Data-Driven Prognostics of Lithium-Ion Rechargeable Batteries Using Bilinear Kernel Regression

Charlie Hubbard, John Bavlsik, Chinmay Hegde and Chao Hu

Iowa State University

Applications

- Electric/hybrid vehicles

- Mobile devices

- Medical devices

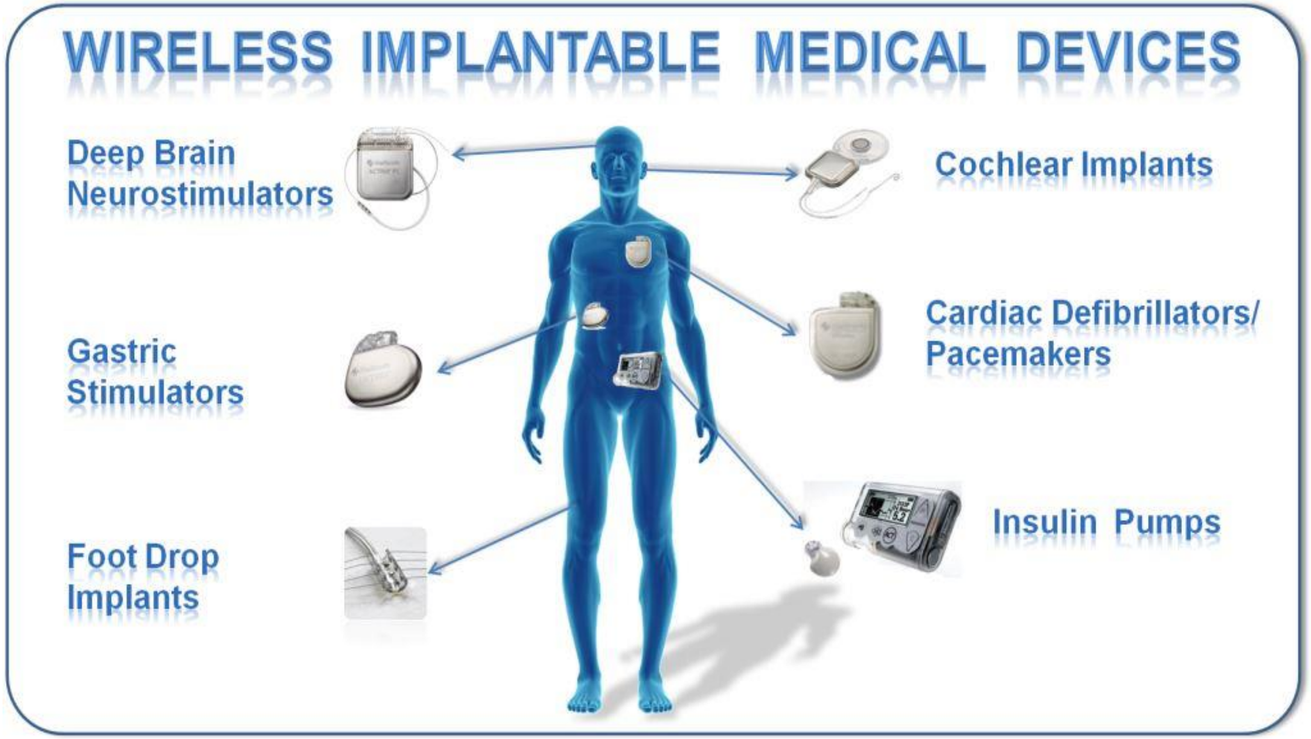


Image credit: http://uhaweb.hartford.edu/; https://rahulmittal.files.wordpress.com; http://www.carsdirect.com/

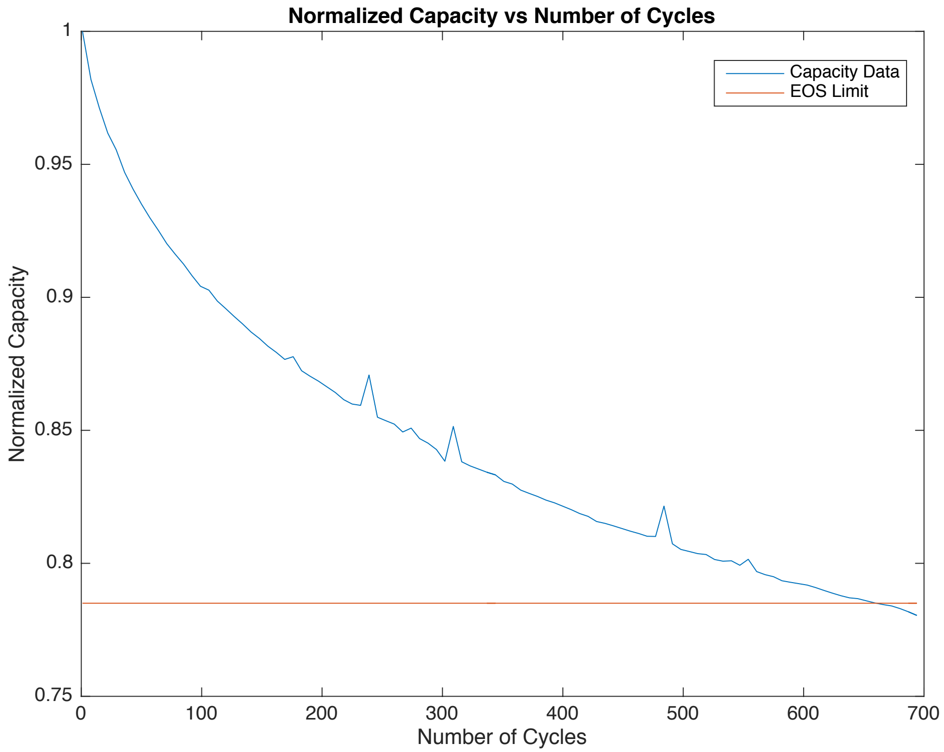
Health of an Li-Ion Battery

- Available signals
    - Current
    - Voltage
    - Internal temperature
- Battery-health indicators
    - State of Charge (SOC)
    - State of Health (SOH)
    - Remaining Useful Life (RUL)

Image credit: www.ionarchive.com

Remaining Useful Life Estimation

- Remaining Useful Life (RUL)
  - End of Service limit determined by application
  - Measured in charge/discharge cycles



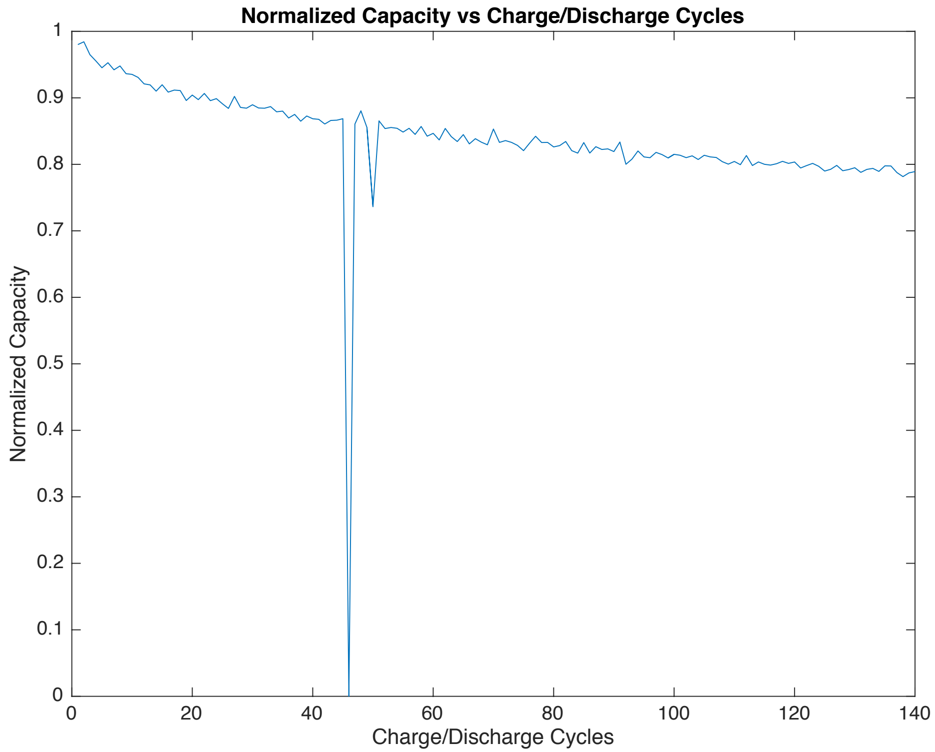
Methods for SOH/RUL Estimation

- Model-based
  - Assume knowledge of physical system
  - Fit data to model of system
    - Si, et. all 2013
    - Gebraeel & Pan 2008
- Data-driven
  - Assume no knowledge of physical system
  - Machine learning/pattern recognition methods determine model
    - Neural Networks – Liu, et. all 2010;
    - Relevance Vector Machines – Hu, et. all 2015;

Data Deficiencies

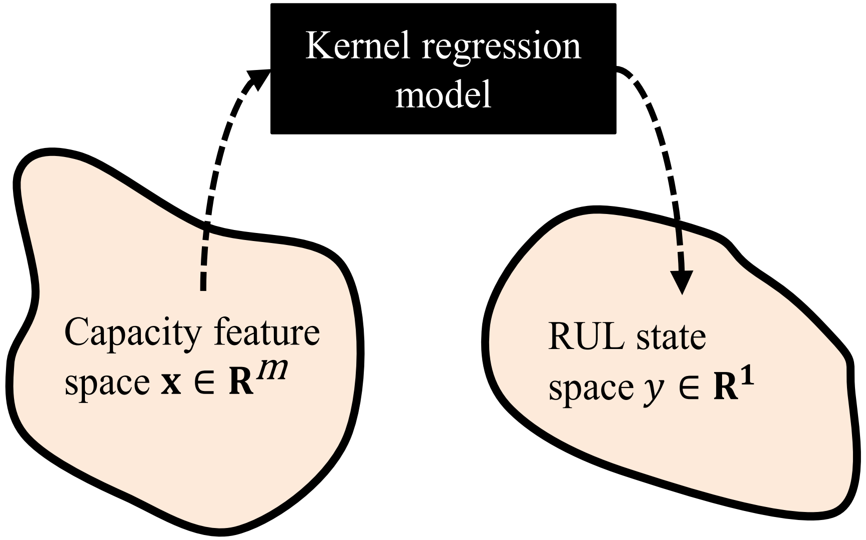
- Types of deficiencies
  - Missing data
  - Noise

- Handling of deficiencies
  - Remove spurious data points
  - Allow model to ignore them



Our Approach

- Data-driven
- Least-squares regression based
  - Add additional regression step to find (and remove) noise in training data
  - Add kernel to transform data
  - Add regularization to encourage sparsity



Least-squares Regression

- Feature vector: x
  - Contains n most recent capacity readings
- Data matrix: $X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$

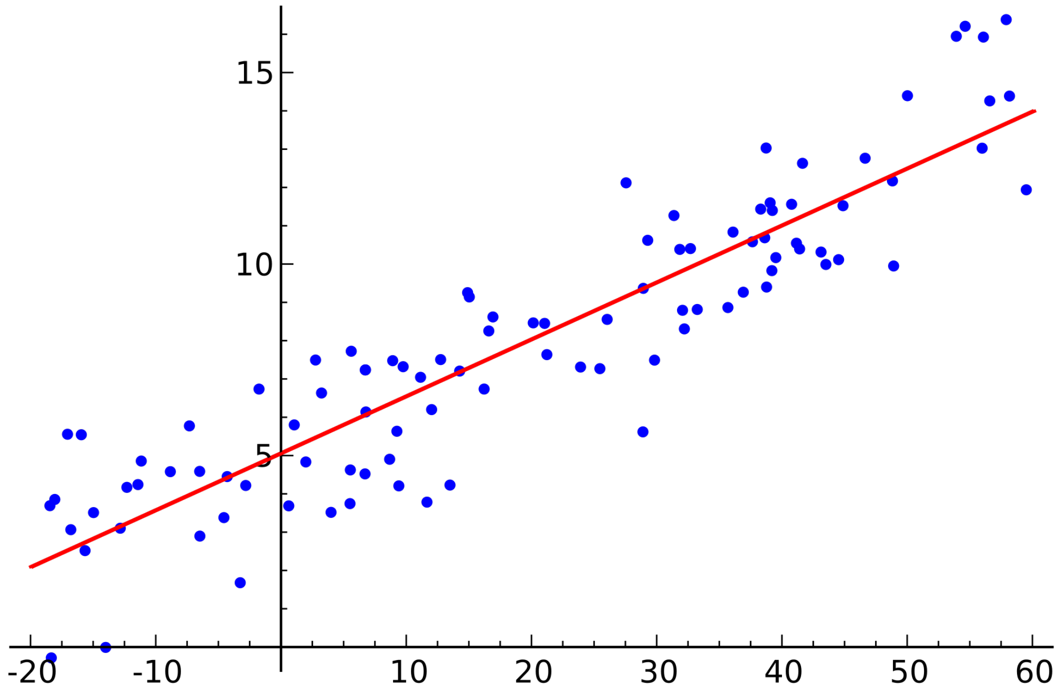
- Empirically determined RUL: y
- RUL vector: $Y = \begin{matrix} y_1 \\ \vdots \\ y_m \end{matrix}$

Least-squares Regression

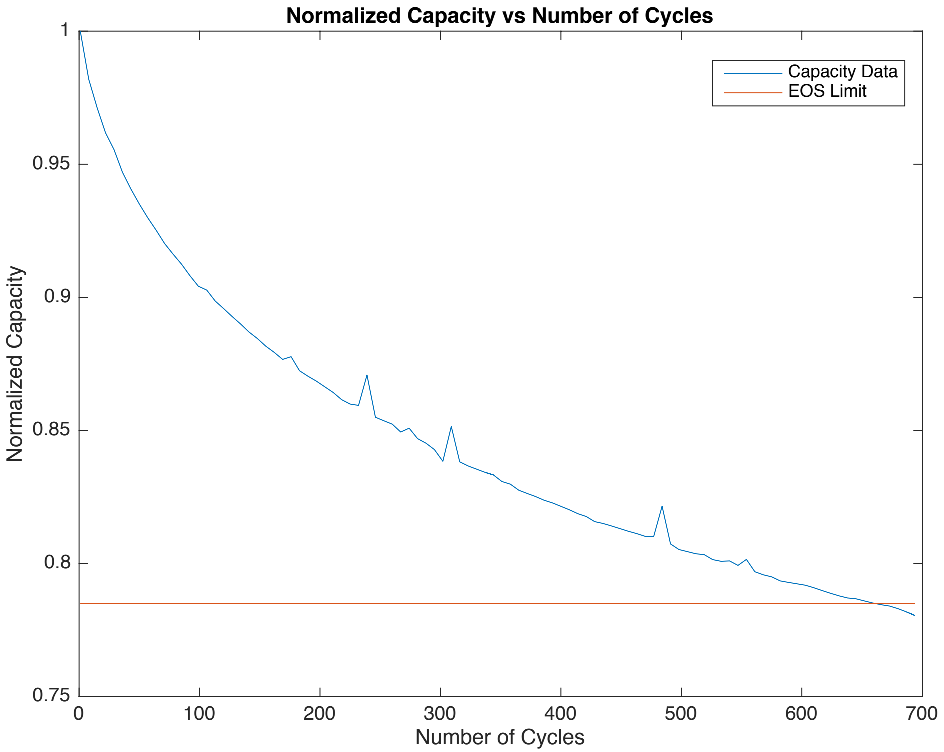
- Given X and Y, solve:

$$\min_{w} \|Y - Xw\|_2^2$$

- Prediction vector: $\widehat{w}$
- Predicted RUL: $\hat{y} = \langle x, \widehat{w} \rangle$



Kernel Least-Squares Regression



- Replace X with kernel matrix, K
  - Transform data to higher-dimensional space
  - Allows us to fit linear predictor, w to non-linear data
- Kernel choice is flexible
  - Dependent on data set
  - Gaussian Kernel is a common choice

Gaussian Kernel Matrix, K

- i^{th} row of K given by:

$$K_i = \exp(-\frac{\|x - x_i\|_2^2}{r^2})$$

- Add a vector of ones as the first column of K
  - Allow of non-zero “y-intercept” for linear function

$$K_{i,1} = \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}$$

$$K_{i,(j\neq 1)} =$$

$$\begin{pmatrix} \exp(-\frac{\|x_1 - x_1\|_2^2}{r^2}) & \cdots & \exp(-\frac{\|x_1 - x_{m+1}\|_2^2}{r^2}) \\ \vdots & \ddots & \vdots \\ \exp(-\frac{\|x_m - x_1\|_2^2}{r^2}) & \cdots & \exp(-\frac{\|x_m - x_{m+1}\|_2^2}{r^2}) \end{pmatrix}$$

Kernel Least-Squares Regression

- Given X and Y, obtain $\widehat{w}$ by solving:

$$\min_{w} \ \|Y - Kw\|_2^2$$

- Given feature vector x, compute: $k(x)\ where\ k(x)_1 = 1,$

$$k(x)_j = \exp\left(-\frac{\|x - x_j\|_2^2}{r^2}\right), j = \{2, \dots, m+1\}$$

- Predicted RUL: $\hat{y} = \langle k(x), \widehat{w} \rangle$

Kernel Regression and LASSO

Overfitting

- Kernel regression can over-fit to training data
- Forcing w to be sparse can prevent over-fitting
  - w will be allowed to have only a few non-zero entries
  - Only the most representative points will be used to calculate $\hat{y}$

Enforcing Sparsity

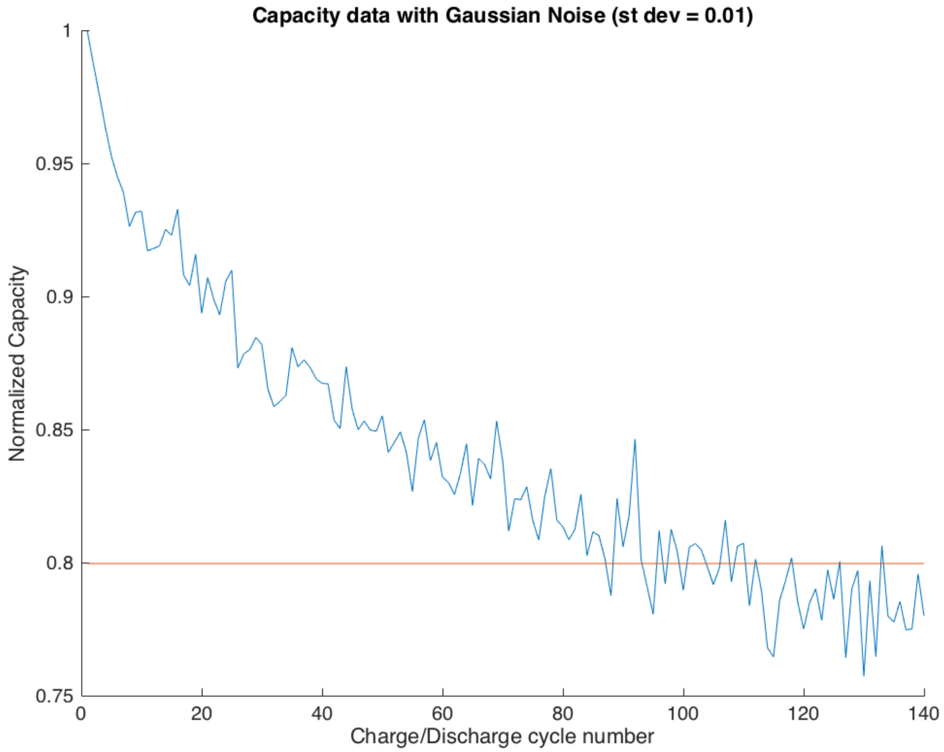
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Rewrite regression problem as:

$$\min_{w} \ \|Y - Kw\|_2^2 + \lambda\|w\|_1$$

Noise and Bilinear Regression

- Noise appears in almost every data set

- Writing $K = K_{true} + E$ we look to find and remove errors from K to obtain more accurate predictions



Bilinear Regression

- Kernel regression with LASSO (noiseless training data):

$$\min_{w} \|Y - K_{true}w\|_2^2 + \lambda\|w\|_1$$

- With noise in training data: $K - E = K_{true}$

$$\min_{w,E} \|Y - (K - E)w\|_2^2 + \lambda\|w\|_1$$

- Include a regularization term on E

$$\min_{w,E} \|Y - (K - E)w\|_2^2 + \lambda\|w\|_1 + \tau\|E\|_p^p$$

Algorithm – Bilinear Regression

Setup

- **INPUTS:** Training data $\{(\mathbf{x}_i, y_i)\}$,

  $i=1, 2, \ldots, n$.
- **OUTPUTS:** Estimated kernel prediction vector $\widehat{\mathbf{w}}$.
- **PARAMETERS:** Optimization parameters λ and τ, kernel bandwidth r, number of iterations T

Algorithm

Initialize: $\widehat{\mathbf{w}_0} \leftarrow 0, \widehat{E_0} \leftarrow 0, t \leftarrow 0.$

Compute: the kernel matrix K.

While $t < T$ **do:**

$t \leftarrow t + 1$

Set $\bar{K} = K - \widehat{E_t}$. Solve:

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

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

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

Record prediction error:

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

Find t^* that minimizes $Prederr(t)$.

Output: $\widehat{\mathbf{w}} \leftarrow \widehat{\mathbf{w}}_{t^*}$

Data Set

Test Cells

- 8 test cells
- Normalized capacity data from 700 charge/discharge cycles
  - Cells were cycled from 2002-2012
  - Weekly discharge rate

Noise Addition

- Imbued with Gaussian noise
  - Zero mean
  - std. dev. σ = {0, 0.005, 0.010, 0.015}

Experiment Setup

Cross Validations and Error Metric

- Leave-one-out cross validations (CVs)
- Root mean squared error (RMSE)

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

Test Procedure

- Feature vectors: three most recent capacity readings
- For each noise level σ
  - Perform two 8-fold CVs

Results

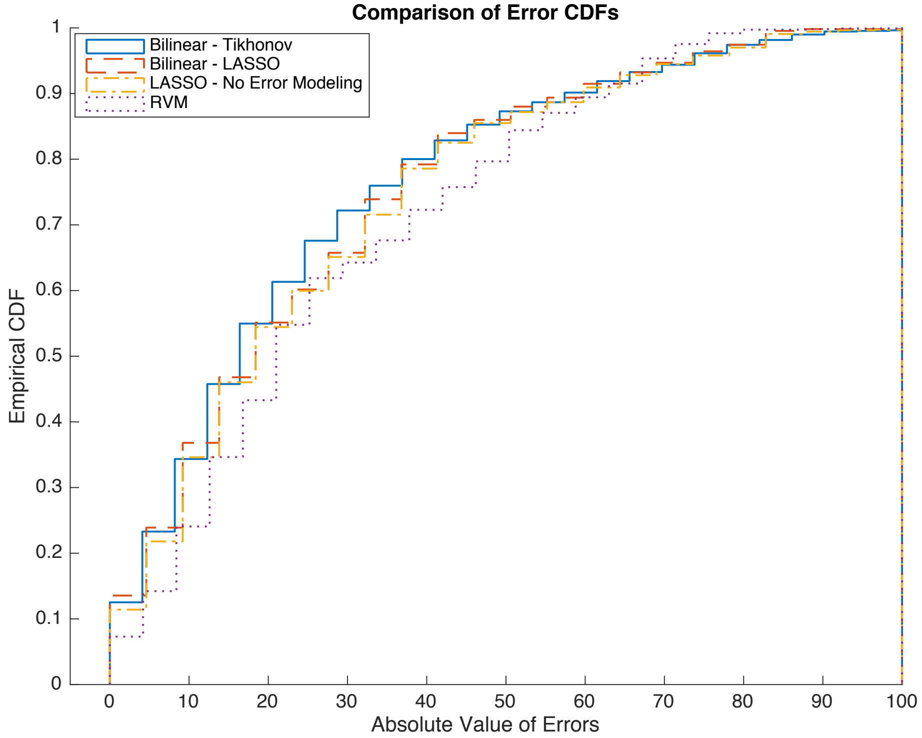
Prediction Method	RMSE						
	Noise in Training Data (std. dev of Gaussian noise)				Noise in Test and Training Data (std. dev of Gaussian noise)		
	0	0.005	0.01	0.015	0.005	0.01	0.015
Lasso	31.24	33.052	34.72	50.42	42.33	61.53	83.67
Bi-Lasso	30.26	31.62	33.28	46.22	40.96	60.16	82.40
Bi-Tikhonov	29.57	30.92	32.80	48.88	40.57	60.58	83.52
RVM	30.91	32.67	36.16	47.67	41.50	60.22	82.58

Results

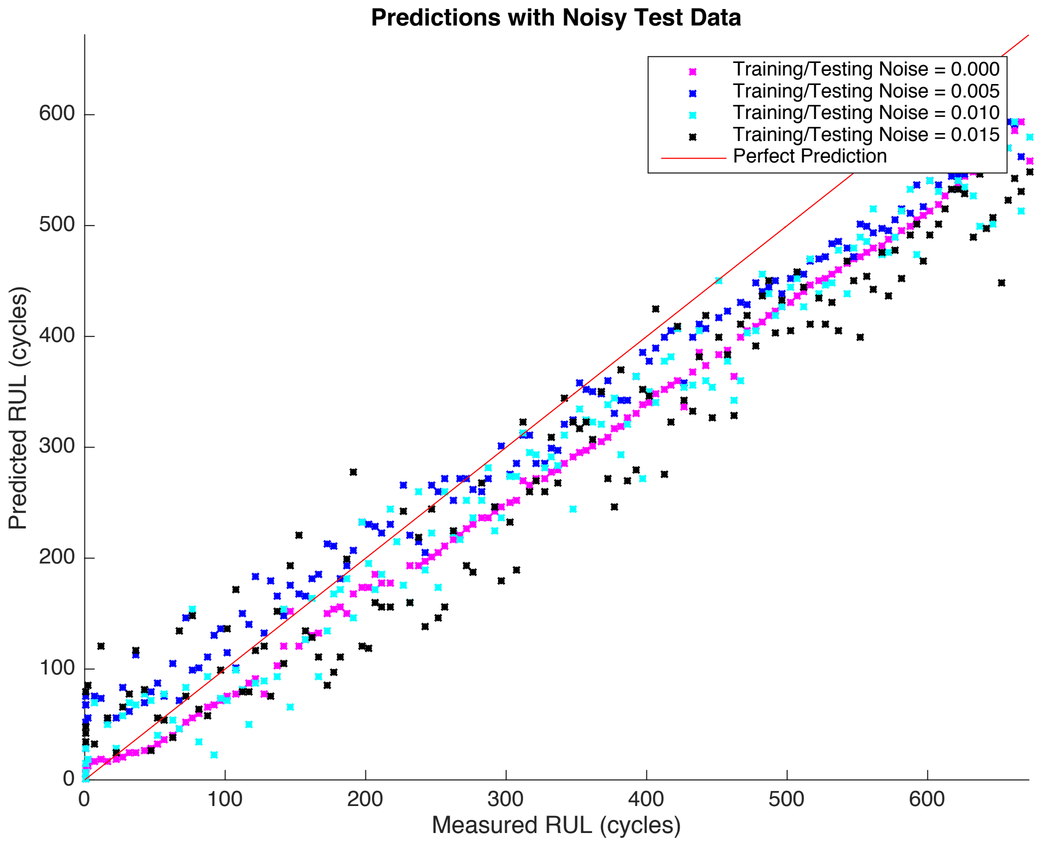
Noise-free test data

	RMSE			
	0	**0.005**	**0.01**	**0.015**
Lasso	31.24	33.052	34.721	50.415
Bi-Lasso	30.259	31.615	33.282	46.222
Bi-Tikhonov	29.572	30.921	32.8	48.88
RVM	30.91	32.67	36.16	47.67

Error CDF's

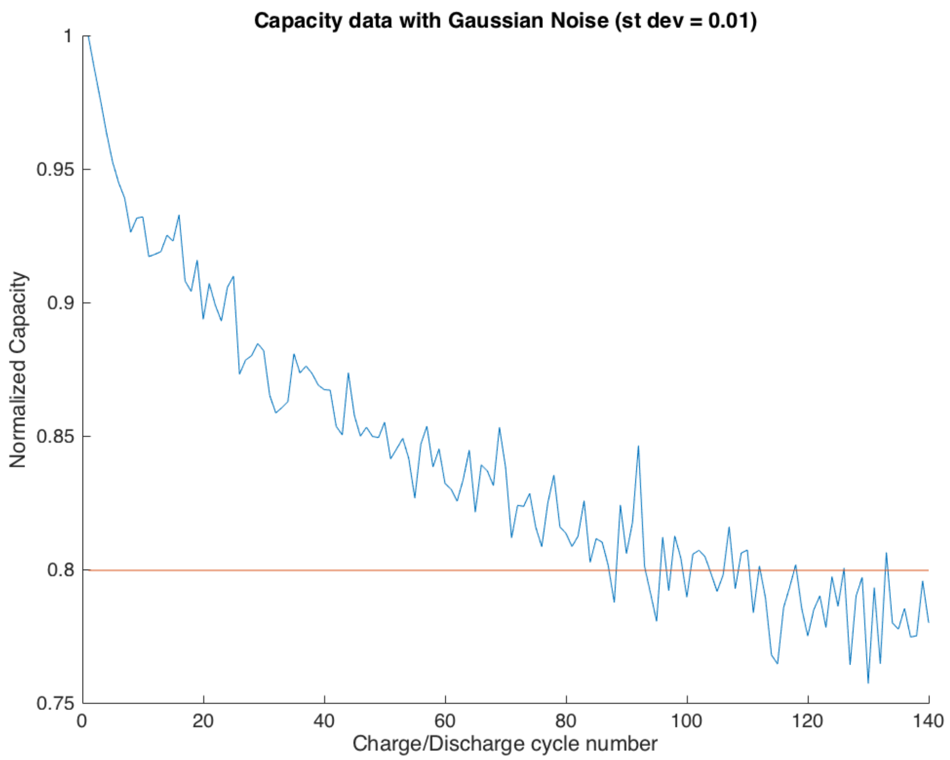


Testing with Noisy Data



Summary

- RUL predictions in batteries can be of critical importance
- Capacity fade data is nonlinear and often noisy
- Our model leverages:
    - Error estimation (and removal)
    - Data transformations
    - Sparse predictions
- Key contribution: Provide accurate RUL estimation in the presence of noisy data



References

- Hubbard, Bavlisk, Hu, Hegde (2016).
  Data-Driven Prognostics of Lithium-Ion Rechargeable Batteries Using Bilinear Kernel Regression. Annual Conference of the Prognostics and Health Management Society 2016.
- 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.
- 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.
- 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.
- 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