Title
Statistical and machine learning-based durability-testing strategies for energy storage

Journal
Joule, 7(5)

ISSN
2542-4785

Authors
Harris, Stephen J
Noack, Marcus M

Publication Date
2023-05-01

DOI
10.1016/j.joule.2023.03.008

Peer reviewed

Statistical and Machine Learning-Based Durability-Testing Strategies for Energy Storage

Stephen J. Harris[*]

Marcus M. Noack[**]

[*]Energy Storage and Distributed Resources Division

[**] Applied Mathematics and Computational Research Division

Lawrence Berkeley National Lab

Berkeley, CA 94720

ABSTRACT

There is considerable interest in developing new energy storage technologies for the electric grid, but economic viability will require that manufacturers provide warranties guaranteeing 15+ years of life. While there are extensive efforts to make early predictions for the expected life of new storage technologies, we argue here that for purposes of pricing warranties and valuing second life potential—considerations that are crucial to whether the technologies can be commercialized—the full failure probability distribution, not just the expected life, is required. We use published battery cycle-life data to suggest efficient statistical and machine learning-based testing and analysis strategies that can rapidly estimate and also take advantage of the failure probability distribution. One approach is a Weibull analysis, which can (a) reduce the number of testing machine-hours required for setting a warranty, (b) quickly determine whether a new technology is better than a baseline technology, and (c) estimate the maximum intensity of testing acceleration that does not change the failure mode. A second approach is driven by the idea that all measured data — such as capacity or energy as a function of time or cycle number — is valuable and generated by an underlying latent function. The analysis employs a Gaussian process to find the underlying latent function, together with its uncertainties, which can be used to calculate the failure distribution.

INTRODUCTION

Commercial viability for grid storage technologies will require that manufacturers provide warranties guaranteeing performance for thousands of cycles over perhaps 15-20 years[1-5]. Life predictions are difficult because of the statistical nature of failure, typically due to microscopic heterogeneities[6], so extensive testing is required. However, testing energy storage devices requires years and ties up expensive equipment, resulting in a paucity of publicly available data[7-12].

A important recent approach to speed up the prediction of energy storage device lifetimes is the use of machine learning[12-22], which can be combined with the more traditional accelerated testing[23], mechanistic[3], and numerical and physics-based models[24]. Almost always, such efforts aim to predict the expected life (or the expected remaining useful life). However, even a perfect predictive capability for expected life does not help a manufacturer price a warranty. Suppose that a manufacturer's business plan requires that no more than 1% of their devices fail under warranty. This means estimating the number of cycles or years until 1% of the devices fail, a number that cannot be estimated from knowing just the expected life. In fact, the expected lifetime will have to be far longer than 15 years to justify a 15-year warranty. But how much longer?

There is a fundamental asymmetry in the warranty business: manufacturers get no direct benefit from devices that last beyond their warranty period, but they are penalized for those that fail early. Thus, a device with a shorter expected life and a narrow failure distribution may be preferred to one with a longer expected life and a wide failure distribution. Said another way, the left-hand side (LHS) of the failure distribution, not the expected life, is necessary for setting a warranty. For other purposes, such as valuing the potential for second life[25, 26], the RHS of the failure distribution is necessary. In this Perspective we propose that durability research efforts be expanded to include the full failure distribution.

We will use published[10] cycle-life data for 24 nominally identical commercial lithium-ion cells tested under nominally identical conditions in order to evaluate more efficient ways that could have been used to collect and analyze the data required to determine the failure probability density function (pdf). For our purposes, "efficient" means minimizing the required resources, in units of testing machine-hours. We take 2 approaches.

Our first approach uses Weibull analysis[27]. Previous battery studies using Weibull analysis[10, 28-38] have focused almost exclusively on whether and how well the data can be fit to the Weibull function. Our analysis is distinguished from them by using Weibull statistics to demonstrate (a) how the failure pdf can be quickly estimated and (b) ways in which the failure pdf can be put to good economic use. In particular, we will examine the sudden death Weibull protocol, which sacrifices information about the center and the RHS of the failure pdf in order to more rapidly get information about its LHS, necessary for setting a warranty. We will also show how Weibull analysis can be used to quickly determine whether a new technology is better than a baseline technology, and we propose a way to estimate the maximum amount of testing acceleration that can be used without changing the failure mode. We are unaware of other such studies in the energy storage field.

Our second approach is driven by the idea that all measured data is valuable and generated by an underlying latent function. Cell-to-cell variability is interpreted as noise. Thus, we can employ a Gaussian process[39] to find the underlying latent function together with its uncertainties, which will be used to calculate the failure distribution. Stochastic processes have been proposed for failure analysis[40]; we will show how these statistical methods can be applied in the energy storage field.

METHODS

Weibull Analysis

Weibull Function

Since it was introduced[41], 2-parameter Weibull analysis has become the most commonly used technique for analyzing durability data[42]. Theoretically, one reason that the Weibull pdf is so useful is that any situation in which the failure mode can be described by the failure of the weakest link/largest flaw–often the case for materials–can be described by the Weibull distribution. (But see Zok[43].) In part, this is why Weibull analysis can make useful failure forecasts even for small samples with only a few failures[42]; but also in part because it incorporates information about cells that have been tested but did not fail.

The Weibull failure pdf is given by

$$f(t) = (\beta/\eta)(t/\eta)^{\beta-1}e^{-(t/\eta)^{\beta}}$$

where *f(t)* is the fraction of the original cohort that fail on cycle *t*. β is the shape parameter and η is the scale parameter. Integrating, we have

$$F(t) = 1 - e^{-(t/\eta)^{\beta}}$$

where *F(t)* is the cumulative fraction that fails by cycle *t*. Taking the log twice and rearranging, we have

$$\ln(-\ln(1-F(t))) = \beta\ln t - \beta\ln\eta \qquad \text{Equation (1)}$$

so that plotting the LHS of this equation vs ln(*t)* gives a straight-line Weibull plot (if the failure pdf is Weibull), from which β and η can be determined. Weibull analysis only applies to one failure mode at a time; each failure mode has its own Weibull plot. A strongly curved Weibull plot may suggest multiple failure modes, which could be valuable information when trying to determine if accelerated testing is creating new failure modes. Acceleration may be especially important when considering calendar life[44]. However, other causes of curvature are possible.

Two other useful functions include the survival function

$$S(t) = e^{-(t/\eta)^{\beta}}$$

the fraction that survives through cycle *t*; and the hazard function

$$h(t) = f(t)/S(t)$$

the fraction of survivors that fail on cycle *t*.

The shape of the Weibull pdf is determined by the parameter β, and the shape is extremely flexible. For example, $\beta < 1$ gives a pdf that falls faster than exponential and corresponds to infant mortality (almost all failures occur in the first few cycles, so that the hazard decreases with time). $\beta = 1$ gives an exponential pdf and implies a constant hazard (i.e., failure can occur with equal probability at any time, such as by driving over a pothole). $\beta > 1$ implies an increasing hazard, characteristic of wear-out. $1 < \beta < 2.5$ gives an approximately lognormal pdf (e.g., failure is caused by the largest flaw), with smaller values of β in this range corresponding to early wear-out failures, and $\beta = 2$ approximating a Rayleigh distribution, where the hazard increases linearly with time. $\beta = 3.5$ gives a near-Gaussian pdf. (This pdf might occur if there is a critical component or additive, necessary for operation, that gets consumed during use; or if there are many simultaneous failure modes.) Large values of β correspond to rapid, sudden wear-out with narrow distributions (all the devices fail at about the same time). Thus, the value of β depends upon–and gives hints at–the failure mechanism. Typically, and of great importance, when the failure mechanism changes, so does β[27]. Because of the Weibull distribution's flexibility, it is sometimes said that, unlike the typical fitting procedure where the data is fit to a predetermined functional form, Weibull analysis fits the functional form to the data.

Sudden Death

The sudden death Weibull test[45-49], along with several other approaches[50], were developed by manufacturers of bearings[51] and other products that face long and expensive testing procedures for validating new technologies. The main benefit of sudden death is that not all samples are run all the way to failure–in effect sacrificing information about the center and RHS of the failure pdf in order to learn quickly about the LHS. Even a small number of LHS failures may give sufficient information about the failure pdf to set a warranty.

Here we will apply sudden death in hindsight to the cycle-life results of 24 nominally identical commercial cells, 20 of which were tested all the way to failure, while 4 were taken off test without failure at 594 cycles, as we described previously[10]. We will refer to this testing approach as "brute force." These cells had lithium cobalt oxide cathodes and synthetic graphite anodes. They were cycled at room temperature (~25º C) with an Arbin BT2000. On each cycle, the cells were charged in a CCCV (Constant Current Constant Voltage) mode at 1C (4.4 A) constant current up to 4.35V, followed by a constant voltage charge until current dropped below C/40. The cells were then discharged at 10C (44 A) constant current until the terminal voltage decreased to 3 V. Note that a 10C discharge rate accelerated the degradation process, saving considerable testing time while risking that new degradation modes, only relevant at high discharge rates, may well have been introduced. Since the cells were to be used in electric helicopters, where high power is needed for takeoff, the acceleration risk was considered to be worth taking. (Of course, a better way to predict failure quickly is to understand the failure mechanisms well enough so that acceleration is not needed[3].)

A sudden death approach could have been carried out as follows. If we have, for example, 6 testing machines to test the 24 cells, we would choose 6 cells at random and run them simultaneously and identically until one of the cells fails. At this point, all 6 cells in the batch are removed (1 cell has failed, while 5 cells are "suspended."), and a preliminary Weibull analysis is carried out for these 6 cells[52]. The next batch of 6 cells is tested in the same manner, giving an updated Weibull analysis. For 24 cells, this procedure is carried out 4 times, at which point we have 4 failures and 20 suspensions. The cells that failed are likely to have shorter than average lifetimes, while average cells and especially longer-lived cells are less likely to have been tested all the way to failure. Because most or all of our information is about the LHS of the failure pdf, even four failures may provide enough information to set a warranty. Information about the suspended cells consists in knowing that they survived for longer than the shortest-lived cell in their batches, and this information is incorporated into the Weibull analysis[52]. Johnson has shown[50] that the reliability estimated from this sudden death approach is "similar" to that obtained by the brute force approach.

Based on the Weibull parameters obtained, we can estimate the number of cycles until 1%, 2%, and 5% (known as B1, B2, and B5 lives, respectively) of a large group of cells will fail, the information required for setting a warranty. We compare below estimates made with sudden death and brute force approaches.

Gaussian-Process-Driven Lifetime Prediction

One of the main drawbacks of sudden-death testing is that the vast majority of testing data (e.g., current and voltage as a function of cycle number) is ignored; only the failure and suspension cycles are used for the analysis. The proposition in this section is to learn as much as possible from all data about a postulated, data-generating latent function to make informed decisions about battery lifetimes. We assume that the quantity of interest — charge or energy — is obtained by noisy function evaluations (y_i) of an underlying (true but unknown) latent function that maps energy or capacity to a cycle number, i.e. $f(x)$: $X \rightarrow R$. Every new test is interpreted as a new set of function evaluations at fixed cycle numbers $x_i \in X$, $i \in \{1,2,...\}$. We also assume that the joint probability distribution over the latent-function evaluations is normal (Gaussian). In other words, there is a ground-truth (latent) function *f(x)* of energy or capacity vs cycle number, and the measured data values are noisy evaluations of that ground-truth function. With these assumptions in place, we can model the underlying latent model function as the posterior probability density function of a Gaussian (stochastic) process[39] with a domain-informed prior-mean function and a flexible noise model. Other stochastic processes can be and have been employed, most notably the inverse Gaussian process[40, 53, 54]. Other statistical methods that have been proposed to model battery degradation include[15, 18]. For this work, we assume Gaussian noise; however, this is not a requirement for the methodology to work. Measured energy or capacity values are collected as data $D = \{x_i , y_i\}$. This gives us the ability to abstract over cell numbers, testing machines, and testing strategy leading to an agnostic methodology; a stochastic model for battery lifetime is available starting with the very first collected data point and is available throughout the test. The presented approach is agnostic to the schedule of the data acquisition. In other words, the algorithm can be used for equi-temporal data acquisition, random, or any other sample strategy. Theoretically, the method is most successful when maximizing the variety of distances that are being considered between data points, which is achieved if data acquisition happens at random intervals.

We define a normal prior distribution over function values — we use nomenclature N(x |mean, covariance) for a normal distribution over x, given a mean and a covariance —

$$N(\boldsymbol{f} \mid \boldsymbol{m}, \boldsymbol{K})$$

with prior mean $\boldsymbol{m}$ and covariance $\boldsymbol{K}$. The likelihood of the data is given by

$$N(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{V})$$

where $\boldsymbol{V}$ is a diagonal matrix containing the noise for each data point, which is calculated by a parametric noise model. The covariance of the prior is calculated by applying the kernel function k(x1,x2;**h**) to all pairs of data points. Kernel functions are symmetric positive-semidefinite functions[55] that allow us — given a set of hyperparameters **h** — to calculate covariances as a function of positions in the input space — a space of cycle numbers. This is commonly referred to as the kernel trick[33, 55]. Given the kernel, we can extend the prior over points of interest, marginalize over latent function values, and condition on observations to obtain a posterior probability density function

$$\mathcal{N}(\boldsymbol{m} - \boldsymbol{k}^T (\boldsymbol{K}+\boldsymbol{V})^{-1} (\boldsymbol{y}-\boldsymbol{m}), \boldsymbol{k}(\boldsymbol{x_0},\boldsymbol{x_0}) - \boldsymbol{k}^T (\boldsymbol{K}+\boldsymbol{V})^{-1} \boldsymbol{k}),$$

where the first term is the posterior mean, which can be interpreted as a surrogate of the latent function *f*, and the second term is the posterior covariances which provide the associated uncertainties. See Figure 1 for an illustration of the posterior being defined over the input domain.

V is the matrix of the noise. To account for an increasing divergence of function values as testing proceeds for any given cell, we adopt a non-constant (heteroscedastic) parametric (polynomial) noise model

$$\sigma_n{}^2(x) = x h_1 + h_2 .$$

To allow the practitioner to embed their knowledge about the battery and give the algorithms the capability to extrapolate, we also inject a polynomial prior-mean function

$$m(x) = -x^2 \backslash h_3 + 360 ,$$

where the 360 comes directly from the observed data at the beginning of each test, but could also, in practice be defined as an additional hyperparameter. The prior-mean function gives the practitioner the opportunity to inject domain knowledge into the model. Often, the degradation curves exhibit “knees” at unknown locations based on the degradation mechanisms. In that case, the prior-mean function could be defined as a parametric piecewise linear model or a combination of several of those modes. All prior-mean, noise, and kernel hyperparameters are found by maximizing the marginal log-likelihood of the observations.

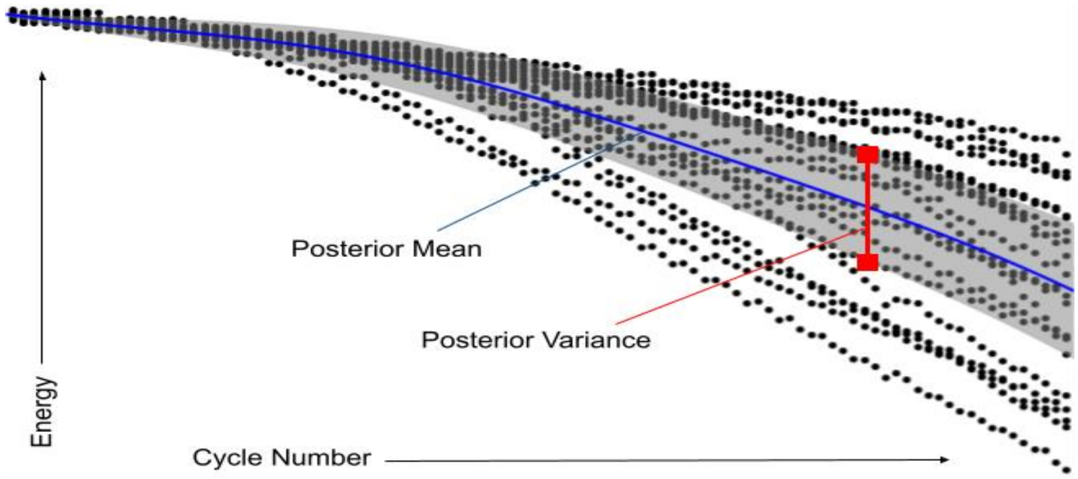

Figure 1 An example of a stochastic model of the underlying latent function cycle number → energy. We use energy instead of capacity as a measure of degradation because energy incorporates changes in both voltage and charge[12]. The data points are displayed as black dots. All data points are used to approximate the data-generating latent function by the GP posterior mean (blue). The grey envelope is the uncertainty associated with the latent function estimate.

The stochastic model can be strategically evaluated along a slice in energy or capacity — depending on the definition of failure, often 80% of the original value — to obtain a failure probability density function. If only the left-hand side of the distribution has to be estimated correctly this approach can be combined with sudden-death strategies. Gaussian processes have been used as a standard tool for design-of-experiment optimization[56-58]. For a combined approach, the spaces (input space for battery chemistry for instance) and the domain of battery degradation (e.g., cycle numbers) could be combined to yield an integrated model. This principle can be extended to include testing protocols (temperature and voltage, for instance).

RESULTS

1. Weibull

Sudden Death Analysis

We are looking retrospectively at failure data for 24 cells[10] that were tested either to failure (20 cells) or to 594 cycles (4 cells), whichever came first. Using all of the data (“brute force”) provides our most accurate estimates for B lives. (The most economical brute force protocol when we have multiple testing machines would be to arrange that every testing machine is operating all of the time, which means immediately replacing any failed cell with a new cell, and running every cell to failure.) Failure was defined as a 20% loss of capacity, but other criteria such as energy loss, or loss in other moments of the voltage discharge curves, could equally well have been used[12].

If we use sudden death and have 2 test machines, we test 2 cells at a time, chosen at random, for 12 batches of 2, and we observe 12 failures and 12 suspensions. With 3 test machines, we test 3 cells at a time for 8 batches, and we observe 8 failures and 16 suspensions; and so on. Thus,

Number of Machines x Number of Batches = total number of cells = 24.

Number of Failures = Number of Batches

With more testing machines, fewer cells are run to failure, reducing the total cost (number of machine-hours) but increasing our uncertainty about the failure pdf. For example, using sudden death with 24 testing machines would mean that we observe only 1 failure, that of the shortest-lived cell. This test would be very quick, but we would have hardly any failure information. Our challenge is to intelligently trade-off cost and uncertainty.

First, we need to recognize that, for any given number of testing machines with sudden death, there can be many different ways to arrange the 24 cells randomly in different batches. For each possible arrangement, different cells will fail or be suspended, which means that the sudden death results depend on the (random) ordering. We therefore used a Monte Carlo approach, carried out as follows. For each of 6 cases—2, 3, 4, 6, 8, and 12 testing machines—we considered possible arrangements for the sudden death tests of the 24 cells. For example, with only 2 testing machines, there are $\left(\frac{24\times 23}{2}\right)$ ways to choose a pair of cells for the first batch, $\left(\frac{22\times 21}{2}\right)$ ways to choose a pair of cells for the second batch, etc. The indices were randomly drawn from a uniform distribution. Since the analysis is done in hindsight, we can calculate the Weibull parameters and the B1, B2, and B5 lives that we would have gotten for each of up to 10,000 arrangements[59]. From the distribution of the calculated B lives we then determined their means and standard deviations. We define the "error" to be the percentage difference between these values and the brute force values, which provide our best estimates. Figure 2 shows the tradeoff between the number of failures (number of batches) and the error.

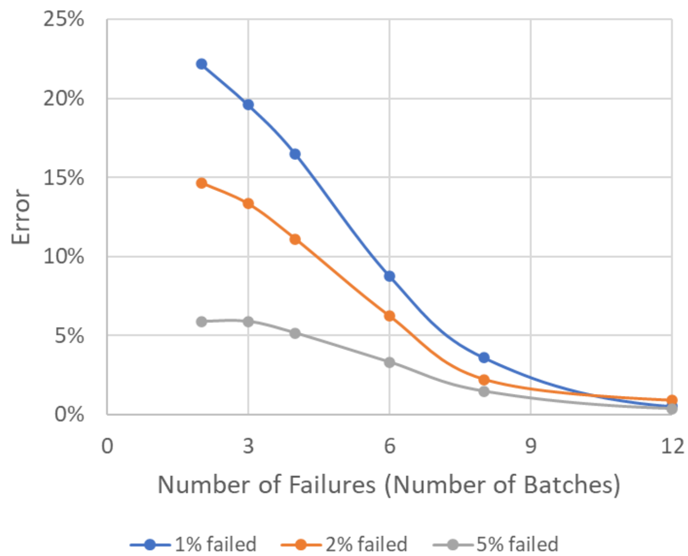

Figure 2. The impact of increasing the number of failures or batches (reducing the number of testing machines) on the error, as defined in the text, in the B1, B2, and B5 lives.

Figure 3 shows the relationship between the number of failures (batches) and the cost, in terms of the number of machine-hours required,

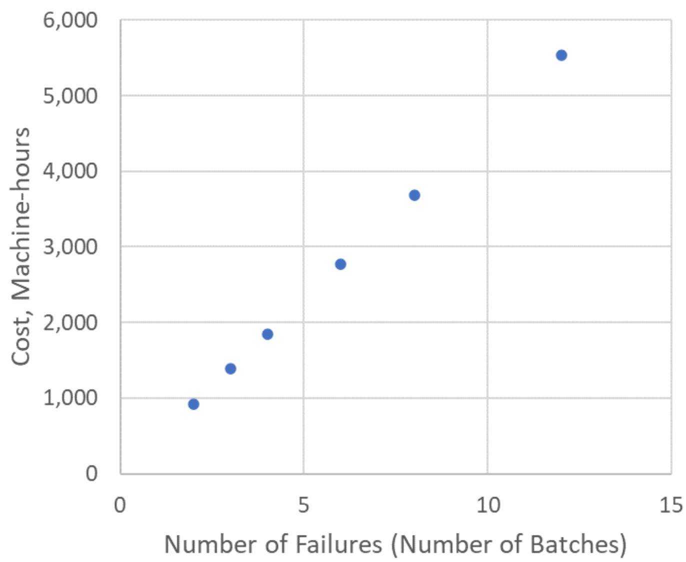

Figure 3. The impact of increasing the number of failures or batches (reducing the number of testing machines) on the cost for testing, in units of machine-hours.

while Fig 4 is a cross-plot that shows the direct tradeoff between cost and error.

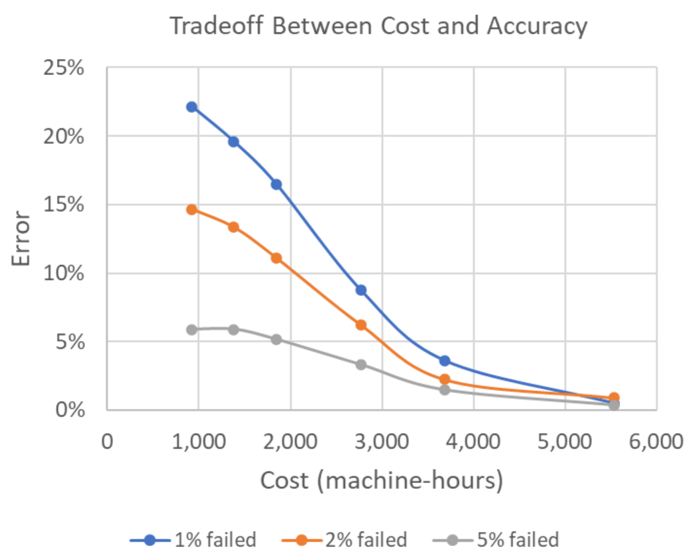

Figure 4. Cross plot showing the tradeoff between cost in machine hours and the expected error in the B1, B2, and B5 lives.

The fractional savings in machine-hours that we achieve by using sudden death compared to the brute force approach is shown in Figure 5, which indicates that the sudden death approach can reduce the cost in machine-hours for this data set by up to

35%, depending on how much uncertainty is tolerated. (Note that if only1 machine is available, then the sudden death and brute force methods are identical.)

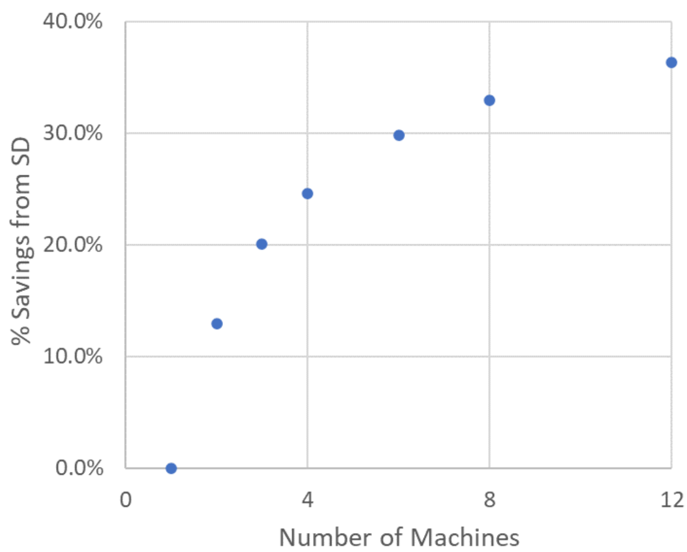

Figure 5. The savings, in units of machine-hours, available from using the sudden death approach.

Weibayes[27, 60]

Suppose that we have made a number of modifications to the battery chemistry or architecture in hopes of increasing B lives. We could repeat the full complement of experiments for each modification, but that might require an unacceptably large number of machine-hours, even using sudden death. Some of the modifications might provide an improvement, but many will not. Rather than trying to get the B lives for each modification, we can ask a question that could be much faster and easier to answer: Can we quickly divide the modified cells into 2 groups, those that are better than the baseline cells and those that are not?

We can answer that question without having to fail a single modified cell if we can assume that β for the modified cells are the same as the known β for the baseline cell. This is roughly equivalent to assuming that the modifications did not change the failure mechanism. The benefit comes because once a value for β is assumed, the Weibull distribution has only 1 parameter, η. This is the Weibayes assumption[27]. (Whether or not this is a good assumption is informed by experience and depends mainly on whether a library for similar systems is available[27].) For example, our brute force Weibull parameters are β = 4.9 and η = 500, with an expected B2 life of 225 cycles. Assume that β = 4.9 for the modified cells. We might decide to call a modified cell "better" if its B2 lifetime has improved by at least 20%, i.e., B2 = 272 cycles. For β = 4.9, this B2 life would correspond to η = 600[52]. For these parameters, if we test 5 modified cells, and there are no failures after 540 cycles, we can claim with 95% confidence that the modified cells are better than the baseline cells (i.e., η is at least 600)[47]. If only a 68% confidence is required, then 440 cycles with no failures will suffice. Rapidly dividing cells into "better than baseline" and "not better than baseline" can save a considerable amount of time, because only cells in the former group need to be considered further. Once the failure pdfs of the improved cells are found, a judgment can be made about whether the assumption of a constant β was justified. An excellent discussion about potential errors associated with the use of Weibayes is available[61].

Accelerated Tests

The idea behind accelerated testing is that by increasing the intensity of the stress (which could be temperature or charging rate, for example) we can induce rapid failures and determine quickly whether one battery type is better than another. Ideally, the performance in an accelerated test can be quantitatively extrapolated to the performance under normal operating conditions. Unfortunately, when a test is accelerated, new failure modes that are not relevant to failure during normal operation may be introduced, in which case accelerated tests may not provide useful information. Weibull analysis may be useful if it can help to determine the maximum amount of acceleration that can be applied without changing the failure mechanism. Experience shows[27, 42] that if the value of β from accelerated and non-accelerated tests are similar, then the failure mode has likely not changed. Thus, by gradually increasing the acceleration intensity, the point where β changes can be identified. To the extent that similar battery types change mechanisms at a similar point, we can test them just below that point and take the maximum advantage of acceleration. Post mortem analysis will be required to validate this proposal, whose value would depend strongly on whether we have a library of failures for different conditions and batteries available. Klein and Basu[62] have published an alternative, much more general, Weibull approach for dealing with accelerated testing with multiple simultaneous failure modes.

Parameter Convergence

Estimates for β and η are obtained after each sudden death failure is obtained, and cumulative estimates are obtained from successive failures. In some cases, this cumulative value may converge before all of the planned tests have been performed. For the example shown in Figure 6 below, taken from ref[10], stopping the tests after 15 failures might have been reasonable. The Gaussian Process discussed below provides a quantitative method for deciding when to stop taking data.

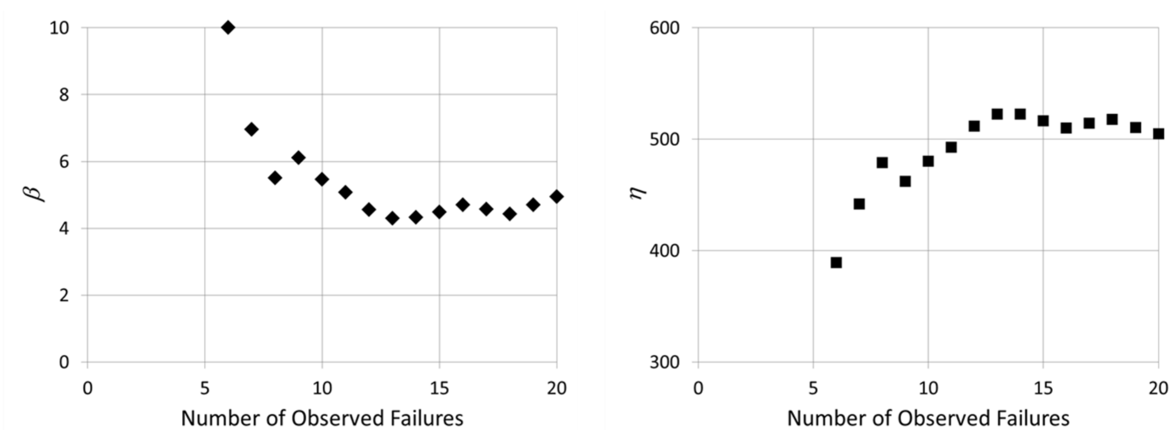

Figure 6. Data from reference[11] showing that the Weibull parameters gradually converge as more failures are observed.

Weibull Analysis: Uncertainties, Caveats, and Comments

We have defined the error introduced by use of sudden death as the percentage discrepancy between the B lives calculated from sudden death and the more accurate B lives calculated from brute force. In this section, we consider additional sources of uncertainty and error. We also provide some caveats with respect to whether or not our analysis might be useful for other types of cells or for cells that are tested under different conditions.

a. The Weibull parameters generated from our “best” data are β = 4.9 and η = 500. But with only 24 tests, we can’t know that these are the parameters for the underlying “true” pdf. How much error is there in the B lives because even our best estimates for β and η come from only 24 cells? We address this question by generating random sets of 24 lifetimes from the above Weibull distribution and calculating the B lives for each set. We find that the standard deviation in the estimated B lives is about 20%. For comparison, if we had tested 100 cells instead of 24, the standard deviation would be about 7%.
b. Each arrangement of the cells in batches gives a different Weibull fit, but only 1 sudden death protocol is run. The standard deviations comparing the B lives among the up to 10,000 Monte Carlo calculations[59] are between 2% and 9%.
c. The Weibull parameters are usually determined either from the slope and intercept of the Weibull plot described in Equation 1 or by maximum likelihood techniques. For the latter case, all possible values of β and η are considered, and for each pair we ask, how likely is it that the observed failures would have been drawn from that pdf? The likelihood is defined as the product of the pdf values at the failure times. In this way, the likelihood is large if most of the failure times are close to the maximum of the *pdf* and few or no failures are observed where the *pdf* is close to zero. There is considerable discussion[27] in the literature about whether the Weibull plot or maximum likelihood yields better values for β and η, but it is straightforward to show that the Weibull plot can give unrealistic estimates for uncertainties in the parameter values, compared to maximum likelihood. As an extreme example, if two samples are run to failure, the Weibull plot has 2 points, and the standard deviations for the slope and intercept are zero, clearly an overly-optimistic result.

Whether or not a Weibull analysis is useful depends on the answers to at least 2 questions:

A. Can the data be fit to a Weibull curve? There are numerous demonstrations in the battery literature that failure data can be fit to a Weibull distribution, even when there is a geometry change[38]. This is not a surprise, since the Weibull distribution is widely used precisely because it so often fits failure data of every sort. To the extent that the data do fit Weibull curves, we have an estimate of the failure pdf, allowing warranties to be priced. The tradeoff that we discussed above between speed and accuracy may depend sensitively on battery type and use case. While we have detailed this tradeoff for our specific cells, the usefulness of the Weibull analysis for a particular case will have to be evaluated based on a library of results.
B. Are there any other benefits to knowing the fitting parameters? Beyond demonstrating the existence of good fits, other fields—gears, bearings, aircraft and automotive components—have taken advantage of specific benefits that may be available once a Weibull fit has been found; for example, the use of Weibayes. Our purpose here has been to propose that analogous benefits might be available in the energy storage field, and for this purpose we have shown how a Weibayes analysis can be used to quickly assess proposed improvements.

We point out that we have not demonstrated a general usefulness of applying a Weibull analysis for energy storage systems, beyond providing the data necessary for setting a warranty. In particular, we have analyzed data from just one set of 24 commercial cells tested under conditions of very fast discharge (10C). We have no evidence for when or whether the value of β that we found would be transferable, for example, to cells tested at different temperatures or duty cycles. Instead, we recommend that a library of β values be generated for different cell types run under different conditions in order to evaluate transferability empirically. Our justification for this recommendation is that it has been found to be useful in many other industries.

A fuller program for finding new cell technologies for grid would search parameter space for improved chemistries and conditions. Such a program would involve both exploitation of existing knowledge (interpolation) and exploration (extrapolation), perhaps using techniques described here[18].

It is not clear to us how to compare the predictive capability of the Weibull analysis described here to the predictive capability of approaches currently taken in the literature. The reason is that, to our knowledge, other prediction approaches aim at predicting the remaining useful life of a particular battery, typically after a small number of cycles. Our Weibull analysis aims at estimating the failure pdf of a large population of batteries. It is a statistical approach and does not concern itself with the performance of an individual battery.

2. Gaussian-Process-Driven Lifetime Analysis

In this test, we are using the same cells we used in the Weibull test above. Since the computational cost of the GP method is much higher, we create only 10 random permutations of the datasets. The testing strategy is similar: we postulate a number of testing machines that can test a number of cells (the batch size) in parallel. In this case, we are postulating 2 testing machines. This would mean 10 test runs for 20 tested cells, for example. From here on we can follow a number of different strategies with decreasing costs but increasing risks:

(1) Comprehensive but costly: After each test run, a GP tries to approximate the underlying latent function and its uncertainties. From here, we can use the uncertainties to decide if more tests have to be conducted. In this scenario, all cells that are tested are tested to failure.

(2) Sudden-Death-Like: Each test run is terminated when the first cell fails. In this case, we will lose out on knowledge that might be gained from the other cells in a test after the failure occurred, but we reduce required resources.

(3) Each test is terminated when the expected information gain for all tested cells in the run reaches a certain threshold. The GP is updated after each test of all cells in the batch.

(4) The GP and therefore the model representation can be updated for every cell separately after every cycle, and individual decisions can be made to stop the testing of a particular cell based on the probability of its performance and potential knowledge gain.

The GP method allows “smart” sudden death in a way that we can estimate the information gain of future points and stop testing when this quantity falls under a certain threshold.

We consider method (1) above and simulate a test of all cells to failure while observing what happens to the predicted failure probability distribution. Figure 7 shows this distribution after a number of tests for 10 different random permutations of the data. It also shows the predicted number of cycles when 5% of cells have failed. In this test, we defined a quadratic prior mean function that can be interpreted as domain knowledge injected into the GP prior. We are also defining a linear noise model. The parameters of the linear noise model (slope and offset) and the prior mean function (a scale), together with other hyperparameters of the kernel function, are found by maximizing the marginal log-likelihood. The key takeaway from Figure 7 is that the distribution of failure pdfs narrows quickly, and confident statements about failure probabilities can be made after only around half of all cells have been tested. In this example, neither the domain awareness (kernel and prior-mean function of the GP) nor the testing methodology is optimized and we still see significant gains compared to standard testing protocols. This makes the case for GP-driven failure probability predictions.

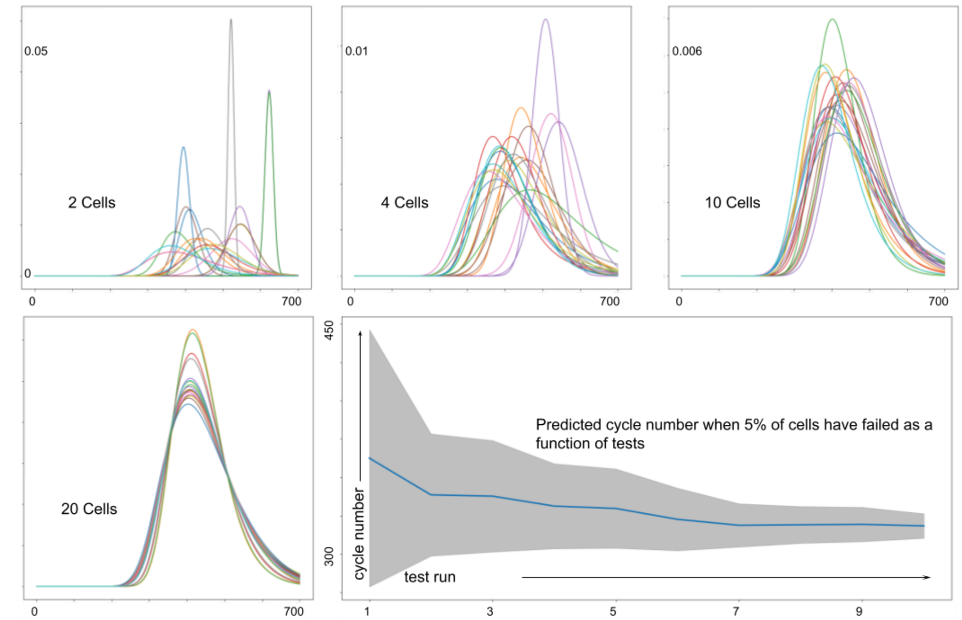

Figure 7. (Top left to bottom right) Failure distributions of 10 random permutations of the data approximated by a Gaussian process after 2, 4, 10, and 20 tested cells, and the associated 5% failure probability as functions of the number of tested cells. The key takeaway is that the failure probability and its standard deviation are rapidly converging after a subset of the cells is tested, demonstrating that it is not necessary to test all cells.

SUMMARY AND CONCLUSIONS

With the explosive growth of electric vehicles well underway, more attention is turning to the need for grid storage. An important barrier to large-scale adoption is the requirement that energy storage devices perform for 15+ years, as discussed in the ROVI Lab Call[6]. Energy storage technologies that are under consideration have rarely if ever been operated for such an extended period under the expected use conditions. If we confine ourselves to predicting the expected life, we do not help a manufacturer aiming to price a 15+ year warranty, considering that the expected life will have to be much longer than 15 years. Similarly, we do not help a utility to estimate the value of a second life when weighing the commercial viability of a specific technology. In this Perspective, we argue that research should broaden beyond the expected life to the full failure pdf.

We have suggested two approaches that reduce the testing resources required to estimate the full failure pdf. One approach assumes that the failure pdf is approximately described by the 2-parameter Weibull function[27]. Using published battery cycling data as an example, we showed that sudden death Weibull could reduce the number of machine-hours required to set a warranty by up to 35%, and we quantified the tradeoff between cost (resource use) and accuracy. We also showed how a version of Weibull analysis (Weibayes) can quickly distinguish new technologies that provide a significant benefit from those that do not. And we suggested how the use of Weibull statistics could optimize accelerated testing by determining the point where the failure mode changes. We plan to test these ideas in future work by analyzing other data sets. The latter two applications (Weibayes and accelerated testing) will require building a library of testing data from many classes of batteries and testing protocols.

The second approach assumes that all data collected during a testing campaign is valuable and created by one underlying latent process with added normally distributed noise. The underlying latent function can then be approximated by the posterior of a Gaussian process (GP) with domain-informed prior mean function and a flexible heteroscedastic noise model.

In the future, we will tailor the GP-driven methodology and combine it with sudden-death strategies to minimize testing time and efforts even further, in the hope of eventually being able to calculate failure distributions with a fraction of the traditionally-collected data.

ACKNOWLEDGMENTS

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

AUTHOR CONTRIBUTIONS

SJH is responsible for the discussion of Weibull statistics. MMN is responsible for the discussion of Gaussian processes.

BIBLIOGRAPHY

1. Gür, T. M., Review of electrical energy storage technologies, materials and systems: challenges and prospects for large-scale grid storage. *Energy & Environmental Science* **2018,** *11* (10), 2696-2767.

2. Yuan, X. Z.; Song, C.; Platt, A.; Zhao, N.; Wang, H.; Li, H.; Fatih, K.; Jang, D., A review of all-vanadium redox flow battery durability: Degradation mechanisms and mitigation strategies. *International Journal of Energy Research* **2019,** *43* (13), 6599-6638.

3. Baure, G.; Dubarry, M., Battery durability and reliability under electric utility grid operations: 20-year forecast under different grid applications. *Journal of energy storage* **2020,** *29*, 101391.

4. Dubarry, M.; Devie, A.; Stein, K.; Tun, M.; Matsuura, M.; Rocheleau, R., Battery Energy Storage System battery durability and reliability under electric utility grid operations: Analysis of 3 years of real usage. *Journal of Power Sources* **2017,** *338*, 65-73.

5. ROVI Rapid Operational Validation Initiative. https://netl.doe.gov/sites/default/files/2022-08/Lab%20Call%20-%20ROVI%20for%20Flow%20Batteries%208-30%20Final.pdf.

6. Harris, S. J.; Lu, P., Effects of Inhomogeneities--Nanoscale to Mesoscale--on the Durability of Li-Ion Batteries. *The Journal of Physical Chemistry C* **2013,** *117* (13), 6481-6492.

7. Dos Reis, G.; Strange, C.; Yadav, M.; Li, S., Lithium-ion battery data and where to find it. *Energy and AI* **2021,** *5*, 100081.

8. Baure, G.; Dubarry, M., Synthetic vs. real driving cycles: A comparison of electric vehicle battery degradation. *Batteries* **2019,** *5* (2), 42.

9. Devie, A.; Baure, G.; Dubarry, M., Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells. *Energies* **2018,** *11* (5), 1031.

10. Harris, S. J.; Harris, D. J.; Li, C., Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells. *Journal of Power Sources* **2017,** *342*, 589-597.

11. Baumhöfer, T.; Brühl, M.; Rothgang, S.; Sauer, D. U., Production caused variation in capacity aging trend and correlation to initial cell performance. *Journal of Power Sources* **2014,** *247*, 332-338.

12. Severson, K. A.; Attia, P. M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M. H.; Aykol, M.; Herring, P. K.; Fraggedakis, D., Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* **2019,** *4* (5), 383-391.

13. Ward, L.; Babinec, S.; Dufek, E. J.; Howey, D. A.; Viswanathan, V.; Aykol, M.; Beck, D. A.; Blaiszik, B.; Chen, B.-R.; Crabtree, G., Principles of the battery data genome. *Joule* **2022**.

14. Gasper, P.; Gering, K.; Dufek, E.; Smith, K., Challenging practices of algebraic battery life models through statistical validation and model identification via machine-learning. *Journal of the Electrochemical Society* **2021,** *168* (2), 020502.

15. Jiang, B.; Gent, W. E.; Mohr, F.; Das, S.; Berliner, M. D.; Forsuelo, M.; Zhao, H.; Attia, P. M.; Grover, A.; Herring, P. K., Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols. *Joule* **2021,** *5* (12), 3187-3203.

16. Berecibar, M., Machine-learning techniques used to accurately predict battery life. Nature Publishing Group: 2019.

17. Fei, Z.; Yang, F.; Tsui, K.-L.; Li, L.; Zhang, Z., Early prediction of battery lifetime via a machine learning based framework. *Energy* **2021,** *225*, 120205.

18. Attia, P. M.; Grover, A.; Jin, N.; Severson, K. A.; Markov, T. M.; Liao, Y.-H.; Chen, M. H.; Cheong, B.; Perkins, N.; Yang, Z., Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **2020,** *578* (7795), 397-402.

19. Cheng, D.; Sha, W.; Wang, L.; Tang, S.; Ma, A.; Chen, Y.; Wang, H.; Lou, P.; Lu, S.; Cao, Y.-C., Solid-State Lithium Battery Cycle Life Prediction Using Machine Learning. *Applied Sciences* **2021,** *11* (10), 4671.
20. Xing, Y.; Williard, N.; Tsui, K.-L.; Pecht, M., A comparative review of prognostics-based reliability methods for lithium batteries. *2011 Prognostics and System Health Managment Confernece* **2011**, 1-6.
21. Olivares, B. E.; Munoz, M. A. C.; Orchard, M. E.; Silva, J. F., Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena. *IEEE Transactions on Instrumentation and Measurement* **2012,** *62* (2), 364-376.
22. Aykol, M.; Gopal, C. B.; Anapolsky, A.; Herring, P. K.; van Vlijmen, B.; Berliner, M. D.; Bazant, M. Z.; Braatz, R. D.; Chueh, W. C.; Storey, B. D., Perspective—combining physics and machine learning to predict battery lifetime. *Journal of The Electrochemical Society* **2021,** *168* (3), 030525.
23. Palacín, M. R., Understanding ageing in Li-ion batteries: a chemical issue. *Chemical Society Reviews* **2018,** *47* (13), 4924-4933.
24. Hu, X.; Xu, L.; Lin, X.; Pecht, M., Battery lifetime prognostics. *Joule* **2020,** *4* (2), 310-346.
25. Martinez-Laserna, E.; Gandiaga, I.; Sarasketa-Zabala, E.; Badeda, J.; Stroe, D.-I.; Swierczynski, M.; Goikoetxea, A., Battery second life: Hype, hope or reality? A critical review of the state of the art. *Renewable and Sustainable Energy Reviews* **2018,** *93*, 701-718.
26. Hossain, E.; Murtaugh, D.; Mody, J.; Faruque, H. M. R.; Sunny, M. S. H.; Mohammad, N., A comprehensive review on second-life batteries: Current state, manufacturing considerations, applications, impacts, barriers & potential solutions, business strategies, and policies. *Ieee Access* **2019,** *7*, 73215-73252.
27. Abernathy, R., *The New Weibull Handbook*. Nancy S. Embury: 2004.
28. Xiong, J.; Wang, S.; Li, X.; Yang, Z.; Zhang, J.; Yan, C.; Tang, A., Mechanical behavior and Weibull statistics based failure analysis of vanadium flow battery stacks. *Journal of Power Sources* **2019,** *412*, 272-281.
29. Jiang, Y.; Jiang, J.; Zhang, C.; Zhang, W.; Gao, Y.; Guo, Q., Recognition of battery aging variations for LiFePO4 batteries in 2nd use applications combining incremental capacity analysis and statistical approaches. *Journal of Power Sources* **2017,** *360*, 180-188.
30. Eom, S.-W.; Kim, M.-K.; Kim, I.-J.; Moon, S.-I.; Sun, Y.-K.; Kim, H.-S., Life prediction and reliability assessment of lithium secondary batteries. *Journal of Power Sources* **2007,** *174* (2), 954-958.
31. Ganjeizadeh, F.; Tapananon, T.; Lei, H.; Bay, C. E.; Hayward, C., Predicting Reliability of Lithium Ion Batteries. *IJERT, Issn* **2014**, 2278-0181.
32. Ossai, C. I.; Raghavan, N., Statistical characterization of the state-of-health of lithium-ion batteries with Weibull distribution function—A consideration of random effect model in charge capacity decay estimation. *batteries* **2017,** *3* (4), 32.
33. Schuster, S. F.; Brand, M. J.; Berg, P.; Gleissenberger, M.; Jossen, A., Lithium-ion cell-to-cell variation during battery electric vehicle operation. *Journal of Power Sources* **2015,** *297*, 242-251.
34. Mouais, T.; Kittaneh, O. A.; Majid, M., Choosing the best lifetime model for commercial lithium-ion batteries. *Journal of Energy Storage* **2021,** *41*, 102827.
35. Zhao, Q.; Jiang, H.; Chen, B.; Wang, C.; Xu, S.; Zhu, J.; Chang, L., Research on State of Health for the Series Battery Module Based on the Weibull Distribution. *Journal of The Electrochemical Society* **2022,** *169* (2), 020523.
36. Chiodo, E.; Lauria, D.; Andrenacci, N.; Pede, G. In *Accelerated life tests of complete lithium-ion battery systems for battery life statistics assessment*, 2016 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), IEEE: 2016; pp 1073-1078.
37. Li, R.; Wu, J.-f.; Wang, H.-y.; Guo, J.-y.; Li, G.-c., Reliability assessment and failure analysis of lithium iron phosphate batteries. *Information Sciences* **2014,** *259*, 359-368.

38. McConohy, G.; Xu, X.; Cui, T.; Barks, E.; Wang, S.; Kaeli, E.; Melamed, C.; Gu, X. W.; Chueh, W. C., Mechanical regulation of lithium intrusion probability in garnet solid electrolytes. *Nature Energy* **2023**, 1-10.
39. Williams, C. K.; Rasmussen, C. E., *Gaussian processes for machine learning*. MIT press Cambridge, MA: 2006; Vol. 2.
40. Peng, C.-Y., Inverse Gaussian processes with random effects and explanatory variables for degradation data. *Technometrics* **2015,** *57* (1), 100-111.
41. Weibull, W., A statistical distribution function of wide applicability. *Journal of applied mechanics* **1951**.
42. Rinne, H., *The Weibull distribution: a handbook*. Chapman and Hall/CRC: 2008.
43. Zok, F. W., On weakest link theory and Weibull statistics. *Journal of the American Ceramic Society* **2017,** *100* (4), 1265-1268.
44. Dufek, E. J.; Tanim, T. R.; Chen, B.-R.; Kim, S., Battery calendar aging and machine learning. *Joule* **2022,** *6* (7), 1363-1367.
45. Pascual, F. G.; Meeker, W. Q., The modified sudden death test: planning life tests with a limited number of test positions. *Planning fatigue experiments and analyzing fatigue data with the random fatigue-limit model and modified sudden death tests* **1997**, 58.
46. Vlcek, B. L.; Hendricks, R. C.; Zaretsky, E. V., Monte Carlo simulation of sudden death bearing testing. *Tribology Transactions* **2004,** *47* (2), 188-199.
47. Jun, C.-H.; Balamurali, S.; Lee, S.-H., Variables sampling plans for Weibull distributed lifetimes under sudden death testing. *IEEE Transactions on Reliability* **2006,** *55* (1), 53-58.
48. Breneman, J. E.; Sahay, C.; Lewis, E. E., *Introduction to reliability engineering*. John Wiley & Sons: 2022.
49. Suzuki, K.; Ohtsuka, K.; Ashitate, M., On a Comparison Between Sudden Death Life Testing and Type II Number Fixed Life Testing-The Precisions and the Testing Times Using the Maximum Likelihood Estimators.". *Journal of the Japanese Society for Quality Control* **1992,** *22*, 5-12.
50. Johnson, L. G., The statistical treatment of fatigue experiments. *ELSEVIER* **1964**.
51. Vlcek, B. L.; Zaretsky, E. V., Rolling-element fatigue testing and data analysis—A tutorial. *Tribology Transactions* **2011,** *54* (4), 523-541.
52. Berner, J. The Weibull Dr. https://applicationsresearch.com/WeibullEase.htm.
53. Fang, G.; Pan, R.; Wang, Y., Inverse Gaussian processes with correlated random effects for multivariate degradation modeling. *European Journal of Operational Research* **2022,** *300* (3), 1177-1193.
54. Zhang, Z.; Shah, A.; Xing, W. W., Battery Degradation Long-term Forecast Using Gaussian Process Dynamical Models and Knowledge Transfer. *arXiv preprint arXiv:2212.01609* **2022**.
55. Noack, M. M.; Sethian, J. A., Advanced stationary and non-stationary Kernel designs for domain-aware Gaussian processes. *arXiv preprint arXiv:2102.03432* **2021**.
56. Noack, M. M.; Doerk, G. S.; Li, R.; Streit, J. K.; Vaia, R. A.; Yager, K. G.; Fukuto, M., Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. *Scientific reports* **2020,** *10* (1), 1-16.
57. Noack, M. M.; Yager, K. G.; Fukuto, M.; Doerk, G. S.; Li, R.; Sethian, J. A., A kriging-based approach to autonomous experimentation with applications to x-ray scattering. *Scientific reports* **2019,** *9* (1), 1-19.
58. Stach, E.; DeCost, B.; Kusne, A. G.; Hattrick-Simpers, J.; Brown, K. A.; Reyes, K. G.; Schrier, J.; Billinge, S.; Buonassisi, T.; Foster, I., Autonomous experimentation systems for materials development: A community perspective. *Matter* **2021,** *4* (9), 2702-2726.
59. Tevetoglu, T. predictr - Predict the Reliability. https://github.com/tvtoglu/predictr.
60. Ojile, J. O.; Teixeira, J. A.; Carmody, C., Mechanical seal failure analysis. *Tribology transactions* **2010,** *53* (4), 630-635.

61. Nicholls, D.; Lein, P. In *Weibayes testing: What is the impact if assumed beta is incorrect?*, 2009 Annual Reliability and Maintainability Symposium, IEEE: 2009; pp 37-42.
62. Klein, J. P.; Basu, A. P., Weibull accelerated life tests when there are competing causes of failure. *Communications in Statistics-Theory and Methods* **1981,** *10* (20), 2073-2100.