Data-Driven Prognostics of Lithium-Ion Rechargeable Battery using Bilinear Kernel Regression

Charlie Hubbard[1], John Bavlsik[2], Chinmay Hegde[1], and Chao Hu[2]

[1] *Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50011, USA*

hubbardc@iastate.edu
chinmay@iastate.edu

[2] *Department of Mechanical Engineering, Iowa State University, Ames, IA, 50011, USA*

jbavlsik@iastate.edu
chaohu@iastate.edu

ABSTRACT

Reliability of lithium-ion (Li-ion) rechargeable batteries has been recognized as of high importance from a broad range of stakeholders, including battery manufacturers, manufacturers of battery-powered devices, regulatory agencies, researchers, and the public. Assessing the current and future health of Li-ion batteries is essential to ensure the batteries operate safely and reliably throughout their lifetime. This paper presents a new data-driven approach for prediction of battery remaining useful life (RUL) in the presence of corruptions (or errors) in capacity features. The approach leverages bilinear kernel regression to build a nonlinear mapping between the capacity feature space and the RUL state space. Specific innovations of the approach include: i) a general framework for robust sparse prognostics that effectively incorporates sparsity into kernel regression and implicitly compensates for errors in capacity features; and ii) two numerical procedures for error estimation that efficiently derives optimal values of the regression model parameters. Results of 10 years' continuous cycling test on Li-ion prismatic cells suggest that the proposed approach achieves robust RUL prediction despite random noise in the capacity features.

Keywords: Bilinear Kernel Regression; Prognostics; Remaining Useful Life; Lithium-Ion Battery

1. INTRODUCTION

Lithium-ion (Li-ion) battery technology has been playing a critical role in realizing wide-scale adoption of hybrid and electric vehicles and show great promise for emerging applications in smart grid and medical devices. Over the past two decades, real-time health diagnostic and prognostic techniques have been developed and deployed in battery management systems (BMSs) to monitor the health condition of a battery in operation (Plett, 2004, 2004; He et al., 2013; Lee et al., 2008; Hu, Youn & Chung, 2012; Xiong et al., 2014; Hu et al., 2015); and to infer, within a maintenance horizon time, the remaining useful life (RUL), i.e., when the battery is likely to fail (Saha & Goebel, 2009; Saha, Goebel, Poll, et al., 2009; Liu et al., 2010; Wang et al., 2013; Dickerson et al., 2015; Hu et al., 2014; Hu et al. 2016). Based on the voltage, current and temperature measurements acquired from the battery, these techniques estimate three performance indicators of the battery: state of charge (SOC), state of health (SOH) and state of life (SOL). Accurate estimation of these parameters provides greater transparency into the current and future health of the battery, more cost-effective maintenance strategies and minimum downtime, and opportunities for battery life extensions.

Research on life prognostics of a general engineered system was conducted with an emphasis on predicting the RUL distribution. In general, three categories of approaches have been developed that enable continuous updating of system health degradation and RUL distribution: (i) model-based approaches (Gebraeel et al., 2005; Luo et al., 2008; Gebraeel & Pan, 2008; Si et al., 2013), (ii) data-driven approaches (Si et al., 2011; Wang et al., 2008; Wang et al., 2012; Hu, Youn, Wang, et al., 2012; Coble and Hines, 2008; Heimes, 2008; Lu et al., 2013), and (iii) hybrid approaches (Goebel et al., 2006; Liu et al., 2012). With the advance of modern sensor systems as well as data storage and processing technologies, the data-driven approaches, which mainly rely on large volumes of sensory data with no stringent requirement on the knowledge about the underlying degradation mechanisms of the system, have recently become popular. A good review of data-driven prognostic approaches was given in (Si et al., 2011). Data-driven prognostic approaches generally require sensory data

fusion and feature extraction, pattern recognition, and for life prediction, interpolation (Wang et al., 2008; Wang et al., 2012; Hu, Youn, Wang, et al., 2012), extrapolation (Coble & Hines, 2012), machine learning (Heimes, 2008), and so on.

Research on life prognostics of Li-ion battery (or battery prognostics) was mainly conducted by researchers in the prognostics and health management (PHM) society (Saha & Goebel, 2009; Saha, Goebel, Poll, et al., 2009; Liu et al., 2010; Wang et al., 2013; Dickerson et al., 2015; Hu et al., 2014). Battery prognostics often begins by estimating the current SOH of a battery in operation based on readily available measurements (i.e., voltage, current and temperature) from the battery (Lu et al., 2013). Capacity and internal resistance are two important SOH indicators that together determine the maximum amount of energy that a fully charged battery can deliver. SOL is a prognostic metric and often used interchangeably with RUL, which refers to the available service time left before SOH of the battery degrades to an unacceptable level. RUL can be measured in either calendar time (e.g., days, weeks, and months) or charge/discharge cycles. A Bayesian framework combining the relevance vector machine (RVM), trained with *sparse Bayesian learning* (SBL) (Tipping, 2001), and a particle filter-based approach was proposed for prognostics of a Li-ion battery based on electrochemical impedance measurements (Saha, Goebel, Poll, et al., 2009). In order to eliminate the reliance of prognostics on impedance measurement equipment, researchers developed various model-based approaches that predict RUL by extrapolating a capacity fade model (Saha & Goebel, 2009; Liu et al., 2010; Wang et al., 2013; Dickerson et al., 2015). An integrated method for capacity estimation and RUL prediction of Li-ion battery was later developed and applied to Li-ion cells for implantable medical devices (Hu et al., 2014). The method employed the coulomb counting approach to estimate battery capacity based on the difference in the SOC values before and after partial charge/discharge. Based on the capacity estimates, a Gauss-Hermite particle filter was used to online update an empirical capacity fade model and project the updated model to an end-of-life (EOL) limit for RUL prediction. More recently, the RVM approach was leveraged to estimate battery capacity by approximating a nonlinear mapping from features (extracted from voltage and current measurements) to capacity (Hu et al., 2015; Hu et al., 2016), and RUL prediction was performed by first fitting linear models to random trajectories of capacity estimates and then extrapolating the models to an EOL limit (Hu et al., 2016).

Despite significant advances in battery prognostics, research innovations are still needed to develop new approaches that can leverage large volumes of data to achieve robust RUL prediction. In particular, the goal is to perform reliable RUL prediction even in the presence of corruptions (or errors) in capacity features. In this paper, a new data-driven approach to RUL prediction is proposed and applied to a Li-ion battery used in implantable medical devices. The new approach fundamentally addresses the issue of input data noise via a new technique known as bilinear kernel regression. Specific innovations of the approach include: i) a general framework for robust sparse prognostics that effectively incorporates sparsity into kernel regression and implicitly compensates for errors in capacity features; and ii) two numerical procedures for error estimation that efficiently derive optimal values of the regression model parameters. We use 10 years' continuous cycling data on eight Li-ion prismatic cells to demonstrate the effectiveness of the proposed approach. Moreover, we compare our proposed bilinear kernel regression framework with previously existing sparse regression approaches, and demonstrate uniformly improved prediction performance. This paper is organized as follows. Section 2 presents the fundamentals of the proposed approach. The approach is applied to a Li-ion battery used in implantable medical devices. Section 3 discusses the experimental results of this application. The paper is concluded in Section 4.

2. TECHNICAL APPROACH

In this study, the capacity of a Li-ion battery cell is viewed as the SOH indicator of the cell. The cell capacity quantifies the maximum amount of charge that the cell can hold. It tends to fade slowly over time, and typically decreases 1.0% or less in a month with regular use. Given the capacity values estimated by an existing estimation algorithm, we are interested in predicting the remaining useful life (RUL) of the cell, i.e., how long the cell is expected to last before the capacity fade reaches an unacceptable level. This section is dedicated to describing the proposed data-driven approach for doing so. Section 2.1 defines the problem of data-driven prognostics considered in this study and discusses the application of kernel regression to solve this problem; Section 2.2 presents the fundamentals of a classical sparse regression technique, namely the *Least Absolute Selection and Shrinkage Operator* (LASSO), and discusses its application to RUL prediction when capacity estimates and RUL responses are error-free; and Section 2.3 describes the fundaments of a robust sparse regression technique, namely *bilinear kernel regression*, and discusses its application to RUL prediction with errors in the capacity estimates.

2.1. Fundamentals

Kernel regression is a non-parametric regression technique that establishes a set of identical weighted functions, called local kernels, from the training data points, and a training process is employed to adjust the weights of the kernels to achieve the best-fit line at these data points. In the context of battery prognostics, a kernel regression algorithm takes the (estimated) capacity values of a battery cell as the inputs, and produces the (predicted) RUL as the output. In this regard, kernel regression approximates the complex mapping from the capacity feature ($\mathbf{x}$) space to the RUL state (y) space (see Fig. 1).

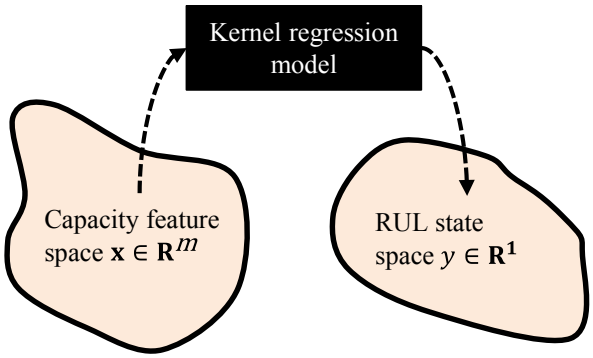

Figure 1. Approximation of mapping from capacity feature space to RUL state space by kernel regression.

Assume that we are given a set of training data $\{(\mathbf{x}_i, y_i)\}$, i =1, 2, …, n, consisting of n samples from an arbitrary distribution D. Here, $\mathbf{x}_i$ represents the data features, each represented by an m-dimensional vector consisting of the m most-recently calculated capacity estimates. Moreover, y_i represents the (measured or true) RUL values for the corresponding capacity estimates. Our goal is to investigate a *purely data-driven* machine learning approach that predicts the RUL from the capacity estimates. The approach employs a nonlinear kernel regression model of the form:

$$y(\mathbf{x}) = \sum_{i=1}^{n} w_i \kappa(\mathbf{x}, \mathbf{x}_i) + w_0 \qquad (1)$$

where $\mathbf{x}$ is a (test) feature vector, y is the predicted RUL, $\mathbf{w} = (w_0, \ldots, w_n)^T$ represent the kernel weights, and $\kappa(\mathbf{x},\mathbf{x}_i)$ is a suitable kernel function. The choice of kernel κ is somewhat flexible, but the key thing to note is that it is centered on the training point $\mathbf{x}_i$. A typical kernel used in nonlinear prediction applications is the Gaussian kernel function:

$$\kappa(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{r^2}\|\mathbf{x} - \mathbf{x}_i\|_2^2\right) \qquad (2)$$

where r is a pre-chosen parameter called the kernel bandwidth. We use this kernel function in all our experiments below. The goal of nonlinear kernel regression is to learn the optimal model (parameterized by the weight vector w) that provides the best prediction performance. Numerous algorithms for learning nonlinear prediction functions have been proposed in the machine learning literature, including singular value decomposition (SVD)-based approaches, stochastic gradient descent, and kernel least-squares (Trefethen & Bau, 1997).

While kernel methods are known to provide very good prediction performance, they are often prone to overfitting to the training data and their performance can degrade on unseen test samples. Following the principle of *Occam's Razor*, machine learning algorithms for nonlinear prediction often attempt to learn a *simple* model that best explains the data. From a computational standpoint, these algorithms learn prediction models by solving a regularized problem that balances two competing objectives (training error versus model complexity). Again, numerous prediction algorithms that exploit such regularization assumptions have been developed in the literature. One approach that has been explored in detail in the PHM literature is the *Sparse Bayesian Learning* (SBL) approach (Tipping, 2001) that constructs a nonlinear regression model, known as the RVM, for online estimation of battery capacity (Hu et al., 2015; Hu et al., 2016). The RVM solves a Bayesian inference problem by imposing a sparse regression model on the optimal prediction weight vector, i.e., only a small subset of the coordinates of the optimal $\mathbf{w}$ are permitted to be nonzero. Tests reveal that such sparsity-based regularization methods yield better generalizability to unseen test samples, and also offer improved interpretability in terms of prediction performance.

2.2. RUL Prediction using the LASSO

We first propose an alternative sparsity-regularized approach for RUL prediction. Our approach is based on (now classical) optimization formulation in sparse regression called the *Least Absolute Selection and Shrinkage Operator* (LASSO). First, using the capacity estimates $\{\mathbf{x}_i\}$, we construct the design matrix Φ of size $n \times (n+1)$, where:

$$\varphi_{ij} = 1 \text{ for } j = 1, \text{ and } \varphi_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_{i-1}) \text{ for } j = 2, \ldots, n+1 \qquad (3)$$

Next, we arrange the corresponding RUL measurements as a response vector $\mathbf{y} = (y_0, \ldots, y_n)^T$. Finally, we define a non-negative real valued parameter λ that controls the tradeoff between the goodness of prediction fit and the sparsity of the prediction vector.

Having defined these quantities, we now obtain a prediction vector by solving the convex optimization problem:

$$\hat{\mathbf{w}} = \arg\min \ \lambda\|\mathbf{w}\|_1 + \|\mathbf{y} - \Phi\mathbf{w}\|_2^2 \qquad (4)$$

Here, $\|.\|_p$ denotes the ℓ_p-norm of a given vector. The choice of the tradeoff parameter λ is dataset-dependent; higher values encourage greater sparsity (i.e., fewer nonzero coefficients) in the prediction vector, and vice versa. In our experiments below (see Section 3), we chose the parameters based on leave-one-out cross-validation (LOOCV).

The optimization problem in Eq. (4) can be solved using any of a number of off-the-shelf methods for convex programming, including cutting-plane methods, interior-point methods, and second-order cone programming. Since the datasets that we consider are medium-to-large scale (see Section 3.2), interior-point methods are too slow for our problem and therefore we limit our study to *first-order* iterative convex programming methods that only use (sub)gradient information while making progress towards the optimal solution. Specifically, in our experiments we use

spectral projected gradient (SPGL1), which has been developed in the context of compressive sensing (Van Den Berg & Friedlander, 2008) for solving large-scale sparse optimization problems.

The LASSO can be viewed as a close relative of the RVM. Indeed, a Bayesian interpretation of the LASSO demonstrates that the solution to a LASSO problem is, in fact, the maximum *a* posteriori (MAP) estimate of the parameters **w**, when the prior $p(\mathbf{w})$ is specified by a multi-dimensional Laplace probability density function. Several studies have shown that convex optimization methods such as the LASSO exhibit typically faster convergence (in terms of number of iterations) than Bayesian inference approaches (Roth et al. 2008). However, in contrast to Bayesian methods, our LASSO-based formulation does not produce a full posterior distribution of the prediction parameters. In our experiments below, we compare the LASSO with the SBL approach (Tipping, 2001).

2.3. RUL Prediction using Bilinear Kernel Regression

Fundamentals of Bilinear Kernel Regression

Until now, we have discussed regression approaches for prediction that (implicitly) assumes that the training samples (capacity estimates as well as RUL responses) are error-free. However, in reality, measurements (i.e., cell voltage, current and temperature) are rarely pristine. Whether due to human, instrumentation, or computation errors, it is very likely that capacity estimates are corrupted. Corruptions can occur due to noise in the measurements, owing to faulty sensor operation or variability in the power load and temperature conditions, or errors by a capacity estimation algorithm. Corruptions can also occur due to outliers, owing to sensor failure or human errors.

Standard regression methods do not account for the possibility of such corruptions, and the consequence is that the inferred prediction model can be grossly incorrect, leading to unpredictable results while testing on new unseen data points. Our hypothesis is that we can build improved RUL prediction models if we explicitly capture and account for errors in the training data. We propose a unified optimization formulation for prediction of battery RUL from capacity estimates that addresses this hypothesis.

First, we propose a mathematical representation of corrupted observations as follows. Suppose the (estimated) capacity measurements available to the regression method are given by:

$$\mathbf{z}_i = \mathbf{x}_i + \mathbf{u}_i \tag{5}$$

where $\mathbf{u}_i$ is a vector of noise values whose dimension equals the feature dimension, and whose values are generated from some probability distribution. Consequently, we use the measurements $\mathbf{z}_i$ to construct a (contaminated) kernel matrix K using Eq. (3). The relation to the "true" kernel matrix is given by:

$$K = \Phi + E \tag{6}$$

where E is an error matrix. The two modes of corruption that we consider are both special cases of Eq. (6). For the additive noise model, we assume that the capacity estimates are contaminated with independent Gaussian noise, i.e., each estimate is perturbed by a small independently chosen random variable from a normal distribution. Up to a first-order approximation, the effect of such contamination can be modeled by assuming that the entries of E are i.i.d. samples from a Gaussian distribution with some variance σ^2. For the outlier noise model, we assume that the capacity measurements are contaminated with sparse (but unbounded) noise, i.e., a randomly chosen fraction of the observations are arbitrary distorted. Up to a first-order approximation, the effect of such contamination can be modeled via a sparsity assumption on the error matrix E.

Given the (contaminated) kernel matrix K and the measured (or true) RUL values **y**, we solve a generalization of the optimization problem in Eq. (4) by jointly estimating both the optimal prediction vector as well as the error matrix:

$$\begin{aligned}&(\widehat{\mathbf{w}}, \widehat{E})\\ &= \arg\min\ \lambda\|\mathbf{w}\|_1 + \tau\|vec(E)\|_p^p\\ &+ \|\mathbf{y} - (K - E)\mathbf{w}\|_2^2\end{aligned} \tag{7}$$

Here, the *vec*() operator vectorizes the contents of an arbitrary matrix in column-major order. The norm parameter p is set to be either 1 or 2 depending on the noise model; the case $p = 2$ models additive noise and encourages dense estimates of the error, while the case $p = 1$ models outlier noise. As before, the parameter λ controls the sparsity of the final solution, while the parameter τ controls the norm of the aggregate errors.

In theory, the solution to Eq. (7) will produce a sparse prediction vector $\widehat{\mathbf{w}}$ that fits a kernel regression model to the "denoised" kernel matrix $K - E$. The denoising is implicit since we simultaneously remove the noise in the kernel matrix as well as estimate the prediction model. We note that unlike Eq. (4), the optimization problem in Eq. (7) is no longer convex. In particular, the squared-error loss term in Eq. (7) is a bilinear function of the optimization variables **w**, E. Therefore, Eq. (7) is an instance of penalized bilinear regression. Variants of bilinear regression have been previously explored in the machine learning literature in (Herman & Strohmer, 2010). In particular, a similar optimization problem is proposed to develop robust versions of the LASSO that are less susceptible to outlier errors in the training data (Chen et al., 2013). To the best of our knowledge, the application of this method to battery-life prognostics has not been attempted. In our experiments below (see Section 3.4), we see that accounting for the errors in the

measurements leads to improvements over the standard LASSO in all test cases, sometimes by a large amount.

Algorithm for Bilinear Kernel Regression

Since, the optimization problem in Eq. (7) is non-convex, off-the-shelf solvers for nonlinear convex optimization cannot be directly used to solve this problem. However, due to the bilinear nature of the prediction error term in the objective function, we observe that the problem *is* convex, provided we fix either one of the variables (**w** or *E*) and optimize over the other variable. This motivates the following, natural two-step iterative procedure based on alternating minimization:

Step 1 Suppose we fix *E*. Then, minimizing the objective function in Eq. (7) over all possible prediction vectors **w** reduces to a variant of the original LASSO formulation. This sub-problem can be solved using convex optimization methods such as SPGL1.

Step 2 Suppose we fix **w**. Then, minimizing the objective function in Eq. (7) over all possible error matrices *E* reduces to an ℓ_p-regularized least squares problem. The ℓ_p-norm is convex for both $p = 1$ and $p = 2$. For $p = 1$, we can solve the sub-problem using a modification of SPGL1. For $p = 2$, the sub-problem can be reduced to ordinary penalized least-squares (also known as Tikhonov regularization (Tikhonov & Arsenin, 1977), and can be solved using standard least-squares methods such as conjugate gradients.

Table 1 summarizes the overall procedure to solve the optimization problem in Eq. (7). The basic idea is to alternate between solving for *E* and **w**. In the limit of infinitely many iterations, this procedure will converge to a *local* minimum of the objective function in Eq. (7). In practice, we can only execute a finite number of iterations. Therefore, we fix an input parameter *T* (representing the maximum allowable iteration count) and at the end of each iteration, we record the prediction error. The final estimates $\widehat{\mathbf{w}}, \hat{E}$ are declared by determining the iteration index that yielded the minimum objective function. The global optimality of such a method (and for non-convex optimization algorithms in general) cannot be guaranteed, but it serves as an effective heuristic. We leave as an open question the theoretical analysis of the above alternating minimization approach.

3. EXPERIMENTAL RESULTS

The verification of the proposed approach was accomplished by using 10 years' continuous cycling data acquired from eight Li-ion prismatic cells. This section reports the results of this verification. Section 3.1 presents the test procedure along with the cycling performance of the test cells. Section 3.2 gives the implementation details of several different methods. Section 3.3 describes the error metric used to quantify the performance of these methods in RUL prediction. The RUL prediction results are reported in Section 3.4.

Table 1. A pseudocode representation of the proposed approach for RUL prediction of Li-ion battery.

ALGORITHM: *Alternating minimization for regularized bilinear regression.*

INPUTS: Training data $\{(\mathbf{x}_i, y_i)\}$, i=1, 2, …, n.

OUTPUTS: Estimated kernel prediction vector $\widehat{\mathbf{w}}$.

PARAMETERS: Optimization parameters λ and τ, kernel bandwidth r, number of iterations T

1. **Initialize**: $\widehat{\mathbf{w}_0} \leftarrow 0, \widehat{E_0} \leftarrow 0, t \leftarrow 0$.
2. **Compute**: the kernel matrix K using Equation (2).
3. **While** $t < T$ **do:**
   a. $t \leftarrow t + 1$
   b. Set $\overline{K} = K - \widehat{E_t}$. Solve:

$$\widehat{\mathbf{w}}_{t+1} = \arg\min\ \lambda\|\mathbf{w}\|_1 + \|\mathbf{y} - \overline{K}\mathbf{w}\|_2^2$$

   c. Set $\bar{\mathbf{y}} = \mathbf{y} - K\widehat{\mathbf{w}}_{t+1}$. Solve:

$$\hat{E}_{t+1} = \arg\min\ \tau\|vec(E)\|_p^p + \|\bar{\mathbf{y}} + E\widehat{\mathbf{w}}_{t+1}\|_2^2$$

   d. Record prediction error:

$$PredErr(t) = \left\|\mathbf{y} - \left(K - \hat{E}_{t+1}\right)\widehat{\mathbf{w}}_{t+1}\right\|_2^2$$

4. **Find** t^* that minimizes $Prederr(t)$.
5. **Output:** $\widehat{\mathbf{w}} \leftarrow \widehat{\mathbf{w}}_{t^*}$

3.1. Test Procedure and Cycling Data

Li-ion cells were constructed in hermetically sealed prismatic cases between 2002 and 2012 and subjected to full depth of discharge cycling with a nominally weekly discharge rate (*C*/168 discharge) under 37°C (Hu, et al., 2014). The cycling test was conducted with the following parameter settings: (i) the charge rate (I_{CC}) for the CC charge was *C*/6; (ii) the charge cutoff voltage (V_{max}) was 4.075 V; (iii) the time duration ($t_{CV} - t_{CC}$) of the CV charge was 30 min; and (iv) the discharge rate was *C*/150 or a nominally weekly discharge rate. The test attempted to simulate a use condition similar to patient use in medical applications. The weekly rate discharge capacities are plotted against the time on test in Fig. 2. Please note that, for confidentiality reasons, the discharge capacity of a cell in Fig. 2 and in the discussions thereafter is presented after

being normalized by the beginning-of-life (BOL) discharge capacity of the cell. As shown in Fig. 2, 80% of the initial capacity is retained even after 10 years of repeated cycling at an elevated temperature, indicating exceedingly stable cell performance. The cycling data also indicate consistent performance of cells manufactured over a long time period.

3.2. Prognostic Data Generation and Method Implementation

In this experimental study, the cycling data from the eight 2002 cells in Fig. 2 were used to verify the effectiveness of the proposed approach in the RUL prediction. Each feature vector (or data point) $\mathbf{x}_i$ consists of the 3 most-recently measured capacities (i.e., $m = 3$). To focus our discussion on RUL prediction, we did not implement capacity estimation in this study, and instead used measured capacities to construct the feature vectors. Each feature vector, $\mathbf{x}_i$, in the training data set was corrupted with additive noise, n_i, where n_i is a random sample from a zero-mean normal distribution with standard deviation σ taking one of the following: 0.0, 0.005, 0.010 and 0.015. For each σ value, all methods were tested using two 8-fold cross validation (CV) experiments: the first where the test data contained no additive noise, and the second where the test data was corrupted with additive noise from the same normal distribution as the training data. To minimize the effect of randomness in additive noise, the data generation and 8-fold cross validation were repeated 10 times.

A cell is considered to reach the EOL when the measured discharge capacity of the cell fades to 78.5% of its initial discharge capacity (Hu et al., 2014). For any test cell whose measured capacities did not reach this EOL limit, the EOL of the cell was identified through a linear extrapolation of the capacity data from the last six charge/discharge cycles. To detect outliers in RUL prediction data, caused by spurious capacity readings input into the proposed models, a linear fit of the data was performed using the robust regression described by Holland in (Holland et al., 1977). Any residual greater than 15 median absolute deviations was removed from the prediction data and not used in the calculation of error.

The tradeoff parameters, λ in Eq. (4) and λ and τ in Eq. (7), were determined empirically, and for each, the value minimizing the overall root-mean-square (RMS) error (see the definition in Section 3.3) of a $\sigma = 0.01$ CV was used for all trials. For both LASSO and bilinear regression, λ was 36,000. For estim ation of the error matrix, τ_{LASSO} and $\tau_{Tikhonov}$ were 26,000 and 16,000, respectively. The kernel bandwith, r, in Eq. (2) was also empirically determined, and was 0.05 for LASSO and bilinear regression and 0.2 for RVM.

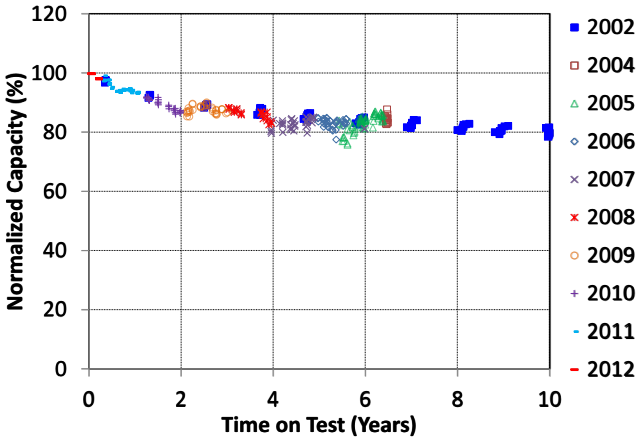

Figure 2. Cycling performance of cells manufactured and cycled between 2002 and 2012 (Hu, et al., 2014).

3.3. Error Metric

The RUL is used as the relevant metric for determining the state of life (SOL) of Li-ion battery. We compare the prediction performance of the proposed methods (LASSO, bilinear regression with Tikhonov Regularization (our proposed algorithm with p=2) to estimate errors, and bilinear regression using LASSO to estimate errors (our proposed algorithm with p=1)) to that of RVM described in (Hu et al., 2015). The accuracy of a method was evaluated by using the k-fold CV. In this study, the complete feature data set $\mathbf{X}$ consists of eight mutually exclusive subsets or folds $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_8$ that were respectively obtained from the eight 2002 cells. In each CV trial, of the eight subsets, one was used as the test set and the other seven subsets were put together as a training set. The CV process was performed eight times (i.e., the total number of CV trials is eight), with each of the eight subsets left out exactly once as the test set. Thus, all the data points in the complete data set were used for both training and testing. Let $\mathbf{I}_l = \{i: \mathbf{x}_i \in \mathbf{X}_l\}$, $l = 1, 2, \ldots, 8$, denote the index set of the feature vectors that construct the subset $\mathbf{X}_l$. The CV root mean square error (RMSE) is computed as the root square of the average error over all the eight CV trials, expressed as

$$RMSE = \sqrt{\frac{1}{U}\sum_{l=1}^{8}\sum_{i\in\mathbf{I}_l}\left(\hat{y}_{\mathbf{X}\setminus\mathbf{X}_l}(\mathbf{x}_i) - y(\mathbf{x}_i)\right)^2} \qquad (8)$$

where U is the number of feature vectors for the CV, $\hat{y}_{\mathbf{X}\setminus\mathbf{X}_l}$ is the predicted RUL by the method trained with the complete data set $\mathbf{X}$ excluding the subset $\mathbf{X}_l$, and $y(\mathbf{x}_i)$ is the true RUL of $\mathbf{x}_i$. The error formula in Eq. (8) indicates that all the U feature vectors in the complete data set $\mathbf{X}$ are used for both training and testing, and each feature vector is used for testing exactly once and for training seven times.

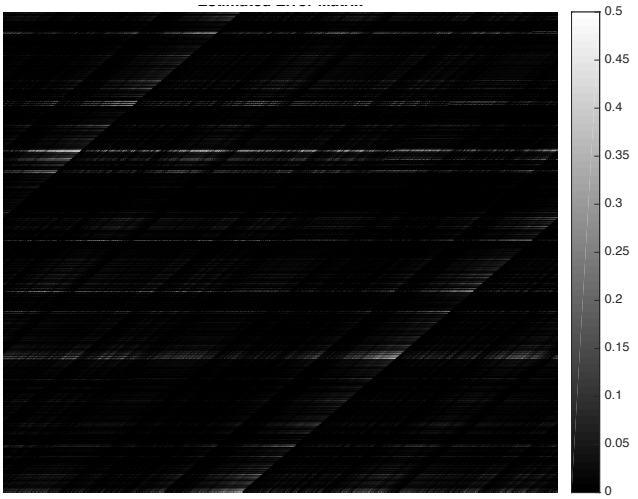

Figure 3. Estimated error matrix created as an intermediate step of bilinear regression with Tikhonov Regularization where $\sigma = 0.01$. Errors are reported as a percentage of the maximum value in the Gaussian kernel, κ.

3.4. RUL Prediction Results

To determine the accuracy of a given prediction method, the errors across all cells of the CV were aggregated and a single RMS value was calculated using Eq. (8). Table 2 summarizes the accuracy of all prediction methods with variable amounts of noise in the training data and test data. With uncorrupted test data, RVM was outperformed by all prediction methods proposed in this paper; LASSO without error estimation was outperformed by both methods incorporating error estimation. Moreover, error modeling with Tikhonov regularization outperformed error modeling with LASSO for the three of the four test cases. When test data is corrupted with small amounts of noise all models provide predictions similar to the error-free predictions in Table 2; greater levels of noise in the test data overwhelms the models ability to make accurate predictions.

Figure 3 demonstrates an example of an estimated error matrix (displayed as a grayscale image) generated in one of the intermediate steps of bilinear regression with Tikhonov regularization. If any data point $\mathbf{x}_i$ is imbued with error, we would expect our estimation of the i^{th} row of our error matrix to be mostly non-zero (i.e. error-filled). This phenomenon is demonstrated in Fig. 3, and by inspection we can infer that the rows with non-zero entries correspond to erroneous data points.

Figure 4 shows estimates of the Cumulative Density Function (CDF) for all prediction methods. The CDF estimate was created using the aggregated absolute value of errors from an 8-cell CV. In this CV, bilinear estimation using Tikhonov regularization outperforms all prediction methods and, as expected, a larger percentage of its errors are small. For all methods explored in this paper over 60% of predictions are within 30 cycles of the true RUL.

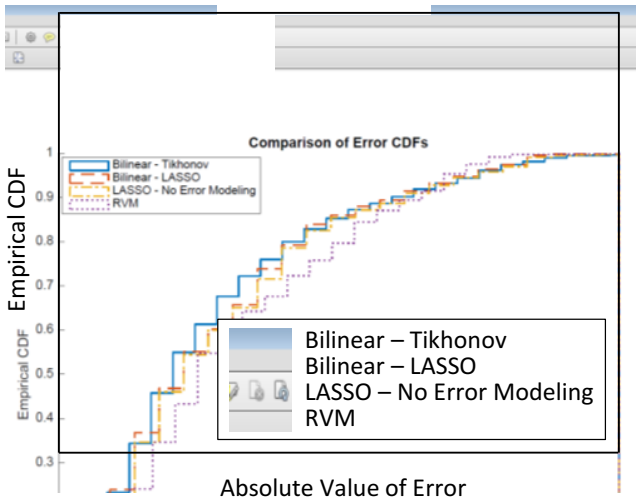

Figure 4. Empirical CDF vs. absolute value of error for each prediction method. Here, $\sigma = 0.01$ for the training dataset; the test data was not corrupted. RMSE – LASSO: 35.3. RMSE – Bilinear-LASSO: 34.5. RMSE – Bilinear-Tikhonov: 33.8. RMSE – RVM: 36.7.

4. CONCLUSION

This paper presents a data-driven approach to online RUL prediction of Li-ion battery by adopting bilinear kernel regression. This approach provides individual users of Li-ion battery-powered devices with estimates of the battery RUL over the whole service life. The RUL allows the users to schedule an optimal replacement near the EOL so that the devices can be used as long as possible, and at the same time, users' safety is not compromised. Our contributions to battery prognostics include the formulation of a general framework for robust sparse prognostics, and the development of two numerical procedures for efficient error estimation. Experiments with 10 years' continuous cycling data verify that the proposed approach achieves more accurate RUL prediction than existing data-driven approaches, and suggest that the proposed method is a promising methodology for the battery prognostics.

It is important to note that the experimental data in Section 3 were obtained from the eight Li-ion cells cycled with a constant discharge rate. Since the fade behavior is fairly consistent among the eight cells (see Fig. 2), a training data set, which carries the information about the fade behavior of 7 training cells, is likely to be capable of capturing the fade behavior of the testing cell. In non-medical applications (e.g., hybrid and electric vehicles, and consumer electronics) where harsher and more inconsistent fade scenarios are often

encountered, the training data set may not fully represent the way a testing cell degrades, and in such cases, the data-driven methods discussed in this paper may produce inaccurate RUL predictions. Our future work will assess the effectiveness of the proposed data-driven methods in the presence of significant cell-to-cell variation in capacity fade as well as investigate the effect of dynamic loading conditions on the accuracy in RUL prediction. Finally, we will also investigate Bayesian-inference algorithms that quantify the uncertainty in the predicted RUL estimates as a function of the noise level.

Table 2. A summary of prediction accuracy of different methods. Predictions were performed using error-free test data and using additive noise where σ is the same for both training and test data. The standard deviation of the ten 8-fold CVs is presented with each RMSE.

Prediction method	Noise in training data (σ)				Noise in training and test data (σ)		
	0	0.005	0.01	0.015	0.005	0.01	0.015
LASSO	31.24 ($\pm$0.00)	33.05 ($\pm$1.26)	34.72 ($\pm$2.33)	50.42 ($\pm$5.73)	42.33 ($\pm$4.77)	61.53 ($\pm$5.54)	83.67 ($\pm$11.23)
Bilinear regression with LASSO	30.26 ($\pm$0.00)	31.62 ($\pm$1.47)	33.28 ($\pm$2.37)	**46.22** ($\pm$5.54)	40.96 ($\pm$4.32)	**60.16** ($\pm$5.11)	**82.40** ($\pm$10.06)
Bilinear regression with Tikhonov regularization	**29.57** ($\pm$0.00)	**30.92** ($\pm$1.49)	**32.8** ($\pm$2.30)	48.88 ($\pm$6.16)	**40.57** ($\pm$4.84)	60.58 ($\pm$4.99)	83.52 ($\pm$11.15)
RVM	30.91 ($\pm$0.00)	32.67 ($\pm$1.00)	36.16 ($\pm$2.44)	47.67 ($\pm$4.24)	41.50 ($\pm$4.65)	60.22 ($\pm$5.77)	82.58 ($\pm$10.96)

REFERENCES

Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge university press.

Chen, Y., Caramanis, C., & Mannor, S. (2013). Robust Sparse Regression under Adversarial Corruption. In *ICML (3)* (pp. 774-782).

Coble, J. B., & Hines, J. W. (2008, October). Prognostic algorithm categorization with PHM challenge application. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on* (pp. 1-11). IEEE.

Dickerson, A., Rajamani, R., Boost, M., & Jackson, J. (2015). *Determining Remaining Useful Life for Li-ion Batteries* (No. 2015-01-2584). SAE Technical Paper.

Gebraeel, N. Z., Lawley, M. A., Li, R., & Ryan, J. K. (2005). Residual-life distributions from component degradation signals: A Bayesian approach. *IiE Transactions*, *37*(6), 543-557.

Gebraeel, N., & Pan, J. (2008). Prognostic degradation models for computing and updating residual life distributions in a time-varying environment. *IEEE Transactions on Reliability*, *57*(4), 539-550.

Goebel, K., Eklund, N., & Bonanni, P. (2006, March). Fusing competing prediction algorithms for prognostics. In *2006 IEEE Aerospace Conference*(pp. 10-pp). IEEE.

He, W., Williard, N., Chen, C., & Pecht, M. (2013). State of charge estimation for electric vehicle batteries using unscented Kalman filtering.*Microelectronics Reliability*, *53*(6), 840-847.

Heimes, F. O. (2008, October). Recurrent neural networks for remaining useful life estimation. In Prognostics and Health Management, 2008. PHM 2008. International Conference on (pp. 1-6). IEEE.

Herman, M. A., & Strohmer, T. (2010). General deviants: An analysis of perturbations in compressed sensing. *IEEE Journal of Selected Topics in Signal Processing*, *4*(2), 342-349.

Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. Communications in Statistics-theory and Methods,6(9), 813-827.

Hu, C., Youn, B. D., & Chung, J. (2012). A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation.*Applied Energy*, *92*, 694-704.

Hu, C., Youn, B. D., Wang, P., & Yoon, J. T. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering & System Safety*, *103*, 120-135.

Hu, C., Jain, G., Schmidt, C., Strief, C., & Sullivan, M. (2015). Online estimation of lithium-ion battery capacity using sparse Bayesian learning. *Journal of Power Sources*, *289*, 105-113.

Hu, C., Jain, G., Tamirisa, P., & Gorka, T. (2014). Method for estimating capacity and predicting remaining useful life of lithium-ion battery. *Applied Energy*, *126*, 182-189.

Hu, X., Jiang, J., Cao, D., & Egardt, B. (2016). Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling. *IEEE*

Transactions on Industrial Electronics, *63*(4), 2645-2656.

Lee, S., Kim, J., Lee, J., & Cho, B. H. (2008). State-of-charge and capacity estimation of lithium-ion battery using a new open-circuit voltage versus state-of-charge. *Journal of power sources*, *185*(2), 1367-1373.

Liu, J., Saxena, A., Goebel, K., Saha, B., & Wang, W. (2010). *An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries*. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION MOFFETT FIELD CA AMES RESEARCH CENTER.

Liu, J., Wang, W., Ma, F., Yang, Y. B., & Yang, C. S. (2012). A data-model-fusion prognostic framework for dynamic system state forecasting.Engineering Applications of Artificial Intelligence, 25(4), 814-823.

Lu, L., Han, X., Li, J., Hua, J., & Ouyang, M. (2013). A review on the key issues for lithium-ion battery management in electric vehicles. *Journal of power sources*, *226*, 272-288.

Luo, J., Pattipati, K. R., Qiao, L., & Chigusa, S. (2008). Model-based prognostic techniques applied to a suspension system. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *38*(5), 1156-1168.

Plett, G. L. (2004). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification. *Journal of power sources*, *134*(2), 262-276.

Plett, G. L. (2004). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation. *Journal of Power Sources*, *134*(2), 277-292.

Roth, V. (2001, August). Sparse kernel regressors. In *International Conference on Artificial Neural Networks* (pp. 339-346). Springer Berlin Heidelberg.

Saha, B., & Goebel, K. (2009). Modeling Li-ion battery capacity depletion in a particle filtering framework. In *Proceedings of the annual conference of the prognostics and health management society* (pp. 2909-2924).

Saha, B., Goebel, K., Poll, S., & Christophersen, J. (2009). Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Transactions on Instrumentation and Measurement*, *58*(2), 291-296.

Si, X. S., Wang, W., Hu, C. H., & Zhou, D. H. (2011). Remaining useful life estimation–A review on the statistical data driven approaches. European Journal of Operational Research, 213(1), 1-14.

Si, X. S., Wang, W., Hu, C. H., Chen, M. Y., & Zhou, D. H. (2013). A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation. *Mechanical Systems and Signal Processing*,*35*(1), 219-237.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Methods for solving ill-posed problems*. John Wiley and Sons, Inc.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research*, *1*(Jun), 211-244.

Trefethen, L. N., & Bau III, D. (1997). Numerical linear algebra (Vol. 50). Siam.

Van Den Berg, E., & Friedlander, M. P. (2008). Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, *31*(2), 890-912.

Wang, D., Miao, Q., & Pecht, M. (2013). Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. *Journal of Power Sources*, *239*, 253-264.

Wang, P., Youn, B. D., & Hu, C. (2012). A generic probabilistic framework for structural health prognostics and uncertainty management. Mechanical Systems and Signal Processing, 28, 622-637.

Wang, T., Yu, J., Siegel, D., & Lee, J. (2008, October). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on* (pp. 1-6). IEEE.

Xiong, R., Sun, F., Chen, Z., & He, H. (2014). A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion olymer battery in electric vehicles. *Applied Energy*, *113*, 463-476.