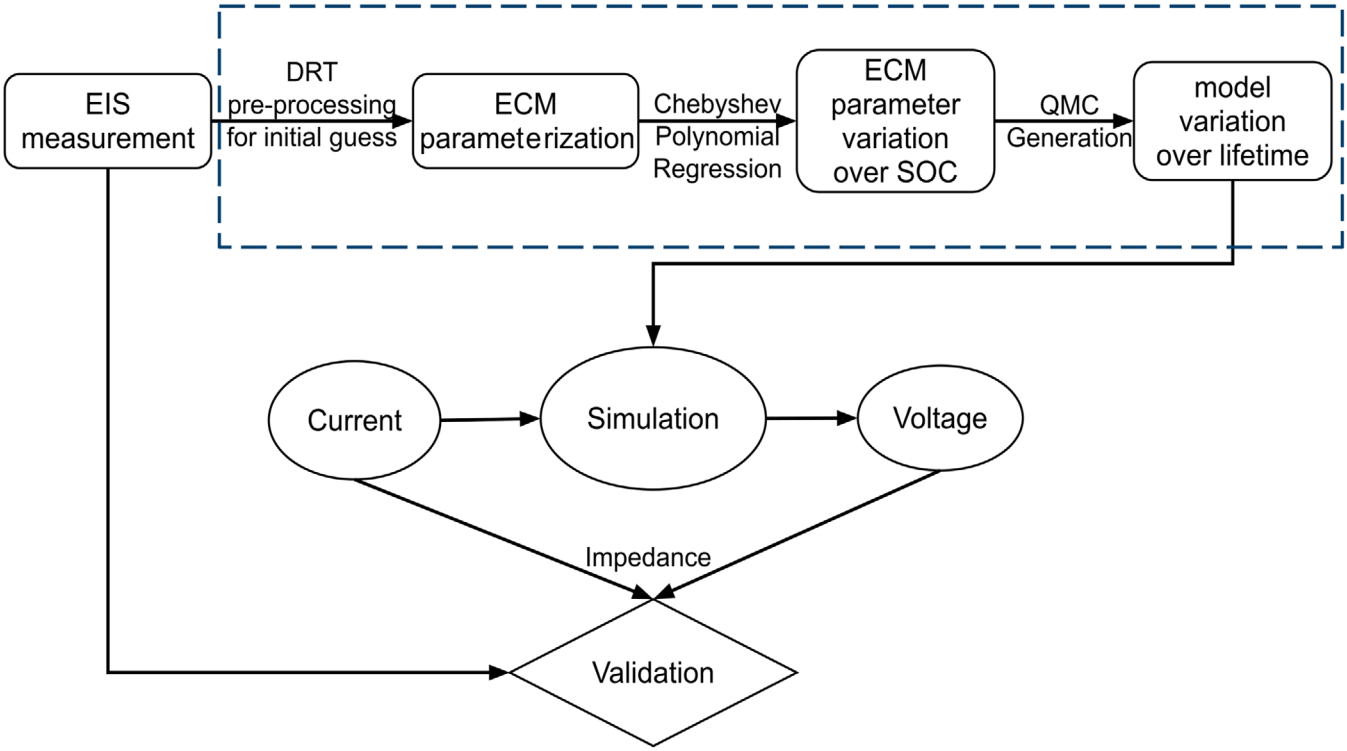

FIGURE 1 | The architecture of the proposed method.

should be generalized in the following form:

$$Z(\omega, \mathrm{SOC}) = R_s(\mathrm{SOC}) + \sum_{k=1}^{3} \frac{R_k(\mathrm{SOC})}{1 + (j\omega\tau_k(\mathrm{SOC}))^{\phi_k(\mathrm{SOC})}} + \frac{\sigma(\mathrm{SOC})}{\sqrt{\omega}}(1-j)\coth(C_{\mathrm{lim}}(\mathrm{SOC})\,\sigma(\mathrm{SOC})). \quad (2)$$

These dependencies can be quantified for each parameter as a function of SOC. However, the relationship between ECM parameters and SOC is often non-linear, making it difficult to find a simple mathematical model that accurately captures the dependencies. Additionally, a non-linear fit of the SOC dependence based on noisy ECM parameter estimates introduces additional sources of error and uncertainty into the parameter estimation process and increases the risk of over-fitting. Overfitting occurs when a model is too complex and captures noise or random variations in the data rather than the true underlying relationships. It leads to poor generalization and inaccurate predictions when the model is applied to new data.

This problem can be mitigated by using an orthogonal function basis, such as Chebyshev polynomials, to project data on the optimal solution. Chebyshev polynomials have the property of minimizing the maximum absolute error over a given interval [9]. This helps to prevent physically unrealistic large values of the fitted polynomials in regions with sparse data, allowing for more accurate fits in areas that are more densely sampled. Therefore, a more robust and precise representation of ECM parameter dependencies on SOC across the entire range is achieved. Additionally, it ensures that the parameterization is equally well-determined across the whole SOC range, reducing the risk of overfitting or underfitting in specific regions.

Unlike simple polynomial regression with monomial basis functions $\{x^k\}$, the use of orthogonal polynomials results in a well-defined least-squares functional approximation,

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x) \quad (3)$$

where a_n is the coefficient associated with the Chebyshev polynomial $T_n(x)$ of order n, and $T_n(x)$ is defined as:

$$T_n(x) = T_n(\cos\theta) = \cos(n\theta), \quad x \in [-1, 1] \quad (4)$$

with weight function $w(x) = \frac{1}{\sqrt{(1-x^2)}}$ in the scalar product [10]. Chebyshev polynomials of the first kind have the advantage of being well-behaved and oscillating within the interval $[-1, 1]$. This characteristic is particularly beneficial because the boundaries of the SOC range, such as 0% or 100%, are rarely reached in practice. By employing Chebyshev polynomials, potential issues related to extrapolation or poor fitting near the boundaries are thus additionally mitigated. For example, the relationship between parameter R_s and SOC, which needs to be mapped to the domain of definition of the Chebyshev polynomials, could be expressed as

$$\tilde{R}_s(\mathrm{SOC}) = a_1 T_1(\mathrm{SOC}) + a_0 T_0(\mathrm{SOC}), \quad \mathrm{SOC} \in [-1, 1]. \quad (5)$$

The goal is to find optimal coefficients a_n to minimize the error between estimated values for, for example, $\tilde{R}_s(\mathrm{SOC})$ and experimentally parameterized values $R_s(\mathrm{SOC})$. Using the discrete orthogonality property, coefficients can be approximated as

$$a_n = \frac{2}{N} \sum_{i=0}^{N-1} f(x_i) T_n(x_i), \quad (6)$$

where N is the total number of data points, $f(x_i)$ is the value of the function being fitted at the data point x_i and $T_n(x_i)$ is the value of the nth Chebyshev polynomial evaluated at x_i. This property simplifies the calculations and makes the convergence of the regression analysis faster, resulting in more efficient parameter estimation.

As a result, the parameters from Equation (2) depending on SOC are expressed as 32-dimensional Chebyshev coefficient vectors instead of 12×17 independent ECM parameter values at each SOH. These Chebyshev coefficients are thus used to create a compact and robust representation of the battery's history over a single charging/discharging cycle. For representing multiple cycles or an ageing procedure, these vectors of coefficients should be concatenated to form a comprehensive trajectory of the battery's entire history. This approach enables a thorough analysis of the battery's behaviour, performance degradation, and ageing effects over time.

2.2 | Model Variation Over Cell Lifetime

Representing SOH changes over multiple charging/discharging cycles through their effects on ECM parameters is challenging. SOH is a specific indicator that quantifies the current condition or health of a system relative to its original or ideal state. It provides a snapshot of the battery's health at a given point in time and it cannot be directly measured but is often inferred based on cycle numbers. However, relying solely on cycle numbers to infer SOH has limitations, especially when abnormalities and subsequent recovery are involved. Sudden abnormalities, such as rapid voltage drops or battery overheating, might temporarily accelerate degradation in batteries but may be followed by partial or nonlinear recovery. Cycle numbers alone may not capture the extent of this recovery. Besides, SOH is a multifaceted and interdependent concept encompassing capacity loss, impedance changes, ageing rates and more. Capturing all these dimensions accurately in a single set of parameters is demanding.

Alternatively, a data-driven approach may be chosen. Machine learning models help to parameterize and represent SOH as multidimensional clusters for quantifying and tracking the health of a battery relative to its initial state. It enables to monitor a lifetime or ageing trajectory and predict the remaining useful life of a battery based on various ageing scenarios. Nevertheless, obtaining reliable and comprehensive experimental data over a long period for machine learning makes a purely data-driven black-box approach time-consuming and resource-intensive. Parameterizing an SOH manifold within a Chebyshev space as described in the preceding section provides a means to amend scarce experimental data with computationally generated, synthetic battery models. As more measured data becomes available during the lifetime of the battery, the synthetic data can be gradually replaced, allowing for continuous refinement and improvement of the statistical models representing the battery's behaviour.

One vector of Chebyshev coefficients represents a battery's SOC-dependent impedance behaviour over a single cycle. Then, SOH is effectively depicted as distinct regions or *clouds* within a broader space covering the realizable Chebyshev vectors. These clouds dynamically shift and evolve as the battery undergoes the ageing process. Each cloud within this space signifies a particular SOH that a battery can manifest. Visualizing SOH in this manner helps in grasping the continuous and non-linear nature of the ageing process. It allows for a more comprehensive understanding of how the battery's health progresses over time, capturing the variability and nuances that come with different ageing patterns. The objective is to determine reasonable values for the variation of these coefficient vectors, effectively broadening their representation to cover specific regions of interest throughout the battery's operational lifespan, while maintaining sufficient resolution between the SOH regions to be able to unambiguously identify them in EIS measurements.

The QMC method generates high-dimensional vectors within a specified interval, aligned with uncertainty estimates of model parameters. By using quasi-random sequences, it samples the parameter space more uniformly and efficiently than traditional random sampling methods. Since Chebyshev coefficients for a single cycle are high dimensional as described above, QMC is particularly suitable for local sampling in this space.

To generate points in the parameter space, one selects a quasi-random sequence, such as the Sobol sequence [11], which has good low-discrepancy properties. The Sobol sequence provides 2^m points in a unit hypercube with d dimensionality $[0,1)^d$, where each point is obtained by bitwise exclusive or (XOR) operations on a set of direction numbers. Then the generated Sobol points are transformed to mimic the normal distribution $N(\mu, \Sigma)$. Therefore, generating a sufficient number of points n through QMC is described as

$$n = 2^m \in [0,1)^d \sim N(\mu, \Sigma). \tag{7}$$

The set of Chebyshev coefficients for the experimental anchor points is established through parameterization based on experimental impedance data, as detailed in Subsection 2.1. Each of the measured reference sets at a certain SOH gives one anchor point. To represent the battery's behaviour across its lifespan, the just described QMC sampling is then applied to generate SOH point clouds around these fundamental anchor points while ensuring appropriate boundary conditions. A simple method assumes high-dimensional spheres around each anchor point, with the radius set to half the distance between neighbouring anchor points, ensuring appropriate separation. This sampling process results in thousands of points that effectively populate the space around the experimental reference sets.

3 | Materials and Methods

3.1 | Electrical Simulation

The generated Chebyshev coefficient vectors at differently aged battery *clouds* are used to construct new ECMs at a variety of SOCs. These ECMs serve as numerical models that represent the battery's behaviour. The ISEA Framework [8], a real-time capable electrical simulation platform, was used to compute voltage responses of each ECM at its respective SOC to oscillating current profiles covering the experimental EIS conditions. The input time-based load profiles are sinusoidal waveforms, and for each time step an Euler scheme is used to calculate the voltage

response based on the differential-algebraic ECM equations [12]. The calculated voltage at every time step builds a voltage profile that corresponds to the current input. The impedance is then calculated as the ratio of voltage to current in the frequency domain.

3.2 | EIS Measurements

To fully parameterize a cell for simulation with the ISEA framework, EIS measurements at multiple SOCs of each cell were acquired using a BioLogic SP-200 potentiostat. To this end, commercial $LiCoO_2$ LIR2023 button cells were charged with 2C using the Constant Current-Constant Voltage (CC-CV) method until a residual charging current of 0.05 C was reached. After a resting period of 10 min, an EIS spectrum in galvanostatic mode between 200 kHz and 10 mHz was measured, followed by a discharge current of 0.1 C for 30 min to ensure approximately 5–7% SOC discharge. Before the next set of EIS measurements, the cell was rested for 1 h until a steady state was reached. This procedure was repeated until the cell was fully discharged.

To exactly determine the SOC values of each subsequent measurement step, the individual discharge capacities were calculated by integrating the discharge current over time. The sum of all capacity values resulted in the 0.1 C discharge capacity of the investigated cell.

Three cells were taken into consideration:

- a *fresh cell* that has seen no use before the parameterization other than calendaric ageing since production (37 mAh).
- an *aged cell* that has been used regularly in real-life applications and therefore has a lower capacity compared to the fresh cell (25 mAh).
- a *dead cell* that has gone through excessive cyclic and calendar aging and shows a significantly lower capacity compared to the fresh cell (16 mAh).

4 | Results and Discussion

4.1 | Approximation of ECM Parameters versus SOC by Chebyshev Polynomials

The determination of Chebyshev polynomials for the ECM parameters, derived from fits of experimental impedance data at various SOC values, is illustrated in Figure 2. The selection of suitable polynomial orders rests on the observed changes in Root Mean Squared Error (RMSE) and coefficient of determination, expressed as R^2.

As pointwise parameterized results demonstrate, the serial resistance R_s increases as SOC decreases, indicating that the battery is discharging [13]. This phenomenon is consistent with previously reported experimental results [14]. The evaluation results quantifying the relationship between SOC and R_s with varying Chebyshev polynomial degrees illustrate that the RMSE decreases, while R^2 increases as the model complexity grows. Selecting the polynomial degree that minimizes RMSE and maximizes R^2 without introducing overfitting requires careful consideration. Additionally, for developing a digital twin model, the objective is to balance accuracy with dimensionality reduction. For the case of R_s, the analysis indicates that RMSE and R^2 stabilize at a relatively low polynomial degree–typically first order. Beyond this point, additional complexity does not provide a meaningful improvement in model accuracy. This stabilization suggests that the relationship between R_s and SOC can be reasonably approximated as linear.

The parameters R_1, τ_1 and ϕ_1 correspond to the low-frequency ZARC element, with τ_1 being the largest time constant among the three ZARC elements. As observed, both R_1 and τ_1 show an increasing trend with SOC, exhibiting an approximately linear relationship. This behaviour is consistent with the role of the time constant $\tau = RC$, where increasing resistance during charging indicates a buildup of charge. This phenomenon may also reflect the formation and growth of the solid electrolyte interface layer, which typically contributes to increased resistance over time. The evaluation of RMSE and R^2 for these parameters suggests that a low-order polynomial, such as a first-order regression, is sufficient to capture their trends with reasonable accuracy.

In contrast, R_3 and τ_3, which represent the high-frequency ZARC element, display a decreasing trend with increasing SOC. Their RMSE and R^2 plots against polynomial order also suggest that a linear regression function is adequate for their representation.

The ZARC element at the middle frequency has more fluctuating values as a function of SOC, requiring higherdegree polynomial functions to accurately capture these dependencies. For instance, R_2 can be effectively described by a second-order polynomial function. However, RMSE and R^2 plots for τ_2 do not demonstrate clear stabilization within the expected polynomial degree range. In this study, the intersection point of these lines is chosen, corresponding to a third-order polynomial, as it strikes a balance between accurately fitting the data and controlling model complexity.

The ϕ parameters exhibit strong fluctuations, necessitating higher degree polynomial functions for adequate representation. Based on the strategy of selecting either stabilizing or intersecting points, ϕ_1 is best represented by a second-order polynomial, ϕ_2 by a third-order polynomial, and ϕ_3 by a first-order polynomial.

The Warburg coefficient σ and limit of capacitance C_{lim} for the diffusion process represent the straight line in the low-frequency region of the Nyquist plot. σ decreases as SOC increases, but conversely, C_{lim} increases with SOC. Both parameters can be expressed reasonably by second-order polynomial functions.

To evaluate the polynomials, the SOC value (in Figure 2 shown, e.g., for SOC = 70% by the dashed red line) is substituted into the defined Chebyshev polynomials and regular polynomial functions to obtain the corresponding parameter values. These values are then used to calculate the new impedance according to Equation (2). A 70% SOC serves as a representative midpoint in the SOC range, where charging and discharging effects are relatively balanced, making it highly relevant for practical

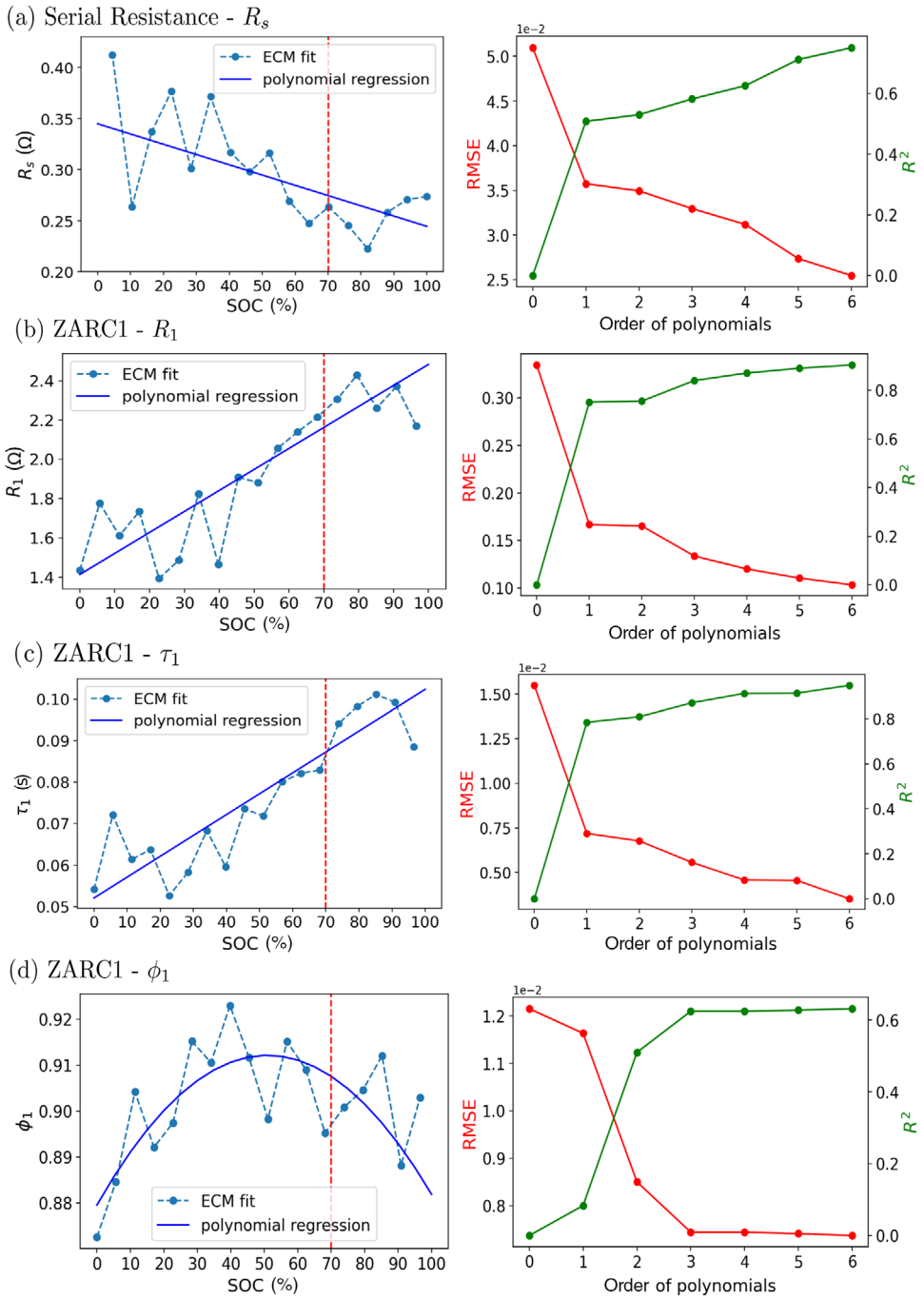

FIGURE 2 | Chebyshev polynomial regression of ECM parameter values as a function of SOC (left), and their RMSE and R^2 for different degrees of Chebyshev polynomial regression (right).

applications requiring optimal battery performance. In Figure 3, the mathematically calculated and experimentally observed EIS spectra, along with their residuals, are compared.

The maximum polynomial order for parameter fitting is constrained to four, ensuring a highly regularized model. Consequently, the differences between two polynomial expansion methods are not substantial. Although both fitting approaches yield visually satisfactory results on the Nyquist plot, the Chebyshev polynomial regression outperforms regular polynomial fitting in terms of residual distribution and overall fit accuracy. The more homogeneous distribution of residuals and the lower RMSE across the entire frequency range highlight its precision and reliability in capturing the impedance behaviour of the system.

While pointwise fitting of experimental EIS data provides valuable insights into battery behaviour at a specific SOC, it becomes impractical for applications requiring continuous or interpolated SOC data. For such scenarios, a global EIS fit offers a more efficient and scalable solution by consolidating information from all SOCs into a unified representation. This approach avoids the redundancies and inconsistencies that may arise when

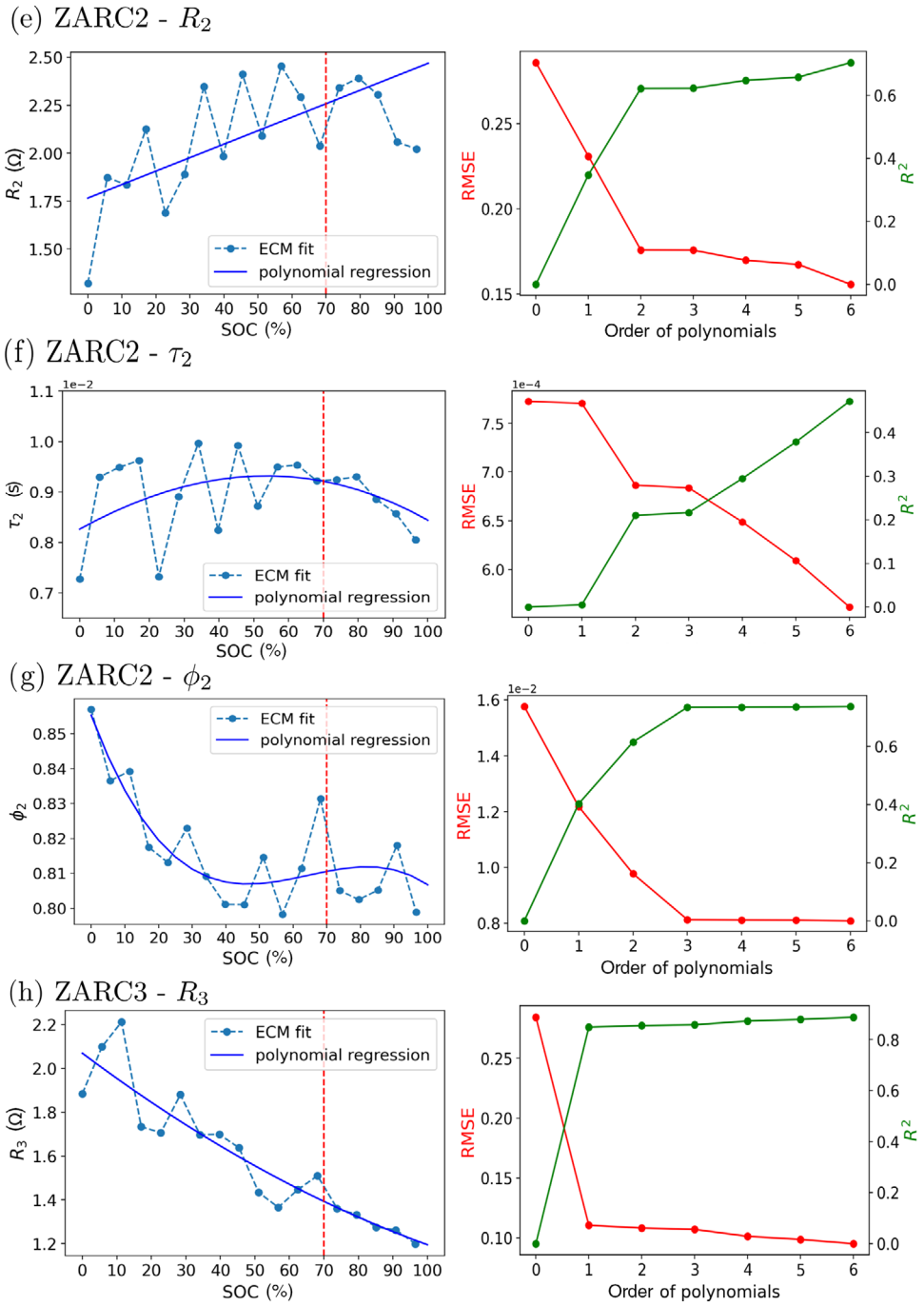

FIGURE 2 | (Continued)

performing separate fits for each SOC and enables smoother transitions between SOCs, making it particularly valuable for dynamic battery management systems and digital twin models. To achieve a robust global EIS fit, selecting appropriate starting parameters for the non-linear fitting process is essential. The Chebyshev coefficients taken from the pointwise approach, as described in this context, offer a reliable foundation for initializing. Details of the optimized Chebyshev coefficients, obtained through direct fitting of the experimental EIS data, are provided in the Supporting Information.

Chebyshev polynomial coefficients provide a capable and dependable alternative for representing traditional ECM parameters across the SOC range. This analysis focuses on an individual spectrum at a specific SOC. The subsequent section will delve into a comparative analysis across multiple spectra.

4.2 | Model Variation Over Lifetime

The Chebyshev coefficients determined above are based on ECM fitting to measured impedance data from a single cycle. Representing an ECM requires 32 coefficients, making visualization of these high-dimensional vectors challenging. Thus the parameter ϕ_1 was taken as an example, which is

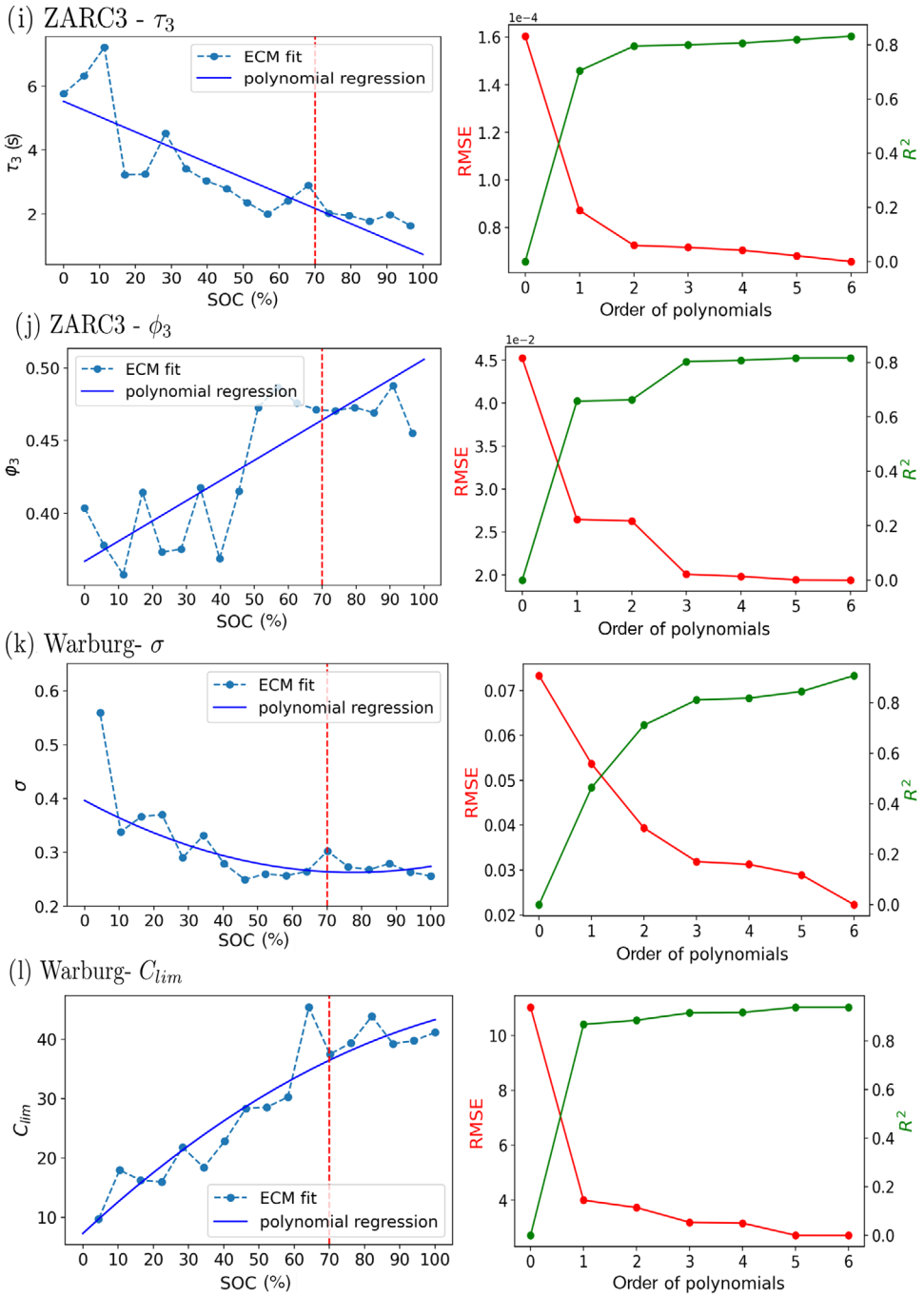

FIGURE 2 | (Continued)

parameterized by a second-order polynomial function with three coefficients, as visualized in Figure 4.

Three fundamental sets of coefficients, derived from experimental data and denoted as a *fresh cell* (red), an *aged cell* (blue) and a *dead cell* (green), serve as anchor points indicated by stars in this three-dimensional space. Using the QMC sampling method, additional coefficients are generated around these anchor points within spherical boundaries, as described in Subsection 2.2, and are visualized as dot clouds.

The generated coefficients serve as the basis for computing ECM parameter values and constructing polynomial functions across the SOC range. These new polynomial functions are required to align with the behaviour of the Chebyshev polynomials represented by the anchor points. If the generated polynomial functions deviate significantly from this behaviour, the corresponding points are discarded. For example, if a parameter value is expected to decrease with increasing SOC but the generated coefficients cause the function to yield values that increase with SOC, these coefficients are considered erroneous and are therefore excluded.

The analysis reveals that the *fresh cell* and *aged cell* groups are located in close vicinity, but still well separated, whereas the *dead cell* group is significantly farther away. It indicates that the healthy

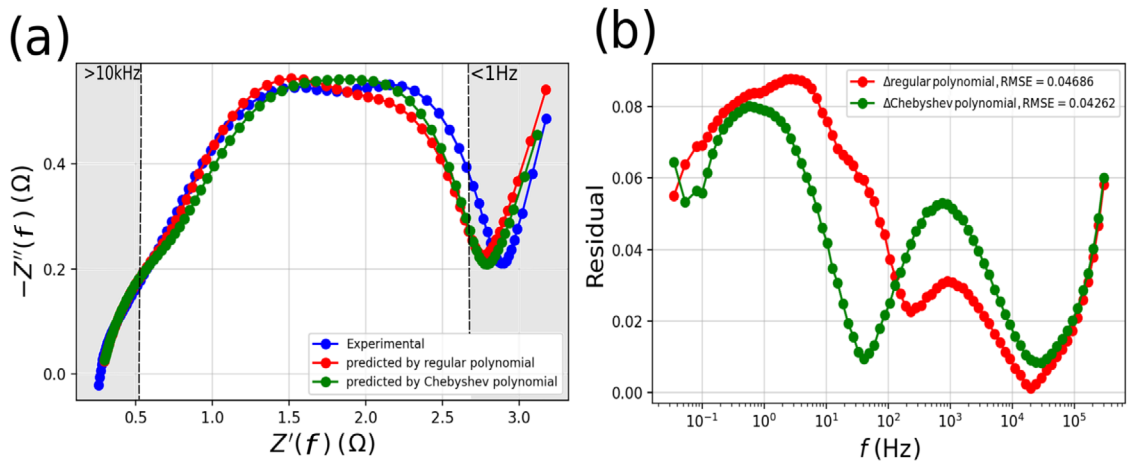

FIGURE 3 | The differences between the observed impedance from EIS measurement (blue) and predicted impedances from regular (red) and Chebyshev (green) polynomial regression models at SOC = 70% (a) and their residuals (b).

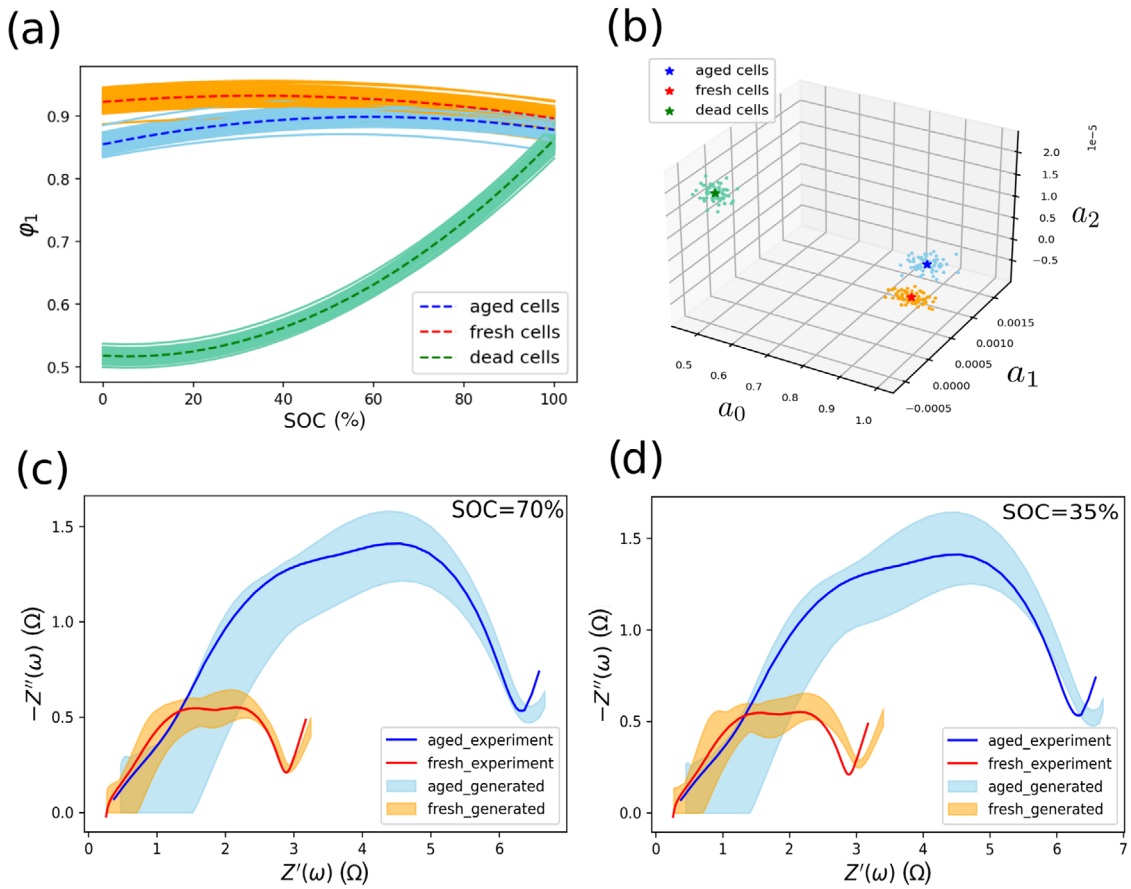

FIGURE 4 | The generated parameter ϕ_1 values as a function of SOC (a) and their Chebyshev representations for aged, fresh and dead cells (b). For better visualization of two closely located *fresh* and *aged* cell clouds, their experimental and generated impedance spectra at SOC = 70% (c) and at SOC = 35% (d) are shown.

batteries and unhealthy batteries are easily discriminated in the Chebyshev space.

Notably, ϕ_1 of *dead cells* increases dramatically with SOC, while the variation is minimal for healthy cells. At higher SOC values, the three sets of curves strongly overlap in Figure 4a, whereas at lower SOC values, they remain well separated. This observation suggests that batteries can be effectively grouped by their lifetime characteristics, even when examining a single parameter. Additionally, implicit SOC normalization should be considered at high SOC. 100% SOC corresponds to the maximum capacity that a cell can achieve at its current SOH. Therefore, while high

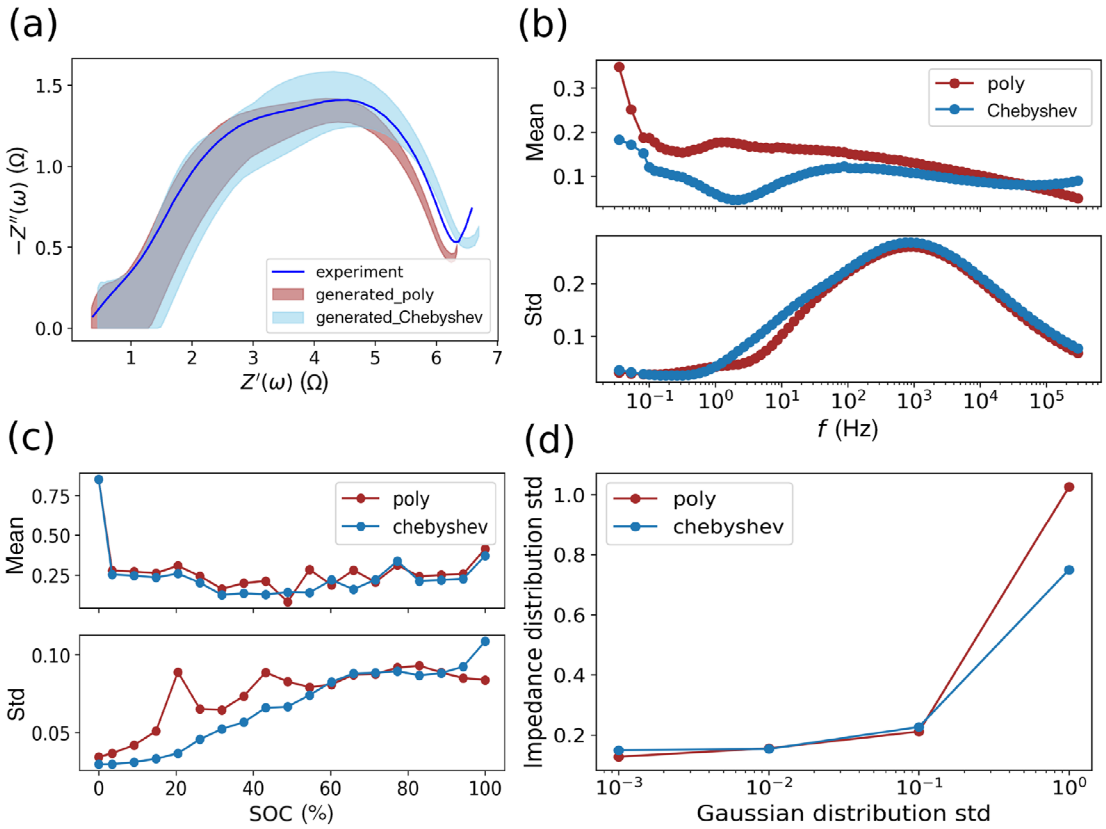

FIGURE 5 | The generated aged cell models based on Chebyshev polynomials and regular polynomials. (a) shows the impedance calculated from the generated parameters of the aged cell cloud at a single SOC; (b) illustrates the mean and standard deviation of residuals when compared to the experimental data; (c) shows residuals over multiple SOCs; (d) is impedance variance along with standard deviation of QMC generation over multiple SOCs.

SOC values reflect roughly the same range of behaviour for *fresh* and *aged cells*, they exhibit significantly different behaviour for *dead cells* due to their lower absolute capacity. An outlier is observed in the *fresh* and *aged cell* clouds, even at low SOC levels, emphasizing the importance of monitoring the generation of outliers in this synthetic data approach as the standard deviation of QMC generation increases.

Additionally, impedance data at SOC = 70% and SOC = 35%, derived from *aged* and *fresh* battery models, are presented in Figure 4c,d. These figures demonstrate significant differences in impedance between the two distinct ageing states while the generated models align closely with their respective anchor point models. This holds true despite the similarity in individual parameter values, as shown in Figure 4a. Furthermore, Figure 2a shows that the original values of internal resistance at SOC = 70% closely match the polynomial-fitted values. Therefore, SOC = 35% was chosen for further investigation, where the fitted internal resistance deviates notably from the original. As illustrated in Figure 4d, this minor variation in internal resistance does not significantly impact the overall behaviour, confirming that the generated models accurately represent the populations of specific ageing states. The distinct separation between *fresh* and *aged* groups and analysis persists, regardless of SOC values, validating the robustness of the modelling approach.

Expanding this approach, QMC generation applied to the rest of parameters builds high-dimensional vectors representing differently aged cells, enabling the storage of large physics-based synthetic datasets for further machine learning approaches (Figure 5). The impedance generated by Chebyshev polynomials closely aligns with the experimental data, consistently outperforming standard polynomials. The overall mean value of Chebyshev polynomials is smaller than that of regular polynomials. Additionally, Chebyshev polynomials exhibit a smoother transition in response to changes in SOC, ensuring a more stable and realistic representation of battery behaviour.

The sensitivity of impedance variation to the standard deviation in QMC generation also highlights that the expansion rate of generated clouds differs between polynomial classes. As the variance of the Gaussian distribution increases, the variance of impedance grows more rapidly with standard polynomials compared to Chebyshev polynomials. In practical terms, this slower expansion rate with Chebyshev polynomials ensures gradual cloud growth, preventing significant overlap between distinct clouds. This controlled scaling supports the reliable separation of different ageing states while maintaining the integrity of the generated datasets.

5 | Conclusions

To reduce the dimensionality of model representation within a single charging/discharging cycle, Chebyshev polynomials effectively quantify the SOC dependence of ECM parameters

in a robust way, which allows a reconstruction of the ECM at any desired SOC value in a concise manner. Each cycle can be represented by a specific set of Chebyshev coefficients that correspond to the fitted parameters at different SOC values during that cycle. These coefficients collectively form the battery's history over that specific cycle.

The QMC method facilitates the efficient generation of high-dimensional Chebyshev coefficient vectors based on few experimental data, allowing for the construction of distinct and meaningful regions of battery health without requiring extensive experiments. This approach provides a structured representation of a battery's health, capturing its progression through distinct categories: *fresh*, *aged* and *dead* battery cells. The use of physics-informed synthetic training data, derived from these sampled models, establishes a solid foundation for further machine learning applications.

Acknowledgements

This project has been financially supported by the Helmholtz AI Cooperation Unit (HAICU), project 'Intelligent, individual battery management using spectroscopy and machine learning' (i2Batman).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. J. T. J. Burd, E. A. Moore, H. Ezzat, R. Kirchain, and R. Roth, "Improvements in Electric Vehicle Battery Technology Influence Vehicle Lightweighting and Material Substitution Decisions," *Applied Energy* 283 (2021): 116269, https://doi.org/10.1016/j.apenergy.2020.116269, https://www.sciencedirect.com/science/article/pii/S0306261920316597.

2. Y. Guo, J. Cai, Y. Liao, J. Hu, and X. Zhou, "Insight Into Fast Charging/Discharging Aging Mechanism and Degradation-Safety Analytics of 18650 Lithium-Ion Batteries," *Journal of Energy Storage* 72 (2023): 108331, https://doi.org/10.1016/j.est.2023.108331, https://www.sciencedirect.com/science/article/pii/S2352152X23017280.

3. Z. Wang, G. Feng, D. Zhen, F. Gu, and A. Ball, "A Review on Online State of Charge and State of Health Estimation for Lithium-Ion Batteries in Electric Vehicles," *Energy Reports* 7 (2021): 5141–5161, https://doi.org/10.1016/j.egyr.2021.08.113, https://www.sciencedirect.com/science/article/pii/S2352484721007150.

4. Y. Yang, R. Wang, Z. Shen, Q. Yu, R. Xiong, and W. Shen, "Towards a Safer Lithium-Ion Batteries: A Critical Review on Cause, Characteristics, Warning and Disposal Strategy for Thermal Runaway," *Advances in Applied Energy* 11 (2023): 100146, https://doi.org/10.1016/j.adapen.2023.100146, https://www.sciencedirect.com/science/article/pii/S2666792423000252.

5. J. S. Edge, S. O'Kane, R. Prosser, et al., "Lithium Ion Battery Degradation: What You Need to Know," *Physical Chemistry Chemical Physics* 23 (2021): 8200–8221, https://doi.org/10.1039/D1CP00359C.

6. S. Piller, M. Perrin, and A. Jossen, "Methods for State-of-Charge Determination and Their Applications," *Journal of Power Sources* 96, no. 1 (2001): 113–120. Proceedings of the 22nd International Power Sources Symposium, https://doi.org/10.1016/S0378-7753(01)00560-2, https://www.sciencedirect.com/science/article/pii/S0378775301005602.

7. C. H. Bartsch, L. J. Author, F. P. Bereck, et al. "Weighted Distribution of Relaxation Time Analysis of Battery Impedance Spectra Using Gaussian Process Regression for Noise Estimation", preprint (2024), https://doi.org/10.26434/chemrxiv-2024-1gxgq.

8. H. Witzenhausen, "Elektrische Batteriespeichermodelle: Modellbildung, Parameteridentifikation und Modellreduktion" (Ph.D. dissertation. RWTH Aachen University, 2017), https://doi.org/10.18154/RWTH-2017-03437, https://publications.rwth-aachen.de/record/687819.

9. N. Karjanto, "Properties of Chebyshev Polynomials," (2020) arXiv: 2002.01342 [math.HO].

10. A. T. Benjamin, L. Ericksen, P. Jayawant, and M. Shattuck, "Combinatorial Trigonometry With Chebyshev Polynomials," *Journal of Statistical Planning and Inference* 140, no. 8 (2010): 2157–2160, Lattice Path Combinatorics and Applications, https://doi.org/10.1016/j.jspi.2010.01.011, https://www.sciencedirect.com/science/article/pii/S0378375810000224.

11. I. M. Sobol', "On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals," *USSR Computational Mathematics and Mathematical Physics* 7, no. 4 (1967): 86–112, https://doi.org/10.1016/0041-5553(67)90144-9, https://www.sciencedirect.com/science/article/pii/0041555367901449.

12. F. Hust, *ISEA Framework Documentation*, https://isea.pages.rwthaachen.de/framework/index.html. (accessed: August 2023).

13. T. A. Kurniawan, A. Natajaya, P. S. Priambodo, and G. Wibisono, "Real Time Monitoring State-of-Charge Battery Using Internal Resistance Measurements for Remote Applications," *Journal of Physics: Conference Series* 1528, no. 1 (2020): 012034, https://doi.org/10.1088/1742-6596/1528/1/012034.

14. D. Wang, Y. Bao, and J. Shi, "Online Lithium-Ion Battery Internal Resistance Measurement Application in State-of-Charge Estimation Using the Extended Kalman Filter," *Energies* 10, no. 9 (2017): 1284. ISSN: 1996-1073, https://doi.org/10.3390/en10091284, https://www.mdpi.com/1996-1073/10/9/1284.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.