Title
Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells

Authors
Harris, Stephen J
Harris, David J
Li, Chen

Publication Date
2017-02-01

DOI
10.1016/j.jpowsour.2016.12.083

Peer reviewed

FAILURE STATISTICS FOR COMMERCIAL LITHIUM ION BATTERIES: A STUDY OF 24 POUCH CELLS

Stephen J. Harris[1], Chen Li[2], David J. Harris[3]

Abstract

There are relatively few publications that assess capacity decline in enough commercial cells to quantify cell-to-cell variation, but those that do show a surprisingly wide variability. Capacity curves cross each other often, a challenge for efforts to measure the state of health and predict the remaining useful life (RUL) of individual cells. We analyze capacity fade statistics for 24 commercial pouch cells, aiming to provide an estimate for the time to 5% failure. Our data indicate that RUL predictions based on remaining capacity or internal resistance are accurate only once the cells have already sorted themselves into "better" and "worse" ones. Analysis of our failure data with normal and with 2- and 3-parameter Weibull probability density functions provide uniformly good fits using a variety of definitions of failure, but we argue against using a 3-parameter Weibull function for our data. *pdf* fitting parameters appear to converge after about 15 failures, allowing failure times and confidence intervals to be estimated from a modest number of tests. We suggest that testing should continue until predesignated confidence intervals for failure have been achieved, rather than using a predesignated number of cycles or a predesignated number of failures. Increased efforts to make batteries with more consistent lifetimes should lead to improvements in battery cost and safety.

---

[1]Materials Science Division, Lawrence Berkeley National Lab, Berkeley, CA; email: sjharris@lbl.gov; phone: 248-245-1830

[2]ZeeAero, Mountain View, CA.

[3]University of Florida, Gainesville, FL.

Introduction

Over the past couple of decades, a number of "next generation" Li-ion batteries and "beyond" Li-ion batteries have been demonstrated in laboratory settings[1-2], but few have been commercialized. For the most part, the critical barrier to their practical application has been durability. There is, therefore, a paramount need to understand and improve battery life[3].

As part of efforts to predict and improve battery life, a large number of models have been proposed[1, 4-10]. These models fall into various categories, such as physics-based[11], data-fitting/pattern recognition[12-15], or hybrids[13], and they have been extremely valuable for identifying important factors—temperature, charging rate, SOC window, etc.—that control life. However, these models, similar to models that predict the life expectancy of a newborn, are at best rough guides to the actual durability of any given cell. This is because failure is generally a statistical process[16-18], where life is determined by "hidden" variables over which we typically exercise little control[19-21]. In batteries these variables could be microscale, such as heterogeneities within particles[22-24]; or mesoscale, such as variation in local porosity[17] or state of charge[25]; or macroscale, such as the location in a pouch[26] or cell-to-cell variation in the time-temperature history in a pack[9, 27].

Although the importance of identifying and controlling such variability is well-recognized[28-29], the lithium battery durability literature sometimes treats failure as deterministic, with an implicit suggestion that variability could be limited if only the macroscopic battery parameters were tightly enough constrained. It is therefore common to see only one or two, replications—or more commonly, no replications—of capacity fade data. Even in cases where variability is explicitly acknowledged, the analyses often rely on experimental data involving only a handful of nominally identical replications[29-34], so that the statistical confidence may not be high. We note in this regard that disagreement in the literature over, for example, whether or not a particular additive increases battery life could be due to insufficiently large samples being studied.

In contrast to the dearth of life variability data with multiple replications in the open literature, there is significant research on how such data should be treated statistically, when it exists[35-40]. In large part this literature focuses on state of health (SOH) and state of available power (SoAP)[41] monitoring and on predicting the remaining useful life (RUL) of a battery or pack[6], either as a single RUL value or as a distribution of predicted RUL values. These analyses often presume the existence of a library of previously measured (offline) data sets from which can be derived correlations between the RUL and the evolution of some operating parameter(s), such as impedance, rate of capacity loss, or open circuit voltage[42]. An important question, then, is whether these libraries involve large enough samples so that confidence in the RUL predictions can be high. We note, for example, that there have been nearly a hundred citations to a data file[33] that has only 3 or 4 replications for each cycling condition.

In one of a relatively few[14-15, 27, 43-45] published studies looking at a large number of individually tested cells, Baumhofer et al[14] analyzed the capacity fade of 48 commercially available carbon-NMC Sanyo/Panasonic UR18650E cylindrical cells. The cells were graded into group C by the manufacturer and were all drawn from the same production lot. Care was taken to ensure that the cells were cycled under identical electrical, mechanical, and temperature conditions. Their results are striking: after an initial several hundred cycles where the cells performed very similarly, the spread in capacity fade became quite wide so that, for example, cells lost 20% of their original capacity in a range between about 900 and 1300 cycles, Figure 1, reprinted here with permission[46]. Of course, batteries that operate in the field, where conditions are not so well controlled, may have still broader degradation rates.[47]

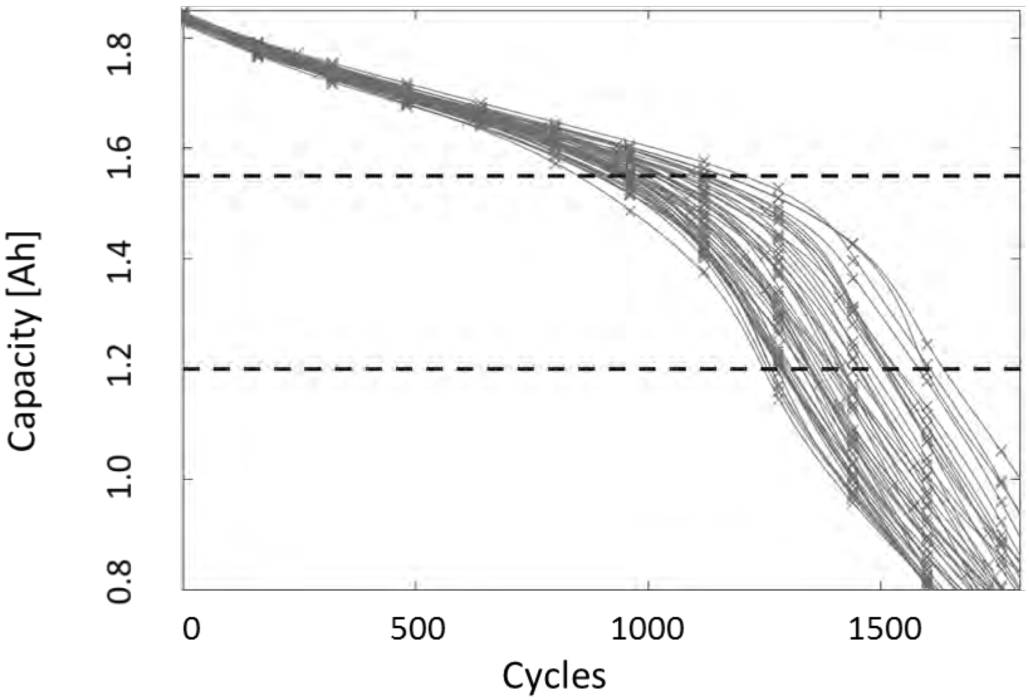

Figure 1. Durability data of Baumhofer et al[14] from 48 nominally identical Panasonic 18650 batteries. The x symbols mark experimental data. The lines are cubic splines connecting the data points. Reproduced with permission[46].

A close inspection of the Baumhofer data shows that the capacity *vs* cycle curves cross each other dozens of times and that, as the authors report, there is little or no relationship between the relative performance of any given cell early and late in the testing. These experimental results imply, at the least, that capacity alone, or capacity derivatives, cannot be readily used to predict the RUL of these cells based on capacities at low numbers of cycles, even for these standardized commercial cells run under idealized conditions. The authors then investigated making predictions of RUL based on 385 "valid" and "significant" cell measurements taken during cycling, from which they selected the 24 (impedance and pulse cell resistance) measurements that best correlated the experimental outcomes. After training, the model's predictions were quite good. Unfortunately, as they point out, there seemed to be no logical reason why these particular 24 measurements should have been chosen, suggesting to the authors that these same parameters might not be particularly informative for another set of batteries or even for these same batteries run under different conditions. Although their work suggests that great care should be taken in assessing the confidence that can be placed in RUL predictions for individual commercial cells, we hasten to point out that making commercial batteries with extremely consistent capacity fade over many years can be accomplished under some conditions[37].

In this work we measure the durability of 24 nominally identical commercial pouch cells. Rather than using the data to try to predict RUL, we focus on the sort of statistical measures that could provide value to battery manufacturers in setting their warranties. For example, we might ask, How confident can we be that 95% of the cells will still have adequate capacity after, say, 250 cycles? We will provide an answer to this question near the end of the paper. This sort of analysis could also be used to address other questions, such as, How many units must be tested for how long, without any failures, to verify that an old failure mode has been eliminated or significantly improved.

There is a reason that some commercial cells last so much longer than others that are nominally identical, but the explanation is obscure. Ultimately, we believe that an increased emphasis on making batteries whose lifetimes are more reproducible—requiring a deeper understanding the variables that make some batteries better than others—will lead to improvements in battery cost and safety.

Experimental

The commercial high power lithium ion cells tested in this work are pouches with a nominal capacity of 4.4 Ah. The active material of the anodes and cathodes are synthetic graphite and LCO (Lithium Cobalt Oxide), respectively. Power capability of these cells is shown in Appendix 1 in Supplemental Information.

24 cells selected randomly from a single batch of cells were cycled at room temperature (~ 25 degrees 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 was employed for these cycle life tests in order to accelerate the degradation process. All cells were compressed in custom-designed cell fixtures during the cycling tests. Special attention was paid to ensure contact resistances between the cell tabs and the fixture leads were small enough so that the high-rate discharge was not affected by any external factors. Both capacity and internal resistance (~ 1 kHz) were measured after each cycle.

Calculations were performed with the Weibull-DR code[48] and with custom *R* code[49]. The raw data from the tests is provided in the Supplemental Information.

Results and Analysis

Results for our 24 cells up to 593 cycles are shown in Figure 2a. (Internal resistance measurements generally follow capacity measurements.) The most striking features of this graph are that (1) the cells perform very nearly identically up to about 150 cycles, after which they diverge markedly; and (2) capacity retention early in the testing says almost nothing about its retention later on, as shown in Figure 2b. Others[14-15, 43-44] have observed these same features. (We point out, however, that by post-selecting just the best-performing 26 cells out of 43 samples tested, Wang et al[50] found relatively smooth behavior.)

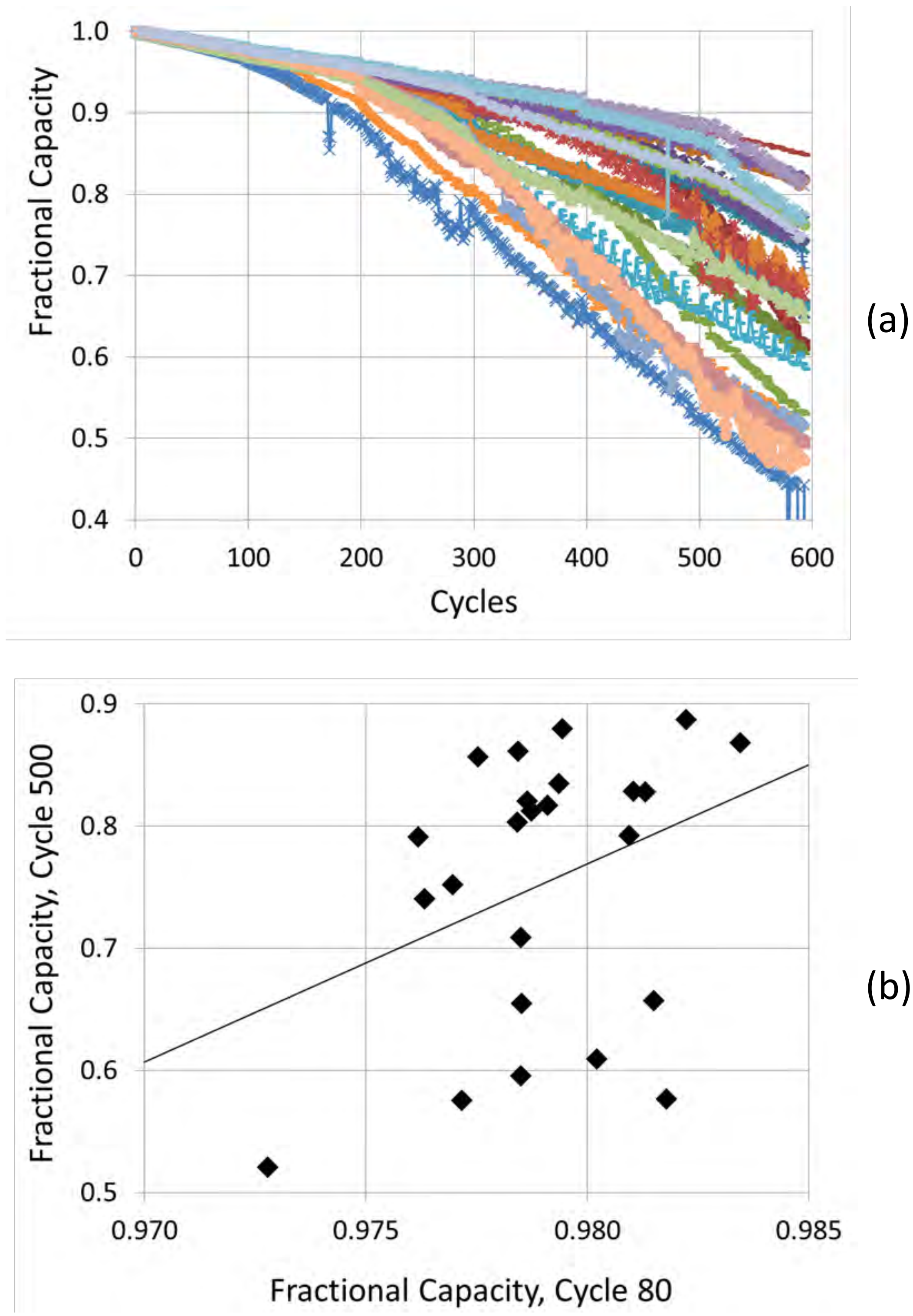

Figure 2. (a) Experimental capacity curves. (b) Correlation between fractional remaining capacity at 80 and 500 cycles. The correlation coefficient for the least squares line is about 0.1.

Statistical Analysis

A statistical analysis of failure[51] can begin by defining four time-dependent (or cycle number-dependent) functions. They are

(1) *f(t)*, the failure probability density function *pdf,* which is the fraction of all of the cells in a population that fail on cycle *t*. (We are assuming that there enough cycles so that the density function is approximately continuous.)

(2) $F(t) = \int_0^t f(t')\, dt'$, the cumulative fraction of cells that have failed by cycle *t*.

(3) *S(t)* = 1 – *F(t),* the survival function, the fraction of cells that survive at cycle *t.*

(4) *h(t)* = *f(t)/S(t),* the hazard function, the fraction of survivors that fail on cycle *t*.

As a simple example, *f(100)* could be the fraction of people who die during their 100'th year, perhaps 0.25%. *F(100)* is the fraction of people who have died by the age of 100, perhaps 99%. *S(100)* is the fraction of people who survive to age 100, 1%. Then *h(100)*, the hazard, is 25%, the fraction of people who survive to age 100 but then die during their 100'th year. (It is quite hazardous to live to a very old age.)

A commonly used failure *pdf* is the normal distribution,

$$f(t) = \frac{1}{\sqrt{2\pi\,\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

where μ is the mean (and median and mode) of the distribution, and σ is its standard deviation. A normal failure *pdf* might be expected, for example, if the cell durability is greatly enhanced by the presence of an additive (whose initial concentration is Gaussian distributed) that is gradually consumed during operation, so that when the additive runs out the battery fails.

Another common functional form taken for *f* in the failure literature is the Weibull function[8, 43-44],

$$\mathbf{S}(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}$$

$$\mathbf{F}(t) = 1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}$$

$$f(t) = \left(\frac{\beta}{\eta}\right)\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}$$

where β is the shape parameter or the Weibull slope, η is the scale parameter, and γ is the location parameter. Similar to a normal analysis, a potentially useful feature of a Weibull analysis is that there may be mechanistic information that can be inferred from the fit. For example, if $\beta = 1$, f reduces to exponential decay, corresponding to a hazard function that is independent of the age of the sample, valid for random failures such as failure caused by driving over a pothole. If $\beta < 1$, f has a rapid initial fall, and the hazard function falls with time, characteristic of infant mortality. For $\beta > 1$, the hazard function rises with time, indicating wear-out. If $\beta \approx 2$, f has a tail to the right and so looks lognormal) (weakest link mechanism), while for $\beta \approx 3.6$, f is nearly normal. In comparing Weibull and normal distributions, we notice that the Weibull distribution has 3 parameters, while the normal distribution has 2 parameters. However, in many analyses γ is set to 0, so that the Weibull distribution also has 2 parameters. Bazant has pointed out[52] that a 3-parameter Weibull has no mathematical basis; only the 2 parameter Weibull comes from the limiting behavior of the smallest of a set of independent, identically distributed random variables ("weakest link") with bounded power law tails[53-55]. Nevertheless, a 3-parameter Weibull function is sometimes used for purposes of improving an empirical fit[8, 21].

We next need to define failure. For many systems, the definition of failure is obvious: the person dies, the car won't move, the river overflows its banks, the part fractures. In this sense, battery failure is different if we want to use capacity to define failure, since its definition is arbitrary[44]. For the moment, we will define a failed battery as one that has lost 20% of its initial capacity, the criterion generally used for electric vehicle batteries. With this definition, our data show that 20 of 24 cells failed during our 593 cycle test, Figure 2a. We refer to the remaining four tests as “suspended”. (The existence of these suspended tests has a significant impact on the analysis.)

A Weibull analysis might involve using median ranks[56] (MR) to estimate the values of F at the measured failure times, followed by an ordinary least squares (OLS) analysis of a linearized version of the Weibull function to determine β and η[21],

$$
\begin{aligned}
&1 - F(t) = e^{-(t/\eta)^{\beta}} \\
&-\ln\left(1 - F(t)\right) = (t/\eta)^{\beta} \\
&\ln\left(-\ln\left(1 - F(t)\right)\right) = \beta \ln t - \beta \ln \eta
\end{aligned}
$$

This approach is referred to as Median Rank Regression (MRR). The 20 experimental failure times and associated median rank estimates[57] for $F(t)$ are shown in Table 1; the 4 suspended tests are indicated at the end. The MRR technique gives excellent fits to the data, with apparently tight constraints on the fitting parameters. For example, using an OLS analysis, we find $\beta = 2.3 \pm 0.15$ (2σ) with an R^2 correlation coefficient of 0.98 for the 3-Weibull fit, Figure 3a. For the 2-Weibull fit we find $\beta = 4.6 \pm 0.4$ (2σ) with an R^2 correlation coefficient of 0.96, Figure 3b. Note that this and all of our analyses include data for all of the samples, even those that seem to be outliers for particular cases. We choose to include outliers because our goal is to describe the full distribution of outcomes..

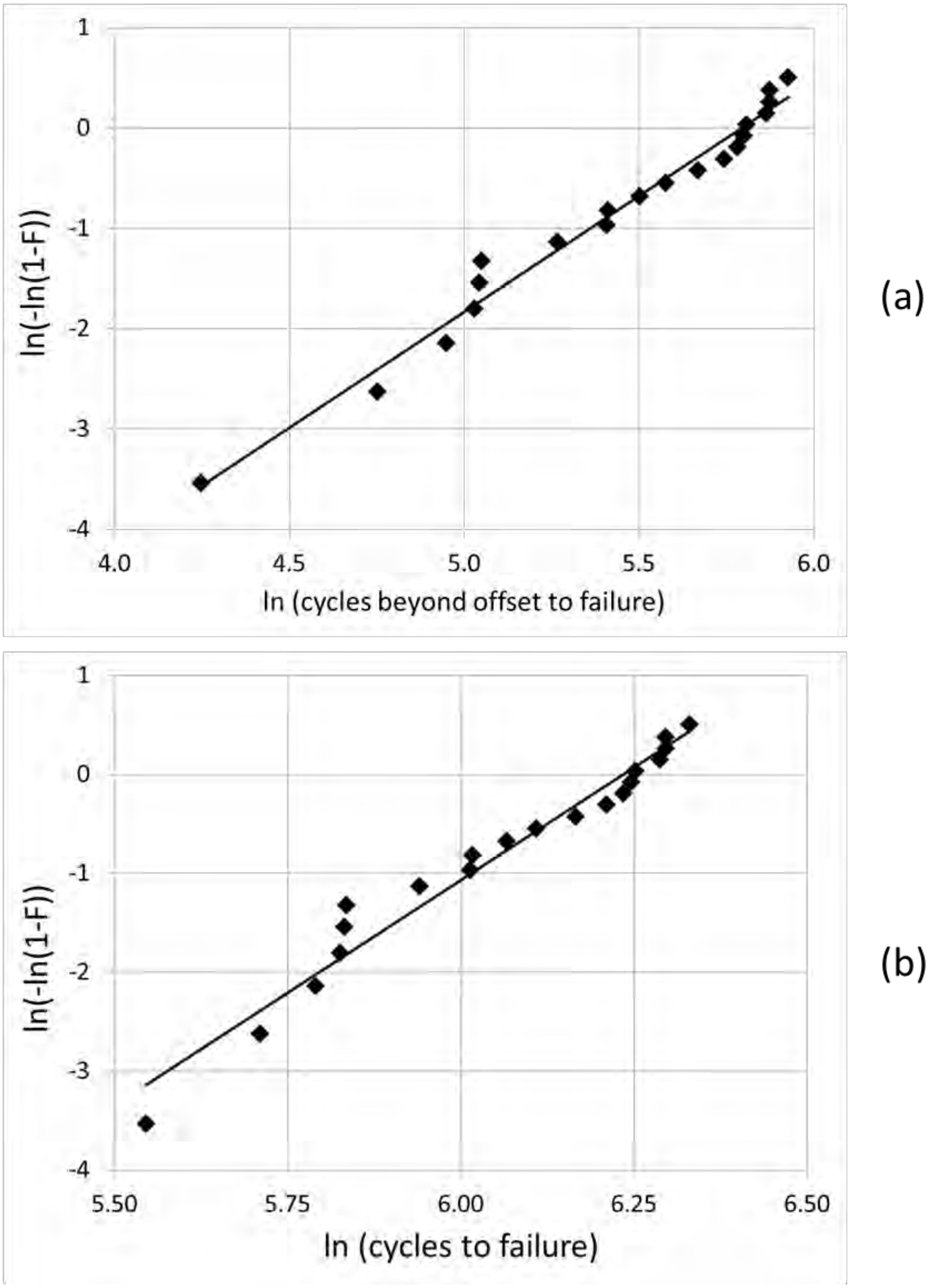

Figure 3. OLS fit using the MRR technique for (a) 3-Weibull and (b) 2-Weibull distributions. The fits appear to provide strong constraints on the slopes, but this perception is overly optimistic.

Unfortunately the quality of the fit gives a significantly over-optimistic sense of how well the parameters are known: the ordinary least squares (OLS) analysis does not provide good confidence intervals on the estimated parameters, as the OLS-based confidence intervals are too short[49] (see Appendix 2 in Supplemental Information). In addition, the MRR technique does not take into account the cycle number where the tests ended (593 in this case) when there are suspended tests. This, in addition to other well-known issues with MRR analysis[56, 58], motivates us to avoid the OLS-based analysis and use the maximum likelihood (ML) technique[48, 56]. This involves, in effect, searching for the values of (β, η), (β, η, γ), or (σ, μ) that maximize the likelihood of the associated *pdf,* where likelihood is defined as the probability that a given combination of parameters would produce the observed vector of failure times. For non-suspended cells, this can be calculated as the product of the *pdf* values at the cycle numbers where we observe failures; for the likelihood associated with suspended trials see [references]. Thus, if most of the observed failure times occur near the peak of a particular *pdf* curve, and if there are no observed failures or suspensions where the *pdf* is very small, then the corresponding parameters have high likelihood.

In Figure 4 we illustrate how we analyze our data with a 3-Weibull distribution by evaluating the likelihood at a grid of 4,000 values of (β, η, γ) triples, so that for each associated Weibull curve we ask how likely is it that the failure times that we measured could have come from that curve. The images show successive 2D cuts through (β, η, γ)-space for 4 values of γ, with the likelihoods color-coded. The maximum likelihood Weibull curve has $\beta \approx 2.3$, $\eta \approx 300$, and $\gamma \approx 185$, in agreement with the MRR analysis. However, it is clear that β is not tightly constrained; Weibull curves with $1.5 < \beta < 4$ can have likelihoods within about a factor of 2 of the maximum likelihood, depending on γ and η. This result, together with the fact that good fits are obtained from (2 parameter) normal and 2-parameter Weibull distributions, suggests that our data do not justify the use of 3 parameters. We emphasize that MRR may provide the best estimates for the Weibull parameters, given the data available, but the uncertainty in those parameter values cannot be readily evaluated from examining linearized Weibull plots.

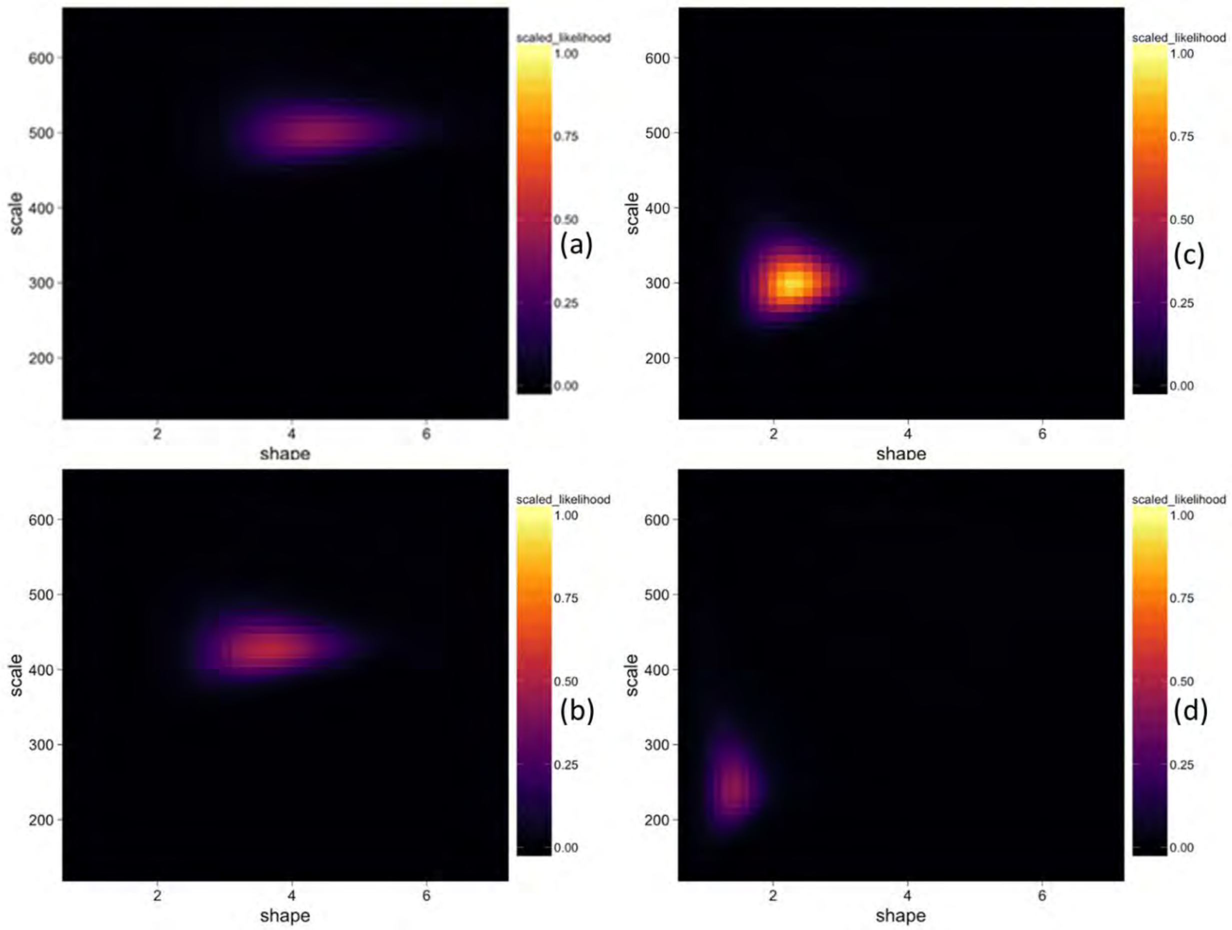

Figure 4. Slices at $\gamma \approx 50$, 120, 190, and 250 that show the relative likelihood of different combinations of shape factor β and scale factor η for each value of γ. A video showing shape and scale likelihoods for intermediate values of γ is included in the Supplementary Materials.

Figure 5 shows the likelihoods associated with different cumulative failure distributions $F(t;\beta,\eta,\gamma)$. We have drawn a horizontal line where F = 5% of the samples have failed; this might represent a critical threshold for a costly battery recall, for example. We note from the figure that the confidence interval is much narrower at 5% failure than at, say, 95% failure, reflecting the fact that we have no data above 593 cycles. (Confidence intervals can be obtained directly from the ML analysis or from other techniques[59-60].)

Fortunately, from a practical point of view, we are normally most interested in small values of F.

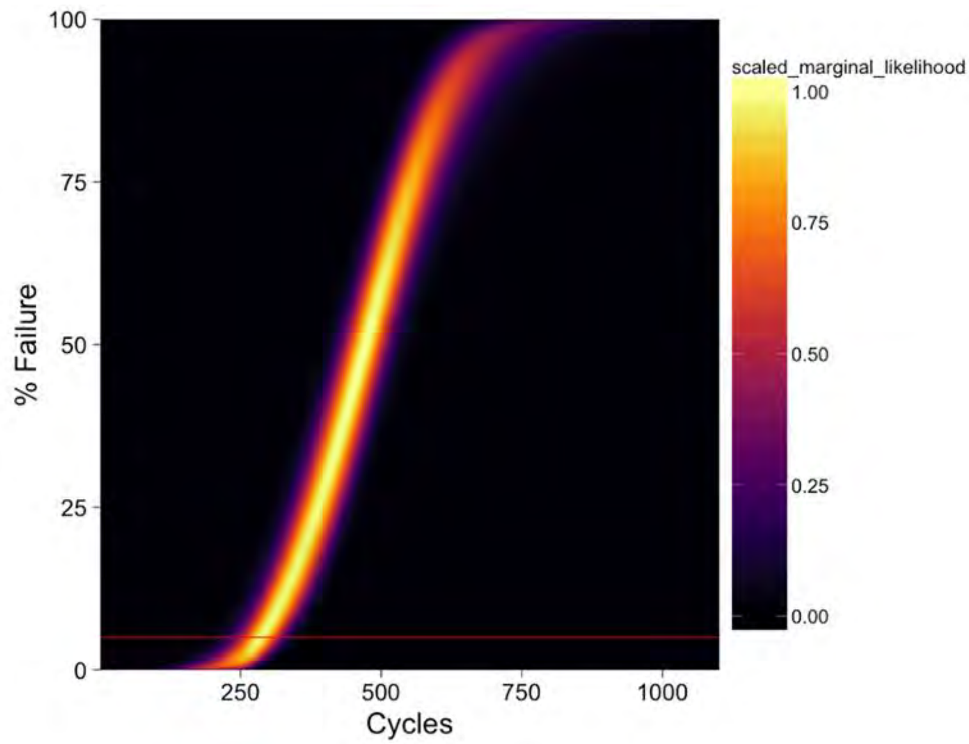

Figure 5. Relative likelihoods of associated with different cumulative failure functions F, obtained from adding up all the likelihoods (pixel intensities from Figure 4) for all values of β,η,γ that produce a given percent failure at a given number of cycles. A line is drawn across the image at 5% failure.

Number of Failures Observed

A focus on low values of F suggests that we ask how much useful information we lose by stopping the testing early. For example, we can calculate the expected number of cycles to 5% failure, as well as the associated confidence intervals, at any point in time, such as after every experimentally observed failure. The data points in Figure 6a show how these quantities vary as a function of the number of observed failures. Initially the confidence intervals show no trends, since the addition of each additional failure can significantly alter the analysis. But by around a dozen failures, trends in the data appear. After 20 failures have been observed, we calculate with 90% confidence that the expected number of cycles for 5% failure is between about 200 and 375 cycles, with an ML estimate at 275.

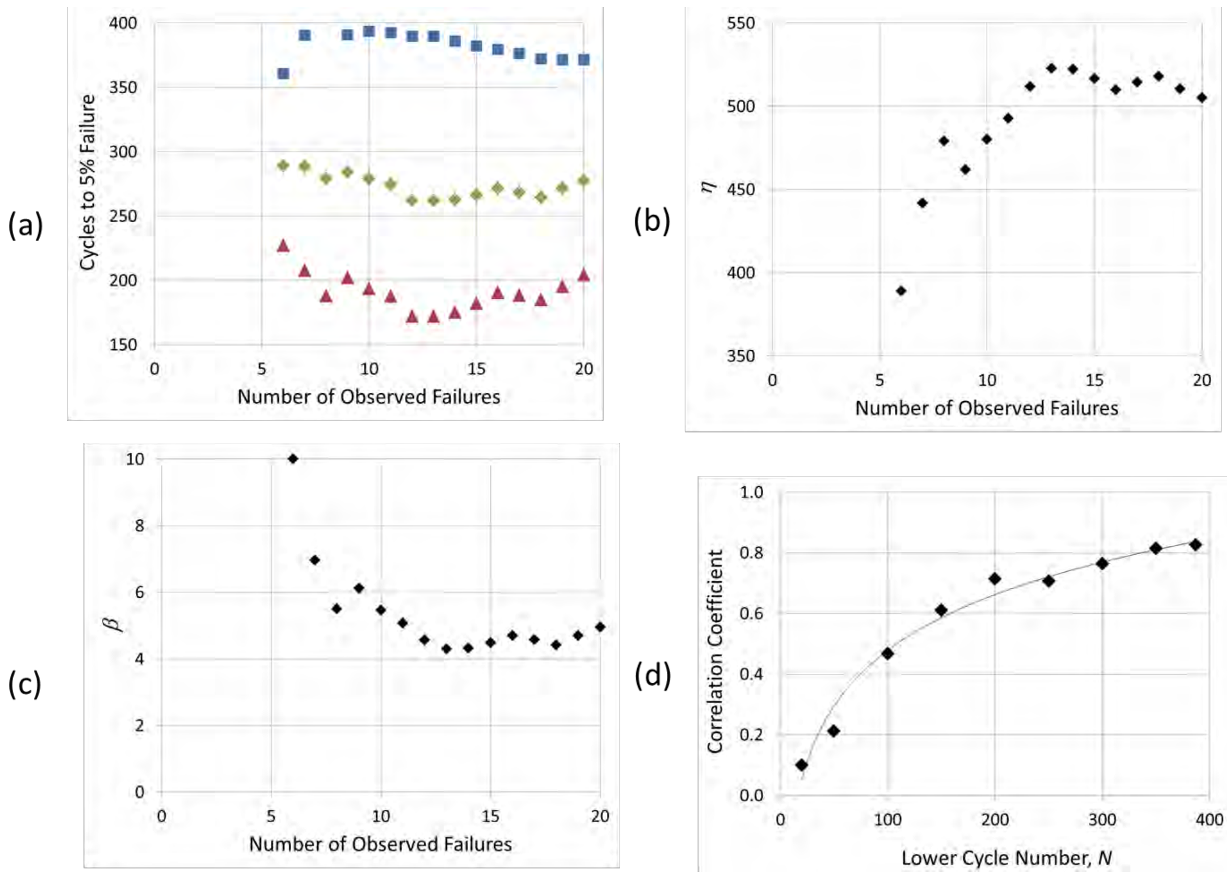

Figure 6. Impact of varying the number of failures observed (type II censoring). (a) Confidence limits for the number of cycles until 5% of the cells fail. Blue squares, 5%; green diamonds, 50%; red triangles, 95%. (b) 2-Weibull fitting parameter η as a function of the number of observed failures. (c) 2-Weibull fitting parameter β as a function of the number of observed failures. (d) Correlation coefficient between performance on cycles N and N+200. A line with the form $y = a + b \ln x$ is shown as a guide to the eye.

Most of the information about the 5% failure rate can be found in the first few failures, as indicated by the fact that the estimates and confidence limits in Figure 6a are relatively stable as the testing continues up to 20 failures. Similarly, figures 6a and 6b suggest that the underlying parameters may also converge after about 10 or 15 failures, although we recognize that the fitted values are approximate (Figure 4). More testing will be required to determine if this apparent convergence is real.

If we can accurately estimate the Weibull parameters from 15 or 20 failures (Figures 6a and 6b), does that mean we can make accurate extrapolations for the performance of individual cells for RUL predictions? To address this question, we carried out an analysis similar to that shown in Figure 2b, but correlating the capacity of each cell on cycle N with its capacity on cycle N+200. Figure 6d shows that the correlation coefficients are poor for $N < 150$ cycles. During this period, all the cells are performing similarly, and there is very little information about their future performance. But by the time N = 350, the correlation coefficient is above 0.8. This increase demonstrates that reasonably accurate predictions of performance 200 cycles ahead are possible, but only after sufficient degradation has occurred to produce significant capacity spread. In other words, once it is clear which cells have good or bad capacity retention properties, their individual performance can be extrapolated further with modest confidence.

Figures 6a and 6b show that even after a small number of failures the Weibull parameters are in the range of their converged values. Schuster et al[45], who measured the spread of capacity values at a time before there were any failures, found $50 < \beta < 116$, corresponding to extremely narrow distributions. As they point out, these somewhat unexpected results are the result of the fact that there is hardly any variation in capacity at early times, consistent with published data[14-15] and Figure 2a. Figure 2b shows that data taken at such early times provides very little predictive power.

Definition of Failure

We pointed out above that our choice of failure at 80% retained capacity is fundamentally arbitrary, although there are good practical reasons why this value was chosen by the auto industry. To determine how our results are affected by our choice of failure criterion[44], we carried out analyses with failure defined as 70%, 75%, 80%, 85%, and 90% retained capacity.

Figure 7 compares the fitted shape parameter β as a function of how we define failure, using 2- and 3-parameter Weibull functions. The most interesting result here is that the 2-Weibull fit returns a value of β that is roughly constant at about 5, while the 3-Weibull fit has β changing by nearly a factor of 2. If β is to have any physical meaning, then its value should hardly be expected to change so much just because we change our definition of failure, suggesting that a 3-Weibull fit will not provide much insight. As with the 80% data above, the fitted 2-Weibull and normal functions are very similar for every definition of failure.

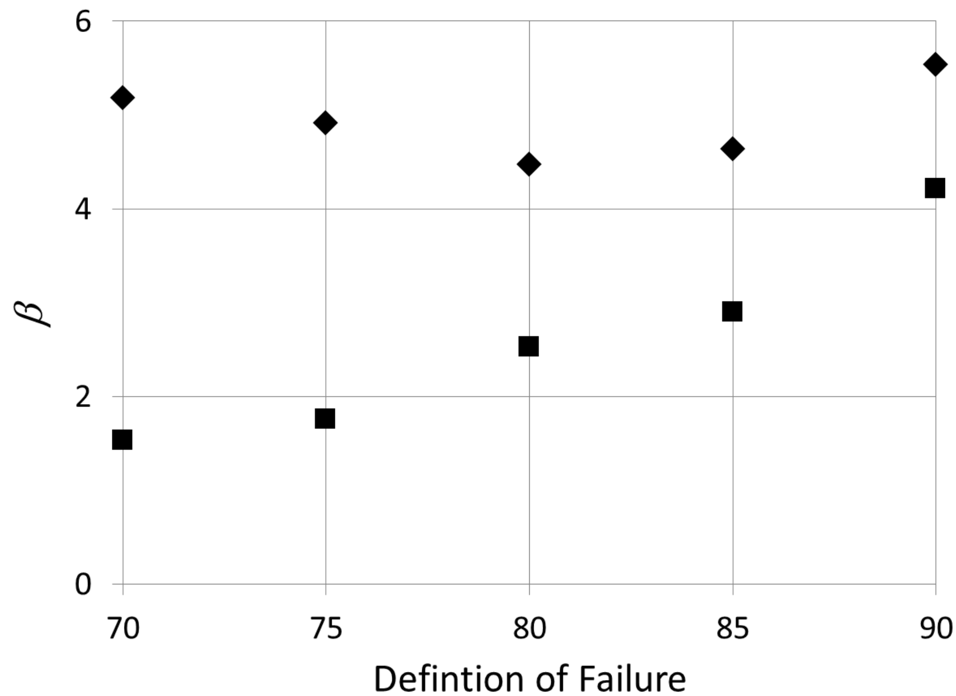

Figure 7. Fitted values for β as a function of how we define failure, where final capacities range between 70% and 90% of the initial capacities. Diamonds, 2-Weibull. Squares, 3-Weibull

Choosing Among Functional Forms

Both Weibull and normal distributions are widely used functions for fitting failure data. In principle, an advantage of using one of these distributions is that the fits may give insight into the failure mechanisms, as described above. Figure 8 shows the best fits to the 2-Weibull, 3-Weibull, and normal distributions. The 2-Weibull and normal distributions are nearly identical, except at high cycles where there are no data. Thus, it is not particularly useful to discuss which functional form fits the data better. The graph shows that the 2-Weibull function can locate itself at almost precisely the same place as the other functions, which have explicit location parameters. Thus, the addition of a (third) location parameter for the 3-Weibull function is more or less redundant: it simply allows the shape and scale parameters to take a wider range of values while still maintaining a good overall fit, as we have seen in Figure 4. Furthermore, the 3-Weibull distribution falls off a cliff below 200 cycles—in, fact it predicts that the probability of a failure below 185 cycles (the value of γ) is precisely zero. Given the failure points shown on the graph, this prediction seems implausible, and in our opinion, disqualifies the 3-Weibull distribution as representing the true failure distribution, no matter how good its fit to the linearized failure data[8], Figure 3a. This situation is typical of what happens when data is over-fit—extrapolations tend to be unreliable.

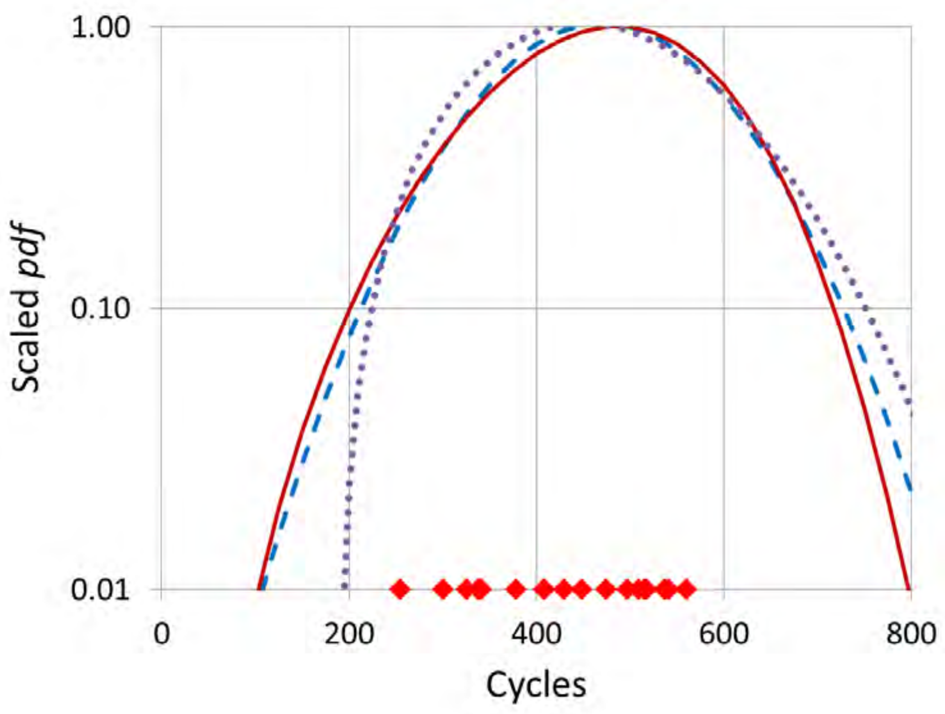

Figure 8. Best fit 2-Weibull (solid red), 3-Weibull (dotted purple), and normal (dashed blue) distributions to our failure data on a log scale, shown as diamonds along the *x*-axis. The 2-Weibull and the normal curves are essentially indistinguishable in the range where we have experimental data (up to 593 cycles). Note that use of a 3'rd parameter γ in a Weibull distribution leads to the predictin that the probability for failure occurring before the γ^{th} cycle is zero. In this case, γ = 185, which appears implausible.

Without more failure data we have no basis for choosing between a 2-Weibull and a normal distribution. It could in principle be valuable to distinguish between them, since there are insights to be gleaned about a failure mechanism that is known to be normal, for example. However, we estimate that this would take at least hundreds of observed failures. (Data on this scale could become available from warranty returns.) However, if the goal of the testing is to set a warranty, it makes no difference which of these two functions we choose: the predicted failure distributions and confidence intervals of the two functions are practically indistinguishable.

Conclusions

We have cycled 24 nominally identical commercial pouch cells under laboratory conditions for almost 600 cycles, during which time their capacities dropped to between 45% and 85% of their initial values. Not only did the distribution of capacities become very wide, but the capacity vs cycle curves crossed each other numerous times. Although our testing protocol was unusual (10C discharge in order to hasten failure), our results qualitatively mirror those of others[14-15, 43-44] and suggest that great care should be exercised when deciding how much faith to put in SOH/RUL predictions based on only a handful of replications or on early performance data. While it is possible that good RUL predictions can be made using other measurements, such as impedance or OCV, we don't believe that there is sufficient data available to warrant such an assumption, especially since capacity and impedance are correlated[61]. The availability of more data sets with relatively large numbers of replications will, we believe, be necessary to resolve such issues.

We show in Appendix 2 that a median rank regression (MRR) analysis can give an unrealistically optimistic sense of how well the Weibull parameters are known. Instead, we have used maximum likelihood techniques to show that wide ranges of Weibull parameters can give reasonable fits to our data, even when the failure data fall closely on a linear Weibull plot. Thus, likelihood-based methods should be preferred over MRR, even when the distribution of failure times looks approximately Gaussian.

Our data provide no rationale for choosing between 2-Weibull and normal distributions as fitting functions, and, indeed, they predict almost identical failure distributions and confidence limits for cycle ranges for which we have data. Apart from theoretical reasons to doubt the value of employing a 3-Weibull distribution[52-55], we find that the best 3-Weibull fit does not add insight and is not physically reasonable for our data, even though the fit is better.

We note that the data presented and discussed[8, 14-15, 44] here all come from commercial pouch and 18650 cells. It would be interesting to know whether cells fabricated in academic labs are more reproducible[62] (because they're handmade[63]) or less reproducible (because they don't have that secret additive) than commercial cells[64]. Future work will explore the differences between cells that fail early and cells that fail late. Our results suggest that looking for failure modes in a single degraded battery, as is commonly done, may be missing a key point: the failure mode(s) of the cell that ended the test at 85% capacity might be quite different from that of the cell that ended the test at 45% capacity. We will also compare the predictive value of the approach suggested in Figure 6d with more sophisticated approaches to predicting RUL. In sum, we believe that an increased emphasis on research aimed at understanding why battery lifetimes are so variable could lead to improvements in battery cost and safety.

Acknowledgments

SJH acknowledges support from the Assistant Secretary for Energy Efficiency, Vehicle Technologies Office of the U.S. Department of Energy (U.S. DOE) under the Advanced Battery Materials Research (BMR) Program. DJH acknowledges support from the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through Grant GBMF4563 to E.P. White. We are grateful for valuable discussions with Prof. Martin Bazant at MIT and important contributions from Steve Howard in the UC Berkeley Statistics Department. We thank Dr. Nansi Xue for assisting with the cell tests.

Table 1. Failure data, including cycle number for each Failure or Suspension, the number of Failures or Suspensions, and the Median Rank value.

CYCLES	F/S	QTY	MED RANK
255	F	1	0.029
301	F	1	0.070
326	F	1	0.111
338	F	1	0.152
340	F	1	0.193
341	F	1	0.234
379	F	1	0.275
408	F	1	0.316
409	F	1	0.357
430	F	1	0.398
449	F	1	0.439
475	F	1	0.480
497	F	1	0.520
509	F	1	0.561
515	F	1	0.602
518	F	1	0.643
537	F	1	0.684
541	F	1	0.725
541	F	1	0.766
560	F	1	0.807
560	S	4	--------

Appendix 1. Rate Data

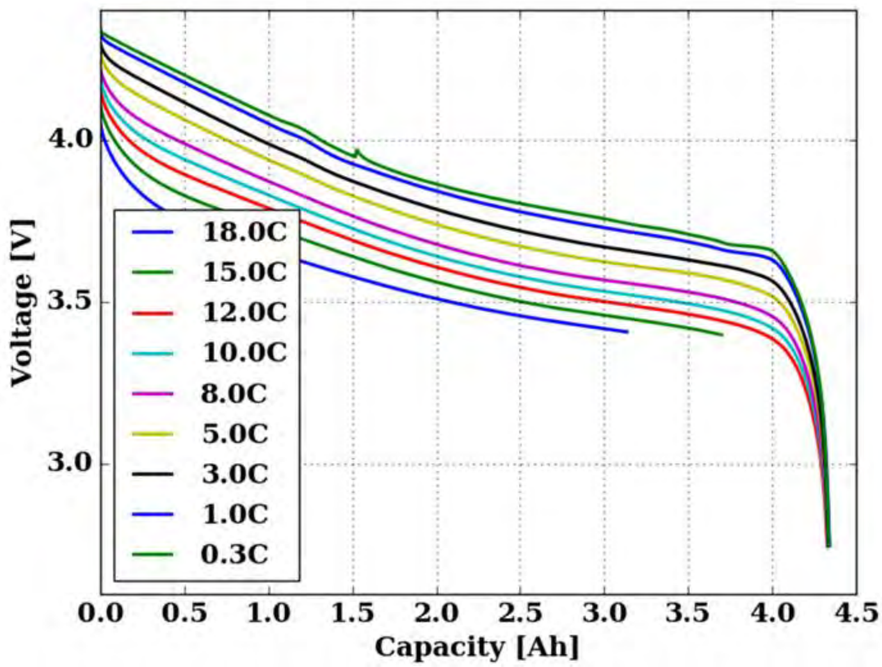

Figure A1. Rate data for our cells

Appendix 2. Problems with OLS-based estimation of Weibull parameters

As outlined above, the simplest method for estimating Weibull parameters involves running a linear regression (ordinary least squares, OLS) on median ranks. This performs well for generating a single estimate of the Weibull parameters. It would seem only natural, then, to use the standard confidence intervals from OLS to estimate uncertainty in these parameters. However, these confidence intervals are too short and fail to achieve the desired coverage level. This happens because the assumptions of OLS are not justified when regressing median ranks.

To demonstrate this, we simulated 25 failure times from a Weibull distribution with η = 250, β = 1.5, and γ = 0. We estimated the shape parameter β using OLS as described and calculated the 95% confidence interval. We repeated this procedure 10,000 times and in only 31% of the replications did the confidence interval contain the true value β = 1.5, far below the nominal 95% level. Maximum likelihood, on the other hand, produced confidence intervals that contained the true value in 95.3% of replications. Figure A2 shows histograms of the estimated parameters and standard errors from the 10,000 replications using OLS; it is clear that the OLS standard errors considerably underestimate the sampling variability in the estimated shape.

Furthermore, the R^2 values from OLS can lead to a false sense of fit to the Weibull distribution. Figure A3 shows histograms of R^2 values from each replication for three different simulations based on: (a) a 2-Weibull distribution with η = 250, β = 1.5; (b) a uniform distribution (density is constant between 1 and 300); and (c) a 50/50 mixture of two 2-Weibull distributions, η = 250, β = 0.8 and η = 250, β = 5.

In the uniform case, the R^2 value is above 0.9 in 80.5% of replications, and even in the mixture case this still happens in 51.1% of replications. If we use R^2 to gauge the model fit (whether the observations seem to fit a true Weibull distribution), we are bound to draw overconfident “yes” conclusions in a wide variety of situations.

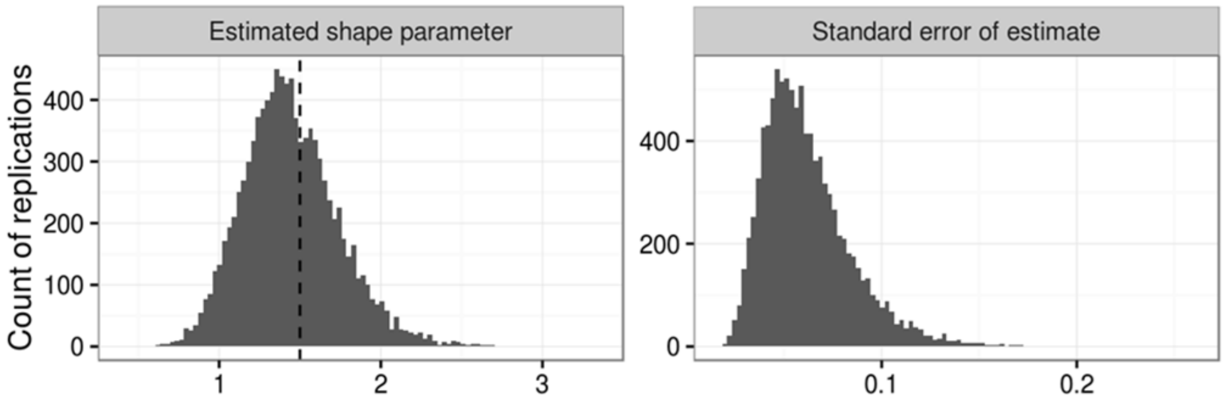

Figure A2: Histograms of estimates and standard errors from OLS-based median rank regression. Dotted line indicates true shape parameter.

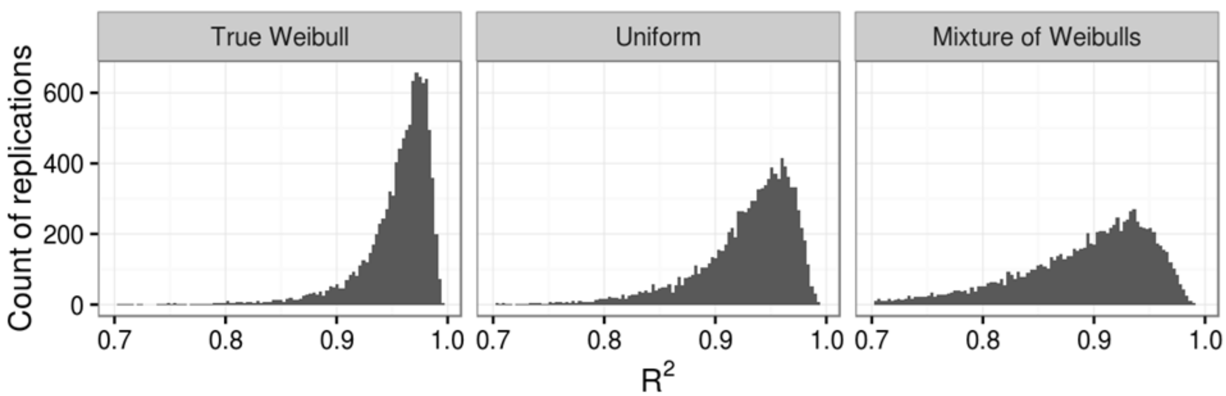

Figure A3. Histograms of R^2 values from OLS-based median rank regression for samples drawn from different underlying distributions. (a) Values from the true Weibull distribution used above. (b) Values drawn uniformly between 1 and 300. (c) Values drawn from a mixture of two very different Weibulls.

Bibliography

1. Fotouhi, A.; Auger, D. J.; Propp, K.; Longo, S.; Wild, M., A review on electric vehicle battery modelling: From Lithium-ion toward Lithium–Sulphur. *Renewable and Sustainable Energy Reviews* **2016,** *56*, 1008-1021.
2. Girishkumar, G.; McCloskey, B.; Luntz, A.; Swanson, S.; Wilcke, W., Lithium− air battery: promise and challenges. *The Journal of Physical Chemistry Letters* **2010,** *1* (14), 2193-2203.
3. Hu, X.; Martinez, C. M.; Yang, Y., Charging, power management, and battery degradation mitigation in plug-in hybrid electric vehicles: A unified cost-optimal approach. *Mechanical Systems and Signal Processing* **2016**.
4. Wu, L.; Fu, X.; Guan, Y., Review of the Remaining Useful Life Prognostics of Vehicle Lithium-Ion Batteries Using Data-Driven Methodologies. *Applied Sciences* **2016,** *6* (6), 166.
5. Delacourt, C.; Safari, M., Mathematical Modeling of Aging of Li-Ion Batteries. In *Physical Multiscale Modeling and Numerical Simulation of Electrochemical Devices for Energy Conversion and Storage*, Springer: 2016; pp 151-190.
6. Berecibar, M.; Gandiaga, I.; Villarreal, I.; Omar, N.; Van Mierlo, J.; Van den Bossche, P., Critical review of state of health estimation methods of Li-ion batteries for real applications. *Renewable and Sustainable Energy Reviews* **2016,** *56*, 572-587.
7. Li, J.; Zou, L.; Tian, F.; Dong, X.; Zou, Z.; Yang, H., Parameter Identification of Lithium-Ion Batteries Model to Predict Discharge Behaviors Using Heuristic Algorithm. *Journal of The Electrochemical Society* **2016,** *163* (8), A1646-A1652.
8. Chiodo, E.; Lauria, D.; Andrenacci, N.; Pede, G. In *Accelerated life tests of complete lithium-ion battery systems for battery life statistics assessment*, Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 2016 International Symposium on, IEEE: 2016; pp 1073-1078.
9. Smith, K.; Wood, E.; Santhanagopalan, S.; Kim, G.; Neubauer, J.; Pesaran, A. In *Models for battery reliability and lifetime*, Battery Congress, April, 2013; pp 15-16.
10. Yuksel, T.; Litster, S.; Viswanathan, V.; Michalek, J. J., Plug-in hybrid electric vehicle LiFePO 4 battery life implications of thermal management, driving conditions, and regional climate. *Journal of Power Sources* **2017,** *338*, 49-64.
11. Zhang, Q.; White, R. E., Capacity fade analysis of a lithium ion cell. *Journal of Power Sources* **2008,** *179* (2), 793-798.
12. Ali, J. B.; Khelif, R.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. In *The use of nonlinear future reduction techniques as a trend parameter for state of health estimation of lithium-ion batteries*, Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2015 16th International Conference on, IEEE: 2015; pp 246-251.
13. Ecker, M.; Gerschler, J. B.; Vogel, J.; Käbitz, S.; Hust, F.; Dechent, P.; Sauer, D. U., Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data. *Journal of Power Sources* **2012,** *215*, 248-257.
14. 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.
15. Park, J. I.; Baek, S. H.; Jeong, M. K.; Bae, S. J., Dual features functional support vector machines for fault detection of rechargeable batteries. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2009,** *39* (4), 480-485.
16. Harris, S. J.; Timmons, A.; Baker, D. R.; Monroe, C., Direct in situ measurements of Li transport in Li-ion battery negative electrodes. *Chemical Physics Letters* **2010,** *485* (4-6), 265-274.

17. Kehrwald, D.; Shearing, P. R.; Brandon, N. P.; Sinha, P. K.; Harris, S. J., Local tortuosity inhomogeneities in a lithium battery composite electrode. *Journal of The Electrochemical Society* **2011,** *158* (12), A1393-A1399.
18. 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.
19. Chandran, K. R., Duality of fatigue failures of materials caused by Poisson defect statistics of competing failure modes. *Nature materials* **2005,** *4* (4), 303-308.
20. Ponson, L., Statistical aspects in crack growth phenomena: how the fluctuations reveal the failure mechanisms. *International Journal of Fracture* **2016,** *201* (1), 11-27.
21. Jin, Y.; Cai, P.; Tian, Q.; Liang, C.; Ke, D.; Wang, G.; Zhai, T., An experimental methodology for quantitative characterization of multi-site fatigue crack nucleation in high-strength Al alloys. *Fatigue & Fracture of Engineering Materials & Structures* **2016**.
22. Lim, J.; Li, Y.; Alsem, D. H.; So, H.; Lee, S. C.; Bai, P.; Cogswell, D. A.; Liu, X.; Jin, N.; Yu, Y.-s.; Salman, N. J.; Shapiro, D. A.; Bazant, M. Z.; Tyliszczak, T.; Chueh, W. C., Origin and hysteresis of lithium compositional spatiodynamics within battery primary particles. *Science* **2016,** *353* (6299), 566-571.
23. Nemani, V. P.; Harris, S. J.; Smith, K. C., Design of bi-tortuous, anisotropic graphite anodes for fast ion-transport in Li-ion batteries. *Journal of The Electrochemical Society* **2015,** *162* (8), A1415-A1423.
24. Harris, S. J.; Deshpande, R. D.; Qi, Y.; Dutta, I.; Cheng, Y.-T., Mesopores inside electrode particles can change the Li-ion transport mechanism and diffusion-induced stress. *Journal of Materials Research* **2011,** *25* (08), 1433-1440.
25. Wang, X. L.; An, K.; Cai, L.; Feng, Z.; Nagler, S. E.; Daniel, C.; Rhodes, K. J.; Stoica, A. D.; Skorpenske, H. D.; Liang, C.; Zhang, W.; Kim, J.; Qi, Y.; Harris, S. J., Visualizing the chemistry and structure dynamics in lithium-ion batteries by in-situ neutron diffraction. *Sci Rep* **2012,** *2*, 747.
26. Cai, L.; An, K.; Feng, Z.; Liang, C.; Harris, S. J., In-situ observation of inhomogeneous degradation in large format Li-ion cells by neutron diffraction. *Journal of Power Sources* **2013,** *236*, 163-168.
27. Paul, S.; Diegelmann, C.; Kabza, H.; Tillmetz, W., Analysis of ageing inhomogeneities in lithium-ion battery systems. *Journal of Power Sources* **2013,** *239*, 642-650.
28. Santhanagopalan, S.; White, R. E., Quantifying Cell-to-Cell Variations in Lithium Ion Batteries. *International Journal of Electrochemistry* **2012,** *2012*, 1-10.
29. Dubarry, M.; Devie, A.; Block, D.; Center, F. S. E. *Cell Emulation and Preliminary Results*; Hawaii Natural Energy Institute: 2016.
30. Arunachala, R.; Makinejad, K.; Athlekar, S.; Jossen, A.; Garche, J. In *Cycle life characterisation of large format lithium-ion cells*, Electric Vehicle Symposium and Exhibition (EVS27), 2013 World, IEEE: 2013; pp 1-9.
31. He, W.; Williard, N.; Osterman, M.; Pecht, M., Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. *Journal of Power Sources* **2011,** *196* (23), 10314-10321.
32. Guo, J.; Li, Z.; Pecht, M., A Bayesian approach for Li-Ion battery capacity fade modeling and cycles to failure prognostics. *Journal of Power Sources* **2015,** *281*, 173-184.
33. Saha, B.; Goebel, K., Battery Data Set. *"Battery Data Set", NASA Ames Prognostics Data Repository (https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/), NASA Ames Research Center, Moffett Field, CA* **2007**.
34. Thomas, E.; Bloom, I.; Christophersen, J.; Battaglia, V., Statistical methodology for predicting the life of lithium-ion cells via accelerated degradation testing. *Journal of Power Sources* **2008,** *184* (1), 312-317.
35. Tang, S.; Yu, C.; Wang, X.; Guo, X.; Si, X., Remaining useful life prediction of lithium-ion batteries based on the wiener process with measurement error. *Energies* **2014,** *7* (2), 520-547.

36. Ng, S. S. Y.; Xing, Y.; Tsui, K. L., A naive Bayes model for robust remaining useful life prediction of lithium-ion battery. *Applied Energy* **2014,** *118*, 114-123.
37. Hu, C.; Jain, G.; Tamirisa, P.; Gorka, T., Method for estimating capacity and predicting remaining useful life of lithium-ion battery. *Applied Energy* **2014,** *126*, 182-189.
38. Mahdi, S., Optimal conditional confidence interval for the shape parameter of a Weibull distribution. *Brazilian Journal of Probability and Statistics* **2003**, 57-74.
39. Ye, Z.-S.; Xie, M., Stochastic modelling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry* **2015,** *31* (1), 16-32.
40. Pinson, M. B.; Bazant, M. Z., Theory of SEI formation in rechargeable batteries: capacity fade, accelerated aging and lifetime prediction. *Journal of the Electrochemical Society* **2013,** *160* (2), A243-A250.
41. Farmann, A.; Sauer, D. U., A comprehensive review of on-board State-of-Available-Power prediction techniques for lithium-ion batteries in electric vehicles. *Journal of Power Sources* **2016,** *329*, 123-137.
42. Hu, X.; Jiang, J.; Cao, D.; Egardt, B., Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling. *IEEE Transactions on Industrial Electronics* **2016,** *63* (4), 2645-2656.
43. 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.
44. Park, J. I.; Park, J. W.; Jung, M.; Huh, Y. H.; Bae, S. J., Cycle life test time reduction by combining different types of acceleration. *Journal of the Society of Korea Industrial and Systems Engineering* **2008,** *31* (4), 153.
45. 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.
46. https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=3a4b9c93-fc06-4690-84f3-1e97524ff572, Journal of Power Sources.
47. Lambert, F., Tesla battery data shows path to over 500,000 miles on a single pack. *Electrek* November 1, 2016. https://electrek.co/2016/11/01/tesla-battery-degradation/.
48. Weibull-DR, http://applicationsresearch.com/WeibullEase.htm.
49. Howard, S., https://gist.github.com/gostevehoward/ff5d09bcb252956f8300524deaf7258d. **2016**.
50. 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 Transactions on Instrumentation and Measurement* **2016,** *65* (6), 1282-1291.
51. Kalbfleisch, J. D.; Prentice, R. L., *The statistical analysis of failure time data*. John Wiley & Sons: 2011; Vol. 360.
52. Bazant, M., Private Communication. **2016**.
53. Le, J.-L.; Bažant, Z. P., Unified nano-mechanics based probabilistic theory of quasibrittle and brittle structures: II. Fatigue crack growth, lifetime and scaling. *Journal of the Mechanics and Physics of Solids* **2011,** *59* (7), 1322-1337.
54. Le, J.-L.; Bažant, Z. P.; Bazant, M. Z., Unified nano-mechanics based probabilistic theory of quasibrittle and brittle structures: I. Strength, static crack growth, lifetime and scaling. *Journal of the Mechanics and Physics of Solids* **2011,** *59* (7), 1291-1321.
55. Fisher, R. A.; Tippett, L. H. C. In *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge Univ Press: 1928; pp 180-190.
56. Genschel, U.; Meeker, W. Q., A comparison of maximum likelihood and median-rank regression for Weibull estimation. *Quality Engineering* **2010,** *22* (4), 236-255.

57. Reliasoft, http://www.weibull.com/GPaper/ranks2_6.htm.
58. Olteanu, D.; Freeman, L. *Technical Report on the Evaluation of Median Rank Regression and Maximum Likelihood Estimation Techniques for a Two-Parameter Weibull Distribution*; Virginia Tech: http://www.stat.vt.edu/research/Technical_Reports/TechReport08-4.pdf.
59. Wittman, D., Fisher Matrix for Beginners. *http://wittman.physics.ucdavis.edu/Fisher-matrix-guide.pdf*.
60. Feigelson, E. G.; Nelson, P. I., Statistical Methods for Astronomical Data with Upper Limits. Univariate Distributions. *The Astrophysical Journal* **1985,** *293*, 192-206.
61. Schuster, S. F.; Brand, M. J.; Campestrini, C.; Gleissenberger, M.; Jossen, A., Correlation between capacity and impedance of lithium-ion cells during calendar and cycle life. *Journal of Power Sources* **2016,** *305*, 191-199.
62. Mohanty, D.; Hockaday, E.; Li, J.; Hensley, D. K.; Daniel, C.; Wood, D., Effect of electrode manufacturing defects on electrochemical performance of lithium-ion batteries: Cognizance of the battery failure sources. *Journal of Power Sources* **2016,** *312*, 70-79.
63. Marks, T.; Trussler, S.; Smith, A.; Xiong, D.; Dahn, J., A guide to Li-ion coin-cell electrode making for academic researchers. *Journal of The Electrochemical Society* **2011,** *158* (1), A51-A57.
64. Dubarry, M.; Vuillaume, N.; Liaw, B. Y., From single cell model to battery pack simulation for Li-ion batteries. *Journal of Power Sources* **2009,** *186* (2), 500-507.