

<https://doi.org/10.1038/s44296-024-00011-1>

Probabilistic machine learning for battery health diagnostics and prognostics—review and perspectives

Adam Thelen¹, Xun Huan², Noah Paulson³, Simona Onori⁴, Zhen Hu⁵ & Chao Hu⁶✉

Diagnosing lithium-ion battery health and predicting future degradation is essential for driving design improvements in the laboratory and ensuring safe and reliable operation over a product's expected lifetime. However, accurate battery health diagnostics and prognostics is challenging due to the unavoidable influence of cell-to-cell manufacturing variability and time-varying operating circumstances experienced in the field. Machine learning approaches informed by simulation, experiment, and field data show enormous promise to predict the evolution of battery health with use; however, until recently, the research community has focused on deterministic modeling methods, largely ignoring the cell-to-cell performance and aging variability inherent to all batteries. To truly make informed decisions regarding battery design in the lab or control strategies for the field, it is critical to characterize the uncertainty in a model's predictions. After providing an overview of lithium-ion battery degradation, this paper reviews the current state-of-the-art probabilistic machine learning models for health diagnostics and prognostics. Details of the various methods, their advantages, and limitations are discussed in detail with a primary focus on probabilistic machine learning and uncertainty quantification. Last, future trends and opportunities for research and development are discussed.

Lithium-ion (Li-ion) batteries have witnessed growing adoption in consumer electronics, electric vehicles (EVs), and grid energy storage systems, largely owing to their excellent energy density and power output. However, continued usage and adverse operating environment drive irreversible chemical reactions and material morphology changes, leading to gradual but inevitable degradation of battery capacity and power over time. Consequently, accurately estimating the state of health (SOH) of Li-ion batteries and predicting their future degradation is crucial to optimizing every part of the battery life cycle—from research and development, to manufacturing and validation, deployment in the field, and reuse and recycling¹.

Some of the earliest research into Li-ion battery health diagnostics and prognostics focus on mathematical modeling of the capacity fade during cycle aging tests^{2–4}. Notably, Bloom et al.⁴ found that the cell capacity could be aptly captured by a modified Arrhenius relationship, which is generally used to describe the rate of a chemical reaction. This capacity fade model excelled in extrapolating battery performance to

new, untested conditions, providing great utility for design and engineering. Not long after, researchers began experimenting with empirical and semi-empirical mathematical models for battery capacity fade modeling to gain better accuracy. In the work of Spotnitz⁵, researchers developed a semi-empirical model for Li-ion battery capacity fade considering reversible and irreversible capacity loss due to solid-electrolyte interphase (SEI) growth on the graphite anode in the cells. Similar work by Broussely et al.⁶ proposed an empirical quadratic equation to model the capacity fade of NMC/Gr Li-ion cells during long-term storage. The quadratic model was primarily developed to capture the effect of SEI growth from electrode-electrolyte reactions during storage. Later, Liaw et al.⁷ demonstrated that empirical models could be used to extrapolate cell resistance increase with thermal aging to update the parameters of a simple equivalent circuit model for capacity estimation. Since these initial seminal works, many more advanced empirical and semi-empirical models have been developed^{8–10}.

¹Department of Mechanical Engineering, Iowa State University, Ames 50011 IA, USA. ²Department of Mechanical Engineering, University of Michigan, Ann Arbor 48109 MI, USA. ³Data Science and Learning and Applied Materials Divisions, Argonne National Laboratory, Lemont 60439 IL, USA. ⁴Department of Energy Science and Engineering, Stanford University, Stanford 94305 CA, USA. ⁵Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn 48128 MI, USA. ⁶School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs 06269, USACT.

✉e-mail: chao.hu@uconn.edu

The great success of empirical and semi-empirical models soon led researchers to investigate alternative methods of modeling battery aging from experimental data. Saha et al.¹¹ were some of the first to use a machine learning (ML) algorithm as part of a framework to model battery capacity fade and predict remaining useful life. The researchers used a relevance vector machine (RVM) (see the section “Relevance vector machine”) to model the exponential growth observed in the cell’s internal resistance with aging. The RVM was used to predict future resistance parameters for an equivalent circuit model that was then used to predict cell capacity. Altogether, the RVM was shown to do an exceptional job at rejecting outliers from the dataset and providing good uncertainty estimates with its predictions. This approach inspired others to further investigate ways of using ML models for battery health diagnostics and prognostics^{12–14}.

Over the past decade, the use of ML for battery health diagnostics and prognostics has expanded substantially. The rapid growth can be attributed in part to the recent advances in ML and deep learning technology, like open-source ML software and datasets, that enable easier modeling of complex data¹⁵. Well-studied applications of ML for battery health diagnostics and prognostics include battery performance simulation and state estimation (primarily state-of-charge (SOC) and power estimation)^{16–19}, SOH estimation and capacity grading^{20–22}, and capacity forecasting and remaining useful life (RUL) prediction^{23–25}. Newer, emerging battery prognostic problems include early lifetime prediction^{26,27}, knee point prediction²⁸, capacity trajectory prediction from early aging data^{29,30}, and initial works investigating the applicability of existing diagnostic and prognostic models to battery aging data collected from the field^{31–33}.

Despite these significant research efforts on battery health diagnostics and prognostics, most ML-focused works have yet to incorporate uncertainty quantification systematically. Here, “uncertainty” refers to the predictive uncertainty of an ML model, such as a neural network, for a training/test sample point that is ideally associated with how confident the model is when predicting at this point³⁴. The idea is that an ML model does not simply produce an output (e.g., an estimate of a cell’s SOH indicator); it also estimates the uncertainty associated with this prediction to the most accurate extent possible. For example, this predictive uncertainty can be in the form of a standard deviation of a Gaussian-distributed output that describes the spread of the probability distribution around the mean prediction. A spread that is too large indicates that the uncertainty level is high enough for the output not to be trusted. In such cases, a human end user may discard this prediction or provide the ML model with additional information to reduce the predictive uncertainty. Predictive uncertainty can be confused with prediction error. The former comes as an uncertainty estimate by an ML model with uncertainty quantification capability and is thus *known*; in contrast, the latter is *unknown* without access to the ground truth. That is why access to predictive uncertainty is important for applications not tolerating large prediction errors well. Ideally, in these applications, we expect predictive uncertainty (known) to be a reliable indicator of prediction error (unknown) on a per-sample basis.

Quantifying predictive uncertainty in ML-based health diagnostics and prognostics becomes especially important given the dynamic and multi-physical nature of Li-ion batteries, where even small variations in manufacturing and testing conditions can significantly change the electrical, thermal, and mechanical performance, resulting in larger cell-to-cell variability³⁵. Furthermore, this inherent cell-to-cell variability becomes even more pronounced as the cells age. Early work by Baumhofer et al.³⁶ investigated the production-caused variation in capacity fade of a group of 48 cells cycled under identical conditions, finding that the lifetimes varied by as much as a few hundred cycles. These results, and many similar studies^{36,32,35,37}, highlight the great need for probabilistic diagnostic and prognostic algorithms that often have to learn from small datasets and extrapolate to the tail-end of the lifetime distribution for a population of cells. Such extrapolations are often associated with large prediction errors, which, although infeasible to quantify without access to the ground truth, can be communicated to the user, to some degree, through high predictive uncertainty and low model confidence. Probabilistic models with properly

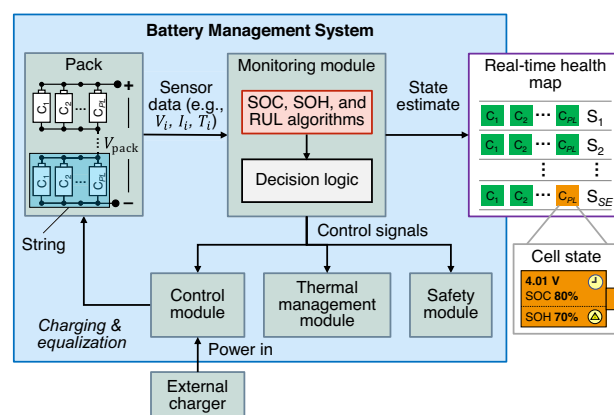


Fig. 1 | An illustrative flowchart showing essential functions of a BMS that manages the power and health of a battery pack. This battery pack comprises *SE* serially connected strings, each of which is composed of *PL* cells connected in parallel.

calibrated uncertainty estimation are paramount for setting warranties on battery-powered devices like consumer electronics and, more recently, EVs, where failing to deliver a promised lifetime due to maintenance/control decision making informed by largely incorrect ML-based diagnostic and prognostic results can cost companies their reputation in addition to the monetary burden associated with honoring warranty repairs.

In practice, quantifying diagnostic and prognostic uncertainty is especially important for large battery packs with many modules, where the capacity of a module consisting of serially connected cells will be limited to the capacity of the worst-performing cell. Thus, probabilistic models (like those discussed in the sections “Probabilistic ML techniques and their applications to battery health diagnostics and prognostics”, “Advanced topics in battery health diagnostics and prognostics”, and “Future trends and opportunities”) that can accurately model worst-cell performance through uncertainty estimates made by learning from a limited dataset are crucial to module and pack development. This cell-to-cell variability poses a direct challenge for battery management systems (BMSs) that need to balance cell voltages to maximize the capacity and power availability of a battery pack. In essence, a BMS is an electronic system consisting of hardware, software, and firmware that is responsible for managing the power and health of a rechargeable battery (e.g., a Li-ion battery cell or pack). Figure 1 outlines some key functions of a BMS in an illustrative flowchart. The BMS in the figure is built for a battery pack consisting of *SE* serially connected strings (or modules), each with *PL* parallelly connected cells. The BMS takes voltage (*V*), current (*I*), and temperature (*T*) measurements from each cell in the pack at regular intervals (e.g., every 1–5 s) and estimates the SOC and SOH of each cell, both of which cannot be directly measured. As will be detailed in the section “Battery health diagnostic and prognostic problems”, SOH estimation is an important battery diagnostic problem. For situations that require knowing how long each cell/module can be used before replacement, the BMS monitoring module also predicts each cell’s RUL and, in some cases, the cell’s SOH trajectory in future cycles. RUL prediction and SOH trajectory prediction are two well-studied battery prognostic problems of significance, as will be discussed in the section “Battery health diagnostic and prognostic problems”. Most importantly, neglecting uncertainty when predicting cell SOC and SOH may lead the BMS to incorrectly balance cell voltages, ultimately reducing the available capacity and power of the pack. It is worth mentioning that SOH estimation and RUL prediction can be computed in the cloud instead of directly at the BMS device, as the SOH and RUL usually need not be updated in real-time.

One major advantage of predictive uncertainty quantification for battery maintenance and control is its value in informing BMS actions during operation. For example, if estimates of cell SOC are highly uncertain, the BMS may limit the overall charge power in order to prevent cells from

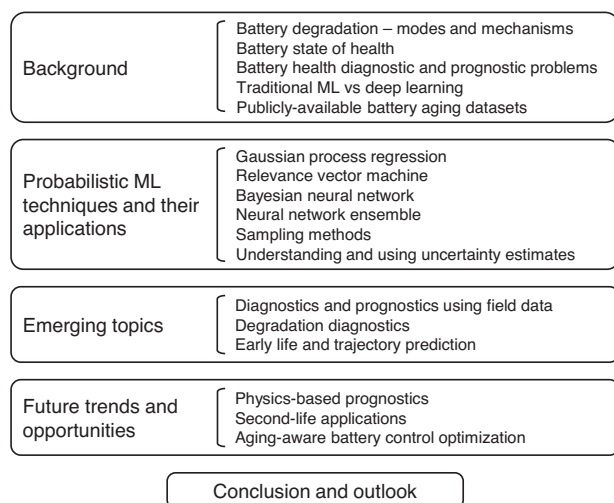


Fig. 2 | Overview and organization of this review paper. Outline of sections and subsections.

entering over-voltage conditions during charging. However, parameterizing models that can accurately quantify predictive uncertainty is challenging because battery datasets are usually limited in size due to the large expenses required to operate thermal chambers for extended periods. Further, it is difficult to replicate real-world operating conditions in the laboratory, and much care is needed to ensure newly parameterized models can accurately quantify prediction uncertainty on field data (see the section “Diagnostics and prognostics using field data”). The trend of small datasets is likely to continue as cells grow larger in size for automotive and grid storage applications. Large-format and high-capacity (>100 Ah) Li-ion battery cells require even more expensive testing equipment to achieve the high C-rates (>3C for a 100 Ah cell requires >300 A continuous current) necessary for aging cells quickly and studying fast-charging protocols—research that is imperative for lowering the “refueling time” of today’s EVs and accelerating the transition to electrified transportation. With costs for cells and test equipment on the rise, calibrating the predictive accuracy and uncertainty of battery diagnostic and prognostic models prior to deployment becomes critically important. Incorrect control decisions based on erroneous predictions and uncertainty may lead to suboptimal performance, damage to battery cells, and in rare cases, thermal runaway that results in catastrophic product loss and endangers the safety of people nearby.

To this end, developing and validating probabilistic battery diagnostic and prognostic models is an essential area of research in the battery community. A handful of reviews on battery health diagnostics/prognostics exist today and can be found here^{38–45}. However, all the reviews to date focus primarily on deterministic ML modeling methods, and do not emphasize existing research that studies probabilistic methods for battery health diagnostics and prognostics. To address this gap, we seek to provide a comprehensive overview of probabilistic modeling and ML for battery health diagnostics and prognostics. After providing an overview of Li-ion battery degradation, we review past and present studies on probabilistic battery health diagnostics and prognostics and discuss their methods, advantages, and limitations in detail. Our review offers unique insights into each of the probabilistic modeling approaches with detailed discussions on the implementation approach and recommendations for future research and development. Figure 2 presents an outline of this review paper. Below are a few key items covered in our review.

1. First, we provide an overview of Li-ion battery degradation, discussing the types, main causes, and resulting effects on cell-level performance and SOH in the sections “Battery degradation—modes and mechanisms” and “Battery state of health”. The classification of battery degradation modes and analysis of their root causes provides relevant background knowledge that motivates the need for battery diagnostic/

prognostic models that can estimate cell health and predict future cell degradation. In the section “Battery health diagnostic and prognostic problems”, we provide a high-level overview of six general problems relevant to battery health estimation and life prediction. Additionally, we highlight the pivotal role that publicly available battery aging datasets have played in facilitating existing research in the area (the section “Publicly available battery aging datasets”).

2. Second, we analyze and compare the advantages and limitations of various probabilistic ML techniques and their application to battery health diagnostics and prognostics (the section “RVM applications to battery diagnostics and prognostics SOH estimation”). This section, uniquely focusing on probabilistic techniques for health diagnostics and prognostics, covers both the methodologies of each technique and examples of its applications to SOH estimation, SOH forecasting, and RUL prediction. This particular emphasis on probabilistic ML is a noteworthy feature of this review that sets it apart from existing reviews on battery health diagnostics and prognostics.
3. Third, we delve into three emerging and “newer” topics in battery health diagnostics and prognostics in the section “Advanced topics in battery health diagnostics and prognostics”. Specifically, this section offers unique insights from three researchers actively working on problems related to battery SOH estimation from field data (the section “Diagnostics and prognostics using field data”), degradation diagnostics (the section “Degradation diagnostics”), and early life and trajectory prediction (the section “Early life and trajectory prediction”). This unique coverage of emerging topics further sets our review apart from existing ones.
4. Fourth and finally, we discuss future trends and research opportunities in physics-based prognostics (the section “Physics-based diagnostics and prognostics”), second-life applications for used Li-ion cells (the section “Second-life applications”), and aging-aware battery control optimization (the section “Aging-aware battery control optimization”). This discussion constitutes the final distinctive element of this review, not commonly found in most other reviews.

Our review paper is concluded in the section “Conclusion”, where we also discuss prospects for future research essential to addressing long-standing challenges in battery health diagnostics and prognostics.

It is worth noting that this review focuses primarily on the application of various probabilistic ML and deep learning methods to unique problems (the section “Battery health diagnostic and prognostic problems”) within the field of battery health diagnostics and prognostics. A limitation of this work is that it does not cover specific challenges related to emerging ML topics, such as hybrid modeling, transfer learning, federated learning, and similar ML-focused concepts.

Background

Battery degradation—modes and mechanisms

Battery degradation is a complex and multi-scale process that varies with cell design and is driven by the way a cell is used. Understanding the fundamental mechanisms of Li-ion battery degradation is essential for effectively modeling and designing around it. Typically, researchers and engineers will conduct lab-based aging experiments to study the effects of different operating conditions on cell aging and SOH, which is most often quantified as a cell’s remaining capacity or internal resistance. Periodic reference performance tests (RPTs) are carried out during aging experiments to assess cell capacity and resistance under standard conditions (usually 25 °C) to help isolate the effect of aging on changes in cell capacity and resistance⁴⁶. Comparing cell SOH measured from RPTs is important because cell capacity and resistance are influenced by temperature, C-rate, voltage limits, among other factors.

Battery aging tests are used to understand how stressors, like time, temperature, and energy throughput, affect the rate of capacity fade and the progression of internal degradation modes^{47,48}. An overview of battery degradation mechanisms, their corresponding modes, and measurable cell-level effects is shown in Fig. 3. Even without cycling, Li-ion batteries lose capacity over time as internal side reactions occur between the

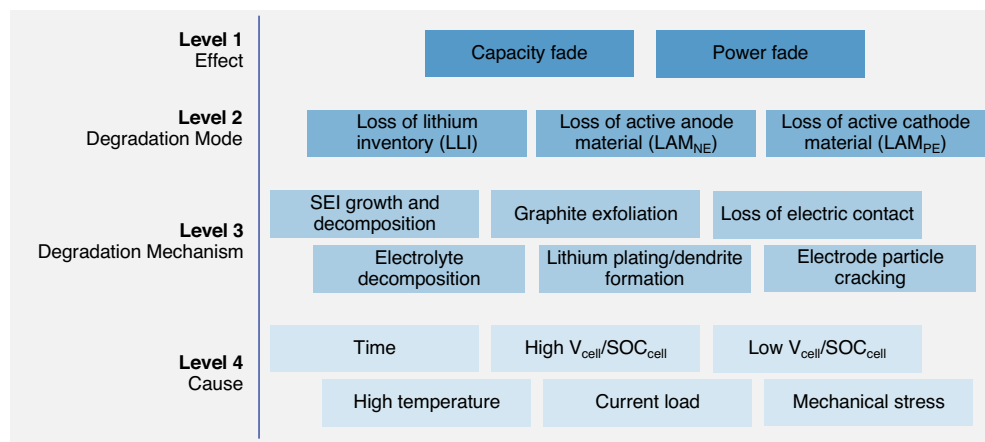


Fig. 3 | Cause and effect of Li-ion battery capacity and power fade. Figure modified from Thelen et al.⁸¹ and based on the original figure in Birkel et al.⁷¹ showing the relationship between cell use/environment (Level 4), the corresponding degradation

mechanisms (Level 3), their connections to the degradation modes (Level 2), and the resulting capacity/power fade (Level 1).

electrolyte and electrode materials. The most prevalent of these side reactions is the formation of the solid electrolyte interphase (SEI) on the graphite anode common in nearly all Li-ion batteries used today⁴⁹. SEI growth is mainly driven by time, but is also influenced by temperature, cell voltage, and cell load⁵⁰. Fortunately, the formation of SEI on graphite anodes is entirely expected and well-studied as it plays a large role in determining a battery's maximum capacity and expected lifetime. It has been widely accepted that capacity fade from the growth of SEI scales with a square-root-of-time ($Q(t) = a \cdot t^{0.5}$) relationship⁹. Further, many researchers have modeled Li-ion battery capacity fade due to SEI formation at various temperatures by scaling the $t^{0.5}$ term using Arrhenius-like equations that model the influence of temperature on reaction rate^{51,52}. Additionally, SEI formation has been shown to be directly related to the SOC a battery is stored at, where higher voltages generally lead to faster reactions and greater capacity fade. However, next-generation battery designs are pursuing new anode materials and may reduce or even eliminate the use of graphite in the anode altogether, thus introducing new degradation mechanisms that will need to be studied and mitigated.

When a Li-ion battery is cycled, more degradation mechanisms arise in addition to the always-present capacity fade from SEI growth and other side reactions. Often during cycling, the SEI growth rate accelerates because the movement of Li-ions in/out of the electrodes causes repeated swelling and subsequently cracking of the already formed SEI, revealing new sites for SEI to form, and ultimately consuming more lithium in the process^{51,53,54}. Like SEI formation, electrode swelling is expected by battery designers, and is a well-studied degradation mode. Li-ion battery degradation from electrode cracking has been shown to be sensitive to the depth of discharge (DOD) and the C-rate the cell is subjected to—where deeper discharge and faster rates increase the rate of capacity fade^{26,51,55}. Capacity fade driven by electrode cracking during cycling has been diagnosed as a primary driver of cycling-driven capacity fade in Nickel-based battery chemistries like nickel-cobalt-aluminum-oxide (NCA)⁵⁵ and nickel-manganese-cobalt-oxide (NMC)^{26,51}. In these studies, researchers found that loss of cathode active material (LAM_{PE}) to be a primary contributor to a cell's capacity fade and was strongly correlated with a cell's eventual lifetime.

Under more extreme conditions such as cold temperatures ($T \ll 10$ °C), high charging C-rates ($I \gg 3C$), or the combination of these conditions, intercalation of Li-ions into the anode and cathode are slowed, causing Li-metal to plate onto the surface of the anode instead of intercalate inside it⁵⁶. Lithium plating poses a great safety risk due to the possibility of a lithium metal dendrite growing large enough to puncture the separator and cause an electrical short circuit. Unlike SEI formation and electrode cracking, lithium plating is not expected to take place inside Li-ion batteries during normal operation. Therefore, much work has been done to detect and model the

lithium plating degradation mechanism so that it can be safely mitigated through design and control strategies. However, lithium plating is a dynamic process that is affected by the cell design (energy density), charge rate, temperature, and SOC, making it challenging to detect and quantify. Research by Huang et al.⁵⁷ demonstrated how differential pressure measurements could be used to detect lithium plating inside cells in real-time during fast charging. Their method holds promise for online monitoring and real-time control of cells operating in the field, but the technology still needs to be demonstrated on the pack level before it might be considered for mass production. Other research by Konz et al.⁵⁸ demonstrated a method for quickly quantifying the lithium plating limits of a cell using standard battery cyclers by measuring the coulombic inefficiency of the cell after cycling at various C-rates. The method performs sweeps over a series of charge rates and SOC cutoffs to map out the lithium plating limits at the tested temperature. The method provides a cheaper and faster approach to mapping the lithium plating limits and designing an optimal fast-charging protocol using experiments instead of the traditional approach of using an electrochemical model of a cell. Regardless of the strategy employed, modeling and mitigating lithium plating is imperative to ensuring the safe and reliable operation of batteries over their lifetimes. Later we will revisit the topic of lithium plating when discussing emerging strategies for prolonging battery lifetime in the section “Aging-aware battery control optimization”.

With researchers pushing for higher energy densities by introducing new materials into batteries, there will always be new degradation mechanisms that present challenges. Recent efforts to increase the capacity of existing Li-ion battery chemistries, like lithium cobalt oxide (LCO), lithium NCA, lithium NMC, and lithium iron phosphate (LFP), by adding silicon (Si) to the graphite anodes has led to a field of research devoted to studying silicon-anode technology. However, the high capacity of silicon as an anode material presents its own set of challenges around swelling and cracking. Silicon-anode batteries are notoriously known for swelling as much as 20% their original thickness, posing a unique set of degradation and packaging challenges⁵⁹. Similarly, high energy density Li-metal batteries pose their own set of unique challenges, mainly related to the reversibility of the metal plating and stripping process on the negative current collector. Likewise, solid-state batteries face challenges related to degradation of the solid electrode/electrolyte interfaces and the materials themselves. On the other hand, low energy density lithium titanium oxide (LTO) anodes are much safer from an abuse perspective, but suffer from extreme gassing which creates bubbles between electrode layers and subsequently delamination which deactivates areas of the electrodes, causing accelerated capacity loss and aging^{60,61}. Less mature batteries, like Li-S and Li-air chemistries face a host of issues with fast capacity fade and poor coulombic efficiency that prevent scaling to production. Readers interested in the challenges surrounding degradation of next-

generation silicon anode, Li-metal batteries, solid-state, LTO, LiS, and Li-Air batteries are referred to these reviews on the topics—silicon anode:^{59,62}, Li-metal:^{63,64}, solid-state:^{65,66}, LTO^{60,61}, Li-S^{67,68}, and Li-Air⁶⁹.

Until these battery chemistries are refined further, applications of battery health diagnostics and prognostics are mainly limited to the laboratory. In light of this, our review primarily focuses on probabilistic ML modeling methods applied to standard Li-ion chemistries. However, it is envisioned that nearly all of the ML-based modeling methods discussed in this paper will be transferable to new battery chemistries to some degree.

Battery state of health

Battery degradation observed during controlled laboratory experiments or normal operation in the field is the result of the interaction and accumulation of various component-level degradation mechanisms like those discussed in the section “Battery degradation—modes and mechanisms”. The most frequently used measures of battery SOH are capacity and resistance because they are directly measurable during aging experiments using periodic RPTs^{46,70}. Resistance and impedance measurements taken at various SOCs are used to quantify the cell’s ability to deliver power and is a crucial battery state for implementing safe management controls. Direct and alternating current (DC and AC) resistance can usually be measured with fast diagnostic pulses (<30 s). However, directly measuring the capacity of cells operating in the field is largely infeasible without significantly interrupting the normal operation of the product to run a long charge/discharge diagnostic test. In practice, the SOH of cells operating in the field must be estimated from the available cell-level electric, thermal, and mechanical data.

More recently, advances in battery modeling and the availability of larger publicly available aging datasets has lead many researchers to further extend the definition of cell SOH to include the three primary degradation modes that drive capacity and power fade: LAM_{PE}, LAM_{NE}, and LLI (see

Fig. 3). Together, these three degradation modes capture the combined effect of the individual degradation mechanisms on cell health and provide better insight into the health of the cell’s major components than do capacity and resistance. For example, identifying that the anode is degrading more quickly than the cathode can help with identifying when a knee-point in the cell’s capacity fade trajectory may occur⁵⁶. Similarly, capacity fade is often complex and path-dependent. For example, the dominant degradation mechanism driving capacity fade during the early life of a battery is typically SEI growth. Later on, other degradation modes, like electrode particle cracking, begin to appear as the cell accumulates more cycles and the electrodes experience repeated swelling and relaxation⁵¹. Quantifying cell SOH through the three degradation modes provides more insight into when and to what degree cell degradation is occurring than simply estimating cell capacity.

While quantifying battery SOH through the various component-level degradation modes is useful in the lab, the same methods are not necessarily useful nor viable for cells operating in the field. Relevant metrics of cell SOH for field units like EVs and consumer electronics are primarily focused around quantifying remaining capacity, resistance, impedance, and any risks of thermal runaway, as these impact the user experience the most. Quantifying battery SOH from field data presents a new set of challenges, since the quality and quantity of diagnostic measurements are heavily influenced by user behavior. For example, it is rare that cells will ever complete full DOD cycles in the field due to BMS limits and cell voltage imbalance. Thus, gathering usable data for SOH estimation becomes a real challenge. Later in the section “Diagnostics and prognostics using field data”, we discuss current research focusing on health diagnostics and prognostics from field data.

Battery health diagnostic and prognostic problems

Figure 4 provides an overview of battery diagnostic and prognostic problems where probabilistic ML techniques can be applied to build regressors

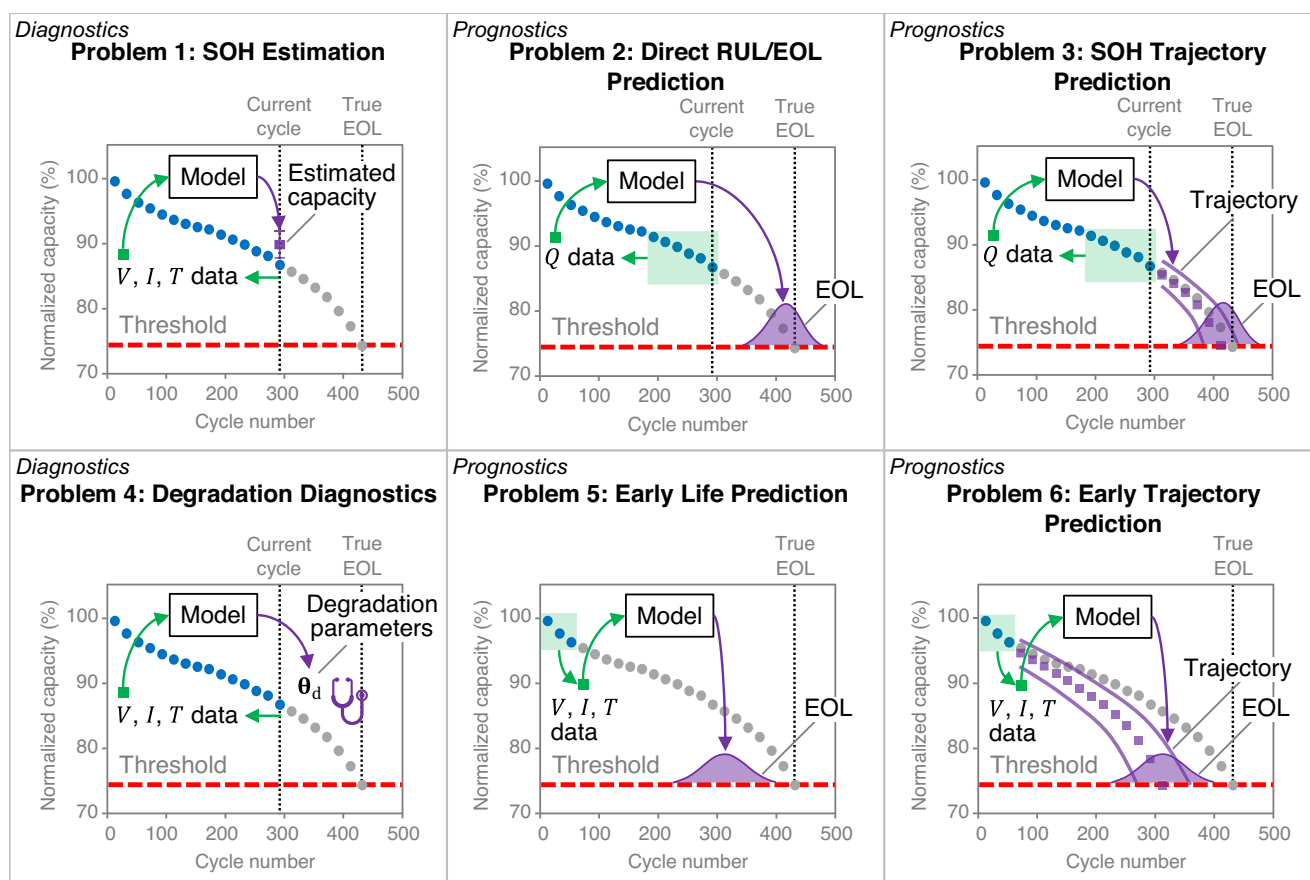


Fig. 4 | An overview of six general problems relevant to battery health diagnostics and prognostics. Here, V, I, T , and Q denote voltage, current, temperature, and capacity, respectively, and θ_d denotes an estimated vector of three degradation parameters defined to quantify three degradation modes.

with uncertainty quantification capability (i.e., the ability of these regressors to quantify the predictive uncertainty in their outputs). We divide the fields of battery health diagnostics and prognostics into six unique problems to highlight the subtle differences in the various research articles published on the topics. Broadly, problems 1 and 4 are classified as diagnostic problems since battery health is estimated at the current cycle. Problems 2, 3, 5, and 6 are classified as prognostic problems since battery health (and/or lifetime) is predicted for future cycles. The six general problems are briefly summarized as follows:

- **Problem 1: SOH estimation** Approaches to this first problem aim to estimate the current battery health, often based on voltage, current, and temperature measurements readily available to a BMS. In practice, it comes down to estimating the capacity and resistance, which together determine a battery's energy and power capabilities. This problem is probably the most extensively studied in the battery diagnostics field, with multiple review papers dedicated to this problem every year.
- **Problem 2: Direct RUL/EOL prediction** Approaches targeting this second problem predict the RUL by training an ML model that directly maps a sequence of most recent capacity observations to RUL. These capacity observations can be either actual capacity measurements via coulomb counting on full charge/discharge cycles or capacity estimates by an algorithm. The idea is to feed this sequence of capacity observations to an ML model, which produces an RUL estimate. In other words, this ML model takes a sequence of capacity observations, consisting of the observation at the current cycle and a few recent past cycles, and produces an RUL estimate, for instance, in the form of a probability distribution when a probabilistic ML model is adopted.
- **Problem 3: SOH trajectory prediction** Unlike SOH estimation (Problem 1), which centers on inferring current health, this third problem focuses on predicting future capacity and resistance, often by examining the degradation trend over a few most recent cycles and extrapolating this trend. Similar to SOH estimation studies, SOH forecasting studies mostly look at capacity forecasting. A simple and popular approach is to take a sequence of capacity observations at the current and recent past cycle and feed these observations as input into an ML model, which may produce a sequence of probabilistic capacity estimates, for instance, the means and standard deviations of the forecasted capacity observations at the next few cycles that all follow Gaussian distributions. These estimates form a capacity degradation trajectory, based on which an end-of-life (EOL) estimate can be derived as the cycle number when this trajectory down-crosses a predefined capacity threshold (typically 80% of the initial capacity for automotive applications). An RUL estimate can be obtained by subtracting the current cycle number from the EOL estimate. Unlike RUL prediction through SOH forecasting, direct RUL prediction, as discussed in Problem 2, skips the step of capacity forecasting and directly maps a capacity sequence to the RUL.
- **Problem 4: Degradation diagnostics** Degradation diagnostics is a sub-problem of SOH estimation focused on diagnosing the degradation modes that drive capacity fade and resistance increase⁷¹. This sub-problem aims to estimate three degradation parameters that measure the degrees of three degradation modes: the loss of active material on the cathode, the loss of active material on the anode, and the loss of lithium inventory. Estimating these three degradation parameters almost always requires access to high-precision voltage and current measurements during a full charge/discharge cycle, but workarounds do exist (see the section "Degradation diagnostics").
- **Problem 5: Early life prediction** This is an emerging prognostic problem where ML models map data from an early life stage to the lifetime (or the EOL cycle). A key step to solving this problem is defining early-life features predictive of the lifetime. A concise review of recent studies attempting to solve this problem will be provided in the section "Early life and trajectory prediction".
- **Problem 6: Early trajectory prediction** This sixth problem is similar to yet more challenging than early life prediction. The added difficulty

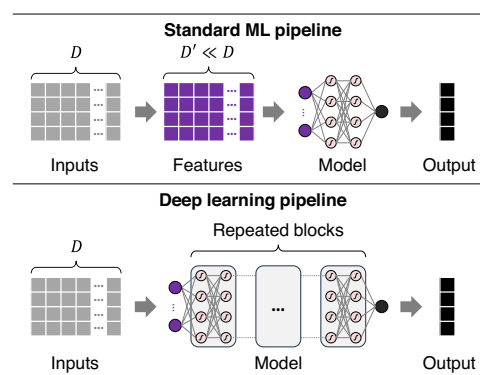


Fig. 5 | A high-level comparison between traditional ML and deep learning. Here, a typical traditional ML pipeline requires engineers to manually identify D' features from a D -dimensional raw input, also known as feature extraction. In contrast, a typical deep learning pipeline does not need manual feature extraction and can automatically learn features with predictive power from data.

comes from the need to predict the entire capacity trajectory rather than a single EOL cycle, as done in early life prediction. In addition to early-life features, capacity fade models are also required to produce a sequence of capacity estimates for any range of cycle numbers.

Traditional ML vs. deep learning

Over the past decade, hundreds, if not thousands, of data-driven approaches have been created for battery health diagnostics and prognostics. These existing approaches can be broadly categorized as traditional ML and deep learning. Figure 5 illustrates the key difference between these two categories. Traditional ML requires manually defining and extracting hand-crafted features. ML models are then built to approximate the often highly non-linear relationship between these input features and the output (or target). Examples of traditional ML algorithms for building these models include regularized linear regressions (e.g., ridge regression, lasso, and elastic net), support vector machines, RVMs, Gaussian process regression (GPR) or kriging, random forests, Bayesian linear regression, gradient boosting machines (e.g., XGBoost and light gradient-boosting machines), k-nearest neighbors, and shallow neural networks.

An input to a traditional ML model can be formulated from voltage and current measurements during a partial charge cycle. This input can be (1) a vector of features extracted from voltage vs. time (V vs. t) and current vs. time (I vs. t) curves^{13,14,72}, (2) a vector of features extracted from an incremental capacity vs. voltage (dQ/dV vs. V) curve^{73,74}, (3) a vector of features extracted from a differential voltage vs. capacity (dV/dQ vs. Q) curve⁷⁵, or (4) any combination of these three vectors⁷⁶. The output can be capacity or resistance for SOH estimation or EOL/RUL for health prognostics. The performance of ML models highly depends on the collective predictive power of these manually extracted features. Additionally, the same set of features that works well on a specific battery chemistry and application often does not transfer to a different chemistry or application. Thus, when dealing with a new chemistry or application, one has to repeat the tedious and time-consuming process of manual feature extraction.

Unlike traditional ML, deep learning can automatically learn high-level abstract features of predictive power from large volumes of data. An obvious benefit is that manual feature extraction is no longer needed and is replaced by "feature learning". However, deep learning approaches have two well-known limitations. *First*, deep neural network models are more prone to overfitting data than shallower neural network models, especially when a training dataset is small (e.g., a few tens to hundreds of training samples). Given the time- and resource-demanding aging tests, most battery health diagnostics and prognostics applications reside in the small data regime⁷⁷. Aging data available for model training are even more limited in an early research and development stage where lifetime prognostics may be applied to accelerate materials design⁷⁸ or charging protocol optimization⁷⁹. As a

result, a deep learning model built for battery diagnostics/prognostics may produce high accuracy on the dataset this model was trained on but may generalize poorly to “unseen” test samples that could fall outside of the training data distribution. These out-of-training-distribution test samples are often called *out-of-distribution* (OOD) samples. A solution to the conflict between what is needed (i.e., big data) and what is available (i.e., small data) is quantifying the predictive uncertainty through probabilistic deep learning³⁴. The uncertainty estimate could serve as a proxy for model confidence, i.e., how confident this model is when making a prediction. The ability to convey model confidence is crucial for safety-critical battery applications, where SOH/lifetime predictions with large errors and no warnings are simply unacceptable. *Second*, deep learning models are inherently “black-box” models whose predictions do not come naturally with an interpretation. A direct consequence is that it is almost impossible to understand why a deep learning model predicts a certain outcome and whether this prediction is reasonable and complies with physics or domain knowledge. Although efforts have been made to achieve varying degrees of interpretability mostly through post-processing⁸⁰, deep learning models are still harder to interpret than simpler traditional ML models, some of which are inherently interpretable⁷⁷.

Traditional ML models are likely to perform better on small training sets ($N < 1000$) than deep learning models. It is not surprising to see battery aging datasets with less than 100 cells tested to their EOL⁴⁸. In such cases, a training set may consist of only $N < 100$ input-output pairs. Simple ML methods such as the elastic net, a regularized linear regression method, random forests (RFs), and Bayesian linear regression are probably the best choices^{27,77,81}. Deep learning approaches, such as deep neural networks, are well-suited for applications where (1) it is reasonably feasible to run large-scale aging test campaigns to generate large training sets ($N \gg 1000$) and (2) model interpretability is not a primary concern.

Publicly available battery aging datasets

Publicly available battery aging datasets have enabled a majority of the research in the battery diagnostics and prognostics field. The University of Maryland’s Center for Advanced Life Cycle Engineering (CALCE)^{82–84} and the National Aeronautics and Space Administration’s (NASA)^{85,86} were among the first organizations to publish publicly available battery aging datasets with greater than 20 cells. The initial work by the team at NASA focused on using an unscented Kalman filter to estimate internal battery states, namely max discharge capacity and internal resistance, and electrochemical model parameters over the course of the cells’ lifetime⁸⁶. Notably, the battery cells in the NASA dataset were subjected to randomized discharge current loads, making the dataset more similar to real-world battery operation and making it more challenging to estimate the cells’ SOH and predict their RUL. The researchers at CALCE first demonstrated RUL prediction on their dataset by using an empirical battery degradation model where the parameters are initialized using Dempster-Shafer theory and updated online using recursive Bayesian filtering⁸². The model was shown to provide accurate non-parametric predictions of battery RUL by evaluating the many Bayesian-filtered model parameters.

While the battery aging datasets from NASA and CALCE were undoubtedly influential for their time, the trend as of late has been to test *more* batteries under *more* operating conditions so that modern machine and deep learning models can be applied⁴⁸. A more recent battery aging dataset from Stanford, MIT, and Toyota Research Institute was used to study the problem of early lifetime prediction (see the section “Early life and trajectory prediction”)²⁷. The researchers then used the early lifetime prediction model in a close-loop optimization algorithm to speed up the process of experimentally searching for a fast charging protocol that maximized a cell’s cycle life⁷⁹. Similar work by a team of researchers at Argonne National Laboratory used a diverse dataset composed of 300 pouch cells with six unique battery cathode chemistries to study the role of battery chemistry and feature selection in early life prediction⁸⁷. Other large datasets include the one from Sandia National Laboratory that was used to study commercial 18650-size NMC-, NCA-, and LFP-Gr cells under different operating

conditions⁸⁸ and the dataset from Oxford⁸⁹ used to study the path-dependency of battery degradation.

A relatively new dataset from the collaborators at Stanford, MIT, the Toyota Research Institute, and the SLAC National Accelerator Laboratory consists of more than 360 21,700-size automotive cells taken from a newly purchased 2019 Tesla Model 3 to study aging under a wide range of cycling conditions⁵⁵. Another large dataset made available this year is the dataset from a research collaboration between Iowa State University (ISU) and Iowa Lakes Community College that contains 251 Li-ion cells cycled under 63 unique cycling conditions²⁶. Both large aging datasets were curated specifically to study ML-based approaches to battery health prognostics and the role of feature generation and engineering in battery lifetime prediction.

Recently, there has been a push to demonstrate battery diagnostic and prognostic algorithms that work on modules and packs operating in the field. One approach to do this is to replicate real-world operating conditions in cell-level laboratory aging experiments. Pozzato et al.³⁰ cycled NMC/Gr + Si 21700 format cylindrical cells using a typical EV discharge profile while periodically characterizing cell health with RPTs. Similarly, Moy et al.⁹¹ cycled 31 cells using synthetically generated autonomous EV discharge profiles based on real-world driving telemetry data. While the datasets are still useful for studying battery degradation under more realistic conditions, they are still synthetic in nature and conducted on cells, making it difficult to understand how the study results translate to real-world packs and modules.

Research into module and pack based battery aging is becoming more prevalent. She et al.⁹² examined telemetry data from electric city buses operated in Beijing, China, finding that incremental capacity features extracted from the voltage readings during constant current charging were predictive of battery health, but changed drastically with the changing seasons (summer, fall, winter, spring) in the city. Similarly, Pozzato et al.³¹ looked at real-world EV data from an Audi E-tron. The team found that DC resistance measured during braking and acceleration along with charging impedance were good predictors of battery SOH. But unfortunately, neither of the aforementioned datasets were made publicly available, and to the best of our knowledge, no other publicly available battery module/pack aging datasets exist.

Publicly available battery aging datasets will continue to play a large role in furthering research in the area of battery diagnostics and prognostics by enabling those without access to battery testers to study battery aging and diagnostic and prognostic modeling. Websites like [Battery Archive](#) are important for sharing and disseminating battery aging data to a wider audience. Additionally, industry-academia collaboration will be key for gathering and disseminating real-world battery module/pack aging data. Access to real-world battery pack aging data will be crucial for studying and developing diagnostic and prognostic algorithms that can work beyond the lab. The next big leap in the battery health diagnostics and prognostics research community will be to understand how models built using lab data perform in the field.

Probabilistic ML techniques and their applications to battery health diagnostics and prognostics

This section introduces a handful of probabilistic ML/deep learning methods for building reliable probabilistic ML pipelines for battery state estimation (see an illustrative flowchart in Fig. 6 in the context of capacity estimation). Following the introduction of each probabilistic ML method, we review the state-of-the-art in applying this method to solve the first three problems (Problems 1–3) on battery diagnostics/prognostics shown in Fig. 4, i.e., SOH estimation, capacity forecasting, and RUL prediction. The other three problems (Problems 4–6) are emerging and will be discussed in “RVM applications to battery diagnostics and prognostics SOH estimation”.

As one proceeds through this section, one will notice that it goes beyond merely addressing the theoretical and battery application aspects of each probabilistic ML technique covered. It also includes specific algorithm examples (e.g., Figs. 7–12) and offers a tutorial-style description of the algorithmic procedures. Our aim is to present easily digestible materials, particularly for newcomers in this field, such as fresh Ph.D. students eager to

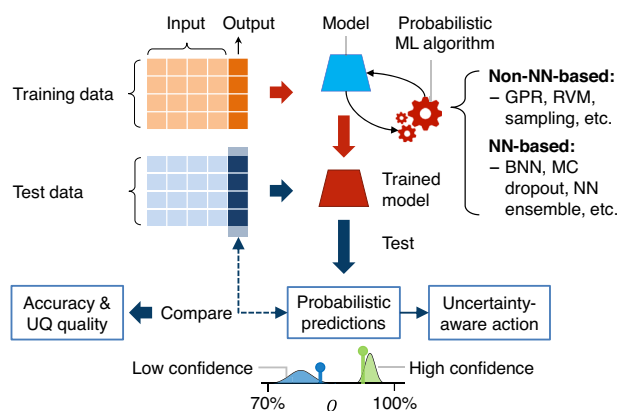


Fig. 6 | A probabilistic ML pipeline for battery state estimation. Among the probabilistic ML algorithms listed here, “GPR” stands for Gaussian process regression (the section “Gaussian process regression”); “RVM” stands for relevance vector machine (the section “Relevance vector machine”); “sampling” stands for sampling methods (the section “Sampling methods”); “BNN” stands for Bayesian neural network (the section “Bayesian neural network”); “MC dropout” stands for Monte Carlo dropout (the section “Bayesian neural network”); and “NN ensemble” stands for neural network ensemble (the section “Neural network ensemble”).

grasp the fundamentals of probabilistic battery diagnostics and prognostics. We also note that there exists a broader spectrum of probabilistic ML methods beyond those discussed in this paper (e.g., see Nemani et al.³⁴); we aim in this paper to highlight a select few that have been most prominently used in, and generally applicable, to the type of problems encountered in battery diagnostics and prognostics.

Before we introduce the probabilistic ML methods, it is meaningful to walk through some key steps when applying probabilistic ML to battery state estimation. As a representative example, Fig. 6 provides a graphical overview of these key steps for capacity estimation. Similar steps can be expected when solving other problems on battery diagnostics and prognostics.

- This pipeline starts with defining an ML model’s input and output. For example, when dealing with capacity estimation by traditional ML models, the input could consist of predictive features extracted from the voltage (V), current (I), and temperature (T) measurements during a partial charge cycle, and the output would be the full capacity (Q) of the cell at that cycle.
- The next step is defining a training and test dataset. An important consideration is that the test dataset should include a decent number (e.g., $\geq 30\%$) of OOD samples to evaluate the generalization performance of a trained ML model. Furthermore, one should avoid randomly assigning samples from the same cell to both training and test datasets. In most cases, all samples from one cell should be exclusively assigned to a training or test dataset. Again, this treatment ensures that the test dataset serves the purpose of evaluating how well a trained ML model generalizes to samples outside of the dataset the model has been trained on.
- The third step is selecting a probabilistic ML algorithm. This section covers three non-neural-network-based algorithms, GPR (the section “Gaussian process regression”), RVM (the section “Relevance vector machine”), and sampling methods (the section “Sampling methods”), and two neural-network-based algorithms, BNN (the section “Bayesian neural network”) and neural network ensemble (the section “Neural network ensemble”). Selecting a probabilistic ML algorithm requires assessing several criteria to ensure the algorithm’s suitability for a specific use case. Several key criteria are listed as follows (see Nemani et al.³⁴ for further details): (1) prediction accuracy (e.g., evaluated by comparing mean predictions with ground truth on a validation dataset split out of a training dataset), (2) quality of uncertainty quantification (i.e., the algorithm’s ability to produce accurate uncertainty estimates), (3) computational efficiency (an

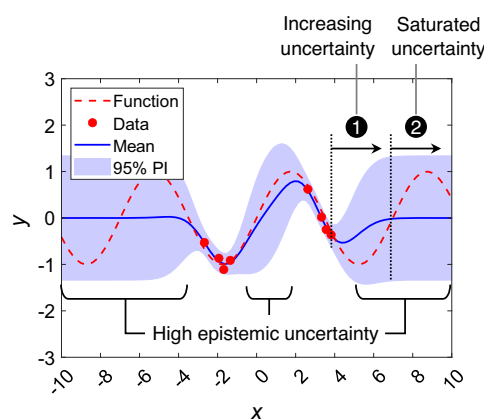


Fig. 7 | An illustration of distance-aware uncertainty estimation by GPR on a 1D regression example. Here, the 95% prediction interval (denoted as “PI” in the legend) represented by the shaded area is computed as $\mu \pm 2\sigma$.

important factor in applications where real-time or near-real-time diagnostics/prognostics may be required), (4) scalability (i.e., the algorithm’s ability to train models based on large volumes of datasets, i.e., in the big data regime), and (5) robustness (i.e., the algorithm’s ability to maintain performance in the presence of outliers, high noise, and adversarial variations in the input data). These criteria may become conflicting objectives that need to be weighed based on the needs and wants of a specific diagnostic/prognostic use case. It is often desirable to experiment with a suite of algorithms and choose one (standalone) or multiple (hybrid) algorithms for a specific use case.

- After selecting the algorithm, one feeds the training data, some observed input-output pairs, into the algorithm. This algorithm then generates a mathematical model that infers something about the underlying process that generated the training data. Using the trained model, one can make probabilistic capacity estimations for cells or their cycle numbers the model has not seen before. Each capacity estimate can be expressed as a probability distribution of a certain type (e.g., Gaussian, log-normal, or exponential) or an empirical probability distribution.
- If one has access to the ground truth for the test data, one can compare the capacity estimates with the observations to derive prediction accuracy metrics, such as the root-mean-square error (RMSE) and mean absolute percentage error, and uncertainty quantification quality metrics, such as the expected calibration error, Area Under the Sparsification Error curve (AUSE), and negative log-likelihood (NLL).

Gaussian process regression

Gaussian process regression methodology

Gaussian process regression (GPR), also known as kriging, is a principled, probabilistic method for learning an unknown function f from a given set of training data comprising N input vectors, $\{\mathbf{x}_i\}_{i=1,\dots,N}$, and N targets, $\{y_i\}_{i=1,\dots,N}$. Here, $\mathbf{x}_i \in \mathbb{R}^D$ is the D -dimensional input feature vector of the i -th training sample, and $y_i \in \mathbb{R}$ is the corresponding one-dimensional output, i.e., a noise-free or noisy observation of f at \mathbf{x}_i . The regression model learned by GPR is non-parametric because this model does not have a predefined functional form. It is common to assume the so-called Gaussian observation model where each observation y_i is an addition of the true function value $f(\mathbf{x}_i)$ and a zero-mean Gaussian noise ε_i :

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, \quad (1)$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. GPR starts by placing a Gaussian process prior on the unknown function f , i.e., $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, where $m(\cdot)$ is the mean function and $k(\cdot, \cdot)$ is the kernel, also known as the covariance function evaluated at \mathbf{x} and \mathbf{x}' ⁹³. For example, if we assemble the N function values

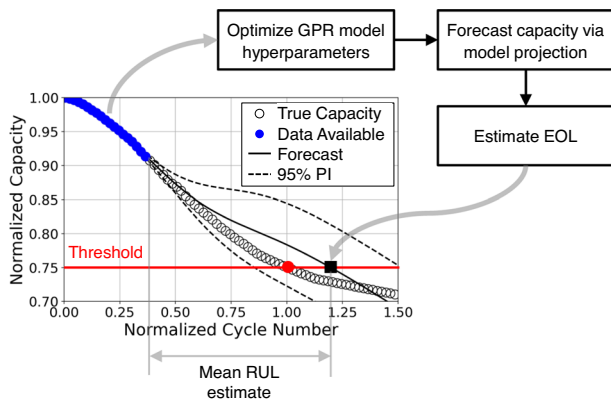


Fig. 8 | Probabilistic capacity forecasting by GPR at a specific cycle. Here, the GPR model is fit to available data and extrapolated to forecast capacity and EOL.

into a vector, $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T$, this vector follows a multivariate (N -dimensional) Gaussian distribution:

$$\mathbf{f} \sim \mathcal{N}(m(\mathbf{X}), \mathbf{K}_{\mathbf{X},\mathbf{X}}), \quad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T \in \mathbb{R}^{N \times D}$ is a matrix assembly of the N training input points, $m(\mathbf{X})$ is a vector of the mean values at these input points, and $\mathbf{K}_{\mathbf{X},\mathbf{X}} \in \mathbb{R}^{N \times N}$ is a covariance matrix that takes the following form:

$$\mathbf{K}_{\mathbf{X},\mathbf{X}} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}. \quad (3)$$

Nonlinearity in the model arises from the kernel $k(\cdot, \cdot)$, which models the covariance between function values at two different input vectors. In practice, we can choose a kernel from many candidates. For example, the probably most popular kernel is the *squared exponential kernel*, also known as the *radial basis function (RBF)* and the *Gaussian kernel*. The squared exponential kernel can be expressed as

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right). \quad (4)$$

where σ_f is the signal amplitude, $\|\cdot\|$ is the L2 norm or the Euclidean norm, the square of which (σ_f^2) defines the signal variance, and l is the length scale. The signal variance (σ_f^2) sets the upper limit of the variance and covariance for the Gaussian process prior (see the covariance matrix in Eq. (2)); the length scale l determines how smooth the approximate function appears (the smaller the length scale, the more rapidly the function changes). These two kernel parameters are two hyperparameters of the GPR model, which, together with the noise standard deviation σ_ϵ , need to be optimized during GPR model training. The squared exponential kernel has been widely used as it is simple and captures a function's stationary and isotropic (dimension-dependent) behavior. Another popular choice is the RBF kernel with *automatic relevance determination (ARD)*⁹⁴, which assigns N different length-scale parameters to the N dimensions rather than using the same parameter as is done by the standard RBF kernel. The resulting RBF-ARD kernel can capture dimension-dependent patterns in the covariance structure.

For notational convenience, we denote the collection of training input-output pairs as a training set, $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, and write the N noisy observations as a vector, $\mathbf{y} = [y_1, \dots, y_N]^T \in \mathbb{R}^N$. For a new, unseen input point \mathbf{x}_* , the predictive distribution of the corresponding observation y_* can be derived based on the conditional distribution of a

multivariate Gaussian as the following:

$$\mu_* = m(\mathbf{x}_*) + k(\mathbf{X}, \mathbf{x}_*)^T (\mathbf{K}_{\mathbf{X},\mathbf{X}} + \sigma_\epsilon^2 \mathbf{I})^{-1} (\mathbf{y} - m(\mathbf{X})), \quad (5)$$

and

$$\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{X}, \mathbf{x}_*)^T (\mathbf{K}_{\mathbf{X},\mathbf{X}} + \sigma_\epsilon^2 \mathbf{I})^{-1} k(\mathbf{X}, \mathbf{x}_*) + \sigma_\epsilon^2, \quad (6)$$

where $k(\mathbf{X}, \mathbf{x}_*) = [k(\mathbf{x}_1, \mathbf{x}_*), \dots, k(\mathbf{x}_N, \mathbf{x}_*)]^T$, denoting a vector of N cross covariances between \mathbf{X} and \mathbf{x}_* .

GPR is a probabilistic ML method most well-known for its distance-aware uncertainty quantification capability. This capability is illustrated in Fig. 7, where a simple sine function is adopted to generate synthetic data after adding zero-mean Gaussian noise. Two observations can be made from this figure. First, high epistemic uncertainty due to a lack of data is associated with test points far from the eight training points. The GPR model seems to produce predictive uncertainty estimates that properly capture the high epistemic uncertainty at these OOD test points. Second, as a test point deviates from the training data distribution (e.g., when x starts to become larger than 4), the predictive uncertainty first increases due to the distance-aware property of GPR and then saturates to a maximum constant (i.e., $\sigma_*^2 \approx \sigma_f^2 + \sigma_\epsilon^2$). The above briefly overviews the math behind GPR and its uncertainty quantification capability. Our recent tutorial on uncertainty quantification of ML models³⁴ provides a more detailed explanation of GPR.

Figure 8 shows how GPR operates to forecast the capacity of future cycles probabilistically at a specific charge/discharge cycle. We first train a GPR model based on the available capacity data (blue points). This GPR model uses an empirical capacity fade model as the prior mean function to capture the known fade trend. The model training optimizes the GPR model's hyperparameters by maximizing the likelihood of observing the capacity data. Intuitively, we fit a GPR regression model to available capacity data and use this model to make predictions for future cycles without capacity data. Because GPR is a probabilistic ML technique, the predictions are in the form of a mean curve, the solid line, and a 95% prediction interval, the dashed lines. So, in the next step, we forecast capacity beyond the current cycle using the trained GPR model, making predictions outside our data distribution. We then estimate the mean EOL when the mean prediction curve down-crosses a predefined capacity threshold or EOL limit. The black square is the mean prediction. We can imagine having a prediction interval around this mean, representing the uncertainty of our EOL prediction. We are often interested in knowing the RUL, i.e., the number of remaining cycles till the EOL limit. Our RUL estimate can be obtained by simply subtracting the current cycle number from the predicted EOL. Since this device was cycled to its EOL, we have the entire capacity trajectory and true EOL. We can compare the prediction and truth to know how well our GPR model does.

Now imagine we repeat the prediction steps in Fig. 8 at every cycle, as the battery is used in the field. Figure 9 shows how the prediction evolves from an early-life cycle to a late-life cycle. This figure panel shows six snapshots of probabilistic capacity forecasting by GPR at six different cycles. As we move along the cycle axis, we have more and more capacity data to train a GPR model and our prediction horizon till the EOL becomes shorter and shorter. As a result, the EOL and RUL predictions become more and more accurate. These predictions converge to the ground truth at around halfway through the lifetime. Also, We can see the prediction interval for the EOL, in general, gets narrower with time, indicating reduced predictive uncertainty, which is also what we expect to see.

GPR applications to battery diagnostics and prognostics SOH estimation

Two notable efforts applying GPR to SOH estimation were made almost simultaneously^{72,95}. Both studies extracted features from the raw voltage vs. time (V vs. t) curve acquired from a charge cycle. Richardson et al.⁷² used the time differences between several equispaced voltage values and their minimum as the input features to a GPR model. These time differences were

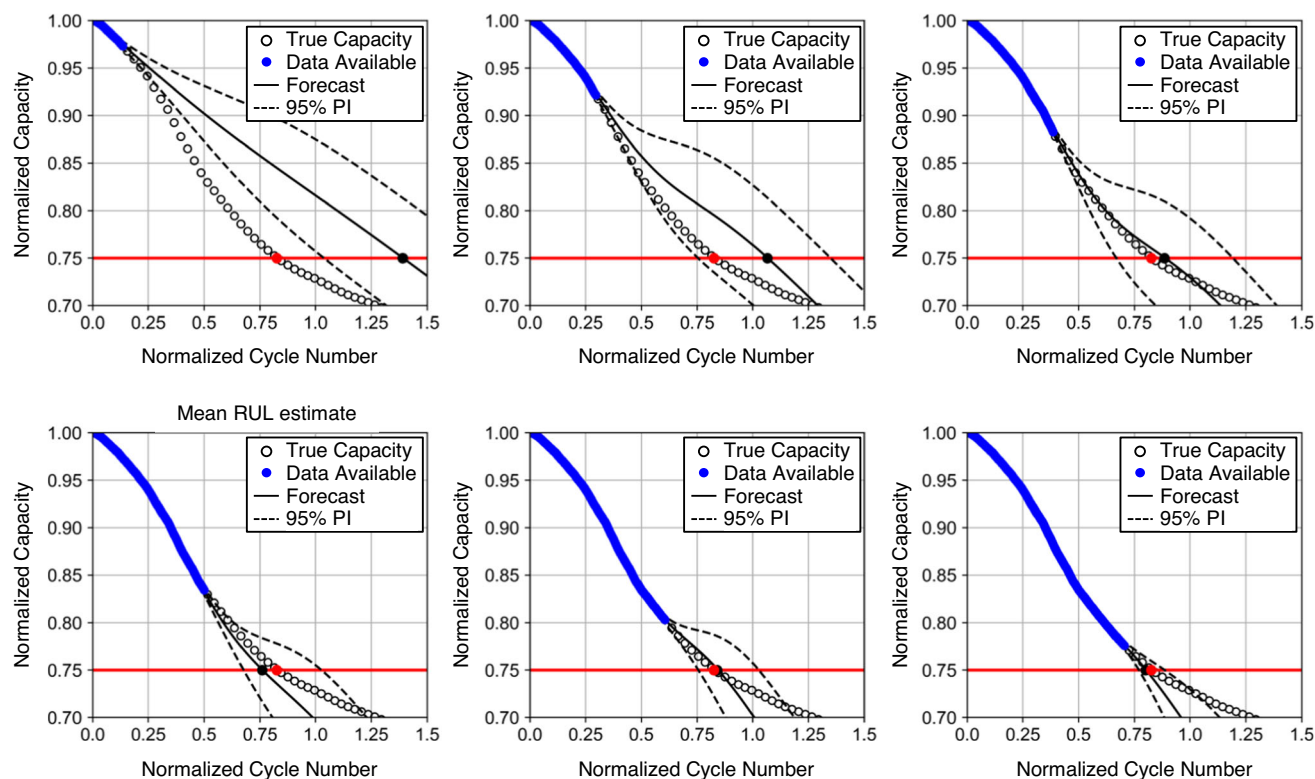


Fig. 9 | Probabilistic capacity forecasting by GPR at multiple cycles over the life span. Here, the GPR model is fit to available data and extrapolated to EOL. The various lengths of data available change the projected trajectory.

computed based on a segment of a full charge curve (after smoothing) within its constant-current (CC) portion. Yang et al.⁹⁵ took a different feature engineering path by extracting time and slope features from a full charge curve consisting of both the CC and constant-voltage portions. It is noted that earlier studies on SOH estimation investigated similar features extracted from a charge curve starting at a partially discharged state^{13,96}. Both studies^{72,95} evaluated their algorithms on a battery aging dataset from the NASA Ames Prognostics Center of Excellence Data Set Repository (e.g., the Randomized Battery Usage dataset⁸⁶ used in ref.⁷²).

These two early applications of GPR to SOH estimation reported two unique properties of GPR: nonparametric regression, making the regression model self-adaptable to data complexity, and uncertainty estimation under a Bayesian framework and with distance awareness, enabling principled quantification of predictive uncertainty and reliable detection of OOD samples. Additionally, GPR is known for its minimal overfitting risk due to using a Bayesian probabilistic framework. These desirable properties may have driven many later studies that investigated the applicability of GPR to SOH estimation when only partial charge curves are available. Two examples of such investigations examined features extracted from the incremental capacity vs. voltage curve during partial charge⁷³ and features extracted from the capacity vs. voltage curve during partial charge⁹⁷. As discussed next, GPR can extrapolate reasonably well when a prior mean function is properly defined. However, GPR only operates well on small datasets and has limited scalability to bigger datasets³⁴.

SOH forecasting and RUL prediction

The first application of GPR in the battery diagnostics and prognostics literature was SOH forecasting, not SOH estimation. It was reported in a comparative study on resistance and capacity forecasting led by a group of researchers at NASA's Ames Research Center⁹⁸. This study compared two regression techniques, polynomial regression and GPR, and one state estimation technique, particle filtering, in forecasting resistance and capacity. This comparative study was an extended version of the probably first-ever publication on battery diagnostics/prognostics, led by the same

group of researchers¹¹, which used a combination of RVM and particle filtering for capacity forecasting. SOH forecasting using GPR has an ~10-year longer history than SOH estimation using GPR. After the first application of GPR to SOH forecasting in the late 2000s, two notable studies attempted to improve the extrapolation performance of GPR, essential to long-term SOH forecasting, by using explicit prior mean functions^{99,100}. Note that the default option for the prior mean function is either zero or a non-zero constant⁹³. An empirical capacity fade model could be used as an explicit mean function, allowing the GPR model to capture the trend of degradation encoded in the capacity fade model^{100,101}.

All the above studies on SOH forecasting assume that use conditions (e.g., charge and discharge C-rates and temperature) are time-invariant. This assumption may not hold in many real-world applications. A more realistic scenario is that these use conditions vary randomly over time but approximately follow a duty cycle pattern. As a follow-up to their earlier study on SOH forecasting¹⁰², Richardson et al. defined a capacity transition model to predict the capacity change during each usage period following a load pattern. A GPR model was built to approximate the relationship between features extracted from a load pattern (input) and the capacity change within this usage period (output). The outcome was the ability to forecast capacity probabilistically under time-varying use conditions. Two more recent studies also examined capacity forecasting under time-varying use conditions, specifically in cases where future charge and discharge C-rates vary significantly with cycle^{103,104}. Similar to the study by Richardson et al.¹⁰², these two more recent studies also consider future use conditions when designing the input to an ML model. Specifically, they used charge and discharge currents in future cycles as part of the ML model input. The difference is that these two studies additionally incorporated the current or recent cell state into the model input. The cell state was characterized by either (1) a combination of features from electrochemical impedance spectroscopy (EIS) measurements in the current cycle and those from voltage and current measurements in the current and some past cycles¹⁰³ or (2) only features from historical voltage and current measurements¹⁰⁴. GPR was not used as the ML algorithm in either study. Instead, an ensemble of

XGBoost regressors was used by Jones et al.¹⁰³ to quantify forecasting uncertainty, while uncertainty quantification was not considered by Lu et al.¹⁰⁴. Overall, it is interesting to see both studies focused on features with physically meaningful connections to future degradation when designing the ML model input. In fact, formulating a meaningful forecasting problem and designing highly predictive input features should be the centerpiece of almost any data-driven SOH forecasting effort.

Finally, it is worth noting that multiple probabilistic ML models can be combined to form a hybrid model for SOH forecasting or RUL prediction. In what follows, we briefly discuss three examples of hybrid modeling involving GPR.

- The first example is the delta learning approach employed by Thelen et al. in their study on battery RUL prediction¹⁰¹. The basic idea of this approach is correcting initial RUL predictions by a GPR capacity forecasting model with a data-driven ML model. The prior mean function of the GPR model was explicitly designed to be an empirical capacity fade model. The approach was demonstrated on three open-source datasets, a simulated dataset, and one proprietary dataset. Initial RUL predictions by GPR capacity forecasting models were considerably under-confident as compared to the GPR delta learning approach with a GPR capacity forecasting model (predictor) and a GPR RUL error correction model (corrector), which was well calibrated on the original dataset. In contrast, the random forest delta learning approach using a probabilistic random forest model as the corrector was over-confident on the original dataset, but exhibited better calibration than the GPR delta learning approach on the simulated OOD dataset.
- Another example is the use of a co-kriging model to forecast capacity degradation by combining two data sources: (1) a high-fidelity source consisting of the capacity measurements from the test cell (whose capacity trajectory beyond the current cycle needs to be predicted) up to the current cycle and (2) a low-fidelity source comprising capacity measurements from other cells of the same or a similar design¹⁰⁵. Similar to the delta learning approach studied by Thelen et al.¹⁰¹, this second study attempted to build a corrective GPR model to compensate for the deviation of an initial GPR model built based on low-fidelity data to depict an “average” degradation trajectory.
- The third example is an effort to modify vanilla GPR models by incorporating electrochemical and empirical knowledge of capacity degradation (i.e., the dependencies of capacity degradation on two cycling condition variables, named temperature and depth of discharge)¹⁰⁶. These two dependencies were captured through the Arrhenius law (temperature) and a polynomial equation (depth of discharge), respectively, and encoded as a compositional covariance function (or kernel) within GPR. Unlike the first two examples, which are purely data-driven, this third example attempted to integrate some physics of degradation into the GPR formulation, which can be treated as a physics-informed probabilistic ML approach.

Relevance vector machine

Relevance vector machine methodology

Suppose we have access to a set of training samples, each sample consisting of an input–output pair, (\mathbf{x}_i, y_i) , $i = 1, \dots, N$, where $\mathbf{x}_i \in \mathbb{R}^D$ is the D -dimensional input features of the i -th training sample, $y_i \in \mathbb{R}$ is the corresponding output (also known as the target or the observation of the state of interest), and N is the number of samples. We are interested in learning a one-to-one mapping from the input (feature) space to the output (state) space. Similar to GPR, RVM also assumes that the observations are samples from a Gaussian observation model. Unlike GPR, which does not assume any functional form of this mapping, RVM approximates this mapping as a parametric, linear kernel regression function¹⁰⁷. This regression function takes the following form:

$$y(\mathbf{x}) = \sum_{i=1}^N \omega_i K(\mathbf{x}, \mathbf{x}_i) + \omega_0 + \varepsilon, \quad (7)$$

where \mathbf{x} is an input feature vector whose target may be unknown and needs to be inferred, $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function comparing the test input \mathbf{x} with each training input \mathbf{x}_i , ω_i is the kernel weight measuring the importance (or relevance) of the i th training sample, ω_0 is a bias term, and ε is a zero-mean Gaussian noise, i.e., $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The bias term and N kernel weights form a $(N + 1)$ -element weight vector, written as $\boldsymbol{\omega} = [\omega_0, \omega_1, \dots, \omega_N]^T$. If we define a design vector consisting of a constant of one and the N kernel functions, i.e., $\boldsymbol{\phi} = [K(\mathbf{x}, \mathbf{x}_1), \dots, K(\mathbf{x}, \mathbf{x}_N)]^T$, we can rewrite Eq. (7) in a convenient vector form, $y(\mathbf{x}) = \boldsymbol{\omega}^T \boldsymbol{\phi} + \varepsilon$. The original RVM formulation follows a hierarchical Bayesian procedure by assuming the $(N + 1)$ weights follow a zero-mean Gaussian prior, whose inverted variances, denoted as $\{\alpha_i\}_{i=0, \dots, N}$, and the inverted noise variance, σ_ε^{-2} , all follow Gamma distributions (hyperpriors).

Training the model in Eq. (7) using sparse Bayesian learning estimates the posterior of the weight vector $\boldsymbol{\omega}$ and noise variance σ_ε^2 via iterative optimization¹⁰⁷. In practice, the posterior for most weights becomes highly peaked at zero, effectively “pruning” the corresponding kernel functions from the trained model. The remaining training samples with non-zero weights are known as *relevance vectors*, typically accounting for a very small portion of the training set (e.g., 5–20%). This unique attribute of sparsity makes the RVM attractive both in terms of generalization performance and test-time efficiency.

Figure 10 illustrates battery capacity estimation by a trained RVM model based on features extracted from voltage and current measurements during partial charge¹³. This SOH estimator possesses two desirable attributes: (1) statistical estimation, i.e., instead of providing a mere point estimate for the SOH parameter, this estimator generates a probability distribution as the parameter estimate; (2) sparsity, i.e., the estimator selectively utilizes only a small subset of feature vectors from the training dataset, known as *relevance vectors*, for real-time health inference (see, for example, the extreme posterior peakness at zero for ω_2 and ω_N). These two attributes offer several advantages for online health inference: (1) statistical estimation enables concurrent estimation of the parameter while quantifying the associated uncertainty, and (2) sparsity enhances the computational efficiency of online inference.

RVM applications to battery diagnostics and prognostics

SOH estimation

As shown in Fig. 10, RVM can be applied to estimate battery capacity from features extracted from readily available voltage and current measurements. Such applications were first attempted in two studies, one focusing on implantable-grade LCO cells¹³ and the other focusing on NMC cells¹⁰⁸. In the former study¹³, five characteristic features, some correlated strongly with capacity, were extracted from a test cell’s voltage vs. time and current vs. time curves at a specific charge cycle that started at a partially discharged state. These features were then fed as input (e.g., \mathbf{x} in Fig. 10) into a trained RVM regression model that produced as output a Gaussian-distributed capacity estimate (e.g., Q in Fig. 10). The sparsity property of RVM made this regression model much smaller than a full-scale model. For example, each cross-validation trial used only <4% of training samples as the relevance vectors to build the final regression model, improving the computational efficiency and generalization of online capacity estimation.

In the later study¹⁰⁸, a feature of predictive power was found to be the sample entropy of a short voltage sequence (time series) measured during a hybrid pulse power characterization test. This feature was concatenated with temperature to form the input vector to an RVM regression model, which outputs a Gaussian-distributed capacity estimate. It is interesting to see the inclusion of use condition parameters (e.g., temperature as reported in Hu et al.¹⁰⁸) in the input of a data-driven ML model. Such a treatment builds *condition awareness* into the ML model, making it applicable under varying use conditions. Similar approaches have been taken in studies on capacity forecasting, as discussed in the section “GPR applications to battery diagnostics and prognostics”.

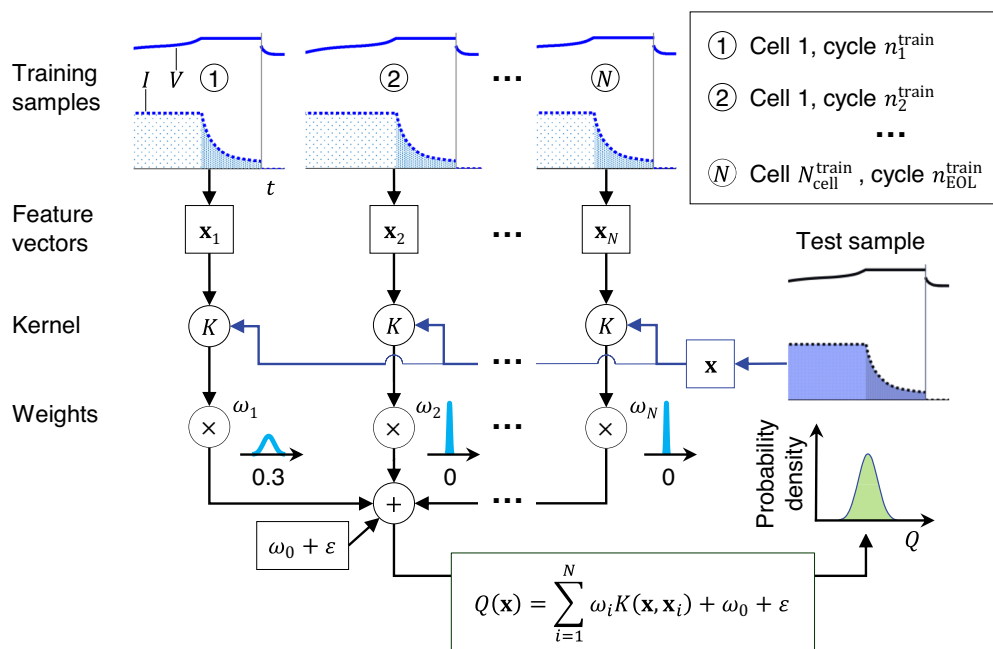


Fig. 10 | Probabilistic capacity estimation by RVM using partial charge curves. Example ML pipeline for capacity estimation using RVM trained on voltage-time and current-time data during cell charging. Figure adapted from Fig. 2 in Hu et al.¹³

It is also widely accepted that a “one-ML-method-fits-all” approach does not work in battery diagnostics and prognostics. In some applications, one ML method may perform better than another regarding prediction accuracy. But, in other applications, accuracy comparisons between these two methods may look very different. Some limited efforts have been made to benchmark different ML methods for SOH estimation. An example is a comparative study on four data-driven ML methods, namely linear regression, support vector machine, RVM, and GPR, using features extracted from capacity vs. voltage curves during discharge¹⁰⁹. These features included the standard deviations of the discharge capacity (Q) and cycle-to-cycle discharge capacity difference (ΔQ), calculated, respectively, from a measured sample of the discharge capacity vs. voltage curve at the current cycle ($Q(V)$) and a measured sample of the discharge capacity difference (between the current cycle and a fixed early cycle) vs. voltage curve ($\Delta Q(V)$).

Like most other studies on battery diagnostics and prognostics, the above comparison exclusively focused on prediction accuracy by looking at error metrics such as RMSE and maximum absolute error. Few research or benchmarking efforts were made to study the quality of uncertainty quantification, i.e., how well an estimate of a model’s predictive uncertainty (known) on a test sample reflects the model’s prediction error (unknown) on this sample³⁴. Additionally, we see that most studies worked with small datasets from limited numbers (mostly <100) of cells. In the small data regime, examining predictive uncertainty is even more important than in the big data regime, simply because (1) small training datasets possess limited representativeness of real-world scenarios, and (2) ML models may generalize poorly to OOD data. Although these existing studies reported high accuracy on small, carefully crafted test datasets mostly acquired from laboratory testing, these accuracy numbers are unlikely to generalize to real-world applications where we would expect wider ranges of and more complex use conditions, higher cell-to-cell variability, and larger measurement noise.

SOH forecasting and RUL prediction

The first application of RVM to battery prognostics was pioneered by a group of researchers at NASA’s Ames Research Center¹¹. The same group of researchers also led the first application of GPR to battery prognostics⁹⁸, as discussed in the section “RVM applications to battery diagnostics and

prognostics SOH estimation”. The role RVM served was identifying mean regression curves on a charge transfer resistance vs. time dataset and electrolyte resistance vs. time dataset, both acquired from EIS. Each mean regression curve was then fitted to a simple two-parameter exponential model to obtain an estimate of the two model parameters. This estimate was treated as an initial ($t = 0$) estimate of the exponential model parameters in a discrete-time state-space model, solved using particle filters for capacity forecasting and RUL prediction. RUL prediction results were shown as an empirical probability distribution that became narrower and more centered at the true RUL as the cycle number where the prediction was made increased. Such plots later became a standard way to visualize results by probabilistic RUL prediction methods^{82,110–112}.

Two later, more direct applications of RVM to battery prognostics were explored with the formulation of two vastly different approaches^{12,113}. Wang et al. performed RVM regression on the capacity vs. cycle number data available to a test cell whose future capacity and RUL were unknown and then fitted a three-parameter variant of the well-known double exponential capacity fade model⁸² to only the capacity (Q) and cycle number (t) data of the relevance vectors¹². Capacity forecasting and RUL prediction were achieved by extrapolating the fitted capacity fade model to a predefined EOL limit. It is important to note that, similar to the first application⁹⁸, the RVM regression model, fitted to an SOH vs. cycle number dataset, was not directly used for capacity forecasting. More specifically, the forecasting was not done by extrapolating the RVM regression model, unlike the capacity forecasting studies using GPR models with empirical fade models as the “built-in” prior mean functions, as discussed in “RVM applications to battery diagnostics and prognostics SOH estimation”. Li et al. took a different approach by formulating capacity forecasting as a time series prediction problem¹¹³. RVM was employed to map the current and several past cycles’ capacity observations to the next cycle’s capacity observation, enabling one-step-ahead prediction. Capacity observations at cycles beyond the next cycle were predicted via iterative one-step-ahead prediction (i.e., marching over time). Again, capacity forecasting was not achieved by extrapolating an RVM regression model fitted to capacity vs. cycle number data. It suggests that simply extrapolating a data-driven regression model without consideration of the capacity

fade trend may not yield reliable capacity forecasting, especially for long-term forecasting.

Bayesian neural network

Bayesian neural network methodology

A neural network f_{NN} makes a prediction for an output variable at an input feature: $\hat{y} = f_{\text{NN}}(\mathbf{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes all tunable parameters of the neural network (e.g., the neural network weight and bias terms). Given training samples (\mathbf{x}_i, y_i) , $i = 1, \dots, N$, the neural network training process seeks to set $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ that minimizes a loss function, commonly the mean squared error:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \left\{ \frac{1}{N} \sum_{i=1}^N [f_{\text{NN}}(\mathbf{x}_i; \boldsymbol{\theta}) - y_i]^2 \right\}. \quad (8)$$

This optimization problem is typically solved via gradient-based algorithms such as stochastic gradient descent^{114,115} or Adam¹¹⁶. The resulting $\boldsymbol{\theta}^*$ is single-valued, and subsequent new prediction using this trained neural network would also be single-valued as well: $\hat{y}_{\text{new}} = f_{\text{NN}}(\mathbf{x}_{\text{new}}; \boldsymbol{\theta}^*)$.

In order to capture the uncertainty in determining $\boldsymbol{\theta}$, one can solve for $\boldsymbol{\theta}$ in a probabilistic manner following the Bayesian framework^{117–119}, and seek the entire *distribution* of plausible $\boldsymbol{\theta}$ values instead of a single-valued “best fit”. Such an approach entails constructing a Bayesian neural network (BNN). In a BNN, $\boldsymbol{\theta}$ is treated as a random variable with an associated probability density function (PDF) representing its uncertainty. When training data become available, the PDF of $\boldsymbol{\theta}$ is updated following Bayes’ rule:

$$p(\boldsymbol{\theta}|\mathbf{x}, y) = \frac{p(\mathbf{x}, y|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x}, y)} = \frac{p(y|\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(y|\mathbf{x})}, \quad (9)$$

where $p(\boldsymbol{\theta}|\mathbf{x}, y)$ is the posterior PDF (updated uncertainty given training data), $p(\boldsymbol{\theta})$ is the prior PDF (initial uncertainty before seeing training data), $p(y|\mathbf{x}, \boldsymbol{\theta})$ is the likelihood PDF, and $p(y|\mathbf{x})$ is the marginal likelihood or model evidence that is constant with respect to $\boldsymbol{\theta}$. Solving the Bayesian inference problem constitutes computing or characterizing the posterior $p(\boldsymbol{\theta}|\mathbf{x}, y)$. Once the posterior becomes available, its uncertainty can be propagated to predictions by first drawing posterior samples $\boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}|\mathbf{x}, y)$ and then evaluating the neural network $\hat{y}_{\text{new}}^{(i)} = f_{\text{NN}}(\mathbf{x}_{\text{new}}; \boldsymbol{\theta}^{(i)})$ for each sample. The set of neural network predictions represent the *posterior-pushforward* distribution that is solely due to the epistemic uncertainty in the neural network parameters. In contrast, the *posterior predictive* distribution would additionally include the aleatory uncertainty from the output observation noise, often portrayed by samples in the form $y_{\text{new}}^{(i)} = f_{\text{NN}}(\mathbf{x}_{\text{new}}; \boldsymbol{\theta}^{(i)}) + \epsilon^{(i)}$. Hence, a distribution of predicted values will be generated to reflect the residual uncertainty in the neural network model parameters.

Solving for the posterior is highly challenging for BNNs due to the high dimensionality of $\boldsymbol{\theta}$ (often thousands to millions in neural networks). Markov chain Monte Carlo^{120–123}, which are classical Bayesian inference algorithms designed to generate samples from the exact posterior, do not scale well to such high dimensions in practice. The exploration-efficient Hamiltonian Monte Carlo (HMC)^{124,125} has been used on some BNNs but usually for smaller cases with hundreds of parameters. Alternatively, variational inference (VI)^{126,127} forms an optimization problem to find the best approximate posterior from a class of parameterized distributions. Let $q(\boldsymbol{\theta}; \boldsymbol{\lambda})$ denote the approximate posterior density parameterized by $\boldsymbol{\lambda}$, VI minimizes the Kullback-Leibler divergence from the true posterior to the approximate posterior:

$$\begin{aligned} \boldsymbol{\lambda}^* &= \arg \min_{\boldsymbol{\lambda}} D_{\text{KL}} [q(\boldsymbol{\theta}; \boldsymbol{\lambda}) || p(\boldsymbol{\theta}|\mathbf{x}, y)] \quad (10) \\ &= \arg \max_{\boldsymbol{\lambda}} \underbrace{\mathbb{E}_{q(\boldsymbol{\theta}; \boldsymbol{\lambda})} [\ln p(y|\mathbf{x}, \boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}) - \ln q(\boldsymbol{\theta}; \boldsymbol{\lambda})]}_{\text{evidence lower bound (ELBO)}}, \quad (11) \end{aligned}$$

where the second equation is simplified to the well-known evidence lower bound that no longer involves the marginal likelihood and can be approximated via Monte Carlo (MC) sampling. By sidestepping posterior sampling with an optimization problem, VI effectively trades off some posterior accuracy for scalability, making it more suitable for BNNs.

The simplest form of VI is the mean-field VI using Gaussians, where $q(\boldsymbol{\theta}; \boldsymbol{\lambda}) = \prod_{k=1}^K q_k(\boldsymbol{\theta}_k; \boldsymbol{\lambda}_k) = \prod_{k=1}^K \mathcal{N}(\boldsymbol{\theta}_k; \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2)$ are set to independent (mean-field) Gaussians¹²⁸, involving the optimization of a $2K$ -dimensional $\boldsymbol{\lambda}$. Gradient information for the VI optimization can also be obtained through back-propagation¹²⁸. However, such a mean-field approach cannot capture parameter correlations and tends to under-predict the uncertainty¹²⁶. While a natural extension is to incorporate a full-covariance instead of the independence assumption, however, tracking all entries of a dense covariance matrix would require $\mathcal{O}(K^2)$ -dimensional $\boldsymbol{\lambda}$, rendering it often too expensive and thus rarely used for BNNs. Other advanced representations of $q(\boldsymbol{\theta}; \boldsymbol{\lambda})$ are possible, for example, via normalizing flows¹²⁹ and transport maps¹³⁰ that parameterize the mapping from the posterior random variable $\boldsymbol{\theta}$ to a standard normal reference random variable.

Stein variational gradient descent (SVGD)¹³¹ also approximates the posterior through an optimization problem but uses particles. SVGD leverages the relationship between the (functional) gradient of objective in Eq. (10) to the Stein discrepancy, the latter which can be approximated using a set of particles. The particles’ positions are then iteratively updated following the gradient-descent direction, transporting them towards the target posterior distribution $p(\boldsymbol{\theta}|\mathbf{x}, y)$. Further enhancements such as Stein variational Newton^{132,133} that makes use of second-order (Hessian) information, and projected SVGD¹³⁴ that finds low dimensional data-informed subspaces, have also been developed.

Lastly, MC dropout is a regularization technique for training deep neural networks¹³⁵ but has been shown to approximate the posterior predictive distribution under a specific Bayesian setup¹³⁶. Adding MC dropout to an existing deterministic deep neural network training infrastructure is very easy and essentially involves generating a set of sparse neural networks by randomly setting some weight parameters to zero. However, MC dropout is not formulated to tackle the Bayesian formulation in Eq. (9), and thus is limited in handling general choices of prior $p(\boldsymbol{\theta})$ and likelihood $p(\mathbf{x}, y|\boldsymbol{\theta})$.

Bayesian neural network applications to battery diagnostics and prognostics

SOH estimation

The use of BNN for battery diagnostics/prognostics has been few in number and largely lacked rigorous analysis of its Bayesian uncertainty quantification. For example, Kim et al.¹³⁷ proposed a knowledge-infused BNN for on-board SOH estimation and RUL prediction of Li-ion batteries in EVs. Their approach incorporated novel domain knowledge by (a) designing impedance-related features based on discharge voltage slopes that have been observed to be correlated with degradation, and (b) introducing into an RNN a knowledge-infused block that uses an empirical double-exponential model for degradation estimation. The RNN was then turned into a BNN via a combination with MC dropout. However, the work had no mentioning of the prior and likelihood, both central for establishing the Bayesian problem formulation in Eq. (9). Elsewhere, Xu et al.¹³⁸ built BNNs to predict the SOH of retired batteries by leveraging unique data from EIS experiments. The paper provided detailed experimental setup, data acquisition, and feature extraction highlighting the use of an equivalent circuit model and ARD. However, similar to the previous work, information regarding the BNN prior and likelihood were missing. While the paper did mention the use of VI for BNN training, it failed to clarify the VI method and what variational families were employed (e.g., if using mean-field Gaussian VI).

SOH forecasting and RUL prediction

In the work of Zhu et al.¹³⁹, the authors used MC dropout to create a general RUL prediction framework, with a demonstration example of battery degradation from a laboratory setting, that also featured an active learning

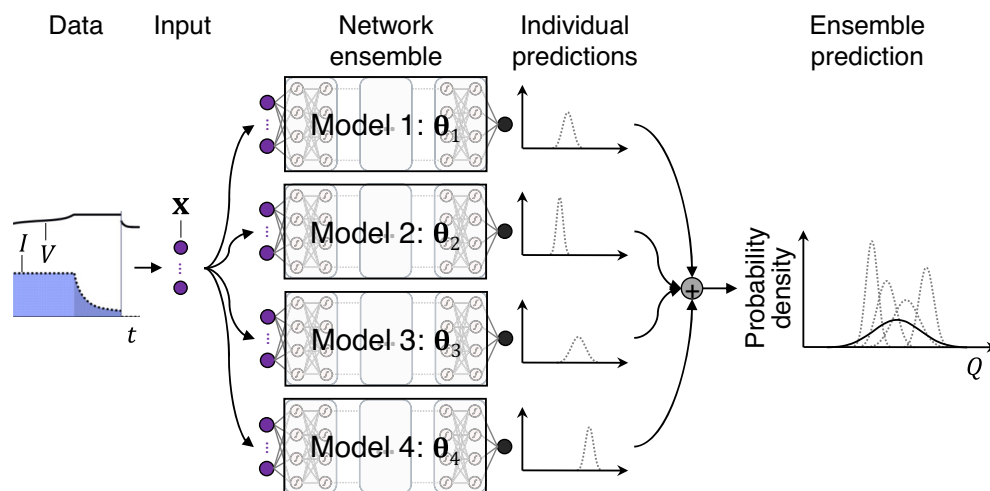


Fig. 11 | A flowchart illustrating the process of constructing a neural network ensemble. In this flowchart, simple averaging is used to combine the Gaussian-distributed predictions by four independently trained neural networks into an ensemble prediction.

procedure for choosing the next sampling point using the posterior predictive variance as the acquisition function. Hong et al.¹⁴⁰ introduced a “first full end-to-end deep learning framework” for predicting Li-ion battery RUL through a dilated CNN architecture that incorporated temporal measurements of battery terminal voltage, current, and cell temperature. The paper then used an explicit ensembling procedure (see next section) as an approximation to BNN.

In contrast to the aforementioned literature, the papers^{141,142} clearly specified the prior and likelihood of their Bayesian setups along with the algorithm for solving the posterior. The former employed mean-field Gaussian VI to build a Bayesian mixture neural network in the form of a hybrid of a Bayesian CNN and LSTM, for predicting the RUL in multiple battery datasets. The latter adopted HMC and VI by backpropagation¹²⁸ to construct BNNs for general RUL prediction without focusing on batteries, and demonstrated instead on an open dataset of turbofan engines.

Overall, research for SOH and RUL prediction is seeing increasing use of BNNs, recognizing the importance of uncertainty quantification in deep learning models that generally tend to be opaque and not interpretable. However, the majority of these BNN works simply cite the connection to MC dropout without mentioning the assumptions and conditions that accompany these off-the-shelf tools. This can be a dangerous practice and lead to incorrectly quantified uncertainties not justified by the data or not intended by the modeler. More careful analysis of the Bayesian results would be warranted, for example, by diagnosing how close the dropout posterior is to the true posterior $p(\theta|\mathbf{x}, y)$. This would require the probing (at least recognition) of *posterior results* (not just posterior-predictive results and not just looking at RMSE of the predictions), which currently are often from BNN literature.

Neural network ensemble

Neural network ensemble methodology

Approaches that combine predictions by multiple ML models to derive a final prediction can be categorized as ensemble learning approaches. The key idea is to introduce diversity among models in the ensemble, encouraging member models to agree more when a test sample falls inside the training data distribution and disagree more when the test sample is OOD. Diversity can be created in many different ways. The sampling methods described in the section “Sampling methodologies” can generally be treated as ensemble learning approaches. For example, bagging (a.k.a. bootstrap aggregating) builds a diverse set of member models in an ensemble by creating random subsets of the original training set and using each subset to train a member model (see Fig. 12 for an illustrative example). These methods allow making probabilistic predictions using deterministic

ML techniques (e.g., the elastic net [end-to-end early prediction paper] and random forest¹⁴).

A recent effort attempted to achieve diversity among neural networks by training multiple neural networks of the same architecture, each with a random (thus different) parameter initialization and a unique sequence of randomly sampled mini-batches, i.e., simply following the standard stochastic gradient descent procedure¹⁴³. The resulting ensemble captures the predictive uncertainty due to observational noise in the target (y) of an *aleatory* nature and insufficient training data of an *epistemic* nature. This recent effort specifically targeted uncertainty quantification of deep neural networks, as they produced state-of-the-art prediction accuracy on many benchmarking problems but had been found to give often over-confident predictions. These predictions, if incorrect, can quickly diminish the value of predictive modeling in safety-critical applications and substantially damage users’ trust in the ML model.

Constructing a neural network ensemble involves (1) training individual neural networks, of which each predicts a Gaussian-distributed output capturing aleatory uncertainty, and (2) aggregating the Gaussian predictions by these individual models as a Gaussian mixture to capture epistemic uncertainty. These two steps are illustrated in Fig. 11 and detailed below in the context of battery capacity estimation.

- **Step 1: Training multiple neural networks with Gaussian output layers** Suppose we have a measured input (\mathbf{x}) from a battery cell during a charge cycle. We are interested in estimating the unknown cell capacity (Q). We independently train multiple (M) neural networks of the same architecture; each training run starts with a random initialization of the network parameters (θ) and operates on randomly sampled mini-batches. Each neural network predicts a Gaussian distribution of \hat{Q} , $\hat{Q} \sim \mathcal{N}(\hat{\mu}(\mathbf{x}; \theta), \hat{\sigma}^2(\mathbf{x}; \theta))$ characterized by two network outputs: the mean $\hat{\mu}(\mathbf{x}; \theta)$ and variance $\hat{\sigma}^2(\mathbf{x}; \theta)$. The predicted variance represents the network-learned observational noise in capacity (Q) measurements (aleatory uncertainty). Network training identifies a local optimum of the neural network parameters θ (e.g., weights and biases), which yields a minimal negative log-likelihood loss derived from the training dataset.
- **Step 2: Aggregating individual predictions as a Gaussian mixture** This second step aggregates the M individual predictions through simple averaging. This model aggregation allows quantifying the parameter (θ) uncertainty arising from insufficient training data. The final ensemble prediction comes from a Gaussian mixture model consisting of the M Gaussian distributions predicted by the member neural networks in the ensemble. The ensemble-predicted Gaussian distribution

takes the following form:

$$p(\hat{Q}(\mathbf{x})) = \frac{1}{M} \sum_{j=1}^M P_{\text{Gauss}}(\hat{Q}; \hat{\mu}(\mathbf{x}; \theta_j), \hat{\sigma}^2(\mathbf{x}; \theta_j)). \quad (12)$$

The M independently trained networks tend to produce more different mean predictions on OOD data than in-distribution data, resulting in larger variance estimates on OOD data¹⁴⁴.

The probabilistic prediction shown in Eq. (12) becomes critical when only a small training dataset is available due to a limited experimental budget, technical constraints, and time. The data-size limitation applies to battery aging tests, given that they are typically costly and labor-expensive to run and may last for many months to years, making it practically difficult to test many cells under a wide range of use conditions. A more detailed description of the neural network ensemble method can be found in the original research paper¹⁴³ and a recent tutorial paper on uncertainty quantification of ML models³⁴.

As a final note, neural network ensemble has been traditionally treated as a non-Bayesian method. A more recent study attempted to connect neural network ensemble and Bayesian methods, such as BNN, by providing evidence that combining predictions from independently trained copies of the same neural network architecture approximates the Bayesian model averaging¹⁴⁵. This approximation may be why neural network ensembles produce improved (less overconfident) estimates of the predictive uncertainty on OOD data¹⁴⁶.

Neural network ensemble applications to battery diagnostics and prognostics SOH estimation

Very few studies attempted to apply neural network ensemble to battery SOH estimation. The most relevant study may be the one done by Shen et al. to estimate cell capacity from a measured sample of the (instantaneous) charge capacity (Q), voltage (V), and current (I) time series during a partial charge cycle¹⁴⁷. These time series measurements were fed as input into multiple deep convolutional neural networks, each of which outputs a capacity estimate. The purpose of ensemble learning was not to quantify predictive uncertainty. Rather, the idea was to combine *deterministic* capacity estimates from multiple neural networks to derive a final deterministic estimate. The weights assigned to the individual neural networks were optimized and often unequal (i.e., unlike the averaging formulation in Eq. (12)). The goal was to ensure prediction accuracy across a wider range of operating conditions than predicting with a single neural network. Ensemble learning, in combination with transfer learning, was validated using a 10-year aging dataset from implantable-grade cells^{13,110,111} and the Randomized Battery Usage dataset from the NASA Ames Prognostics Center of Excellence⁸⁶. The work by Shen et al.¹⁴⁷ is a typical example of a general observation: in most SOH estimation studies that use data-driven ML models, prediction accuracy is often the predominant evaluation criterion and, in many cases, the only criterion. We call for coordinated efforts to promote adding the quality of predictive uncertainty as a standard evaluation criterion to the scope of any future study on SOH estimation and health diagnostics and prognostics in general. This quality can be assessed via means well-established in the ML community, such as calibration curves, sparsification curves, and negative log-likelihood, as summarized in³⁴, and those established in the PHM community, such as the α -accuracy zone¹⁴⁸ and β probability¹⁴.

SOH forecasting and RUL prediction

We generally observe that studies on SOH forecasting and RUL prediction recognize the importance of uncertainty quantification much better than studies on SOH estimation. This observation could be attributed to the consensus that predicting a future state is more challenging than estimating the current state and involves an additional uncertainty source of future operating conditions that are often unknown. Outside the battery

prognostics field, ensembles of probabilistic neural networks have been applied to solve time series prediction problems for general-purpose prognostics. For example, the bearing prognostics work by Nemani et al.¹⁴⁹ built an ensemble of time series predictors, each being a long short-term memory recurrent neural network with a custom Gaussian output layer. A similar group of authors¹⁵⁰ later applied such an ensemble model for the RUL prediction of Li-ion cells in an open-source aging dataset consisting of 169 LFP cells^{27,79}. Like the capacity estimation study by Shen et al.¹⁴⁷, the neural network ensemble by Nemani et al.¹⁵⁰ did not adopt simple averaging as the weighting scheme. Instead, the model weights were optimized to minimize the RUL prediction RMSE on a training dataset. It was found that ensemble learning produced uncertainty estimates more representative of prediction errors over single-model learning. This improvement mostly resulted from increased predictive uncertainty and reduced overconfidence in OOD data, attributable to ensemble diversity, i.e., aggregating Gaussian-distributed outputs of the individual models with different means.

It is interesting to see efforts that combine predictions from deterministic ML models to derive a predictive uncertainty estimate. One such example is the capacity forecasting study considering charge and discharge C-rates that vary randomly from one cycle to the next¹⁰³, as also discussed under “RVM applications to battery diagnostics and prognostics SOH estimation”. The authors trained 10 XGBoost models and used the sample standard deviation of the (deterministic) capacity estimates by these models to quantify the predictive uncertainty. The quality of predictive uncertainty was assessed qualitatively by including a sparsification plot that visualized how the prediction accuracy measured by RMSE decreased by incrementally adding test samples with increasing predictive uncertainty. However, without assessment using quantitative metrics such as the expected calibration error and negative log-likelihood, it is unclear how well the predictive uncertainty (known) can approximate the model prediction error (unknown). Nevertheless, an interesting area of exploration could be uncertainty quantification by an emerging family of *deterministic* methods that only require a single, often deterministic neural network instead of multiple probabilistic neural networks^{151–153}. Benchmarking efforts comparing the more mature probabilistic and emerging deterministic methods will help fill an important knowledge gap in the battery prognostics field.

Counterintuitively, a hybrid method formed by combining a model-based and data-driven method produces lower uncertainty in RUL prediction over the model-based method¹⁵⁴. This uncertainty reduction was achieved by predicting future measurements using an ML model trained on historical data and augmenting the dataset for the model-based method with these predicted future measurements. Given the addition of new data (i.e., the predicted future measurements), uncertainty reduction is not surprising. The lower predictive uncertainty came hand-in-hand with a higher prediction accuracy¹⁵⁴, indicating that the predictive uncertainty was likely a good indicator of the prediction error. This example suggests combining methods or models may increase prediction accuracy (e.g., through data augmentation), together with reduced predictive uncertainty, while combining individual models in a neural network ensemble is not expected to increase prediction accuracy much but yields the benefit of capturing epistemic uncertainty, thereby producing a higher total predictive uncertainty that better represents the prediction error¹⁵⁵.

Sampling methods Sampling methodologies

Unlike Bayesian learning methods, sampling approaches to predictive uncertainty estimation work by evaluating models' fit to different data subsets via repeated sampling and model training. *Bootstrap sampling* is a common method of creating many data subsets by repeatedly sampling a dataset. How the sampling is performed (w/w/o replacement, stratified, etc.) greatly affects the characteristics of the data subsets and, subsequently, the models' fit to them¹⁵⁶. Predictive uncertainty is estimated by fitting models of identical architecture to each of the bootstrapped data subsets and aggregating their predictions—plurality vote for classification and averaging for regression. This method, known as *bagging* and short-hand for “bootstrap

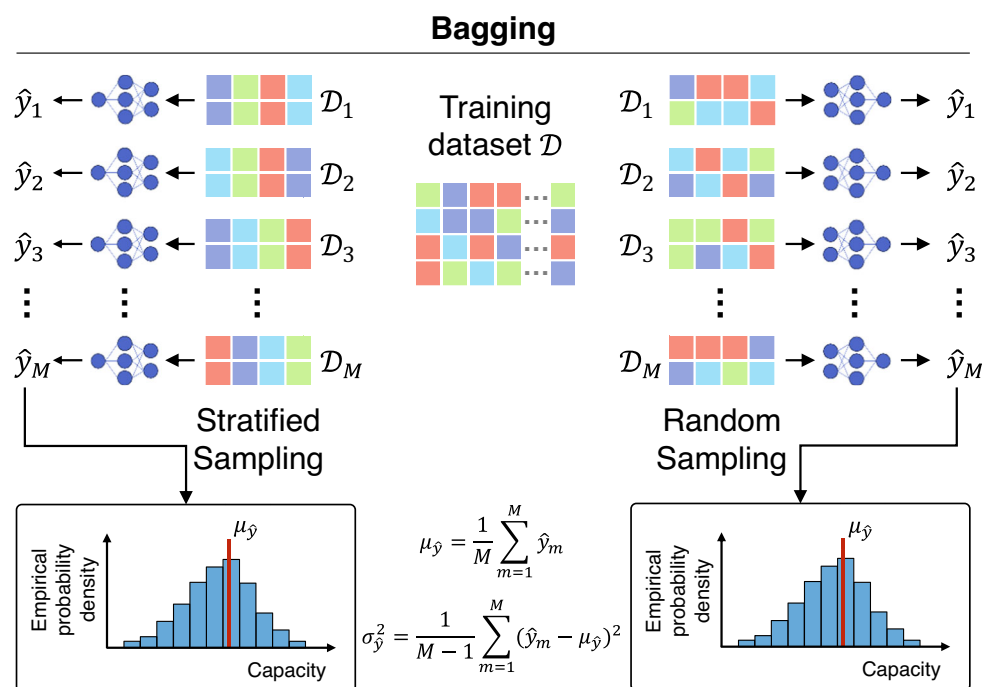


Fig. 12 | Sampling-based ML approach to battery capacity estimation. Example ML pipeline for battery capacity estimation using random and stratified sampling bagging methods.

aggregating”, is designed to estimate predictive uncertainty by capturing the changes in model predictions due to dataset perturbations induced by the random sampling¹⁵⁷. Bootstrap sampling and bagging (a.k.a. bootstrap aggregating) are explained graphically in Fig. 12. Standard bagging uses models of identical architecture and random sampling with replacement. The subset-to-subset variations in the sample selection produce models with varying optimal parameters, leading to diverse sets of predictions. When the properties of the sampled data subsets closely align with the full dataset, the variation in the fitted models is minimal, and the model prediction intervals tend to be small.

Sampling applications to battery diagnostics and prognostics

A common theme among research published on bagging for battery health diagnostics and prognostics is the use of random forest regressors. A random forest regression model is a meta regressor consisting of many binary decision trees that are fit to bootstrap samples of the original dataset¹⁵⁸. The outputs from the many decision trees are averaged to make a mean prediction, and uncertainty can be quantified by examining the spread of the predicted outputs from the individual trees. Random forest models are commonly used in battery health diagnostic and prognostic applications because of their ability to model the nonlinear behavior often observed in battery capacity fade and their probabilistic predictions. Further, random forestmodels have become simple to implement as many publicly available ML model libraries include a random forestimplementation.

SOH estimation

Research on battery capacity estimation using random forestsgenerally focuses on extracting various features from battery capacity-voltage data that correlate strongly with the cell’s capacity. Li et al.¹⁵⁹ conducted aging experiments on two different types of NCM/Gr cells to demonstrate an RF-based capacity estimation algorithm. The researchers extracted IC curves from three different voltage ranges, each spanning roughly 30% of the total SOC range. The researchers found that the voltage range containing a Li+ phase transformation (3.6–3.8 V in NCM/Gr) performed much better than the other two voltage ranges. This was determined to be the case because the phase transformations appear as prominent peaks in the IC curves and the

magnitude and location of the peaks are sensitive to the SOH of the cell. The model leveraged the strong correlation between the IC peaks and cell capacity to achieve an average of 1.3% RMSE across the 23 cells. Similar research by Roman et al.¹⁴ used a random forestmodel for capacity estimation as part of an ML pipeline that first extracted 30 features from battery capacity, voltage, and temperature data before down-selecting using a recursive feature elimination scheme. The selected features were then fed to a random forestmodel to estimate the capacity of various cells tested under both standard and fast-charge conditions. The researchers found that while the random forestmodel had the lowest accuracy of those tested, it was overconfident in its predictions of capacity, as indicated by the predictive uncertainty at some samples being overly small relative to the large prediction errors at these samples. Model predictive uncertainty is closely tied to the sample size of the dataset the model is fit to. In the realm of batteries, most datasets are very small in size due to the high costs associated with testing hundreds or thousands of cells. Specific to batteries, assessing the quality of predictive uncertainty quantification (e.g., through uncertainty calibration) is of great importance and should be investigated further³⁴. Other research on battery capacity estimation for second-life applications by Takahashi et al. investigated using a GPR-based bagging approach¹⁶⁰. The researchers first extracted summary statistics, like mean, variance, and inter-quartile-range, from the CC-constant-voltage part of the charging curve and down-selected them to be used as feature inputs to the GPR models. Then, multiple GPR models were fit to bootstrap data subsets to predict battery capacity.

SOH forecasting and RUL prediction

Researchers have also applied bagging to battery capacity forecasting and RUL prediction. Work by Liu et al.¹⁶¹ developed a bagging approach to RUL prediction using monotonic echo state networks to directly predict RUL from an engineered health index. The health index, calculated as the normalized time spent discharging between two fixed voltage limits, was found to work well for direct RUL prediction. However, it is worth noting that such a feature may be unextractable if the operation of the battery cell is such that it rarely discharges completely, as is the case with EVs and other consumer electronics. Research by Jiao et al.¹⁶² investigated a bagging approach to RUL

prediction for cells cycled under drive-cycle aging protocols. Their approach leveraged random forest decision trees that directly mapped to RUL using resistances extracted from EIS spectra along with standard IC and DV curves measured during discharge. They observed that their bagging approach produced a more accurate model than any single model tested, further highlighting the power of bagging.

Notably, sampling methods can also be used with more traditional modeling methods, like standard curve fitting and algebraic reduced-order models, to estimate the uncertainty in model parameters and enable probabilistic trajectory predictions for simple models that are not inherently probabilistic. Algebraic reduced-order models are among the first methods used to model and predict battery capacity fade^{3,7}. After aging cells in long-term storage and cycling tests, trajectory equations are fit to the normalized cell capacity measurements to model the capacity fade as a function of time and cycles/energy throughput. Trajectory equations of the form $Q = 1 - a \cdot t^b$, where Q is cell capacity, t is time, and a and b are fittable parameters, are commonly used because of their flexibility in fitting many different trajectory shapes and their loose ties to describing physical modes of capacity fade. For example, the exponent $b = 0.5$ is commonly used to model the diffusion-limited process in which capacity is lost with SEI formation on graphite^{3,52}. In many of the models reported in the literature, the capacity fade rate parameter, a , is typically a function that is time- and cycle-independent and captures the effect of the cell operating conditions, like temperature, DOD, SOC, C-rate, pressure, or any other measurable property, on the observed capacity fade rate. Common function forms used to model cell capacity fade rate are Arrhenius- and Tafel-like in nature^{163–168}. Arrhenius relationships are most often used to model the rate of chemical reactions mainly due to the effect of temperature. Similarly, Tafel relationships are used to model the rate of electrochemical reactions considering temperature and the electrochemical potential of materials.

Reduced-order models fit to battery aging datasets are excellent candidates to leverage sampling methods for uncertainty estimation. The various operating conditions and intrinsic cell-to-cell variability (e.g., due to manufacturing tolerances and uncertainty in material properties arising from within-batch and batch-to-batch property variations) produce capacity fade trajectories with varying lifetimes. Researchers Smith, Gasper, and colleagues from the National Renewable Energy Laboratory have published numerous articles on the topic of algebraic life models for forecasting battery SOH and predicting RUL with uncertainty via bagging^{52,169,170}. In the work by Gasper et al.¹⁶⁹, the researchers compared an ML-identified reduced-order aging model to one identified by a human expert. The human expert-identified model included Arrhenius- and Tafel-like expressions to capture the influence of the operating conditions on cell capacity fade. The ML-identified model was discovered through a symbolic regression method that iteratively tests linear combinations of physical descriptors, like $1/T$, $1/T^2$, $1/T^3$, etc., against the dataset to determine the algebraic form of the capacity fade rate submodel as a function of the aging stressors, $a = f(T, \text{SOC}, \dots, \text{etc.})$. Both the human expert- and ML-identified models were compared in terms of absolute accuracy and predictive uncertainty, where predictive uncertainty was estimated using bagging. By repeatedly sampling the aging dataset and fitting reduced-order models, the authors identified distributions for each model parameter and used the numerous sets of parameters to simulate many capacity trajectories. Predictive uncertainty was then quantified through the spread of the many simulated capacity trajectories, finding that the ML-identified model had roughly half the mean absolute error of the human expert-identified model and showed three times lower predictive uncertainty, with much more accurate predictions of cell capacity. In another paper by Gasper et al.⁵², they explained how uncertainty quantification via bootstrapping is an important tool during the model-form selection phase. The authors went on to demonstrate how model parameter uncertainty can be very large when certain test conditions are left out of the dataset, or when too many fittable parameters are included in the chosen model, making it difficult to identify good values for all model parameters. Altogether, many researchers have demonstrated the value of dataset sampling methods like bagging as excellent tools for enabling uncertainty

estimates on traditionally non-probabilistic models and as a means to assess model fitness through estimating model parameter distributions. In the next section, we provide insight into the effects of different bagging approaches for predictive uncertainty estimation for battery health diagnostics and prognostics.

Thoughts on sampling methods for battery data

While bagging is typically performed using standard bootstrap sampling with replacement, varying the sampling strategy can effectively capture different sources of uncertainty in a dataset. Specific to battery health diagnostics and prognostics, stratified sampling can be used to avoid quantifying model predictive uncertainty due to the different operating conditions (or lack thereof) in a battery aging dataset. Standard stratified sampling is performed by first dividing a dataset into various “strata” based on different attributes they share¹⁵⁶. For example, common battery strata are the aging experiment test conditions like temperature, C-rate, and DoD. Then, data subsets are randomly sampled ensuring an equal number of samples come from each stratum. Stratified and random sampling are compared in Fig. 12. The stratified sampling method works well for creating balanced datasets, but is infeasible when a few strata have far fewer samples than the others, as it limits the overall size of the data subsets that can be formed. An alternative approach is to perform unbalanced stratified sampling that ensures at least a single sample from each strata is included. However, depending on the application, this can create models that are biased towards stratum with more samples, and should be done carefully.

Models fit to stratified data subsets will never be expected to extrapolate to unseen operating conditions or other test attributes because at least one battery from each stratum will be included in the data subsets. This effectively eliminates any uncertainties associated with model extrapolation, as all operating conditions are known to the model. Stratified sampling is best used for quantifying the impact of cell-to-cell manufacturing variability, as it is well known that even identically manufactured cells will perform differently when tested under the same conditions^{26,27,35,87}.

An opposite of stratified sampling is leave-out sampling. This approach to sampling specifically excludes one or many strata from each data subset for the purpose of assessing predictive uncertainty as a function of the strata¹⁷¹. Leave-out sampling is similar to standard cross-validation (CV) in that specific data are left out each iteration, but generally, CV leaves out a much greater number of data each iteration and has fewer total iterations. For example, a typical fivefold CV requires that 20% of the entire dataset be left out each iteration and only five total iterations are performed. In contrast, leave-out sampling typically includes >80% of the entire dataset each iteration and sampling is performed with replacement. Leave-out sampling is useful for determining battery test conditions that are essential to accurate model parameterization. By repeatedly leaving out different strata and assessing model accuracy on a balanced validation dataset, one can map out each stratum’s importance to the model fitting process and identify a subset of strata that are essential for accurate model parameterization. Take, for example, a battery capacity estimation model that gets fit to a dataset containing data ranging from 25 to 50 °C. If asked to predict cell capacity for a sample tested at –20 °C, the model would significantly overpredict the capacity because it had no training data from low-temperature strata to inform it that battery capacity significantly declines at low temperatures due to decreased reaction kinetics and increased charge transfer resistance. In this case, leave-out sampling would be able to identify that low-temperature data is crucial to accurately parameterizing the model, and should be prioritized in future testing.

Understanding and using predictive uncertainty estimates from ML models

Understanding the different types of uncertainty that exist and what they quantify is crucial to properly using probabilistic diagnostic and prognostic models for decision making. Discussed previously in the section

“Introduction”, applications like designing serially connected modules or setting warranties are typically concerned with worst-cell behavior. Under these conditions, it is ideal to use a probabilistic ML model that can estimate the full population-wise uncertainty when making predictions, so that the tail-end of the distribution can be accurately quantified.

However, there exist many different probabilistic ML models (see the section “Probabilistic ML techniques and their applications to battery health diagnostics and prognostics”) that each differ in the type of uncertainty they estimate (aleatory, epistemic) and how the uncertainty is quantified (Gaussian, non-parametric, bagging, simple average, etc.). Further, the documentation accompanying the various ML modeling libraries for MATLAB, python, etc., are inconsistent in their documentation making it unclear exactly what uncertainty is being predicted. To this end, this section aims to explain the various types of predictive uncertainty and generally discuss how they can be quantified.

Most publicly available ML models output the total predictive uncertainty—usually a probability mass function for classification or a variance for regression. The total predictive uncertainty can be qualitatively decomposed into *aleatory* and *epistemic* uncertainty, each owing to unique uncertainty sources¹⁷².

- Aleatory uncertainty quantifies the inherent stochastic nature of an input, output, or the dependency between the two. It is irreducible by nature and stems from sources like manufacturing process variability, inconsistency of material properties, and variations in experimental test conditions¹⁷³. Aleatory uncertainty, sometimes referred to as *data uncertainty*, persists in a dataset even if more samples are collected, making it irreducible. Types of aleatory uncertainty most frequently studied in Li-ion batteries are generally electric in nature and include variations in cell capacity, resistance, impedance, and aging. Uncertainty associated with battery performance is considered aleatory uncertainty because testing more cells will not reduce the measured variability since it stems from differences in materials and manufacturing processes. However, testing more samples enables one to more accurately quantify the magnitude of aleatory uncertainty in electrical performance³⁵.
- Epistemic uncertainty arises from an incomplete understanding or model representation of the data and is thus reducible. Common sources of epistemic uncertainty are model simplification, model-form selection, computational assumptions like numerical discretization, and model parameter uncertainties¹⁷⁴. Epistemic uncertainty can generally be further classified into model-form and parameter uncertainty³⁴. Model-form uncertainty arises due to the various simplifications and assumptions made to simplify the model training and inference process. This type of uncertainty is prevalent in battery health diagnostics and prognostics since directly modeling the physics of degradation is largely infeasible, and thus most models assume some simpler form that approximates the observed physics by empirically modeling data, e.g., reduced-order models^{51,169}. Model-form uncertainty can generally be reduced by increasing model complexity or by directly modeling the underlying physics. Model parameter uncertainty can be reduced by collecting more training data with better accuracy and under more conditions, or by increasing the fidelity of the data measurements.

While it is beneficial to understand the origins of aleatory and epistemic uncertainties, it is difficult to individually quantify them in practice. Instead, the predictive uncertainty output of probabilistic ML models captures the combined effects of aleatory and epistemic uncertainties, quantified as the total predictive uncertainty.

Bayesian models, like RVM (the section “Relevance vector machine”) and GPR (the section “Gaussian process regression”), quantify predictive uncertainty through a posterior mean and covariance matrix. Together, the mean and the diagonal of the covariance matrix can be used to construct a predictive Gaussian distribution for each input sample¹⁷⁵. The predictive uncertainty of GPR typically

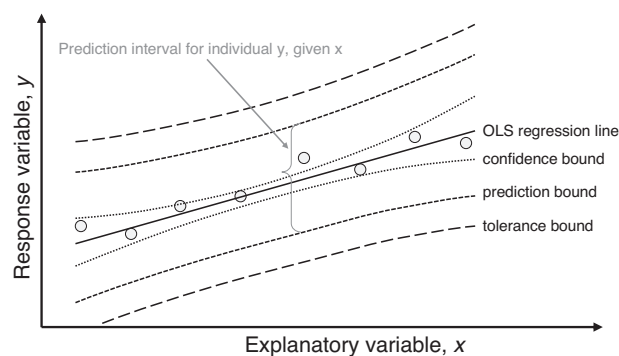


Fig. 13 | Regression illustrating various forms of model uncertainty. Example illustrating confidence, prediction, and tolerance intervals calculated for an ordinary least squares regressor with a single input feature (x) and target value (y).

captures the aleatory uncertainty of the data fairly well, but does not capture model-form and parameter uncertainty since the model is non-parametric in nature. If, for example, a GPR was used with an underlying trend function, it may do a better job at capturing the model-form uncertainty as the final fit considers the fit of the trend function to the data in the presence of noise. Another limitation of Bayesian models is that they inherently assume the spread in the predicted distribution is symmetric about the mean (Gaussian distribution), which may not always be valid depending on the application.

On the other hand, sampling approaches to uncertainty estimation quantify uncertainty through aggregating predictions from the many bootstrapped models and constructing a non-parametric predictive distribution for each input sample. Bagging approaches to uncertainty estimation produce a non-parametric distribution that may or may not be Gaussian. Final predictions are generally made using the mean or median of the predictive distributions. Sampling approaches to uncertainty estimation are flexible making different combinations of sampling methods and models to capture different uncertainties. For example, a neural network ensemble (the section “Neural network ensemble”) will capture aleatory uncertainty and parameter uncertainty well, but will not capture model-form uncertainty since the model structure is not empirically-based and the feature learning process is generally unregulated. On the other hand, an algebraic reduced-order capacity trajectory model that inherently assumes the battery’s capacity fade follows an algebraic trajectory will do a much better job capturing model-form and parameter uncertainty. However, as discussed in the section “SOH forecasting and RUL prediction”, the type of sampling method used (stratified, random, leave-out, etc.) plays a large role in the uncertainty that is captured.

Regardless of the probabilistic ML method used, understanding and quantifying a model’s predictive uncertainty is a crucial step in conveying prediction results. Quantifying model predictive uncertainty is generally done using statistical intervals derived from the predictive distributions. Selecting a statistical interval is application-dependent and depends on the main interest at hand, typically one of quantifying the location of a distribution, the spread of a distribution, or calculating an enclosure interval that captures a portion of the total population. Three main types of statistical intervals exist: confidence interval, prediction interval, and tolerance interval. Explanations of each of these statistical intervals, and their applications to battery health diagnostics and prognostics are outlined below.

- *Confidence intervals* are used for quantifying the precision of a distribution parameter – typically the mean of a distribution. Confidence intervals are frequently used to calculate an upper and lower bound on a distribution mean where the true value of the mean will fall within the calculated range with specified probability p (usually $p = 0.95$). Shown in Fig. 13, the confidence interval is the smallest of the three intervals

because it captures only the uncertainty in the predicted mean value. The range of the confidence interval is closely related to the model-form uncertainty, and because of this, confidence intervals are useful for assessing model parameter uncertainty. As the sample size approaches infinity, the confidence interval collapses to the true value of the mean.

- *Prediction intervals* are used for quantifying the range that a single future sample from the population will fall within. For example, after training a GPR model on a small dataset to predict a battery cell's capacity, one might like to construct an interval that, with a high degree of confidence, will contain the true capacity values for a future cell. Shown in Fig. 13, the prediction interval is wider than the confidence interval because it captures both the uncertainty in the model parameters (epistemic uncertainty) and the data uncertainty (aleatory uncertainty).
- *Tolerance intervals* are used for quantifying the range that present and all future samples from the population will fall within. In battery cell health diagnostics and prognostics, tolerance intervals are useful for establishing an range that will contain a specified portion of the cell population, providing insight for engineers and manufacturers on the predicted performance of *all* future cells given the results of limited testing from a small batch. Shown in Fig. 13, the tolerance interval is the widest of the three because it captures the additional uncertainty due to having an incomplete dataset.

Calculating a statistical interval differs based on the type of predictive distribution, either parametric (like a Gaussian distribution from RVM or GPR) or non-parametric (like the distribution from a bagging approach or an ensemble model). For parametric distributions like the Gaussian, parametric statistical intervals are straightforward to calculate. The primary interval of interest for Gaussian models is generally a prediction interval. Given we have trained a model on a dataset, prediction intervals tell us with a certain probability where newly tested battery samples will fall. A two-sided 95% prediction interval is calculated as follows:

$$[PI_l, PI_u] = \hat{\mu} \pm Z\hat{\sigma}, \quad (13)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the mean and standard deviation from the predictive distribution, Z is the standard Z -statistic and is equal to 1.96 in this instance, and $[PI_l, PI_u]$ are the lower and upper interval bounds, respectively.

For non-parametric distributions, like those generated from sampling-based methods for uncertainty estimation, the user generally has two options for calculating statistical intervals: (1) approximate the distribution using a known parametric distribution, like Gaussian or log-normal, and calculate the corresponding statistical interval accordingly, or (2) calculate a non-parametric statistical interval. Calculating non-parametric statistical intervals is typically preferred, as assuming a distribution has major implications when trying to understand the tail-ends of the population, e.g., for estimating the worst battery cell performance in a pack. Two-sided non-parametric statistical intervals are not necessarily symmetric, owing to the skewed predictive distribution. Non-parametric statistical intervals are calculated using order statistics, where the upper and lower interval bounds are determined by excluding a calculated number of samples from each tail-end of the predictive distribution and setting the bounds at the lower/upper edges of the remaining samples. The correct number of samples to remove from each end of the predictive distribution is influenced by the desired confidence level p , the number of samples in the predictive distribution N , and, in the case of tolerance intervals, the desired population coverage. The size of the interval is highly dependent on the number of samples in the predictive distribution. However, practically speaking, one can simply increase the number of bootstrap samples to a large number (>1000) to reduce the size of the intervals and achieve higher confidence levels with a narrower interval. Readers interested in calculating one- and two-sided distribution-free statistical intervals are referred to Chapter 5 of the book by Meeker et al.¹⁵⁶

Advanced topics in battery health diagnostics and prognostics

Diagnostics and prognostics using field data

The general practice in battery diagnostic and prognostic algorithm design is to use cell data collected in the laboratory under predefined load conditions at controlled temperatures (bottom-up approach) and regard battery capacity as the variable of choice to measure battery health, measured periodically via capacity tests¹⁷⁶. Yet, battery capacity is an elusive health metric to estimate when monitoring a battery system used in the field³¹. Battery health diagnostic and prognostic algorithms are deployed to operate on BMSs in real-time and expected to provide accurate health estimates over the entire lifespan of the battery system. One of the main limitations of this approach is that laboratory data can only serve as a small representation of real-world field battery operation and does not reflect application-specific behavior. For instance, in EV applications, battery data will have geographical climate and time-of-day usage dependency, as well as driver-specific behavior³¹. Most importantly, laboratory data do not (and cannot) provide the richness of history-dependent usage trajectories. When SOH algorithms are built from lab data, their predictive ability is challenged upon on-board BMS deployment. This is even more true if the SOH algorithm itself, or any of its components, is based on data-driven ML approaches. Indeed, ML models are limited by the quality of the data used to train them and in terms of how representative the data is of the application at hand. Given the high variability of battery usage in the field, ML-based SOH algorithms developed exclusively from lab data are likely to fail. Moreover, features extracted from lab-generated data and used in data-driven diagnostic and prognostic models will be substantially different, in quality and quantity, from features defined and extracted from real-world driving field data³². Lastly, cell-to-cell heterogeneities are responsible for exacerbated battery system degradation over the battery lifespan but are typically not assessed via lab experiments nor accounted for in current BMSs. In particular, real-time operating conditions contribute to the variability between cells in the form of thermal and aging gradients propagating across the battery pack system, which makes the task of assessing battery health and predicting remaining useful life in real-time more challenging. Such a task cannot solely rely on lab-based offline designed health algorithms.

In light of these challenges, researchers have begun investigating real-world battery usage data for health diagnostics and prognostics. In a recent publication³¹, authors analyzed one-year worth of battery pack data to define health and performance indicators directly learned and extracted from actual driving and charging signal segments. The proposed features were found to be quite different from features previously proposed in the literature to estimate health and predict the remaining lifespan^{22,27}. The newly extracted features were found to leverage quantities such as resistance calculated during braking or acceleration events and impedance during charging. However, these domain knowledge-based features are also strongly dependent on driving styles, meaning that they would need to be learned on a per-user basis using domain adaptation or similar transfer learning methods. Lu et al.¹⁷⁷ investigated a domain adaptation method to enable seamless transfer of an SOH estimation model from one battery chemistry to another. The researchers were able to train a model that worked well when transferred to a new battery chemistry by extracting domain-invariant features. An algorithm like this one, or similar algorithms proposed in the literature^{178–181} could conceivably be adapted to enable a model built using lab data to provide good diagnostic and prognostic accuracy in the field.

Other examples of SOH estimation from field data include work by Song et al.¹⁸², where they used a deep neural network to learn relevant features from the historical data of 700 EVs. While not openly available, the dataset was collected by the Shanghai Electric Vehicle Public Data Collecting, Monitoring and Research Center for the purpose of optimizing EV usage in the city and has been used by several other research teams to date^{183–185}. Similar work by She et al.¹⁸⁶ investigated incremental capacity features as input to a radial basis function deep neural network for SOH estimation of electrified city bus battery packs.

Examples of battery health prognostics using field data are more rare since collecting sufficient data for lifetime prediction generally requires many years, since after all, most EVs are warranted for 8 years and 100,000 mi¹⁸⁷. In our search, we found only a few papers investigating prognostics from field data. Deng et al.¹⁸⁸ built a sequence-to-sequence model with adaptive error correction from a GPR model to predict future capacity fade of 20 EVs operating for over two years using only the first 3 months of data. The model was very accurate with 1.6% average prediction error, but the small size of the dataset (only 20 packs) and similar operating conditions make it difficult to determine the model's accuracy in various conditions and longer duration. Other work by Zhang et al.¹⁸⁹ used much more data for building a battery prognostic model: two datasets comprising lab aging tests (a few hundred cells) and one dataset comprising data from 7296 PHEVs. The researchers tracked battery aging stressors like SOC, temperature, and throughput using a histogram-based approach and quantified the stressors using summary statistics like mean, variance, skewness, kurtosis, and higher order moments. The extracted feature pool was then reduced using cross-correlation analysis between the features and the target variable of cell capacity. After training a global model on the datasets, the authors implemented an online adjust factor that tunes the global model to an individual battery pack. The adjustment factor is calculated on a rolling basis using the difference between the global model prediction and the observed capacity data, improving accuracy considerably compared to just using the global model for prediction on all vehicles.

Going forward, it will be crucial for academia to collaborate with industry to share data from field units for the purpose of improving diagnostic and prognostic algorithms. Furthermore, developing intelligent data-driven performance forecasting/prediction models for real-time deployment requires re-examining the BMS design paradigm to account for the integration of field operating conditions to allow domain-knowledge learning³³.

Degradation diagnostics

Mentioned earlier in the section “Battery health diagnostic and prognostic problems”, degradation diagnostics is a subproblem of SOH estimation that focuses on methods to non-destructively diagnose internal degradation modes that drive capacity fade and resistance increase⁷¹. Figure 3 gives an overview of the three degradation modes commonly used to quantify cell-level capacity and power fade: loss of active material on the cathode (LAM_{PE}), the loss of active material on the anode (LAM_{NE}), and the loss of lithium inventory (LLI). These three degradation modes are commonly used to quantitatively describe the combined effects of certain groups of degradation mechanisms present in the cell, i.e., the degradation modes come from grouping degradation mechanisms based on their resulting effects on cell-level performance. Research by Birkel et al.⁷¹ experimentally verified the effects of each degradation mode on the full-cell OCV curve, providing a quantitative link between the two for the first time. This was achieved by constructing coin cells with electrodes of different sizes to simulate loss of active materials and lithium inventory. The degradation modes were then quantified by using half-cell OCV data to reconstruct the full-cell OCV curve, where the relative position and size of the positive/negative electrode half-cell curves quantify the degradation modes. This work showed that the degradation modes could be accurately quantified by examining full-cell OCV data, albeit through a lengthy and cumbersome curve fitting process that requires access to high-precision measurements of full- and half-cell voltage, current, and capacity during slow and complete cycles.

In light of this, researchers began investigating methods of automating the diagnosis process using various ML-based techniques. In the work by Tian et al.¹⁹⁰, researchers trained a convolution neural network to estimate a cell's offline OCV curve using data collected from partial charging cycles throughout the day. The trained model could then be used online in place of low-rate full-DOD cycling tests to estimate full-cell OCV, thus enabling and significantly speeding up the process of online degradation diagnostics. Along the same lines, papers^{191,192} demonstrated methods of estimating full-

cell OCV curves by fitting pristine half-cell OCV data to partial charging curves. While accurate, the methods still required a significant duration of charging data (Schmitt et al.¹⁹² required 20–70% SOC) in order to guarantee accuracy in the range of 2% error on capacity estimation.

While the researchers in refs.^{190–192} focused on using cell data to reconstruct OCV curves as an intermediate step in diagnosing degradation modes, the following works aimed to directly correlate cell data with the degradation modes^{193,194}. Han et al.¹⁹³ proposed using membership functions to quantify the areas under the peak locations of the full-cell differential capacity curve (dQ/dV) and correlating these capacities to loss of lithium inventory and loss of negative electrode materials. Costa et al.¹⁹⁴ focused on transforming the full-cell incremental capacity and differential voltage measurements into a 2D image of the cell's state that were then fed into a convolution neural network that directly diagnosed the degradation modes. The model was proven to work well on multiple cell chemistries (LFP, NMC, and NCA), achieving an average of 2% error. These works have undoubtedly shown that cell-level data can be directly correlated with internal degradation modes. Yet, few have studied the transferability of the model to new operating conditions outside the datasets used.

The following few works focused on just that: methods of building generalizable ML models for degradation diagnostics^{81,195,196}. To make the models more generalizable, the following works prioritized incorporating synthetic data from physics-based simulations of cell degradation into their models. Dubarry et al.¹⁹⁵ developed a method relying on an offline database that contained cell full-cell OCV curves and their corresponding degradation modes, simulated using half-cell data from full-cells. The degradation modes of an online cell were quantified by measuring its incremental capacity curve and matching it to the database, interpolating if the curve does not match any of the database entries exactly. While accurate, the database is generally too large to be implemented onboard BMS or other devices. Thelen et al.⁸¹ took a different approach, and instead trained a machine learning model to act as a generalizable aging “database” by training it using a combination of experimental and simulated aging data. Once trained, the model could be used online to directly estimate a cell's degradation modes using the full-cell incremental capacity curve as input. Other work by Ruan et al.¹⁹⁶ took a similar approach, and trained a deep learning model from a large body of simulated aging data, demonstrating that the degradation modes are inherently correlated, and these correlations can be exploited to improve diagnostic accuracy. Altogether, these methods present a significant leap forward in the ability to non-destructively diagnose unobservable degradation modes in Li-ion batteries.

While all the methods discussed so far have focused on degradation diagnostics using electrical measurements (voltage, current, capacity, etc.), there do exist other methods using alternative data streams. Prosser et al.¹⁹⁷ demonstrated a zero-dimensional cell heat generation model that could accurately diagnose the cell's degradation modes in-operando. What's impressive about this work is that the pouch cells were subject to active cooling through the cell tabs, demonstrating the method can be adapted to cells inside modules and packs.

Altogether, the field of degradation diagnostics has come a long way, and we envision these methods and techniques will carry over into battery prognostics, enabling forecasting and life prediction with respect to cell degradation modes in addition to the typical capacity/resistance.

Early life and trajectory prediction

Recognizing the practical value of probabilistic forecasting of SOH evolution and probabilistic predictions of RUL, researchers have recently begun to develop early prediction models with quantified uncertainty. The majority of these papers employ current and voltage information collected early in the battery life to predict RUL or other quantities of interest, as illustrated in Fig. 14. In one of the earliest examples, Fermin-Cueto et al. demonstrated the classification of battery life into categories and the prediction of the number of cycles until the battery capacity exhibited accelerating degradation, termed roll-over or knee-point, both with uncertainty¹⁹⁸. For the classification task, they employed support vector machines to predict

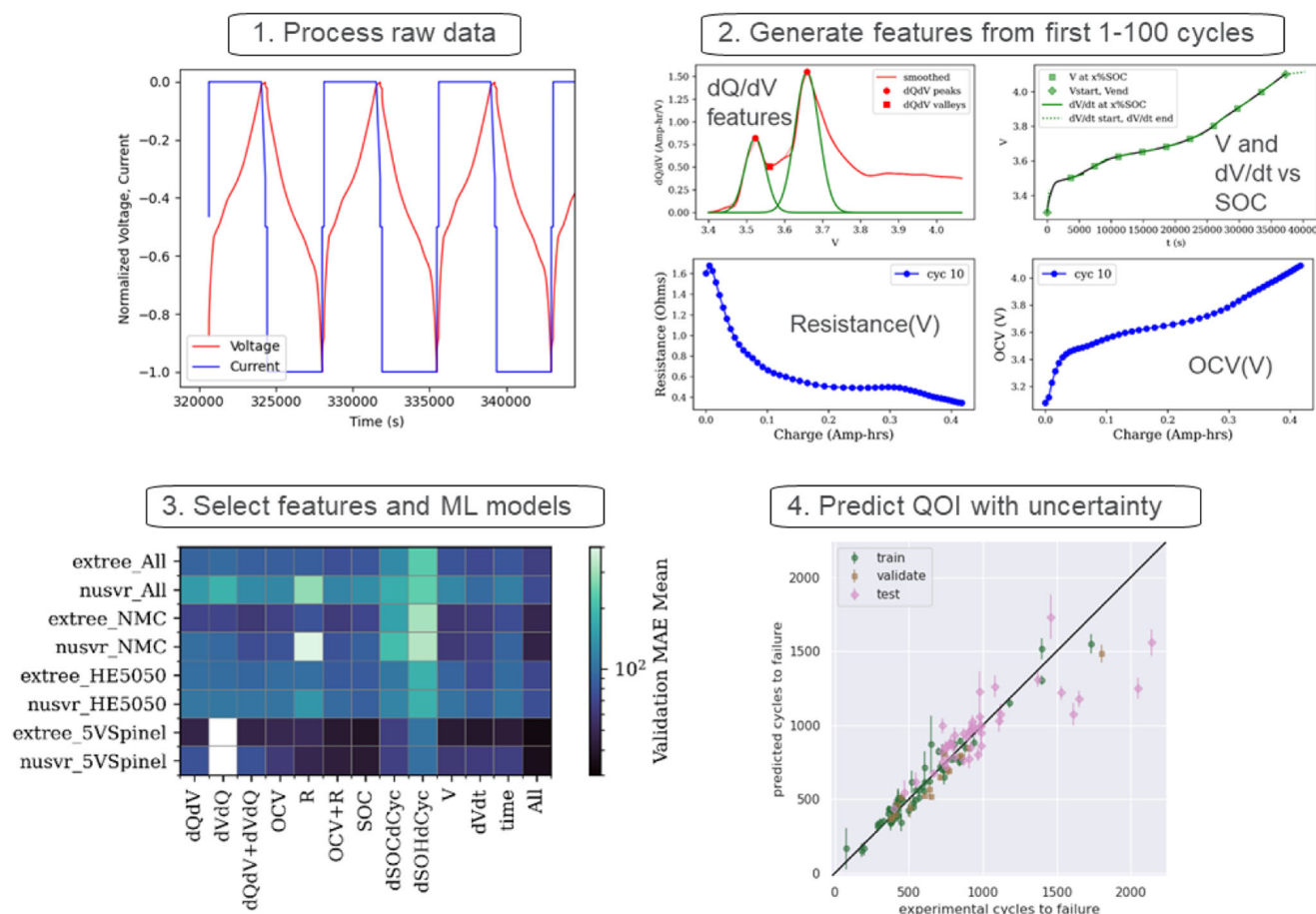


Fig. 14 | A schematic illustrating battery life prediction from preliminary current and voltage information. “QOI” in Step 4 stands for quantity of interest. Figure elements borrowed from Paulson et al.⁸⁷

whether the cells corresponded to low, middle, or long life categories using data from only the first 3–5 cycles. Class probabilities were estimated using the approach of Platt, where the support vector machine outputs were recalibrated via logistic regression¹⁹⁹. Knee-points were predicted using the first 50 cycles via RVMs, with conformal prediction intervals for uncertainty. Both tasks employed the 124 LFP cell fast-charging dataset released by Severson et al. with their pioneering 2021 paper²⁷. Several years later, the same group demonstrated predictions of capacity and internal resistance degradation curves with uncertainty via XGBoost (an ensemble decision tree approach) predictions of the knee-onset and knee-point and the capacities (or resistances) at which they occur, in addition to the EOL—again using only the first 50 cycles²⁰⁰.

As mentioned in the section “SOH forecasting and RUL prediction”, NREL has published several articles employing ML-selected arithmetic relationships to predict changes in battery SOH under a variety of calendar- and cyclic-aging conditions with bootstrapped uncertainty estimates^{52,169,170}. Notably, this approach provides reasonable extrapolations into later life for a variety of chemistries, cell formats, C-rates, and temperatures due to the semi-physical nature of the models selected by the symbolic regression approach employed and enabling deconvolution of aging mechanisms. This approach does not require preliminary cycling data for a given cell to predict SOH evolution, but does require extensive accelerated testing information spanning conditions for that specific cell type. Techno-economic analyses demonstrated the outside impact of prediction uncertainty on energy-storage system lifetime.

Rieger et al. demonstrated prognostics of capacity degradation trends for LFP, NMC, and NCA battery chemistries using between 20 and 100 preliminary cycles as context for the prediction²⁰¹. In contrast to previously mentioned approaches, this work employed a deep learning architecture

capable of making non-parametric predictions of future degradation trends. Specifically, ensembles of recurrent neural networks were used. Final predictions with uncertainty were obtained by combining ten trajectories sampled from the mean and variance outputs from each of five neural networks trained with randomly initialized weights. Analysis of the uncertainty revealed that the model was slightly overconfident in its predictions, but correctly assigned high uncertainties to longer-lived cells that were less represented in the training set.

Future trends and opportunities

Physics-based diagnostics and prognostics

Physics-based models of Li-ion batteries are powerful tools for simulating cell electric, thermal, and mechanical performance. However, they are typically parameterized only on newly manufactured battery cells, relegating them to applications that do not consider cell aging. While in theory, physics-based models can be re-parameterized using data collected from aged cells, doing so would require a large and costly aging test campaign. In light of this, researchers have found other ways to leverage the accuracy and extrapolation capabilities of physics-based models in prognostic frameworks.

One method of leveraging physics-based models for battery health diagnostics and prognostics is to use them as an intermediate step toward SOH estimation or RUL prediction. An example of this approach is illustrated in Fig. 15, where the physics of Li-ion battery degradation is used as an intermediate step to estimate cell capacity and predict cell RUL. Ideally, it is thought that using a physics-based model as an intermediate step helps to include additional information regarding the physics of battery operation and degradation that may not be immediately learnable by a traditional ML model from the available data. Lui et al.²⁰² used this approach to predict the

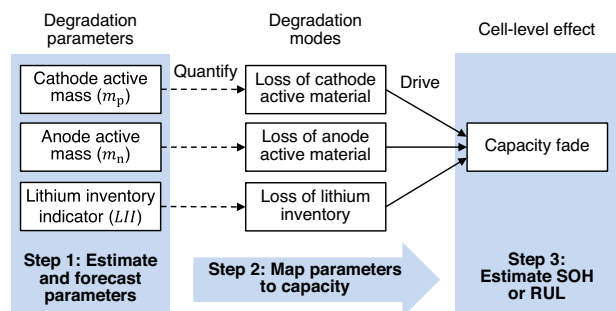


Fig. 15 | Example of a physics-based approach to battery capacity estimation. Cell degradation parameters (see the section “Battery degradation -- modes and mechanisms”) are estimated and used in a physics-based model to estimate full-cell capacity fade. Based on work of Lui et al.²⁰²

RUL of implantable-grade Li-ion battery cells aged under various temperatures and C-rates. Instead of directly extrapolating the observed capacity-fade trajectories to predict RUL, the researcher first fit a physics-based half-cell model to the measured full-cell OCV curves to obtain rough estimates of the cell’s present active masses (LAM_{PE} , LAM_{NE}) and lithium inventory (LLI) (see the section “Battery degradation—modes and mechanisms”). This half-cell model, proposed by Honkura et al.²⁰³ and popularized in later works^{204,205}, uses experimental data from anode and cathode half-cells to simulate the full-cell voltage vs. capacity curve, typically under a very slow charge/discharge rate (e.g., $I < C/20$) to approximate the full-cell OCV curve. Then, Lui et al.²⁰² used bounded empirical capacity fade models to extrapolate the degradation parameter values. Estimating cell capacity was then achieved by running the physics-based half-cell model in reverse—inputting the degradation parameters and receiving an estimate of the full-cell OCV curve and the capacity. This method worked well since it was found that many of the degradation parameter trajectories were nearly linear, making their future trajectories easily extrapolated. For many cells in the dataset, the degradation parameter trajectories were nearly linear but combined to produce nonlinear degradation in the full-cell capacity fade curve. Discussed in detail by Attia et al.⁵⁶, degradation modes have various rates of progression, that when combined together in the full-cell environment, interact to produce nonlinear capacity fade and often times knee points in the capacity trajectory. Attia et al.⁵⁶ delve into details regarding the so called “internal state trajectories” of various degradation modes like electrolyte additive depletion, lithium plating, and resistance growth due to active material loss, that drive measurable capacity loss at the cell-level.

Kohtz et al.²⁰⁶ took a similar approach to battery prognostics, which was categorized as a physics-informed ML approach. The researchers used a physics-based model to estimate the thickness of SEI on a Li-ion cell’s anode as an intermediate step to capacity estimation. Instead of using the physics-based model in the prediction process, the authors trained a GPR surrogate model (GPR model #1) to approximate the battery physics and learn the mapping between a partial segment of the cell’s voltage curve and the corresponding SEI thickness for various temperatures, C-rates, and SOH as simulated by the physics-based model. Next, another GPR model (GPR model #2) was trained to learn the mapping between a cell’s SEI thickness and its discharge capacity. The two GPR models are used in series to make a final prediction of cell capacity, first predicting SEI thickness from a partial voltage segment using GPR model #1 and then predicting cell discharge capacity using GPR model #2. Compared to directly predicting cell capacity from a partial voltage segment, the physics-informed approach to capacity estimation showed significantly lower error, mostly attributed to the extra knowledge of SEI thickness infused into the GPR models.

Other methods of using physics-based models for battery prognostics include using the physics-based models to generate simulation data to use for traditional ML model training^{81,193,195} and online updating the parameters of the physics-based models using measurements of the cell²⁰⁷. In addition to purely physics-based approaches to prognostics, the field of

physics-informed ML has flourished recently with numerous articles leveraging physics knowledge to inform traditional ML approaches for battery health diagnostics^{206,208–210} and prognostics^{207,211}.

Recent work by Pannala et al.²¹² developed a physics-based aging model that linked SEI growth, Li plating, and electrode particle fracture degradation modes to irreversible cell thickness growth, resistance increase, and capacity loss. The method parameterized a single particle model of an NCM111 battery cell and a group of degradation-mode-specific aging models using data collected from RPTs during laboratory aging tests. The model was found to accurately capture and predict changes in cell resistance, capacity, and thickness for a variety of C-rates and DODs. Further, the physics-based nature of the model provides insight into a cell’s remaining lithium inventory (LLI) and remaining positive and negative active electrode materials (LAM_{PE} , LAM_{NE}) with aging.

However, we believe there is still a great opportunity to further leverage physics-based models for battery prognostics. In particular, leveraging physics-based models to track and identify battery aging stressors from field data (see the section “Diagnostics and prognostics using field data”) is a novel idea we have yet to see investigated. Further, it is becoming more feasible to deploy lightweight physics-based models on-board EVs as their computing systems are upgraded to accommodate the demand from driver-assistance systems. In online scenarios, fusing sensor data, physics, and ML poses a real solution to accurate battery prognostics in the field.

Second-life applications

In recent years, the market share of EVs has witnessed remarkable exponential growth. Starting from a modest 4% in 2020, EVs now account for 14% of all vehicles sold as of 2022⁴⁷. It is projected that EVs could constitute 18% of total vehicle sales in the near future⁴⁷. The rise in EV adoption holds promising potential for mitigating greenhouse gas emissions in the transportation sector. However, along with these environmental benefits, there arises a pressing concern regarding the exponentially increasing volume of retired Li-ion batteries, imposing a significant challenge to environmental protection and sustainability efforts. While recycling techniques allow for the recovery of valuable battery materials from retired batteries, within the framework of a circular economy, it is widely believed that it is not economically optimal to directly recycle all the retired EV batteries. This is mainly due to the following two primary considerations:

1. In the automotive industry, a commonly adopted practice for battery retirement is to replace them when their capacity falls below 80% of their nominal value²¹³. This leaves most retired batteries with a significant portion of their initial capacity that can be utilized by other industries that require less capacity/power performance for energy storage.
2. Empirical data show that battery pack degradation often stems from the failure or reduced capacity of a small number of cells within the pack²¹⁴. This means that a majority of cells within a pack tend to have capacities greater than the pack’s EOL condition, making them excellent candidates for reuse in new applications.

The above stipulations have led to immense efforts to integrate *remanufacturing* and *repurposing* into the broader circular economy of EV batteries²¹⁵. The ultimate goal is to prolong the service life of retired batteries, affording them a valuable “second life” before they ultimately undergo recycling for raw material collection and disposal.

- *Remanufacturing* is the process of identifying failed or significantly aged cells within a battery pack and replacing them with new cells or used cells that have been tested and found to meet specifications by OEMs^{215,216}.
- *Repurposing* is the practice of giving retired EV batteries a second life in a diverse range of applications. Typical applications include grid energy storage systems, off-grid stationary storage, and recreational vehicles, all of which can function with cells of lesser capacity and power capability.

Roles of probabilistic ML for second-life batteries

Retired EV batteries maintain significant value and functionality when retired at the typical 70–80% initial capacity. However, it remains practically challenging to determine the suitability of a cell/module/ pack for remanufacturing or repurposing. This section discusses the future role probabilistic ML methods will play in the *evaluation* step of a Li-ion battery's life cycle—with a primary focus on SoH estimation (the section “Background”) and lifetime prediction (the section “Probabilistic ML techniques and their applications to battery health diagnostics and prognostics”). Examples are below.

1. **Degradation modeling:** As detailed in the section “Battery degradation—modes and mechanisms”, Li-ion batteries exhibit numerous aging mechanisms, and the relationship among them is inherently complicated. Given that the safety performance of Li-ion batteries is significantly affected by their aging path and underlying aging mechanisms²¹⁷, comprehending the degradation of EV batteries during their first-life use is crucial in assessing the safety suitability of retired batteries for their intended second-life applications. For instance, capacity degradation may manifest as a loss of Lithium inventory, a phenomenon stemming from various degradation mechanisms such as lithium plating/dendrite, electrolyte decomposition, and SEI decomposition, among others. In cases where dendrite growth is the root cause, growing dendrite could potentially penetrate the separator, causing an internal short circuit, triggering thermal runaway, and, in more severe scenarios, resulting in a fire in the second-life application²¹⁸. Identifying the primary degradation mechanism in the first life can significantly facilitate the safety assessment for subsequent second-life applications. Furthermore, gaining insights into degradation mechanisms from the first-life application could also help predict the potential degradation mechanisms in the second-life application, thus allowing for proactively predicting the service life of the batteries in their second-life applications. The very complicated degradation mechanisms, along with their complex interactions, present substantial hurdles when it comes to degradation modeling and identification. Despite advancements in physics-based degradation modeling, numerous physical phenomena remain unsolved. Leveraging hybrid modeling techniques, which seamlessly integrate probabilistic ML models with physics-based models, offers a promising avenue for addressing these knowledge gaps. Such an approach enables the modeling of un-modeled physics using experimental data, which simultaneously allows for quantifying the predictive uncertainty in the degradation modeling.
2. **SOH estimation:** SOH is a widely accepted indicator for battery screening, guiding the selection of suitable second-life applications²¹⁹. When a battery cell/module's SOH falls significantly below a certain threshold (see Fig. 16), this cell/module can be directly recycled. Conversely, when the SOH remains considerably high, surpassing the EOL criteria for EV applications, the battery can undergo remanufacturing and be repackaged for continued EV use. As depicted in Fig. 16, when the SOH falls within the intermediate range, the battery can be repurposed for second-life applications, such as grid storage and off-grid stationary storage, depending on the battery's specific SOH level. Therefore, SOH estimation is an indispensable step for second-life applications. As discussed in the section “Probabilistic ML techniques and their applications to battery health diagnostics and prognostics”, numerous SOH estimation techniques have been developed in recent years. Among these, data-driven and hybrid approaches have demonstrated their effectiveness in accurately assessing SOH. Specifically, probabilistic ML-based methods, such as the GPR-based approach (see the section “GPR applications to battery diagnostics and prognostics”), provide a distinct advantage. These methods not only produce a mean estimate of the SOH, but they also provide a prediction interval representing the predictive uncertainty. This probabilistic aspect enables decision makers to balance potential benefits and associated risks when choosing the most suitable second-

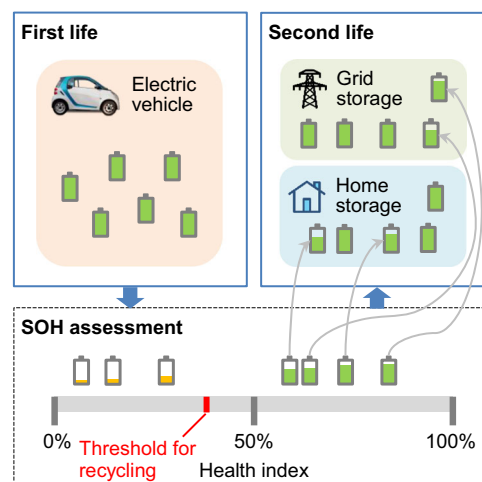


Fig. 16 | High-level assessment of Li-ion batteries for second-life applications. Cell SOH assessment for second-life applications.

life applications for retired EV batteries. The above discussion does not apply to cases where battery repurposing facilities can directly measure the SOH of a retired EV cell/module, e.g., by running a full charge/discharge cycle, and thus do not require SOH estimation using probabilistic or non-probabilistic ML techniques. These cases are expected to become less common, attributable to the rapidly growing volume of batteries from EVs reaching the end of their life over the next decade. Simultaneously, there is an escalating demand for rapidly assessing battery SOH in mass production settings, where ML techniques with uncertainty quantification capability are likely to play a major role.

3. **RUL prediction and economic benefits analysis:** Let us recall that the primary objective of employing second-life batteries lies in realizing the full economic potential of retired EV batteries. However, the economic viability of selecting retired batteries for specific second-life applications hinges on several factors. These factors include but are not limited to the efficacy of the battery management system, the costs associated with dismantling, and the costs involved in repackaging. Should the expenses of repurposing or remanufacturing surpass the benefits generated by the second-life batteries, it becomes economically unfeasible to extend the service life of retired batteries into their second life. Therefore, a comprehensive analysis of the economic benefits associated with second-life batteries becomes pivotal in evaluating the economic viability of a particular second-life application. A cornerstone technique for conducting such an economic benefits analysis is the RUL prediction of a battery for its second-life application. By analyzing how long the battery can continue to function effectively in the second life, we can assess the long-term benefits or cost savings that second-life batteries can contribute. However, it is worth noting that various uncertainty sources are present in the RUL prediction of second-life batteries, such as degradation mechanism uncertainty, lack of sufficient second-life degradation data, and uncertainty in second-life operation conditions, among others. Given the multifaceted nature of these uncertainty sources, the adoption of probabilistic ML models becomes imperative when predicting RUL for second-life applications. Such models are suitable for handling and quantifying these uncertainty sources, enhancing the accuracy and reliability of RUL prediction for the second-life application. Furthermore, integrating probabilistic RUL prediction into the decision-making process allows decision makers to consider and factor in various uncertainty sources within the economic analysis model when deciding on the second-life application.

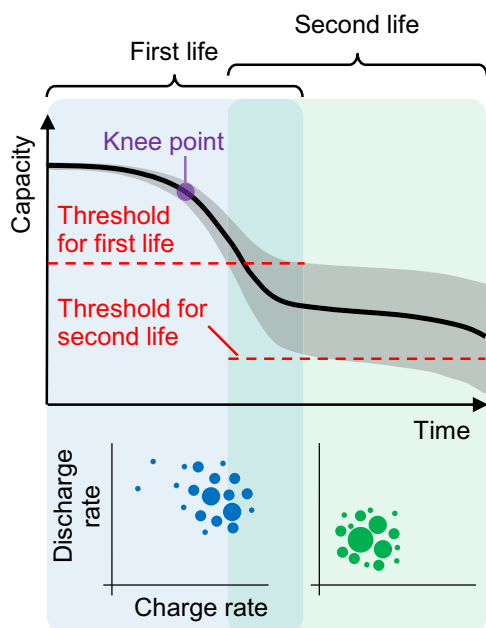


Fig. 17 | Overview of battery usage and capacity fade in first- and second-life applications. An illustration of an aging trajectory over two lives and a knee point in the second life.

Challenges in second-life battery applications

Realizing the economic and environmental potential of second-life batteries faces various challenges. These primary challenges can be summarized as follows (Fig. 17).

1. *Degradation modeling in the second life:* Empirical models, such as stochastic processes and regression models, can be established to predict the first-life degradation of EV batteries based on sensor data acquired from EVs. These empirical models, however, cannot be directly extrapolated for the degradation prediction in the second life, simply because the system may have been reconfigured for the second-life application, and the use conditions in the second life may significantly deviate from those in the first life. Solving this challenge requires establishing strong connections from the first-life field data to the second-life degradation models. Such connections may leverage physics-informed simulation models and the second-life (lab) test data from degradation experiments.
2. *RUL prediction for second-life applications:* The RUL prediction for second-life applications is challenging for several reasons.
 - First, existing prognostic approaches designed for general engineered systems have been successful, in part, in predicting the RULs of these systems. However, these approaches are mostly application-specific and are not robust across different applications, and are thus difficult to be directly applied for lifetime prognostics of retired EV batteries in their second lives.
 - Second, the knee point is where the battery aging transitions from a predominately linear degradation pattern into a nonlinear degradation region with a rapid capacity drop²²⁰. This critical knee point may occur during the first life, as illustrated in Fig. 17, and could also be experienced in second-life application. The exact occurrence time of the knee point is dependent on the cell chemistry, cell design, and usage patterns of the first- and second-life applications. Predicting the knee point is of great importance yet very challenging. Oftentimes, cell manufacturers make efforts in design and manufacturing to ensure an extremely low chance of having a knee point in first-life applications, essentially pushing the knee point into second-life applications. A higher likelihood of having a knee point makes the RUL prediction for the second-life application much more difficult than for the first-life application.

- Third, as mentioned above, degradation data for second-life applications is usually not sufficient. This poses significant challenges for commonly employed data-driven RUL prediction algorithms.

Possible solutions to the challenges

We believe the following research directions could provide potential solutions to the aforementioned challenges.

- *Physics-informed probabilistic ML for RUL prediction.* Physics-informed ML is an emerging concept in the field of failure prognostics²²¹. Incorporating physical laws or domain knowledge into ML models has the potential to substantially reduce the required amount of degradation data for failure prognostics. Furthermore, the synergy between physics-informed ML and the probabilistic ML methods discussed above enables the quantification of the predictive uncertainty due to the lack of sufficient degradation training data in the second-life application. The resulting physics-informed probabilistic ML models may possess the strengths of both learning paradigms. Specifically, physics-informed ML may enhance the extrapolation capability of the degradation model for failure prognostics, while probabilistic ML could quantify the inherent uncertainty in the prediction arising from such extrapolation.
- *Battery passport.* As pointed out in Thelen et al.²²², an alternative solution lies in the concept of *Battery passport*, introduced through a public-private collaboration platform called the Global Battery Alliance in November 2020²²³. The passport encompasses all relevant information about the battery from its initial production to its ultimate repurposing or recycling stage. Such a wealth of information could greatly facilitate the estimation of the SOH and prediction of RUL in the second-life application. A similar effort to harmonize battery data collection and reporting standards is the Battery Data Genome²²⁴. By unifying battery data reporting standards and creating more open-access databases, research breakthroughs will be more likely and electrification progress will accelerate.
- *Battery digital twins.* Similar to the Battery Passport concept, another possible and promising solution is the creation of digital twins tailored to individual batteries²²². The personalized digital replicas of batteries offer the potential for both rapid degradation diagnostics and accurate RUL prediction during their second-life applications.

Aging-aware battery control optimization

Aging-aware battery control optimization aims to regulate battery degradation to either ensure a minimum product lifetime (for ensuring a warranty is met) or extend the product's lifetime as much as possible, possibly at the expense of the user's experience²²⁵. Here, we review three main ways in which battery aging can be controlled to extend battery lifetime: adaptive discharging, adaptive charging, and thermal management.

Adaptive discharging

One method of controlling battery degradation and optimizing lifetime is through active discharge control. Limiting the power draw on the battery during discharge can extend the run time and slow the cycling-driven aging. Limiting the discharge power effectively limits the maximum C-rate the battery experiences, which is critical for reducing swelling and diffusion-induced stresses on the electrodes and preventing electrode degradation^{26,51}. This approach to aging-aware control is shown in Fig. 18, where a power-limited control strategy is compared to standard uncontrolled battery operation. In Fig. 18, the power-limited discharge strategy achieves a much longer run time because the maximum discharge rate is capped. Another effective discharge-based method of reducing battery degradation is by limiting the maximum DOD. Full DOD cycling has been shown to significantly accelerate battery aging, especially on next-generation Si-anode batteries⁵⁹. Limiting the maximum DOD of the battery can prevent significant degradation, thereby prolonging the battery life and, in some cases (when the battery has a shorter expected lifetime than other components), the product that uses the battery as the power source.

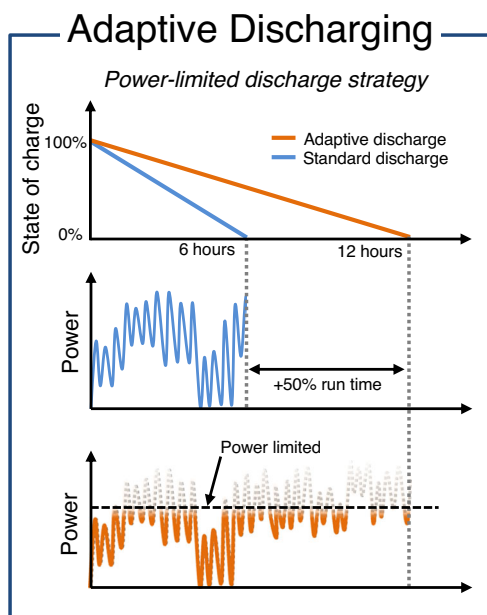


Fig. 18 | A comparison between two adaptive-discharge battery aging control strategies. With standard control strategies, power draw is not limited, and run time is uncontrolled (blue curves). With the adaptive discharge control strategy (orange curves), the maximum discharge power is limited to a safe range to extend the run time and reduce cycle-driven battery aging due to high C-rates.

Unfortunately, discharge-based methods of regulating battery aging have quickly fallen out of favor with engineers because these methods significantly impact the user experience. Probably the most infamous example of a discharge-based strategy to extend a product's lifetime was Apple's attempt to extend the life of users' iPhones by reducing the microprocessor clock speed via a software update. Reducing the processor clock speed effectively reduced the average power required of the battery, which successfully increased the phone's run time between charges (or the recharge interval), but at the expense of the phone's responsiveness to user inputs²²⁶. This practice was extremely unpopular with users and ultimately led to class-action lawsuits against Apple, which were eventually settled in late 2020. This event showed that a majority of users would simply prefer their devices continue to perform like new, even if it means the run time between charges significantly shortens as the battery ages.

Given the challenge of implementing a discharge-based battery aging control strategy without significantly affecting the user experience, engineers and researchers have mostly ceased researching the topic. We see this idea perpetuated in most battery diagnostic and prognostic modeling work reviewed in this paper – most of the works use voltage, capacity, current, and temperature data extracted during charging since battery discharge is mostly uncontrolled and application-dependent. We expect this trend to hold for the foreseeable future and expect future aging-aware control research to focus on charging and thermal management strategies.

Adaptive charging

Depending on the charging speed of interest, charging-based strategies for reducing battery aging can generally be grouped into two categories: slow charging and fast charging.

Slow charging

Research on slow-charging control strategies for reducing battery degradation focuses on a few key competing factors that affect the optimal solution, namely 1) charging during low-cost electricity intervals, 2) charging over long periods to reduce battery self-heating and avoid high temperatures, 3) charging near the end of the available charge time to avoid storage at high SOC, and 4) reducing high-power charging/discharging to

grid²²⁷. Hoke et al.²²⁸ used a reduced-order algebraic battery aging model to quickly evaluate the projected lifetime (years to 80% SOH) of four hybrid EV charging strategies: charge on plug in, charge at midnight, charge as late as possible, and charge minimizing electricity costs. Additionally, the charging strategies were evaluated for different charging powers to account for the effect of C-rate on battery aging. The authors showed that their co-optimization method, which minimized both battery degradation and electricity costs, produced projected battery lifetimes far exceeding the other strategies. Notably, their results showed that the strategy of simply waiting to charge until as late as possible in the charging window and using a low charging rate can significantly increase the projected lifetime of the battery. However, the authors did point out that defining the charging window requires knowledge of the user's behavior, which in many cases needs to be learned from monitoring behavior over a long period. Further, in the context of this study, if the user behavior significantly deviates from the pre-defined schedule, they may find that their HEV is not fully charged, as the strategy waits until as early in the morning as possible to charge the vehicle. However, with modern phone applications that enable users to control their vehicles (e.g., the Tesla mobile app), it is not inconceivable that a user could manually override the charging strategy to prepare the vehicle for a planned road trip. As battery diagnostic and prognostic models are refined, we expect the details of aging-aware slow-charging optimization will change, but the general premise of balancing battery degradation against electricity costs, user behavior, and other engineering constraints will persist.

Fast charging

In recent years, fast charging has received significant attention because of its role in quickly enabling battery-powered devices to continue normal operation after 15–25 min charging sessions. The technology is significantly important to the continued adoption of EVs by enabling them to “refuel” near the same speeds as traditional internal combustion engines that pump liquid fuel. However, significantly more degradation occurs during fast charging than slow charging due to the extreme C-rates and temperatures the battery cells experience. Under high C-rates, Li-ion batteries are prone to experience lithium plating, which creates unsafe Li-metal dendrites that can cause an internal short-circuit and risk of fire (see the section “Battery degradation – modes and mechanisms”). As a result, research into aging-aware fast charging generally focuses on developing optimal fast-charge profiles to reduce battery degradation from lithium plating using a variety of experimental and simulation-based methods.

An optimal fast-charging profile will balance 1) the charge time (affects the user experience to an extent), 2) the available charger power (modern EV fast chargers are typically limited to 350 kW), 3) battery temperatures (to avoid thermal runaway or maintain EV cabin comfort), and 4) battery aging, among other engineering constraints. Attia et al.⁷⁹ took an experimental approach to fast-charge protocol optimization by cycle aging a large batch of cells (>45 cells) with various 2-step fast charging profiles and then using a Bayesian optimization algorithm to suggest new fast charging protocols for testing, sequentially working towards an optimal fast charging profile for the cell design. Instead of waiting for the cells to reach their EOL, which might take many hundreds of cycles, the researchers used an early life prediction model (see the section “Early life and trajectory prediction”) to predict the lifetimes of the cells after only 100 cycles, increasing the rate at which profiles could be evaluated. The high-throughput experimental testing approach used by Attia et al.⁷⁹ was essentially probing the cell's lithium-plating limits while simultaneously considering cell self-heating from the high C-rates. The method proved successful, and the authors were able to extend cell life by an average of 180 cycles over previously published fast-charge protocols.

Different from the approach taken by Attia et al.⁷⁹, which focused on optimizing a fast-charging protocol using cell-level aging performance, Konz et al.⁵⁸ looked at fast charging on the component level. They demonstrated a quick and efficient method of experimentally determining a cell's lithium-plating onset SOC for a given temperature and C-rate by repeatedly cycling graphite half-cells made from the cell's components. By mapping out the specific C-rates, temperatures, and SOC conditions under

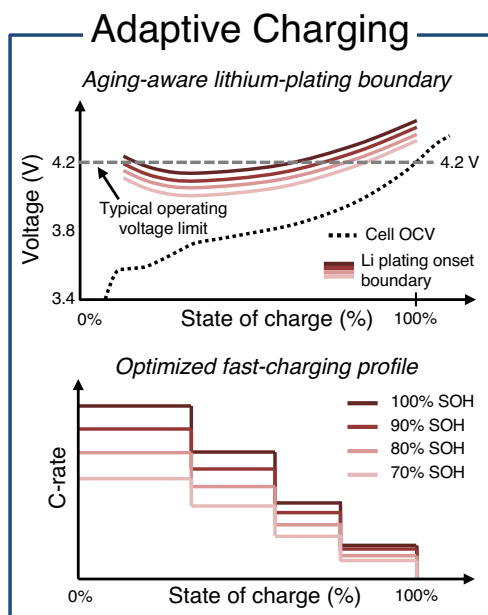


Fig. 19 | Top, illustration of a voltage-based lithium-plating boundary that decreases in voltage as the cell ages due to the increased likelihood of lithium plating in aged battery cells. The OCV of an NMC/Gr cell is shown for reference. *Bottom*, example optimal fast charging profiles for various battery SOHs and designed to charge the cell as fast as possible without the cell voltage exceeding the corresponding lithium-plating voltage limits. Figure based on work by Konz et al.²²⁹

which lithium plating occurs in a cell, one can create an optimal fast-charging profile that applies the maximum C-rate without exceeding a margin of safety around the identified lithium plating limit. This approach to fast-charging profile design is very flexible because the optimal profile can be dynamically determined based on the cell's operating temperature in its intended application. The key advantage of the author's method over existing ones reported in the literature is that it does not require high precision coulometry equipment and can be done quickly on standard battery cyclers. The method works by performing an SOC-sweep to measure the cell's coulombic efficiency at a given temperature and C-rate over varied SOCs. If done correctly, one can identify the lithium-plating onset as the point in the CE vs. SOC curve where the CE begins to significantly decline from 100%. A low CE is an indication of irreversible capacity loss due to lithium plating. While the method in Konz et al.⁵⁸ shows promise for significantly improving the speed at which a cell's lithium-plating limits can be mapped out, the number of tests quickly becomes overwhelming as soon as one wants to adapt the profile as cells age. To capture the effect of cell aging when calculating the optimal fast-charging profile, SOC-sweep cycling tests would need to be performed on aged graphite half-cells, requiring an extensive DOE of aging tests where cells are pulled off at various levels of SOH to conduct SOC-sweeps.

An alternative approach to developing fast-charging profiles considering aging is to use physics-based simulation. In a follow-up paper by Konz et al.²²⁹, the team of researchers conducted a meta-analysis of lithium-plating onset conditions by simulating thousands of unique fast-charging current profiles using a pseudo-2D electrochemical-thermal model of an NMC532/Gr cell. The researchers were able to map out an upper voltage limit over an SOC range that, if exceeded, the cell is likely to experience lithium plating. An example of the lithium-plating voltage limit and a corresponding optimal fast-charging profile is shown in Fig. 19. The main advantage of finding a purely voltage-based lithium-plating limit is its flexibility—voltage is a response to an applied current, making voltage-based lithium-plating limits agnostic to the fast-charging current profile used. To study the impact of aging on the lithium-plating onset, the researchers modified parameters in the electrochemical model to simulate

the effects of aging. For example, they simulated loss of active electrode materials (LAM_{PE} , LAM_{NE} , see the section "Battery degradation—modes and mechanisms") by decreasing the value of the active material fraction parameter in the model. Other simulated aging mechanisms included electrode expansion, decreased charge-transfer kinetics, and loss of lithium inventory. The researchers extensively simulated various fast-charging current profiles with P2D model parameters sampled from distributions spanning the expected range corresponding to cells with between 85 and 100% SOH. As expected, the lithium-plating voltage limit was found to decrease with cell aging, meaning it is more likely for aged cells to experience lithium plating if the fast charging profile is not modified to account for cell aging. While the results are certainly interesting, the study was largely exploratory in nature and did not offer many actionable insights for engineers and practitioners since aging was simulated by randomly sampling the P2D model parameters, which largely ignores the path-dependence of aging typically observed in aging tests and from field data 6. While it is profound that the authors were able to demonstrate the lithium-plating voltage limits decrease with cell aging, more work needs to be done to quantify *how much* the lithium-plating limits need to decrease for a given battery SOH.

Effectively designing aging-aware fast-charging profiles requires accurately estimating the various degradation mechanisms inside the cell. A combination of capacity, power, and degradation mode estimation⁸¹ is required for characterizing cell SOH and prescribing the proper adjustments to the fast-charging profile. Future research in this area will focus on learning from field battery data uploaded to the cloud how best to adjust fast-charging profiles on a per-user basis—essentially personalizing the charging experience to users individually, as no two batteries will age the same. Research into federated learning methods for training/deploying client-specific ML models will be essential to building optimized aging-aware battery control mechanisms without compromising user's privacy²³⁰. Additionally, new methods for quickly detecting lithium-plating and designing optimal fast-charging profiles that work for next-generation battery chemistries are needed. Last, new physics-based simulation techniques will need to be developed to more accurately simulate battery aging under real-world conditions, possibly leveraging ML to account for cell aging variability (see the section "Physics-based diagnostics and prognostics").

Thermal management

The last aging-aware battery control strategy we discuss is thermal management. We highlight thermal management and general thermal modeling of batteries as important future research areas because of their significance in industry-focused product design. Factors like cell/pack packaging, cell form-factor, and regional climate differences (temperature, humidity, pressure, solar radiance) have a great effect on cell temperature within a product, significantly affecting the overall product's life. In general, most Li-ion batteries are sensitive to temperature and have a small ideal operation window where they perform their best and experience minimum temperature-induced degradation effects, typically in the range 0–40 °C. However, we note that different battery chemistries, particularly next-generation batteries, can have various optimal temperature ranges, and so here we discuss thermal management strategies in a general sense of trying to maintain a lithium-ion battery's temperature in its optimal range. Figure 20 shows approximate Li-ion battery cycle life as a function of temperature for various charging C-rates. In this example, there exists a stable window between 0 and 35 °C where cycle life remains mostly constant with temperature, indicating the optimal operation window. Below 0 °C, Li-ion batteries are likely to experience lithium plating (see the section "Battery degradation—modes and mechanisms"), and above 40 °C, the rate of SEI formation and other side reactions are significantly increased, decreasing capacity and cycle life rapidly.

Generally speaking, active thermal management of Li-ion batteries is only feasible on larger battery-powered systems like grid energy storage, EVs, and HEVs since smaller battery-powered electronic devices like phones and laptops lack the space for air/liquid heat exchangers and

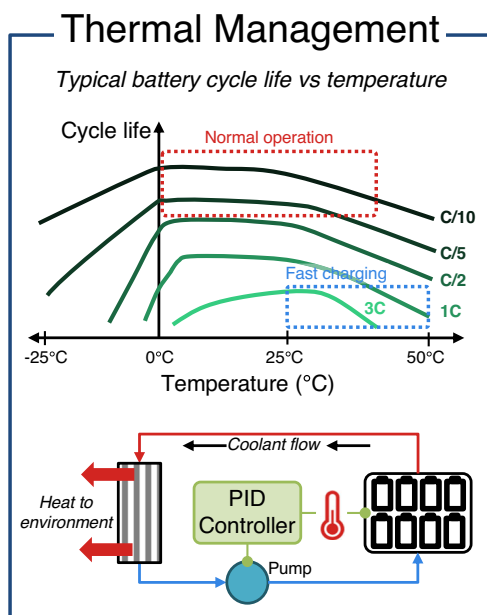


Fig. 20 | *Top*, approximate Li-ion battery cycle life as a function of temperature with regions highlighted for typical operation in portable electronics and EVs. *Bottom*, an example battery pack cooling strategy that uses liquid coolant to pull heat away from the pack and a heat exchanger to reject the heat to the environment.

compressor systems. Instead, engineers and designers can usually reduce battery temperatures in small electronic devices by simply increasing the energy density of the cells. Given the same power load from the device but with a larger battery, the effective C-rate the cell experiences is lower, reducing self-heating and improving aging performance²³¹.

On large battery systems, battery thermal management is carried out using air, liquids, or refrigerants. Typically, at least one surface of each cell in a pack is exposed, allowing air or liquid to flow over the cell surface, pulling heat away from the cell and dumping it into the surrounding environment. A simple coolant system loop for removing heat from a battery pack and rejecting it to the environment is shown in Fig. 20. Optimizing a thermal control system design and strategy requires understanding how battery aging changes with temperature. Reduced-order aging models, such as from these papers^{51,169}, are excellent for control strategy optimization since they can simulate battery aging under various temperatures with minimal computational overhead. While average cell temperatures dominate storage aging, minimum and maximum temperatures play a large role in cycling-driven aging. Coupling battery aging models with thermal models to simulate cell internal heating and heat transfer to the coolant is imperative for proper design optimization. Existing research using battery thermal models to devise thermal control strategies suggests that thermal management system and cell co-design will lead to more optimal battery performance in EVs^{231,232} by reducing cell temperatures, improving heat transfer away from cells, and shortening fast-charging times. Efficient co-design is best achieved using battery digital twin models^{221,222} that couple diagnostic, prognostic, thermal, electrical, and form-factor models to comprehensively simulate multiple battery cells and packs of various designs.

While sufficient battery thermal system optimization and control can be achieved with existing diagnostic and prognostic models, there remain significant challenges that will require further research. Presently, standard practice is to build battery diagnostic and prognostic models for a specific cell design using data collected in the lab. Nearly all the research papers discussed in this review build battery-specific aging models that cannot be used to predict aging for batteries of different designs, packaging configurations, or form-factors. This approach to modeling is inflexible, time-consuming, and not relevant to industry where product design constraints like cell packaging, active cooling, and power requirements change

frequently as new features and subsystems are added to the product. For example, Keyser et al.²³¹ simulated pouch cells with different terminal locations (terminals on the same side vs opposite side) and found considerable differences in cell internal heating. Other work by Gasper et al.¹⁷⁰ found that a large-format (>50 Ah) cell's aspect ratio (area over thickness) significantly affects its thermal resistance and self-heating, demonstrating form factor has a significant effect on cooling system design. These two studies highlight the challenge of building battery diagnostic and prognostic models that can effectively extrapolate to new cell designs, form-factors, and use conditions. Further, regional differences in outdoor air temperatures, solar radiance, and humidity drive aging variability in cells, making uncertainty quantification in simulations essential for drawing accurate conclusions about battery design and control strategies. In light of this, we urge researchers to focus on developing diagnostic and prognostic modeling methods that enable engineers and practitioners to quickly assess the impact of various product design changes on battery aging so that battery-product co-design can be achieved and products can be further optimized. A promising path is physics-based battery diagnostic and prognostic models, like those discussed in the section "Physics-based diagnostics and prognostics". New battery digital twin models that comprehensively model all aspects of battery performance and aging using a combination of physics and machine learning will be paramount to the future development of battery-powered systems like EVs, consumer electronics, and future aviation efforts^{222,233}.

Conclusion

Modeling battery degradation is essential for optimizing every aspect of the battery life cycle. From research and development in the lab, to optimizing a fast-charging protocol for aged cells in the field, probabilistic battery diagnostic and prognostic models are core to the continued deployment and success of battery technology. In this work, we reviewed existing and emerging research into probabilistic ML for battery diagnostics and prognostics, emphasizing and highlighting seminal research focusing on the combination of accurate battery health modeling considering uncertainty. Altogether, our review has outlined the great need for more research into uncertainty quantification for battery prognostic models to solve problems related to a lack of data for modeling due to high testing costs, inherent cell-to-cell performance and aging variability stemming from manufacturing and testing limitations, and the sheer severity of consequences arising from poor maintenance and control of battery cells in consumer devices. As research in this area continues to mature, it is envisioned that probabilistic ML models will play a crucial role in creating safe, reliable, and long-lasting battery systems. To this end, we see several long-standing challenges that need to be further investigated by the research community:

1. Publicly available battery aging datasets are crucial for accelerating the development of probabilistic battery diagnostic and prognostic models. Existing datasets have been instrumental in furthering research in the field (the section "Publicly available battery aging datasets"), however, they primarily consist of cell-level aging data collected in a lab, largely ignoring the important influence of packaging, cooling systems, and time-varying operating conditions on aging. Collaboration between industry and academia to gather and disseminate high-quality cell/module/pack aging data will be crucial for continued research in the coming years.
2. Hybrid ML and physics-based modeling will play a large role in designing the battery-powered systems of tomorrow. There is a great opportunity to develop new physics-based diagnostic and prognostic models and physic-informed ML methods that provide greater accuracy and insight into degradation modes than exist today. We see ML as an important tool for identifying physics-based relationships in battery data collected from the field, and informing the design and development of truly physics-based battery aging models that can provide far greater accuracy than we have today.
3. Last, developing coupled thermal, electrical, mechanical, and aging models will be key to optimizing all aspects of cell design. Such models

enable the possibility of battery/product co-design where the battery form-factor, packaging, cooling, and control algorithms are all optimized considering a set of unified engineering constraints like cost, volume, weight, energy, operating conditions (regional climate, driving habits), and more. Collaborations between engineering disciplines will be crucial to successfully developing the coupled battery digital twin models of the future.

With future infrastructure and transportation trending toward electric power, batteries will continue to play a pivotal role in our society. The path ahead for future battery research is certainly challenging, but ultimately will be achievable through interdisciplinary collaboration between academic researchers, industry engineers, and regulatory bodies.

Data availability

The training and test datasets analyzed to generate Fig. 7 are available from the corresponding author, Chao Hu, upon reasonable request. The battery aging dataset analyzed to generate Figs. 8 and 9 is not publicly available due to confidentiality reasons. The battery aging dataset analyzed to generate Fig. 14 was discussed in an earlier published article⁸⁷ and is available from the article's corresponding author, Noah Paulson, upon reasonable request.

Received: 1 November 2023; Accepted: 11 February 2024;

Published online: 03 June 2024

References

- Harper, G. et al. Recycling lithium-ion batteries from electric vehicles. *Nature* **575**, 75–86 (2019).
- Wright, R. B. et al. Calendar-and cycle-life studies of advanced technology development program generation 1 lithium-ion batteries. *J. Power Sourc.* **110**, 445–470 (2002).
- Ramadass, P., Haran, B., White, R. & Popov, B. N. Mathematical modeling of the capacity fade of Li-ion cells. *J. Power Sourc.* **123**, 230–240 (2003).
- Bloom, I. et al. An accelerated calendar and cycle life study of Li-ion cells. *J. Power Sourc.* **101**, 238–247 (2001).
- Spotnitz, R. Simulation of capacity fade in lithium-ion batteries. *J. Power Sourc.* **113**, 72–80 (2003).
- Broussely, M. et al. Aging mechanism in Li ion cells and calendar life predictions. *J. Power Sourc.* **97**, 13–21 (2001).
- Liaw, B. Y., Jungst, R. G., Nagasubramanian, G., Case, H. L. & Doughty, D. H. Modeling capacity fade in lithium-ion cells. *J. Power Sourc.* **140**, 157–161 (2005).
- Wang, D., Yang, F., Tsui, K.-L., Zhou, Q. & Bae, S. J. Remaining useful life prediction of lithium-ion batteries based on spherical cubature particle filter. *IEEE Trans. Instrum. Meas.* **65**, 1282–1291 (2016).
- Attia, P. M., Chueh, W. C. & Harris, S. J. Revisiting the t0.5 dependence of SEI growth. *J. Electrochem. Soc.* **167**, 090535 (2020).
- Miao, Q., Xie, L., Cui, H., Liang, W. & Pecht, M. Remaining useful life prediction of lithium-ion battery with unscented particle filter technique. *Microelectron. Reliability* **53**, 805–810 (2013).
- Saha, B., Goebel, K., Poll, S. & Christophersen, J. Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Trans. Instrum. Meas.* **58**, 291–296 (2008).
- Wang, D., Miao, Q. & Pecht, M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. *J. Power Sourc.* **239**, 253–264 (2013).
- Hu, C., Jain, G., Schmidt, C., Strief, C. & Sullivan, M. Online estimation of lithium-ion battery capacity using sparse Bayesian learning. *J. Power Sourc.* **289**, 105–113 (2015).
- Roman, D., Saxena, S., Robu, V., Pecht, M. & Flynn, D. Machine learning pipeline for battery state-of-health estimation. *Nature Mach. Intell.* **3**, 447–456 (2021).
- Zhang, L. et al. A review on deep learning applications in prognostics and health management. *IEEE Access* **7**, 162415–162438 (2019).
- Xiong, R., Cao, J., Yu, Q., He, H. & Sun, F. Critical review on the battery state of charge estimation methods for electric vehicles. *IEEE Access* **6**, 1832–1843 (2017).
- Hannan, M. A., Lipu, M. H., Hussain, A. & Mohamed, A. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations. *Renew. Sustain. Energy Rev.* **78**, 834–854 (2017).
- Plett, G. L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background. *J. Power Sourc.* **134**, 252–261 (2004).
- Plett, G. L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. state and parameter estimation. *J. Power Sourc.* **134**, 277–292 (2004).
- Ungurean, L., Cârstoiu, G., Micea, M. V. & Groza, V. Battery state of health estimation: a structured review of models, methods and commercial devices. *Int. J. Energy Res.* **41**, 151–181 (2017).
- Berecibar, M. et al. Critical review of state of health estimation methods of Li-ion batteries for real applications. *Renew. Sustain. Energy Rev.* **56**, 572–587 (2016).
- Hu, X., Che, Y., Lin, X. & Onori, S. Battery health prediction using fusion-based feature selection and machine learning. *IEEE Trans. Transport. Electr.* **7**, 382–398 (2022).
- Hasib, S. A. et al. A comprehensive review of available battery datasets, RUL prediction approaches, and advanced battery management. *IEEE Access.* **9**, 86166–86193 (2021).
- Liao, L. & Köttig, F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Trans. Reliability* **63**, 191–207 (2014).
- Wang, S. et al. A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries. *Energy Rep.* **7**, 5562–5574 (2021).
- Li, T., Zhou, Z., Thelen, A., Howey, D. & Hu, C. Predicting battery lifetime under varying usage conditions from early aging data. *arXiv preprint arXiv:2307.08382* (2023).
- Severson, K. A. et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **4**, 383–391 (2019).
- Fernín-Cueto, P. et al. Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. *Energy and AI* **1**, 100006 (2020).
- Li, W. et al. One-shot battery degradation trajectory prediction with deep learning. *J. Power Sourc.* **506**, 230024 (2021).
- Liu, J., Thelen, A., Hu, C. & Yang, X.-G. An end-to-end learning framework for early prediction of battery capacity trajectory. In *Proc. Annual Conference of the PHM Society*, vol. 13 (2021).
- Pozzato, G. et al. Analysis and key findings from real-world electric vehicle field data. *Joule* **7**, 2035–2053 (2023).
- Geslin, A. et al. Selecting the appropriate features in battery lifetime predictions. *Joule* **7**, 1956–1965 (2023).
- Sulzer, V. et al. The challenge and opportunity of battery lifetime prediction from field data. *Joule* **5**, 1934–1955 (2021).
- Nemani, V. et al. Uncertainty quantification in machine learning for engineering design and health prognostics: a tutorial. *Mech. Syst. Signal Process.* **205**, 110796 (2023).
- Dechent, P. et al. Estimation of Li-ion degradation test sample sizes required to understand cell-to-cell variability. *Batter. Supercaps* **4**, 1821–1829 (2021).
- Baumhöfer, T., Brühl, M., Rothgang, S. & Sauer, D. U. Production caused variation in capacity aging trend and correlation to initial cell performance. *J. Power Sourc.* **247**, 332–338 (2014).
- Harris, S. J., Harris, D. J. & Li, C. Failure statistics for commercial lithium ion batteries: a study of 24 pouch cells. *J. Power Sourc.* **342**, 589–597 (2017).
- Hu, X., Xu, L., Lin, X. & Pecht, M. Battery lifetime prognostics. *Joule* **4**, 310–346 (2020).

39. Wang, Y. et al. A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. *Renew. Sustain. Energy Rev.* **131**, 110015 (2020).
40. Sui, X. et al. A review of non-probabilistic machine learning-based state of health estimation techniques for lithium-ion battery. *Appl. Energy* **300**, 117346 (2021).
41. Aykol, M. et al. Perspective—combining physics and machine learning to predict battery lifetime. *J. Electrochem. Soc.* **168**, 030525 (2021).
42. Ge, M.-F., Liu, Y., Jiang, X. & Liu, J. A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries. *Measurement* **174**, 109057 (2021).
43. Zhang, Y. & Li, Y.-F. Prognostics and health management of lithium-ion battery using deep learning methods: a review. *Renew. Sustain. Energy Rev.* **161**, 112282 (2022).
44. Che, Y., Hu, X., Lin, X., Guo, J. & Teodorescu, R. Health prognostics for lithium-ion batteries: mechanisms, methods, and prospects. *Energy & Environmental Science*, **16**, 338–371 (2023).
45. Zhao, J. et al. Battery prognostics and health management from a machine learning perspective. *J. Power Sourc.* **581**, 233474 (2023).
46. Dubarry, M. & Baure, G. Perspective on commercial Li-ion battery testing, best practices for simple and effective protocols. *Electronics* **9**, 152 (2020).
47. Chen, Y. et al. A review of lithium-ion battery safety concerns: the issues, strategies, and testing standards. *J. Energy Chem.* **59**, 83–99 (2021).
48. Dos Reis, G., Strange, C., Yadav, M. & Li, S. Lithium-ion battery data and where to find it. *Energy AI* **5**, 100081 (2021).
49. Smith, A., Burns, J. C., Zhao, X., Xiong, D. & Dahn, J. A high precision coulometry study of the SEI growth in Li/graphite cells. *J. Electrochem. Soc.* **158**, A447 (2011).
50. Das, S., Attia, P. M., Chueh, W. C. & Bazant, M. Z. Electrochemical kinetics of sei growth on carbon black: Part II. modeling. *J. Electrochem. Soc.* **166**, E107–E118 (2019).
51. Smith, K., Gasper, P., Colclasure, A. M., Shimonishi, Y. & Yoshida, S. Lithium-ion battery life model with electrode cracking and early-life break-in processes. *J. Electrochem. Soc.* **168**, 100530 (2021).
52. Gasper, P., Gering, K., Dufek, E. & Smith, K. Challenging practices of algebraic battery life models through statistical validation and model identification via machine-learning. *J. Electrochem. Soc.* **168**, 020502 (2021).
53. Takahashi, K. & Srinivasan, V. Examination of graphite particle cracking as a failure mode in lithium-ion batteries: a model-experimental study. *J. Electrochem. Soc.* **162**, A635 (2015).
54. Ruess, R. et al. Influence of ncm particle cracking on kinetics of lithium-ion batteries with liquid or solid electrolyte. *J. Electrochem. Soc.* **167**, 100532 (2020).
55. van Vlijmen, B. et al. Interpretable data-driven modeling reveals complexity of battery aging. *Chemrxiv.org* (2023).
56. Attia, P. M. et al. "knees" in lithium-ion battery aging trajectories. *J. Electrochem. Soc.* **169**, 060517 (2022).
57. Huang, W. et al. Onboard early detection and mitigation of lithium plating in fast-charging batteries. *Nat. Commun.* **13**, 7091 (2022).
58. Konz, Z. M. et al. High-throughput Li plating quantification for fast-charging battery design. *Nat. Energy* **8**, 450–461 (2023).
59. Zuo, X., Zhu, J., Müller-Buschbaum, P. & Cheng, Y.-J. Silicon based lithium-ion battery anodes: a chronicle perspective review. *Nano Energy* **31**, 113–143 (2017).
60. Zhang, H. et al. Li₄Ti₅O₁₂ spinel anode: fundamentals and advances in rechargeable batteries. *InfoMat* **4**, e12228 (2022).
61. He, Y.-B. et al. Gassing in Li₄Ti₅O₁₂-based batteries and its remedy. *Sci. Rep.* **2**, 913 (2012).
62. Feng, K. et al. Silicon-based anodes for lithium-ion batteries: from fundamentals to practical applications. *Small* **14**, 1702737 (2018).
63. Albertus, P., Babinec, S., Litzelman, S. & Newman, A. Status and challenges in enabling the lithium metal electrode for high-energy and low-cost rechargeable batteries. *Nat. Energy* **3**, 16–21 (2018).
64. Xia, S., Wu, X., Zhang, Z., Cui, Y. & Liu, W. Practical challenges and future perspectives of all-solid-state lithium-metal batteries. *Chemistry* **5**, 753–785 (2019).
65. Janek, J. & Zeier, W. G. Challenges in speeding up solid-state battery development. *Nat. Energy* **8**, 230–240 (2023).
66. Lewis, J. A., Tippens, J., Cortes, F. J. Q. & McDowell, M. T. Chemo-mechanical challenges in solid-state batteries. *Trends Chem.* **1**, 845–857 (2019).
67. Raza, H. et al. Li-s batteries: challenges, achievements and opportunities. *Electrochem. Energy Rev.* **6**, 29 (2023).
68. He, J. & Manthiram, A. A review on the status and challenges of electrocatalysts in lithium-sulfur batteries. *Energy Storage Mater.* **20**, 55–70 (2019).
69. Luntz, A. C. & McCloskey, B. D. Nonaqueous Li–air batteries: a status report. *Chem. Rev.* **114**, 11721–11750 (2014).
70. Barré, A. et al. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. *J. Power Sourc.* **241**, 680–689 (2013).
71. Birkl, C. R., Roberts, M. R., McTurk, E., Bruce, P. G. & Howey, D. A. Degradation diagnostics for lithium ion cells. *J. Power Sourc.* **341**, 373–386 (2017).
72. Richardson, R. R., Birkl, C. R., Osborne, M. A. & Howey, D. A. Gaussian process regression for in situ capacity estimation of lithium-ion batteries. *IEEE Trans. Ind. Inform.* **15**, 127–138 (2018).
73. Li, X., Wang, Z. & Yan, J. Prognostic health condition for lithium battery using the partial incremental capacity and Gaussian process regression. *J. Power Sourc.* **421**, 56–67 (2019).
74. Li, X., Yuan, C., Li, X. & Wang, Z. State of health estimation for Li-ion battery using incremental capacity analysis and Gaussian process regression. *Energy* **190**, 116467 (2020).
75. Wang, L., Pan, C., Liu, L., Cheng, Y. & Zhao, X. On-board state of health estimation of LiFePO₄ battery pack through differential voltage analysis. *Appl. Energy* **168**, 465–472 (2016).
76. Berecibar, M. et al. Online state of health estimation on nmc cells based on predictive analytics. *J. Power Sourc.* **320**, 239–250 (2016).
77. Attia, P. M., Severson, K. A. & Witmer, J. D. Statistical learning for accurate and interpretable battery lifetime prediction. *J. Electrochem. Soc.* **168**, 090547 (2021).
78. Sendek, A. D. et al. Machine learning modeling for accelerated battery materials design in the small data regime. *Adv. Energy Mater.* **12**, 2200553 (2022).
79. Attia, P. M. et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**, 397–402 (2020).
80. Wang, F. et al. Explainability-driven model improvement for SOH estimation of lithium-ion battery. *Reliability Eng. Syst. Saf.* **232**, 109046 (2023).
81. Thelen, A. et al. Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries. *Energy Storage Mater.* **50**, 668–695 (2022).
82. He, W., Williard, N., Osterman, M. & Pecht, M. Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. *J. Power Sourc.* **196**, 10314–10321 (2011).
83. Xing, Y., Ma, E. W., Tsui, K.-L. & Pecht, M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectron. Reliability* **53**, 811–820 (2013).
84. Williard, N., He, W., Osterman, M. & Pecht, M. Comparative analysis of features for determining state of health in lithium-ion batteries. *Int. J. Progn. Health Manag.* **4**, 2013 (2013).
85. Bole, B., Kulkarni, C. & Daigle, M. Randomized battery usage data set. *NASA AMES prognostics data repository* **70**, (2014).

86. Bole, B., Kulkarni, C. S. & Daigle, M. Adaptation of an electrochemistry-based Li-ion battery model to account for deterioration observed under randomized use. *6*, (2014).
87. Paulson, N. H. et al. Feature engineering for machine learning enabled early prediction of battery lifetime. *J. Power Sourc.* **527**, 231127 (2022).
88. Preger, Y. et al. Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *J. Electrochem. Soc.* **167**, 120532 (2020).
89. Raj, T., Wang, A. A., Monroe, C. W. & Howey, D. A. Investigation of path-dependent degradation in lithium-ion batteries. *Batter. Supercaps* **3**, 1377–1385 (2020).
90. Pozzato, G., Allam, A. & Onori, S. Lithium-ion battery aging dataset based on electric vehicle real-driving profiles. *Data Brief* **41**, 107995 (2022).
91. Moy, K., Ganapathi, D., Geslin, A., Chueh, W. & Onori, S. Synthetic duty cycles from real-world autonomous electric vehicle driving. *Cell Rep. Phys. Sci.* **4**, (2023).
92. She, C., Wang, Z., Sun, F., Liu, P. & Zhang, L. Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network. *IEEE Trans. Ind. Inform.* **16**, 3345–3354 (2019).
93. Rasmussen, C. E. et al. Gaussian Processes for Machine Learning, vol. **1** (Springer, 2006).
94. Neal, R. M. Bayesian Learning for Neural Networks, vol. 118 (Springer Science & Business Media, 2012).
95. Yang, D., Zhang, X., Pan, R., Wang, Y. & Chen, Z. A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. *J. Power Sourc.* **384**, 387–395 (2018).
96. Hu, C. et al. Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery. *Appl. Energy* **129**, 49–55 (2014).
97. Deng, Z., Hu, X., Li, P., Lin, X. & Bian, X. Data-driven battery state of health estimation based on random partial charging data. *IEEE Trans. Power Electron.* **37**, 5021–5031 (2021).
98. Goebel, K., Saha, B., Saxena, A., Celaya, J. R. & Christophersen, J. P. Prognostics in battery health management. *IEEE Instrum. Meas. Mag.* **11**, 33–40 (2008).
99. Liu, D., Pang, J., Zhou, J., Peng, Y. & Pecht, M. Prognostics for state of health estimation of lithium-ion batteries based on combination gaussian process functional regression. *Microelectron. Reliability* **53**, 832–839 (2013).
100. Richardson, R. R., Osborne, M. A. & Howey, D. A. Gaussian process regression for forecasting battery state of health. *J. Power Sourc.* **357**, 209–219 (2017).
101. Thelen, A. et al. Augmented model-based framework for battery remaining useful life prediction. *Appl. Energy* **324**, 119624 (2022).
102. Richardson, R. R., Osborne, M. A. & Howey, D. A. Battery health prediction under generalized conditions using a gaussian process transition model. *J. Energy Storage* **23**, 320–328 (2019).
103. Jones, P. K., Stimming, U. & Lee, A. A. Impedance-based forecasting of lithium-ion battery performance amid uneven usage. *Nat. Commun.* **13**, 4806 (2022).
104. Lu, J. et al. Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning. *Energy Storage Mater.* **50**, 139–151 (2022).
105. Valladares, H. et al. Gaussian process-based prognostics of lithium-ion batteries and design optimization of cathode active materials. *J. Power Sourc.* **528**, 231026 (2022).
106. Liu, K., Hu, X., Wei, Z., Li, Y. & Jiang, Y. Modified gaussian process regression models for cyclic capacity prediction of lithium-ion batteries. *IEEE Trans. Transport. Electr.* **5**, 1225–1236 (2019).
107. Tipping, M. E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**, 211–244 (2001).
108. Hu, X., Jiang, J., Cao, D. & Egardt, B. Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling. *IEEE Trans. Ind. Electron.* **63**, 2645–2656 (2015).
109. Deng, Z. et al. General discharge voltage information enabled health evaluation for lithium-ion batteries. *IEEE/ASME Trans. Mechatron.* **26**, 1295–1306 (2020).
110. Hu, C., Jain, G., Tamirisa, P. & Gorka, T. Method for estimating capacity and predicting remaining useful life of lithium-ion battery. *Appl. Energy* **126**, 182–189 (2014).
111. Hu, C., Ye, H., Jain, G. & Schmidt, C. Remaining useful life assessment of lithium-ion batteries in implantable medical devices. *J. Power Sourc.* **375**, 118–130 (2018).
112. Zhang, Y., Xiong, R., He, H. & Pecht, M. G. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Veh. Technol.* **67**, 5695–5705 (2018).
113. Li, H., Pan, D. & Chen, C. P. Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine. *IEEE Trans. Syst. Man. Cybernet. Syst.* **44**, 851–862 (2014).
114. LeCun, Y. A., Bottou, L., Orr, G. B. & Müller, K.-R. Efficient BackProp. In Montavon, G., Orr, G. B. & Müller, K.-R. (eds.) *Neural Networks: Tricks of the Trade*, 9–48 (Springer-Verlag Berlin Heidelberg, 2012).
115. Robbins, H. & Monro, S. A Stochastic Approximation method. *Ann. Math. Stat.* **22**, 400–407 (1951).
116. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic gradient descent. *ICLR: International Conference on Learning Representations* (2015). arXiv:1412.6980v9.
117. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics (Springer New York, 1985).
118. Bernardo, J. M. & Smith, A. F. M. *Bayesian Theory* (John Wiley & Sons, New York, NY, 2000).
119. Sivia, D. S. & Skilling, J. *Data Analysis: A Bayesian Tutorial*, 2nd edn. (Oxford University Press, New York, NY, 2006).
120. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice* (Chapman & Hall, New York, NY, 1996).
121. Andrieu, C., de Freitas, N., Doucet, A. & Jordan, M. I. An introduction to MCMC for Machine Learning. *Mach. Learn.* **50**, 5–43 (2003).
122. Robert, C. P. & Casella, G. *Monte Carlo Statistical Methods*. (Springer New York, New York, NY, 2004).
123. Brooks, S., Gelman, A., Jones, G. & Meng, X.-L. (eds.) *Handbook of Markov Chain Monte Carlo* (Chapman & Hall/CRC, 2011).
124. Neal, R. M. MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*, 113–162 (2011).
125. Betancourt, M. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434* (2017).
126. Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. Variational inference: a review for statisticians. *J. Am. Stat. Assoc.* **112**, 859–877 (2017).
127. Zhang, C., Butepage, J., Kjellstrom, H. & Mandt, S. Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 2008–2026 (2019).
128. Blundell, C., Cornebise, J., Kavukcuoglu, K. & Wierstra, D. Weight uncertainty in neural networks. In *Proc. 32nd International Conference on Machine Learning*, vol. 37, 1613–1622 (2015).
129. Rezende, D. J. & Mohamed, S. Variational inference with normalizing flows. In *Proc. 32nd International Conference on Machine Learning, ICML 2015*, vol. 2, 1530–1538 (2015).
130. Marzouk, Y., Moselhy, T., Parno, M. & Spantini, A. Sampling via measure transport: an introduction. In *Handbook of Uncertainty Quantification*, 1–41 (Springer International Publishing, 2016).
131. Liu, Q. & Wang, D. Stein variational gradient descent: a general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2378–2386 (Barcelona, Spain, 2016).

132. Detommaso, G., Cui, T., Spantini, A., Marzouk, Y. & Scheichl, R. A Stein variational Newton method. In *Advances in Neural Information Processing Systems*, 9169–9179 (2018).
133. Leviyev, A., Chen, J., Wang, Y., Ghattas, O. & Zimmerman, A. A stochastic Stein Variational Newton method. *arXiv preprint arXiv:2204.09039* 1–17 (2022). 2204.09039.
134. Chen, P. & Ghattas, O. Projected Stein variational gradient descent. *Adv. neural inf. process. syst.* **33**, 1947–1958 (2020).
135. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
136. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proc. 33rd International Conference on Machine Learning, ICML 2016*, vol. 3, 1651–1660 (2016).
137. Kim, S. W., Oh, K. Y. & Lee, S. Novel informed deep learning-based prognostics framework for on-board health monitoring of lithium-ion batteries. *Appl. Energy* **315**, 119011 (2022).
138. Xu, Z., Li, H., Yazdi, M., Ouyang, K. & Peng, W. Aging characteristics and state-of-health estimation of retired batteries: an electrochemical impedance spectroscopy perspective. *Electronics*, **11**, 3863 (2022).
139. Zhu, R., Chen, Y., Peng, W. & Ye, Z. S. Bayesian deep-learning for RUL prediction: an active learning perspective. *Reliability Eng. Syst. Saf.* **228**, 108758 (2022).
140. Hong, J., Lee, D., Jeong, E. R. & Yi, Y. Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning. *Appl. Energy* **278**, 115646 (2020).
141. Zhang, S., Liu, Z. & Su, H. A Bayesian mixture neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Transport. Electr.* **8**, 4708–4721 (2022).
142. Benker, M., Furtner, L., Semm, T. & Zaeh, M. F. Utilizing uncertainty information in remaining useful life estimation via Bayesian neural networks and Hamiltonian Monte Carlo. *J. Manuf. Syst.* **61**, 799–807 (2021).
143. Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* **30**, (2017).
144. Fort, S., Hu, H. & Lakshminarayanan, B. Deep ensembles: a loss landscape perspective. *arXiv preprint arXiv:1912.02757* (2019).
145. Wilson, A. G. & Izmailov, P. Bayesian deep learning and a probabilistic perspective of generalization. *Adv. Neural Inf. Process. Syst.* **33**, 4697–4708 (2020).
146. Ovadia, Y. et al. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. *Adv. Neural Inf. Process. Syst.* **32** (2019).
147. Shen, S., Sadoughi, M., Li, M., Wang, Z. & Hu, C. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Appl. Energy* **260**, 114296 (2020).
148. Saxena, A. et al. Metrics for evaluating performance of prognostic techniques. In *Proc. International Conference on Prognostics and Health Management*, 1–17 (IEEE, 2008).
149. Nemani, V. P., Lu, H., Thelen, A., Hu, C. & Zimmerman, A. T. Ensembles of probabilistic lstm predictors and correctors for bearing prognostics using industrial standards. *Neurocomputing* **491**, 575–596 (2022).
150. Nemani, V., Thelen, A., Hu, C. & Daining, S. Degradation-aware ensemble of diverse predictors for remaining useful life prediction. *J. Mech. Des.* **145**, 031706 (2023).
151. Van Amersfoort, J., Smith, L., Teh, Y. W. & Gal, Y. Uncertainty estimation using a single deep deterministic neural network. In *Proc. International Conference on Machine Learning*, 9690–9700 (PMLR, 2020).
152. Liu, J. et al. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. *Adv. Neural Inf. Process. Syst.* **33**, 7498–7512 (2020).
153. Mukhoti, J., Kirsch, A., van Amersfoort, J., Torr, P. H. & Gal, Y. Deterministic neural networks with inductive biases capture epistemic and aleatoric uncertainty. *arXiv preprint arXiv:2102.11582* (2021).
154. Liao, L. & Köttig, F. A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Appl. Soft Comput.* **44**, 191–199 (2016).
155. Pozzato, G. & Onori, S. Combining physics-based and machine learning methods to accelerate innovation in sustainable transportation and beyond: a control perspective. In *Proc. American Control Conference (ACC)*, 640–653 (IEEE, 2023).
156. Meeker, W. Q., Hahn, G. J. & Escobar, L. A. *Statistical Intervals: a Guide for Practitioners and Researchers*, Vol. 541 (John Wiley & Sons, 2017).
157. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
158. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
159. Li, Y. et al. Random forest regression for online capacity estimation of lithium-ion batteries. *Appl. Energy* **232**, 197–210 (2018).
160. Takahashi, A., Allam, A. & Onori, S. Evaluating the feasibility of batteries for second-life applications using machine learning. *Iscience* **26**, (2023).
161. Liu, D., Xie, W., Liao, H. & Peng, Y. An integrated probabilistic approach to lithium-ion battery remaining useful life estimation. *IEEE Trans. Instrum. Meas.* **64**, 660–670 (2014).
162. Jiao, Z. et al. A lightgbm based framework for lithium-ion battery remaining useful life prediction under driving conditions. *IEEE Trans. Ind. Inform.* (2023).
163. Schmalstieg, J., Käbitz, S., Ecker, M. & Sauer, D. U. A holistic aging model for Li (NiMnCo) O₂ based 18650 lithium-ion batteries. *J. Power Sourc.* **257**, 325–334 (2014).
164. Belt, J., Utgikar, V. & Bloom, I. Calendar and PHEV cycle life aging of high-energy, lithium-ion cells containing blended spinel and layered-oxide cathodes. *J. Power Sourc.* **196**, 10213–10221 (2011).
165. Rumberg, B., Epding, B., Stradtman, I., Schleder, M. & Kwade, A. Holistic calendar aging model parametrization concept for lifetime prediction of graphite/nmc lithium-ion cells. *J. Energy Storage* **30**, 101510 (2020).
166. Smith, K. et al. Life prediction model for grid-connected Li-ion battery energy storage system. In *Proc. American Control Conference (ACC)*, 4062–4068 (IEEE, 2017).
167. Schimpe, M. et al. Comprehensive modeling of temperature-dependent degradation mechanisms in lithium iron phosphate batteries. *J. Electrochem. Soc.* **165**, A181 (2018).
168. Naumann, M., Schimpe, M., Keil, P., Hesse, H. C. & Jossen, A. Analysis and modeling of calendar aging of a commercial LiFePO₄/graphite cell. *J. Energy Storage* **17**, 153–169 (2018).
169. Gasper, P., Collath, N., Hesse, H. C., Jossen, A. & Smith, K. Machine-learning assisted identification of accurate battery lifetime models with uncertainty. *J. Electrochem. Soc.* **169**, 080518 (2022).
170. Gasper, P. et al. Degradation and modeling of large-format commercial lithium-ion cells as a function of chemistry, design, and aging conditions. *J. Energy Storage* **73**, 109042 (2023).
171. Abdar, M. et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf. Fusion* **76**, 243–297 (2021).
172. Der Kiureghian, A. & Ditlevsen, O. Aleatory or epistemic? Does it matter? *Struct. Saf.* **31**, 105–112 (2009).
173. Gal, Y., Hron, J. & Kendall, A. Concrete dropout. *Adv. Neural Inf. Process. Syst.* **30** (2017).
174. Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach. Learn.* **110**, 457–506 (2021).

175. Kendall, A. & Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* **30** (2017).
176. Ha, S., Pozzato, G. & Onori, S. Electrochemical characterization tools for lithium-ion batteries. *J. Solid State Electrochem.* 115646 (2023).
177. Lu, J., Xiong, R., Tian, J., Wang, C. & Sun, F. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nat. Commun.* **14**, 2760 (2023).
178. Shu, X. et al. A flexible state-of-health prediction scheme for lithium-ion battery packs with long short-term memory network and transfer learning. *IEEE Trans. Transport. Electr.* **7**, 2238–2248 (2021).
179. Tan, Y. & Zhao, G. Transfer learning with long short-term memory network for state-of-health prediction of lithium-ion batteries. *IEEE Trans. Ind. Electron.* **67**, 8723–8731 (2019).
180. Ye, Z. & Yu, J. State-of-health estimation for lithium-ion batteries using domain adversarial transfer learning. *IEEE Trans. Power Electron.* **37**, 3528–3543 (2021).
181. Ye, Z., Yu, J. & Mao, L. Multisource domain adaption for health degradation monitoring of lithium-ion batteries. *IEEE Trans. Transport. Electr.* **7**, 2279–2292 (2021).
182. Song, L., Zhang, K., Liang, T., Han, X. & Zhang, Y. Intelligent state of health estimation for lithium-ion battery pack based on big data analysis. *J. Energy Storage* **32**, 101836 (2020).
183. Shi, Y. et al. A real-world investigation into usage patterns of electric vehicles in shanghai. *J. Energy Storage* **32**, 101805 (2020).
184. Qin, Y. et al. Charging patterns analysis and multiscale infrastructure deployment: based on the real trajectories and battery data of the plug-in electric vehicles in shanghai. *J. Cleaner Prod.* **425**, 138847 (2023).
185. Bao, L. et al. Spatiotemporal clustering analysis of shared electric vehicles based on trajectory data for sustainable urban governance. *J. Clean. Prod.* **412**, 137373 (2023).
186. She, C., Wang, Z., Sun, F., Liu, P. & Zhang, L. Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network. *IEEE Trans. Ind. Inform.* **16**, 3345–3354 (2020).
187. Andwari, A. M., Pesiridis, A., Rajoo, S., Martinez-Botas, R. & Esfahanian, V. A review of battery electric vehicle technology and readiness levels. *Renew. Sustain. Energy Rev.* **78**, 414–430 (2017).
188. Deng, Z. et al. Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles. *Appl. Energy* **339**, 120954 (2023).
189. Zhang, Y., Wik, T., Bergström, J., Pecht, M. & Zou, C. A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data. *J. Power Sourc.* **526**, 231110 (2022).
190. Tian, J., Xiong, R., Shen, W. & Sun, F. Electrode ageing estimation and open circuit voltage reconstruction for lithium ion batteries. *Energy Storage Mater.* **37**, 283–295 (2021).
191. Yang, S. et al. A voltage reconstruction model based on partial charging curve for state-of-health estimation of lithium-ion batteries. *J. Energy Storage* **35**, 102271 (2021).
192. Schmitt, J., Rehm, M., Karger, A. & Jossen, A. Capacity and degradation mode estimation for lithium-ion batteries based on partial charging curves at different current rates. *J. Energy Storage* **59**, 106517 (2023).
193. Han, X. et al. A comparative study of commercial lithium ion battery cycle life in electrical vehicle: aging mechanism identification. *J. Power Sourc.* **251**, 38–54 (2014).
194. Costa, N., Sanchez, L., Ansean, D. & Dubarry, M. Li-ion battery degradation modes diagnosis via convolutional neural networks. *J. Energy Storage* **55**, 105558 (2022).
195. Dubarry, M. et al. State of health battery estimator enabling degradation diagnosis: model and algorithm description. *J. Power Sourc.* **360**, 59–69 (2017).
196. Ruan, H., Chen, J., Ai, W. & Wu, B. Generalised diagnostic framework for rapid battery degradation quantification with deep learning. *Energy AI* **9**, 100158 (2022).
197. Prosser, R., Offer, G. & Patel, Y. Lithium-ion diagnostics: the first quantitative in-operando technique for diagnosing lithium ion battery degradation modes under load with realistic thermal boundary conditions. *J. Electrochem. Soc.* **168**, 030532 (2021).
198. Fermín-Cueto, P. et al. Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. *Energy AI* **1**, 100006 (2020).
199. Platt, J. Probabilities for sv machines. In: Smola A. J., Bartlett P. J., Schuurmans D., Schölkopf B., eds. *Advances in Large Margin Classifiers* (1999).
200. Ibraheem, R., Strange, C. & dos Reis, G. Capacity and internal resistance of lithium-ion batteries: Full degradation curve prediction from voltage response at constant current at discharge. *J. Power Sourc.* **556**, 232477 (2023).
201. Rieger, L. H. et al. Uncertainty-aware and explainable machine learning for early prediction of battery degradation trajectory. *Digit. Discov.* **2**, 112–122 (2023).
202. Lui, Y. H. et al. Physics-based prognostics of implantable-grade lithium-ion battery for remaining useful life prediction. *J. Power Sourc.* **485**, 229327 (2021).
203. Honkura, K., Honbo, H., Koishikawa, Y. & Horiba, T. State analysis of lithium-ion batteries using discharge curves. *ECS Trans.* **13**, 61 (2008).
204. Dubarry, M., Truchot, C. & Liaw, B. Y. Synthesize battery degradation modes via a diagnostic and prognostic model. *J. Power Sourc.* **219**, 204–216 (2012).
205. Dahn, H. M., Smith, A., Burns, J., Stevens, D. & Dahn, J. User-friendly differential voltage analysis firmware for the analysis of degradation mechanisms in Li-ion batteries. *J. Electrochem. Soc.* **159**, A1405 (2012).
206. Kohtz, S., Xu, Y., Zheng, Z. & Wang, P. Physics-informed machine learning model for battery state of health prognostics using partial charging segments. *Mech. Syst. Signal Process.* **172**, 109002 (2022).
207. Nascimento, R. G., Corbetta, M., Kulkarni, C. S. & Viana, F. A. Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis. *J. Power Sourc.* **513**, 230526 (2021).
208. Li, W. et al. Physics-informed neural networks for electrode-level state estimation in lithium-ion batteries. *J. Power Sourc.* **506**, 230034 (2021).
209. Tian, J., Xiong, R., Lu, J., Chen, C. & Shen, W. Battery state-of-charge estimation amid dynamic usage with physics-informed deep learning. *Energy Storage Mater.* **50**, 718–729 (2022).
210. Lin, Y.-H., Ruan, S.-J., Chen, Y.-X. & Li, Y.-F. Physics-informed deep learning for lithium-ion battery diagnostics using electrochemical impedance spectroscopy. *Renew. Sustain. Energy Rev.* **188**, 113807 (2023).
211. Shi, J., Rivera, A. & Wu, D. Battery health management using physics-informed machine learning: online degradation modeling and remaining useful life prediction. *Mech. Syst. Signal Process.* **179**, 109347 (2022).
212. Pannala, S., Movahedi, H., Garrick, T. R., Stefanopoulou, A. G. & Siegel, J. B. Consistently Tuned Battery Lifetime Predictive Model of Capacity Loss, Resistance Increase, and Irreversible Thickness Growth. *J. Electrochem. Soc.* **171**, 010532 (2024).
213. Ahmadi, L., Young, S. B., Fowler, M., Fraser, R. A. & Achachlouei, M. A cascaded life cycle: reuse of electric vehicle lithium-ion battery packs in energy storage systems. *Int. J. Life Cycle Assess.* **22**, 111–124 (2017).

214. Standridge, C. R. et al. Remanufacturing, repurposing, and recycling of post-vehicle-application lithium-ion batteries. Technical Report, Mineta National Transit Research Consortium (2014).
215. Hua, Y. et al. Sustainable value chain of retired lithium-ion batteries for electric vehicles. *J. Power Sourc.* **478**, 228753 (2020).
216. Shahjalal, M. et al. A review on second-life of Li-ion batteries: prospects, challenges, and issues. *Energy* **241**, 122881 (2022).
217. Hua, Y. et al. Toward sustainable reuse of retired lithium-ion batteries from electric vehicles. *Resour. Conserv. Recycl.* **168**, 105249 (2021).
218. Han, X. et al. A review on the key issues of the lithium ion battery degradation among the whole life cycle. *ETransportation* **1**, 100005 (2019).
219. Basia, A., Simeu-Abazi, Z., Gascard, E. & Zwolinski, P. Review on state of health estimation methodologies for lithium-ion batteries in the context of circular economy. *CIRP J. Manuf. Sci. Technol.* **32**, 517–528 (2021).
220. Hu, X. et al. A review of second-life lithium-ion batteries for stationary energy storage applications. *Proc. IEEE* **110**, 735–753 (2022).
221. Thelen, A. et al. A comprehensive review of digital twin—part 1: modeling and twinning enabling technologies. *Struct. Multidiscip. Optim.* **65**, 354 (2022).
222. Thelen, A. et al. A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives. *Struct. Multidiscip. Optim.* **66**, 1 (2023).
223. Alliance, G. B. The global battery alliance battery passport: giving an identity to the ev's most important component. *Glob. Batter. Alliance* (2020).
224. Ward, L. et al. Principles of the battery data genome. *Joule* **6**, 2253–2271 (2022).
225. Woody, M., Arbabzadeh, M., Lewis, G. M., Keoleian, G. A. & Stefanopoulou, A. Strategies to limit degradation and maximize li-ion battery service lifetime-critical review and guidance for stakeholders. *J. Energy Storage* **28**, 101231 (2020).
226. Allyn, B. Apple agrees to pay 113 million to settle batterygate case over iPhone slowdowns. *NPR* (2020).
227. Keil, P. et al. Calendar aging of lithium-ion batteries. *J. Electrochem. Soc.* **163**, A1872 (2016).
228. Hoke, A., Brissette, A., Smith, K., Pratt, A. & Maksimovic, D. Accounting for lithium-ion battery degradation in electric vehicle charging optimization. *IEEE J. Emerg. Select. Top. Power Electron.* **2**, 691–700 (2014).
229. Konz, Z. M., Weddle, P. J., Gasper, P., McCloskey, B. D. & Colclasure, A. M. Voltage-based strategies for preventing battery degradation under diverse fast-charging conditions. *ACS Energy Lett.* **8**, 4069–4077 (2023).
230. Lu, H., Thelen, A., Fink, O., Hu, C. & Laflamme, S. Federated learning with uncertainty-based client clustering for fleet-wide fault diagnosis. *Mech. Syst. Signal Process* **210**, 111068 (2024).
231. Keyser, M. et al. Enabling fast charging—battery thermal considerations. *J. Power Sourc.* **367**, 228–236 (2017).
232. Smith, K. & Wang, C.-Y. Power and thermal characterization of a lithium-ion battery pack for hybrid-electric vehicles. *J. Power Sourc.* **160**, 662–673 (2006).
233. Sripad, S., Bills, A. & Viswanathan, V. A review of safety considerations for batteries in aircraft with electric propulsion. *MRS Bull.* **46**, 435–442 (2021).
- acknowledges financial support from the U.S. Department of Energy under Grant No. 680593. S.O. acknowledges partial financial support from the Precourt Institute for Energy at Stanford University. Z.H. acknowledges financial support from the U.S. National Science Foundation under Grant no. CMMI-2301012. C.H. received financial support from the U.S. National Science Foundation under Grant No. ECCS-2015710. The opinions, findings, and conclusions presented in this article are solely those of the authors and do not necessarily reflect the views of the sponsors that provided funding support for this research.

Author contributions

A.T., C.H., and Z.H. devised the original concept for the review paper. A.T. and C.H. were responsible for the introduction, background on battery degradation and SOH estimation, defining the battery diagnostic and prognostic problems, and explaining the differences between traditional ML and deep learning. C.H. was responsible for reviewing the methodologies and applications of GPR, RVM, and neural network ensemble. X.H. was responsible for reviewing and discussing Bayesian neural networks. A.T. was responsible for reviewing sampling methods for uncertainty estimation and the section on understanding and using predictive uncertainty estimates. A.T. was responsible for the discussion on publicly available battery aging datasets. S.O. was responsible for the review and discussion of SOH estimation using field data. N.P. was responsible for the review and discussion on early life and trajectory prediction and contributed to the review of GPR for SOH forecasting and RUL prediction. A.T. was responsible for reviewing future trends and opportunities in physics-based prognostics and aging-aware battery control optimization. Z.H. and C.H. were responsible for reviewing future trends and opportunities in second-life applications of Li-ion batteries. A.T. and C.H. were responsible for concluding remarks. All authors read and approved the final manuscript. All correspondence should be addressed to Chao Hu (chao.hu@uconn.edu).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Chao Hu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Acknowledgements

A.T. received no financial support for this work. X.H. is supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0021397. N.P.