

Model Calibration in Latent Response Space using Principal Component Analysis



Lorraine Guerin¹, Kyle Neal², Joshua Mullins³, Benjamin Schroeder³, Scott Roberts³

1. ATA Engineering, Inc.
2. Vanderbilt University
3. Sandia National Laboratories

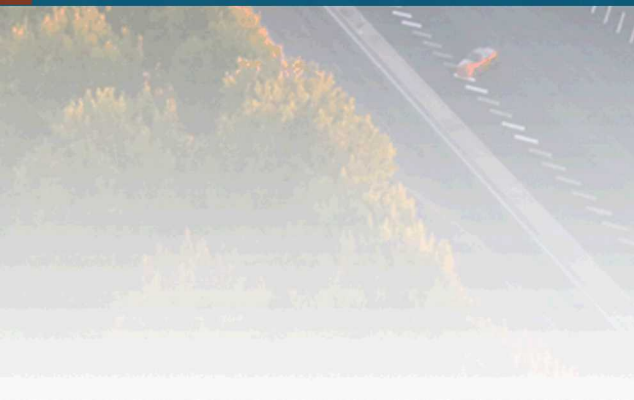
ASME V&V Symposium
May 20-22, 2020

SAND xxxxxxxx

- Motivation:
 - Field data and the challenges it poses to model calibration
- Principal component analysis (PCA):
 - What is it?
 - How can it improve model calibration with field data outputs?
- Example problem
- Conclusion



Field data and its challenges

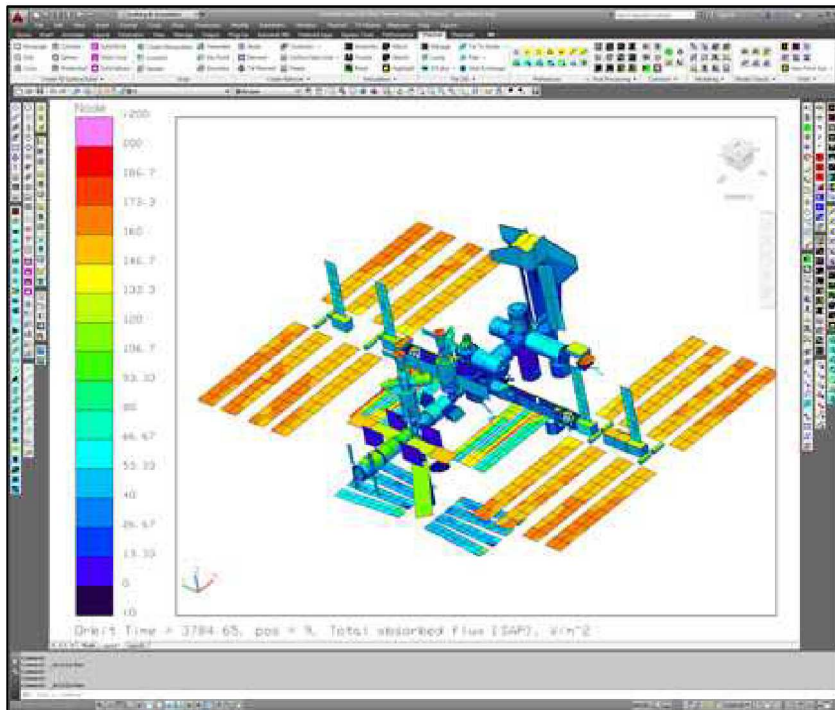


Field data output is common for computational models

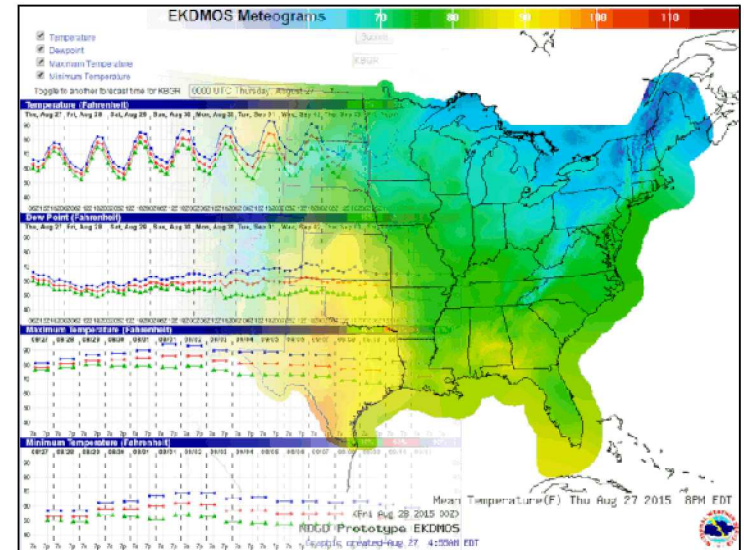
➤ Field data, in the form of spatial, temporal, or spatio-temporal outputs, is common for many computational models.

➤ Examples:

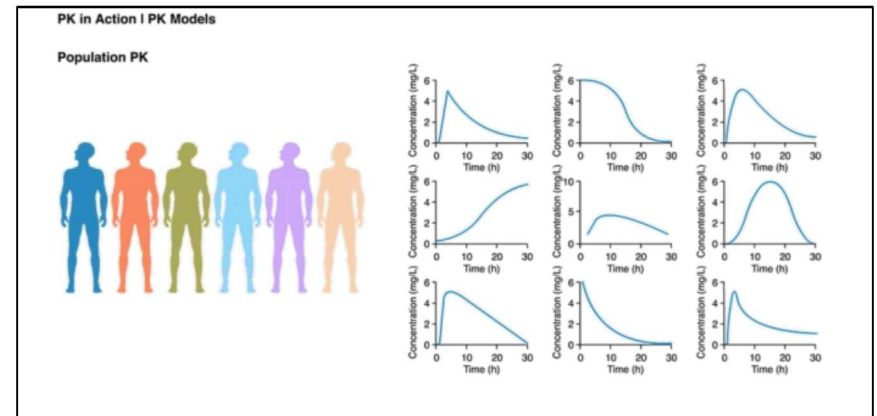
Thermal models of the ISS and other spacecraft [1]



Statistical weather forecasting [2]

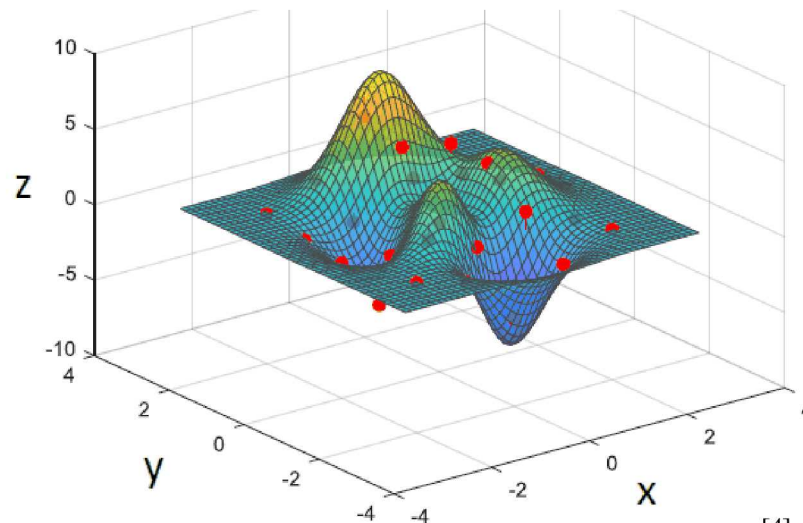


Population pharmacokinetics modeling [3]



Models with field data outputs can be challenging to calibrate

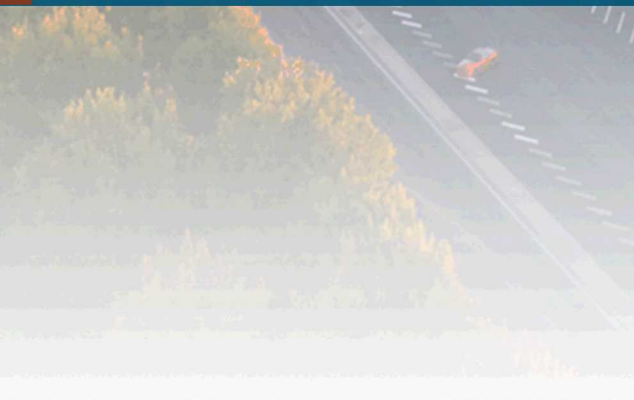
- Objective function formulation
- Identifying a unique solution
- Surrogate model formulation
- Likelihood function formulation and computation
 - High dimensionality of output space
 - Correlation in covariance matrix



[4]



Principal component analysis (PCA)



PCA: what is it and how does it help us?

- Principal component analysis (PCA) is an approach that maps a high-dimensional and highly correlated space onto a set of linearly uncorrelated principal directions, called the latent response space.
- By applying PCA to field data outputs, models can be calibrated in the latent response space instead of physical space:
 - Reduced dimensionality
 - Minimized correlation
- PCA can be performed in two ways:
 - Eigenvalue decomposition of output covariance matrix
 - Singular value decomposition (SVD) of output matrix

Converting simulation data to latent response space

➤ Given an $n \times p$ output matrix, A , performing SVD results in:

$$A_{[n \times p]} = U_{[n \times r]} S_{[r \times r]} V_{[r \times p]}^T$$

- n – number of samples
- p – number of output locations (spatially, temporally, or both)
- r – rank of A ; $r \leq \min(n, p)$
- U, V – column-orthonormal matrices
- S – singular values along diagonal

➤ The rows of A are projected onto k columns of V , resulting in the latent response space, γ :

$$\gamma_{[n \times k]} = A_{[n \times p]} V_{[p \times k]}^*$$

- $V_{[p \times k]}^* = (V_{[k \times p]}^T)^{-1}$
- $k \ll r$

➤ The number of dimensions, or latent variables, k , to retain depends on the desired fraction of variance in the physical output to be represented in the latent response space.

Converting experimental data to latent response space

- In order to calibrate in the latent response space, the simulation and experimental data must be in the same latent response space:

$$\{\gamma_{[n \times k]}\}_{sim} = \{A_{[n \times p]}\}_{sim} \{V_{[p \times k]}^*\}_{sim}$$

$$\{\gamma_{[m \times k]}\}_{exp} = \{A_{[m \times p]}\}_{exp} \{V_{[p \times k]}^*\}_{sim}$$

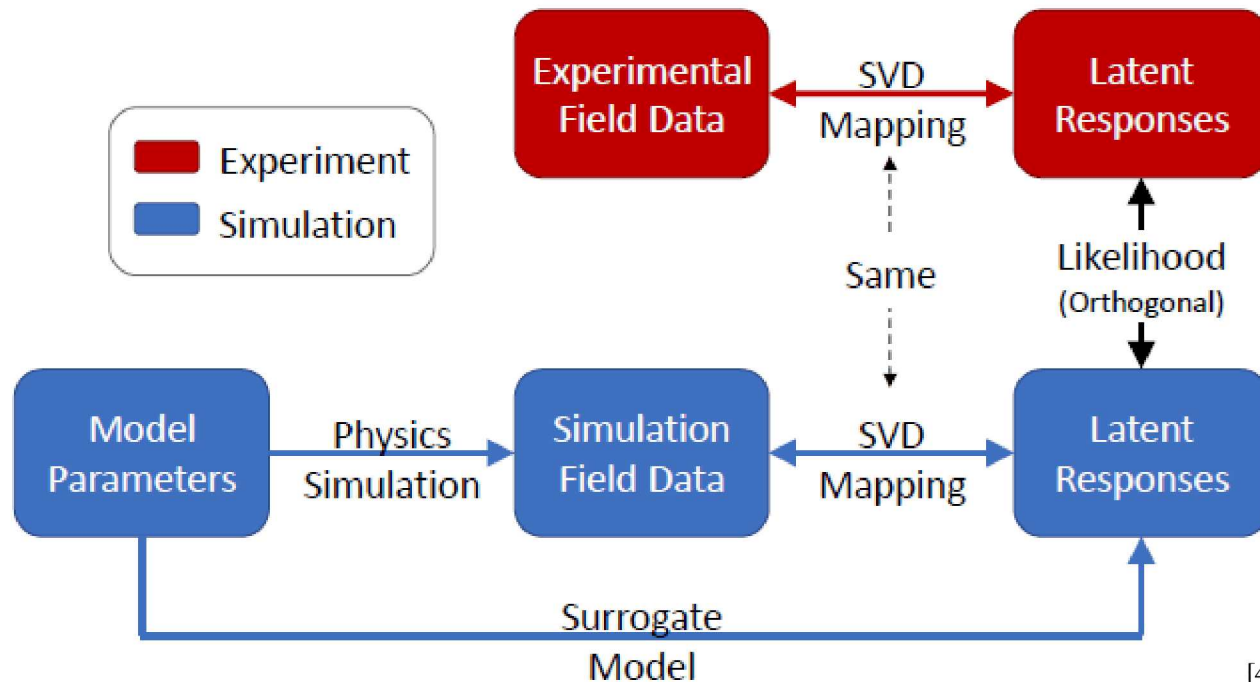
- $V_{[p \times k]}^*$ is created from the simulation data and used to transform both the simulation and experimental data
 - m – number of experimental samples
- To convert back to physical space, the following transformations can be used:

$$\{A_{[n \times p]}^*\}_{sim} = \{\gamma_{[n \times k]}\}_{sim} \{V_{[k \times p]}^{*T}\}_{sim}$$

$$\{A_{[m \times p]}^*\}_{exp} = \{\gamma_{[m \times k]}\}_{exp} \{V_{[k \times p]}^{*T}\}_{sim}$$

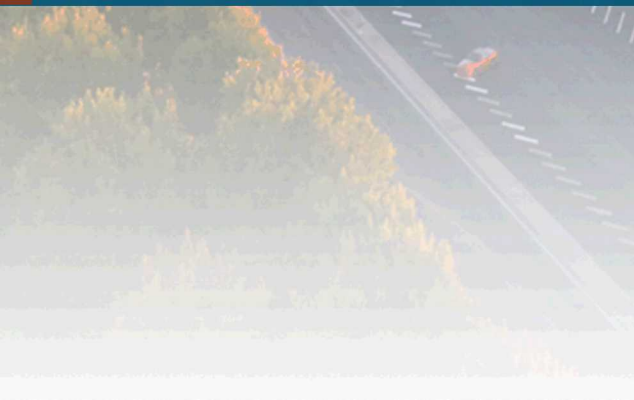
General approach to calibrating in latent response space

- Sample model parameter space
- Run physics simulation to generate field data for each sample
- Transform simulation field data to latent response space using SVD
- Create surrogate models mapping the model parameters to the latent response space
- Transform the experimental field data to the same latent response space
- Perform calibration in latent response space
- Transform back to physical output space and cross-check results via physics simulation





Example problem

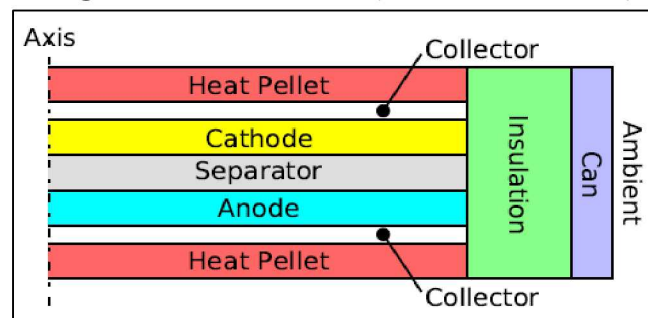


Electrical behavior of a single cell of a thermal battery

➤ Thermally activated batteries use a molten salt electrolyte that remains solid at room temperature. At elevated temperatures, the electrolyte melts and the battery is activated. The battery activation process is complex and involves:

- Heat transfer
- Electrochemical reactions
- Ion and species transport
- Porous flow
- Capillary effects
- Phase change
- Mechanical deformation

Single cell of a thermally activated battery [5]



➤ Sandia's Thermally Activated Battery Simulator (TABS)

- Suite of thermal battery simulations for full battery and single cell configurations
- Characterize thermal, electrochemical, and species behavior over time
- Varying degrees of simulation complexity available (full multi-physics model down to thermal only)

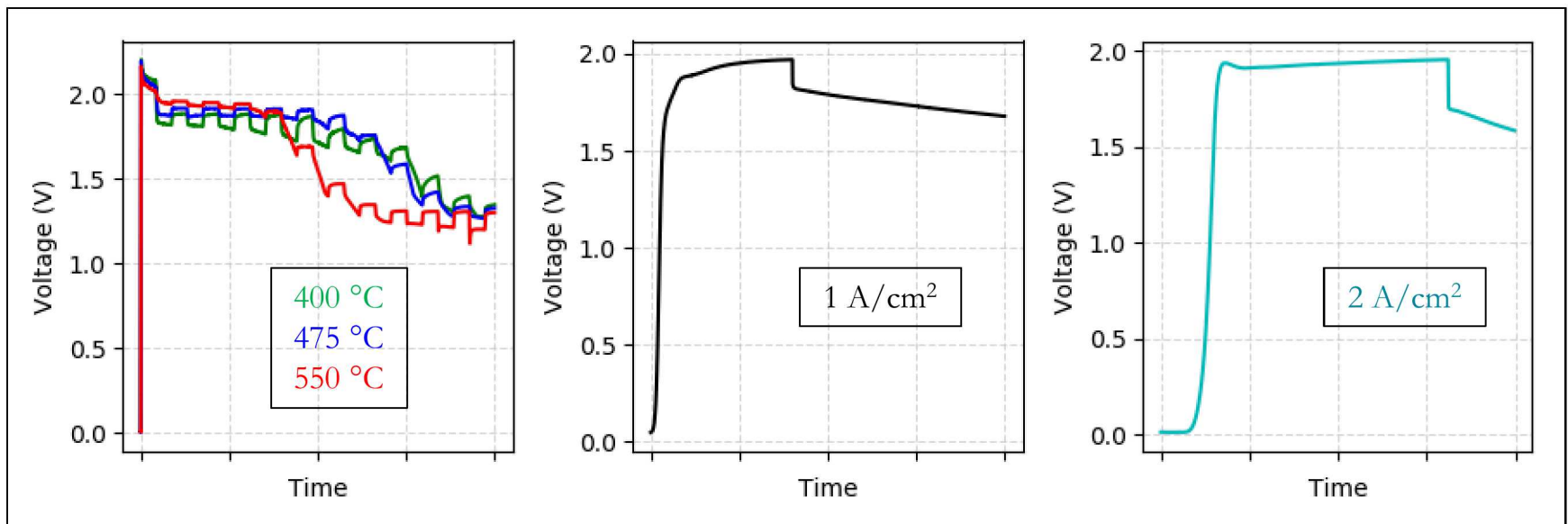
➤ **Model of interest: 1D electrochemical simulation of single cell of a LiSi/FeS_2 thermal battery**

- Model parameters: 23
- Observations: time-dependent voltage traces

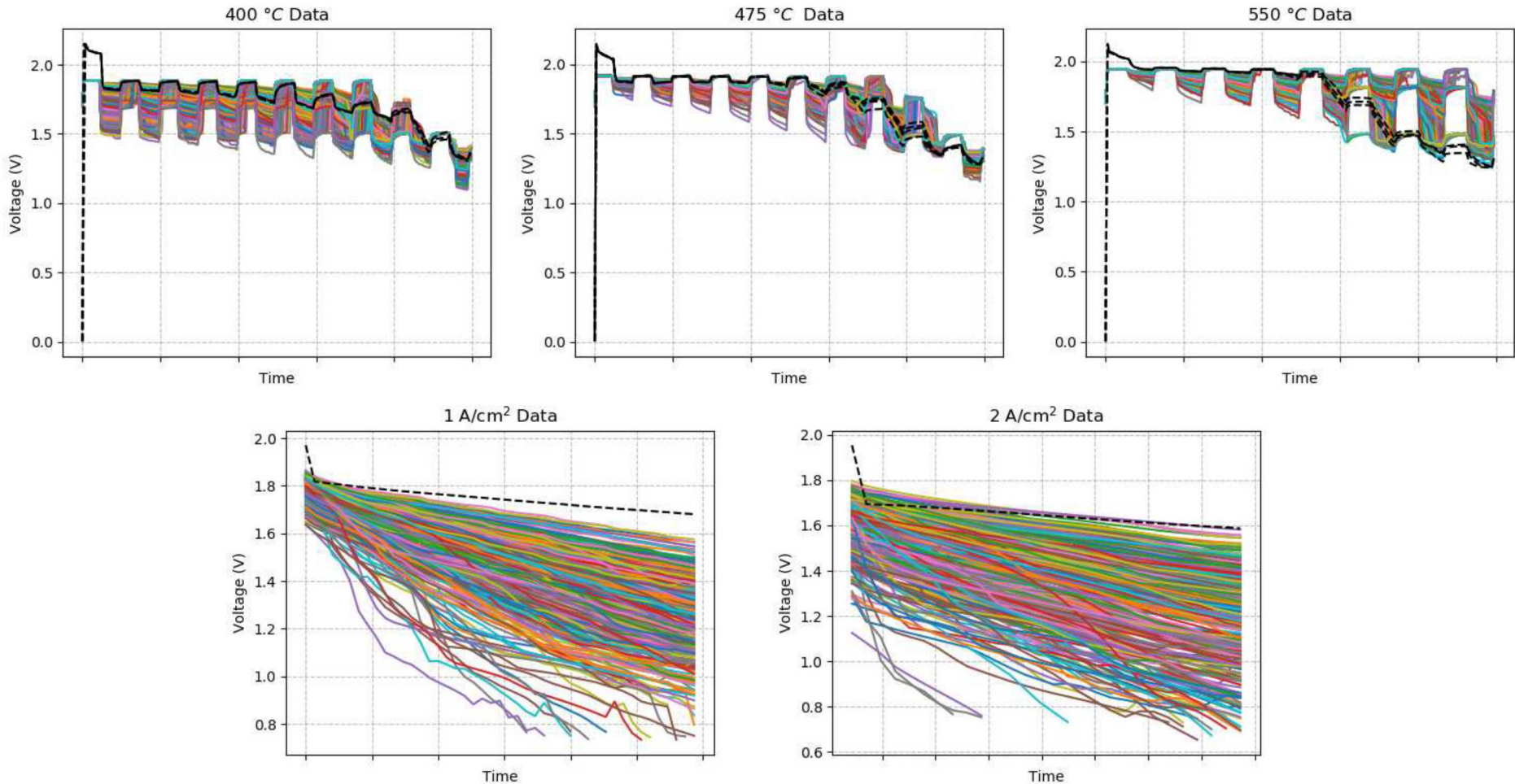
Experimental data: 5 boundary conditions

Case	Thermal BC	Electrical BC	# of Exp. Traces
400 °C Data	Constant, 400 °C	Pulsed	3
475 °C Data	Constant, 475 °C	Pulsed	4
550 °C Data	Constant, 550 °C	Pulsed	3
1 A/cm ² Data	Constant, 525 °C	Constant, 1 A/cm ²	1
2 A/cm ² Data	Constant, 525 °C	Constant, 2 A/cm ²	1

Mean experimental traces:

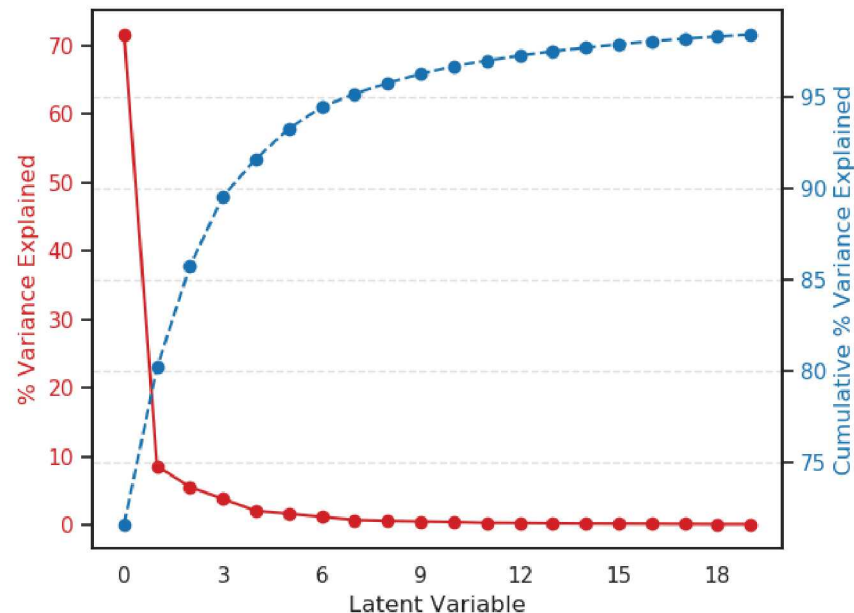


- 320 samples of input parameter space using Latin hypercube sampling (LHS)
 - Experimental data in black



Transform to latent response space and choose number of latent variables to retain

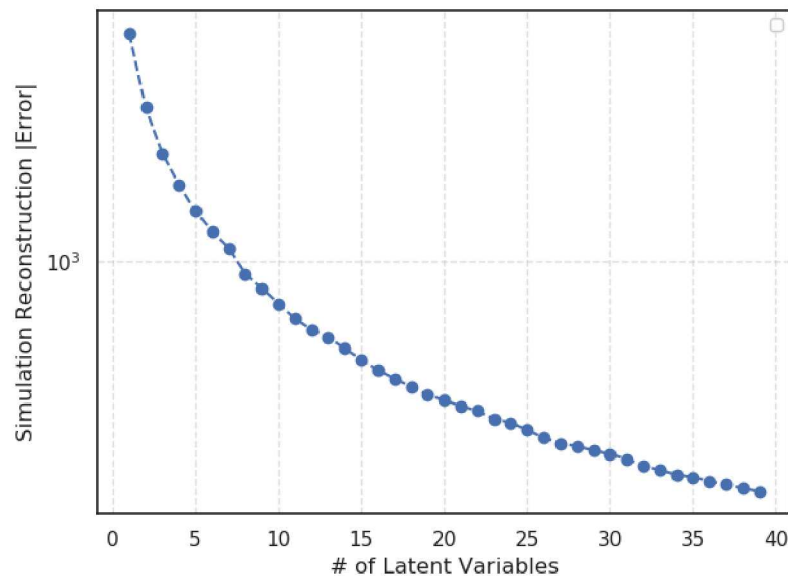
- The number of latent variables for each data set should be based on the minimum number of latent variables required to capture a desired amount of variance in the original dataset
 - The latent variables capture variance in the data in decreasing order, i.e. the first latent variable captures the largest amount of variance in the data
 - Higher latent variables may be mostly capturing noise and can make it difficult to build surrogate models when included
- For the 400 °C data set:



Simulation reconstruction error

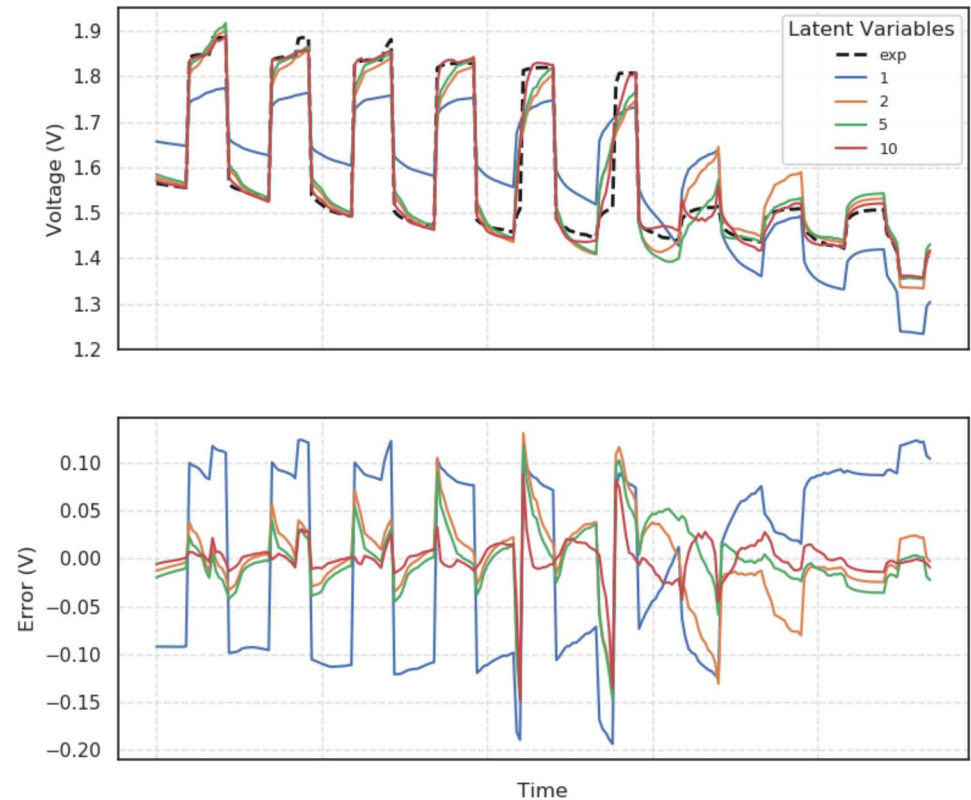
- Transforming back to physical space and comparing the reconstructed traces to the original traces can also help indicate an appropriate number of latent variables to retain during calibration.
- For the 400 °C data set:

Cumulative error for all traces:



Reconstruction error for a single trace:

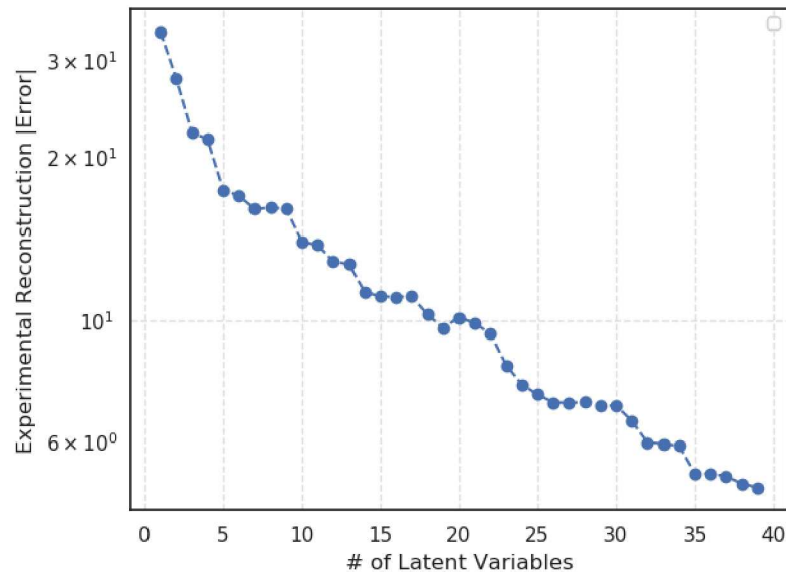
Simulation Reconstruction



Experimental reconstruction error

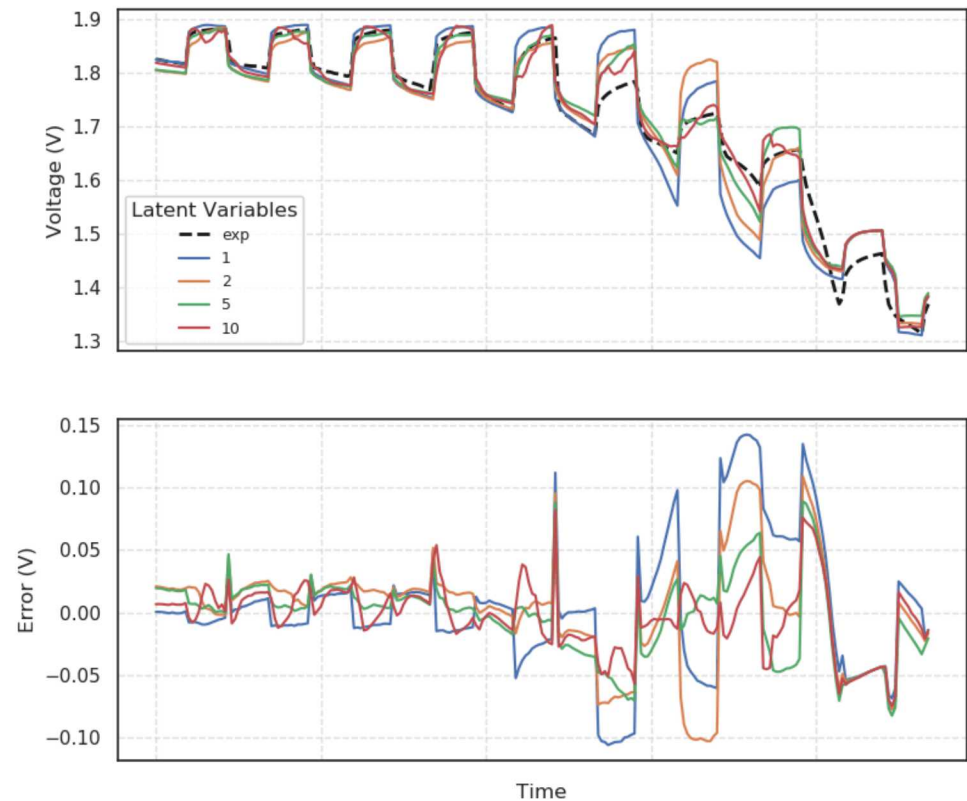
- Similar to the simulation data, it can be helpful to reconstruct the experimental data in physical space to assess an appropriate number of latent variables to retain.
- For the 400 °C data set:

Cumulative error for all traces:



Reconstruction error for a single trace:

Experiments Reconstruction



Calibration in latent response space

➤ Latent variable retained per data set:

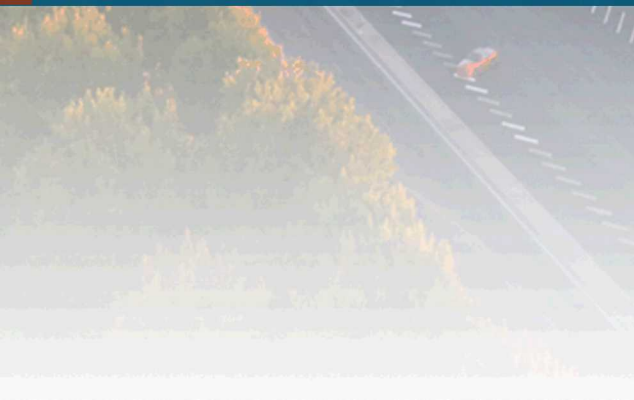
Case	Thermal BC	Electrical BC	# of Exp. Traces	# of Latent Variables
400 °C Data	Constant, 400 °C	Pulsed	3	10
475 °C Data	Constant, 475 °C	Pulsed	4	6
550 °C Data	Constant, 550 °C	Pulsed	3	6
1 A/cm ² Data	Constant, 525 °C	Constant, 1 A/cm ²	1	2
2 A/cm ² Data	Constant, 525 °C	Constant, 2 A/cm ²	1	2

➤ Bayesian calibration:

- Gaussian process surrogate models
- Iterative importance sampling with genetic algorithm [4]

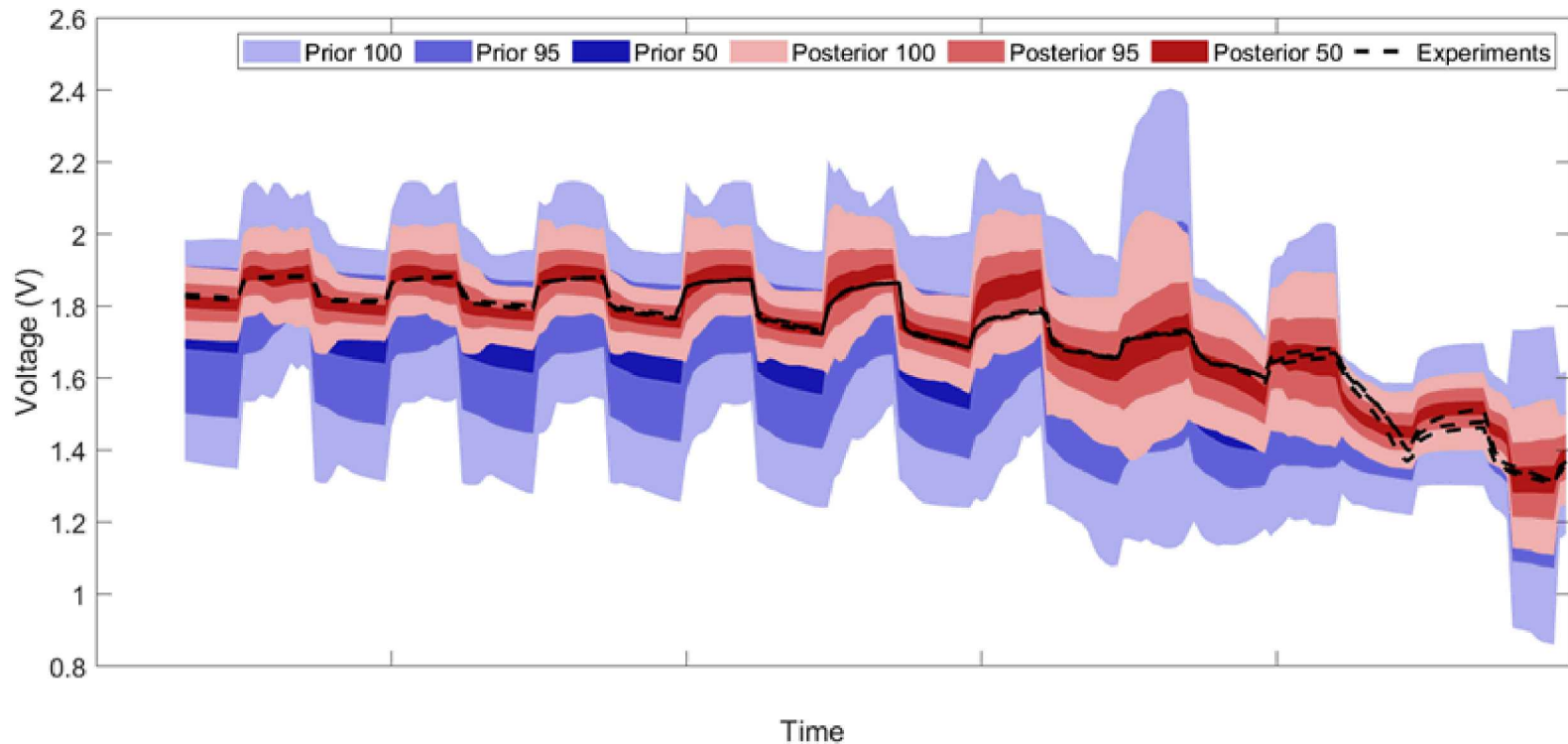


Results



Trace reconstruction

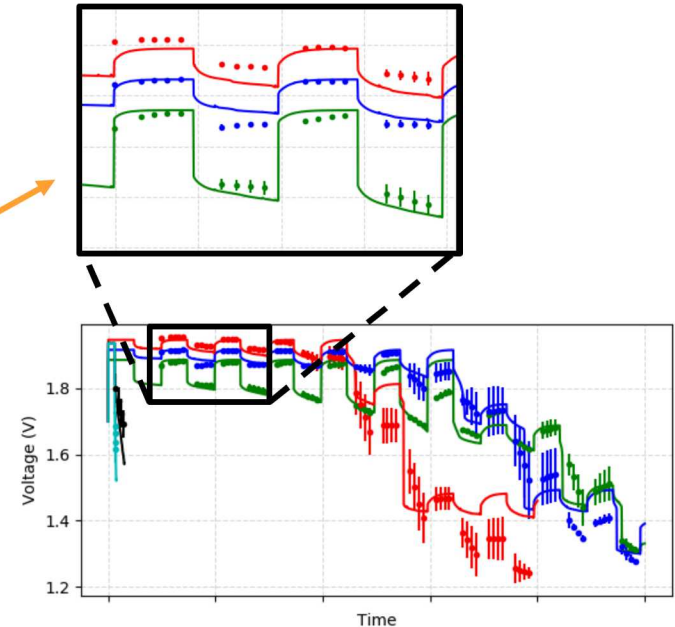
- For each data set, the calibrated traces can be reconstructed in physical output space.
- For the 400 °C data set:



Comparison to other calibration approaches

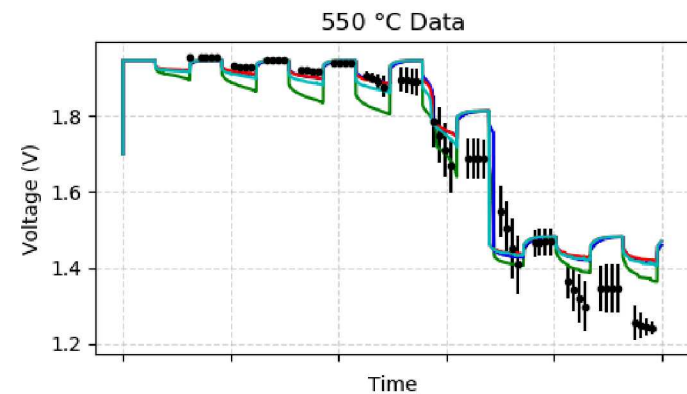
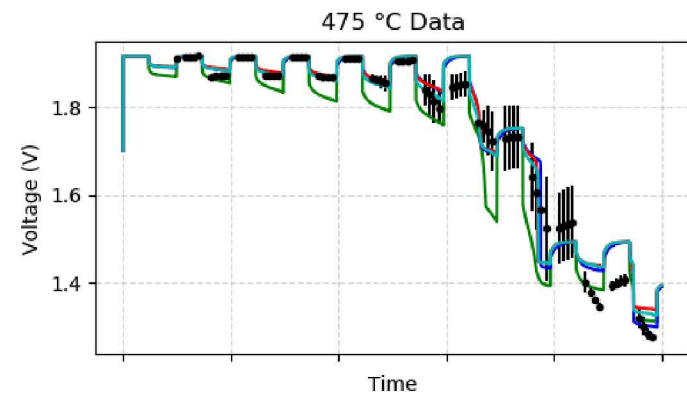
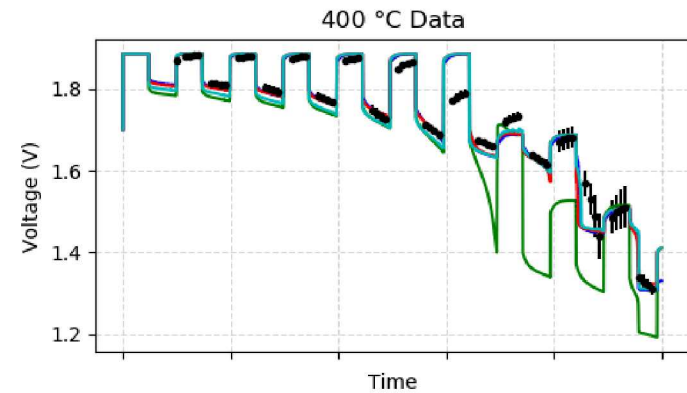
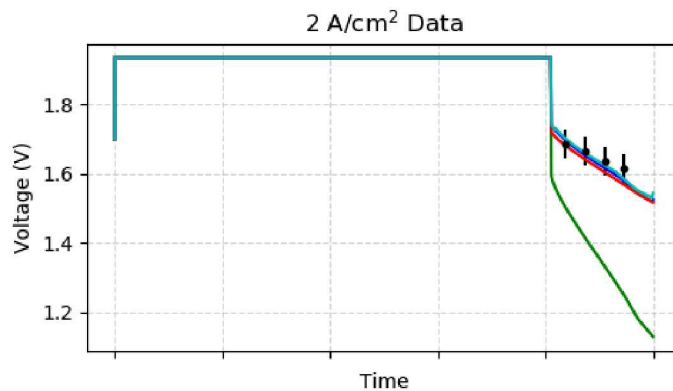
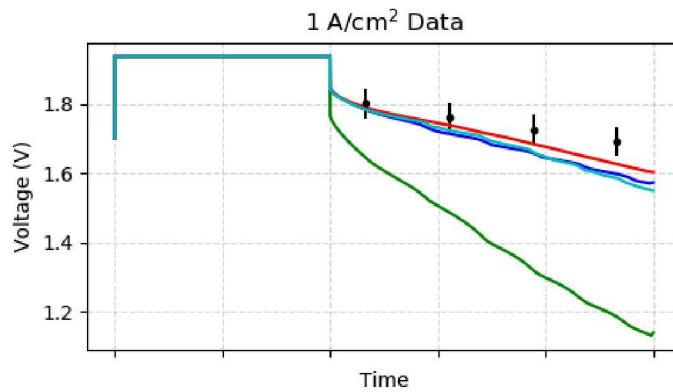
➤ Four sets of results compared:

1. Legacy parameter values:
 - Values present in the model prior to any model calibration
 - Based on literature, internal research, and SMEs
2. Model calibration in physical space:
 - Objective function based on weighted cumulative error term
 - NCSU DIRECT optimization algorithm
3. Model calibration in latent response space:
 - Input parameter set chosen based on maximum likelihood
4. Model calibration in latent response space:
 - Input parameter set chosen based on minimum cumulative weighted error (objective function value)



Comparison to other calibration approaches

0. Experiments
1. Legacy Values
2. Model Calibration in Physical Space
3. Model Calibration in Latent Response Space (Max Likelihood)
4. Model Calibration in Latent Response Space (Min Cumulative Error)



- Calibration efforts using two very different techniques produced similar results
 - This lends credibility to the resulting calibrated model parameter values
 - This lends credibility to the model calibration in latent response space
- Both calibration approaches result in much closer agreement to experimental data than the legacy parameter values
 - This, along with validation approaches, lends credibility to the resulting calibrated model parameters
- Calibrating in latent response space required a minimal level of effort in comparison to calibrating in physical output space
 - Calibrating in physical output space required formulating an appropriate objective function
 - Calibrating in latent response space required selecting an appropriate number of latent variables

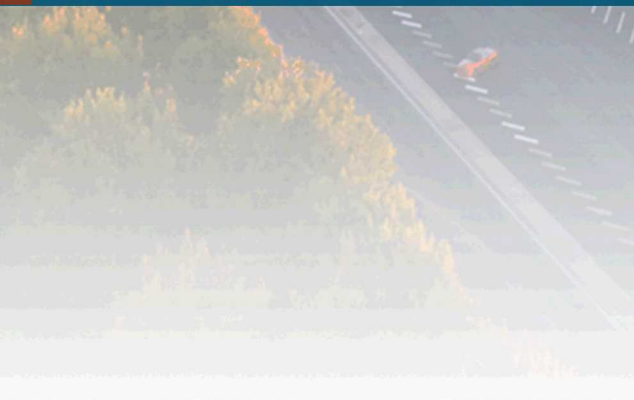
- Introduced an approach utilizing principal component analysis (PCA) to calibrate models with field data output in latent response space
- Applied the PCA-based calibration approach to a 1D electrochemical model of a single cell of a LiSi/FeS₂ thermal battery
 - Model calibration in latent response space produced similar results to model calibration in physical output space but at a much lower level of effort
- **For models with field data outputs, model calibration in latent response space could offer a more robust and efficient calibration approach than calibrating in physical output space.**

- [1] C&R Technologies, “Thermal Desktop: Complete CAD-based Thermal Engineering Tool Suite”, <https://www.crtech.com/products/thermal-desktop>
- [2] National Weather Service, “Ensemble Kernel Density Model Output Statistics (EKDMOS)”, https://www.weather.gov/mdl/ekdmoss_home
- [3] ScienceMedia, “SMi Source Lesson Pharmacokinetics: PK in Action”, https://sciencemedia.com/1/Pharmacokinetics_PK_in_Action/
- [4] K. Neal, “Bayesian Calibration with Spatio-Temporal Data”, IMAC XXXVIII, Feb. 12, 2020.
- [5] T. Voskuilen, “Multi-Physics Cost and Complexity Reduction Strategies for Modeling Thermally Activated Batteries”, 48th Power Sources Conference, June 14, 2018.



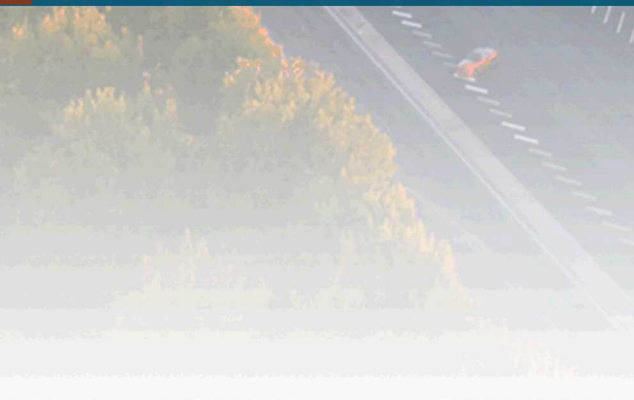
Thank you for your attention!

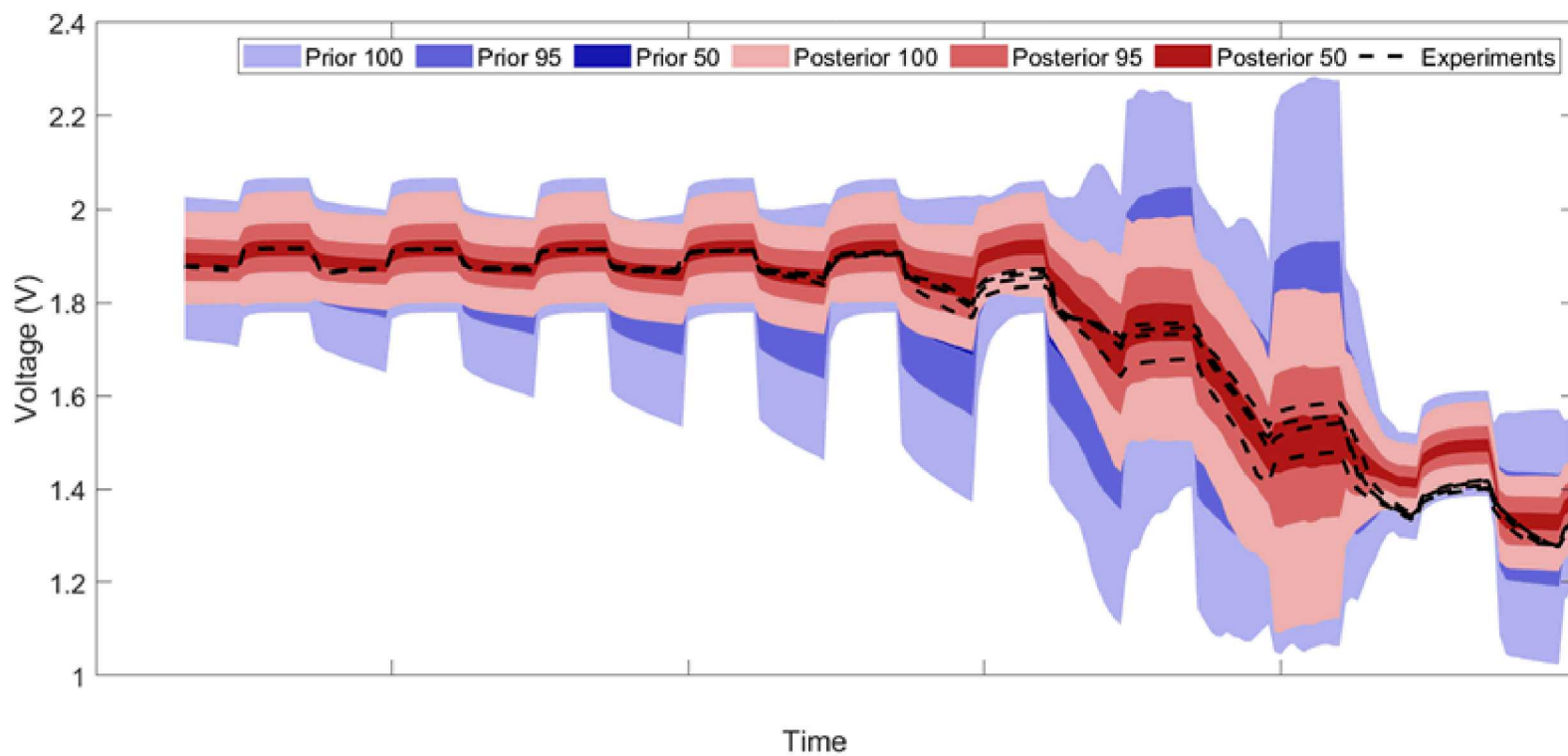
Questions?



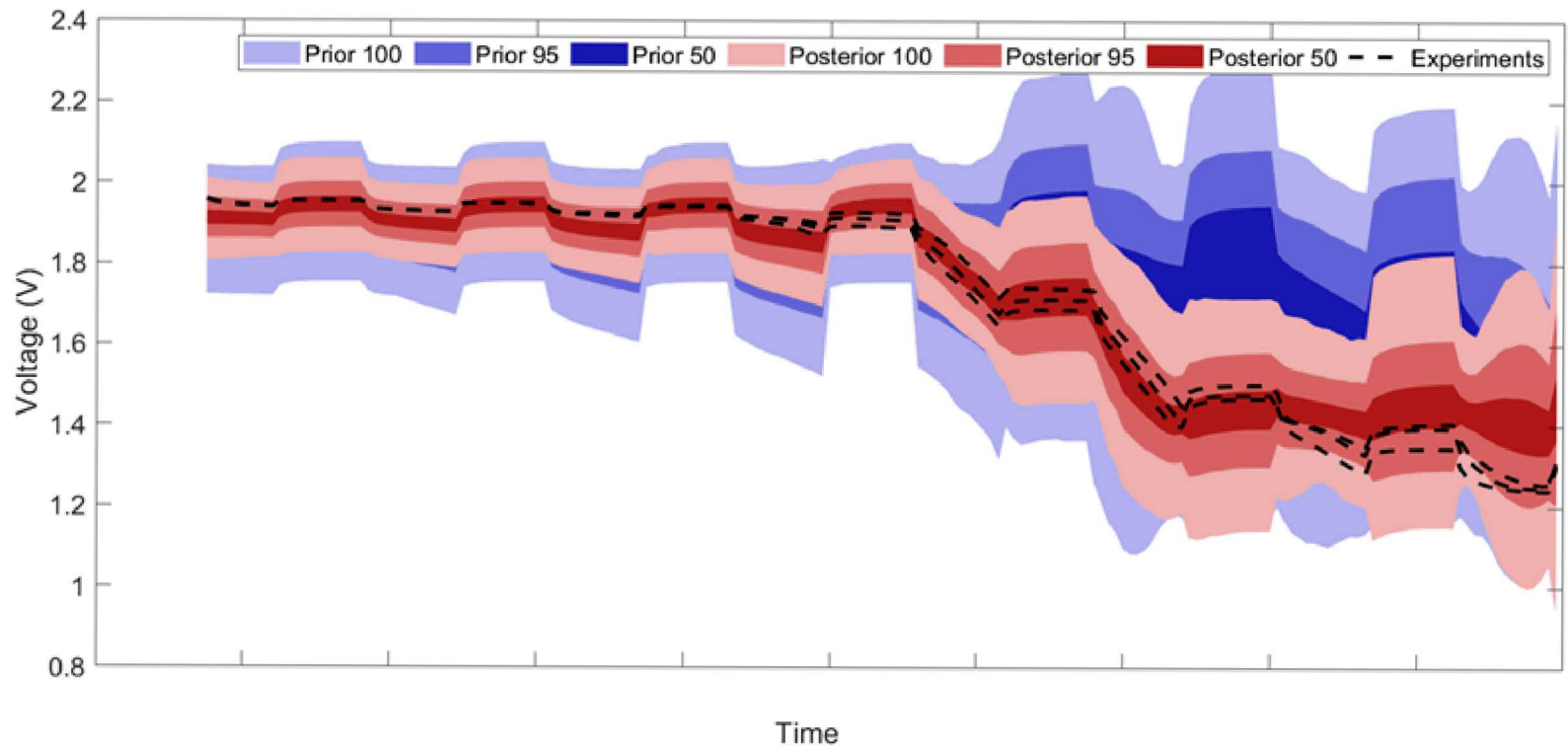


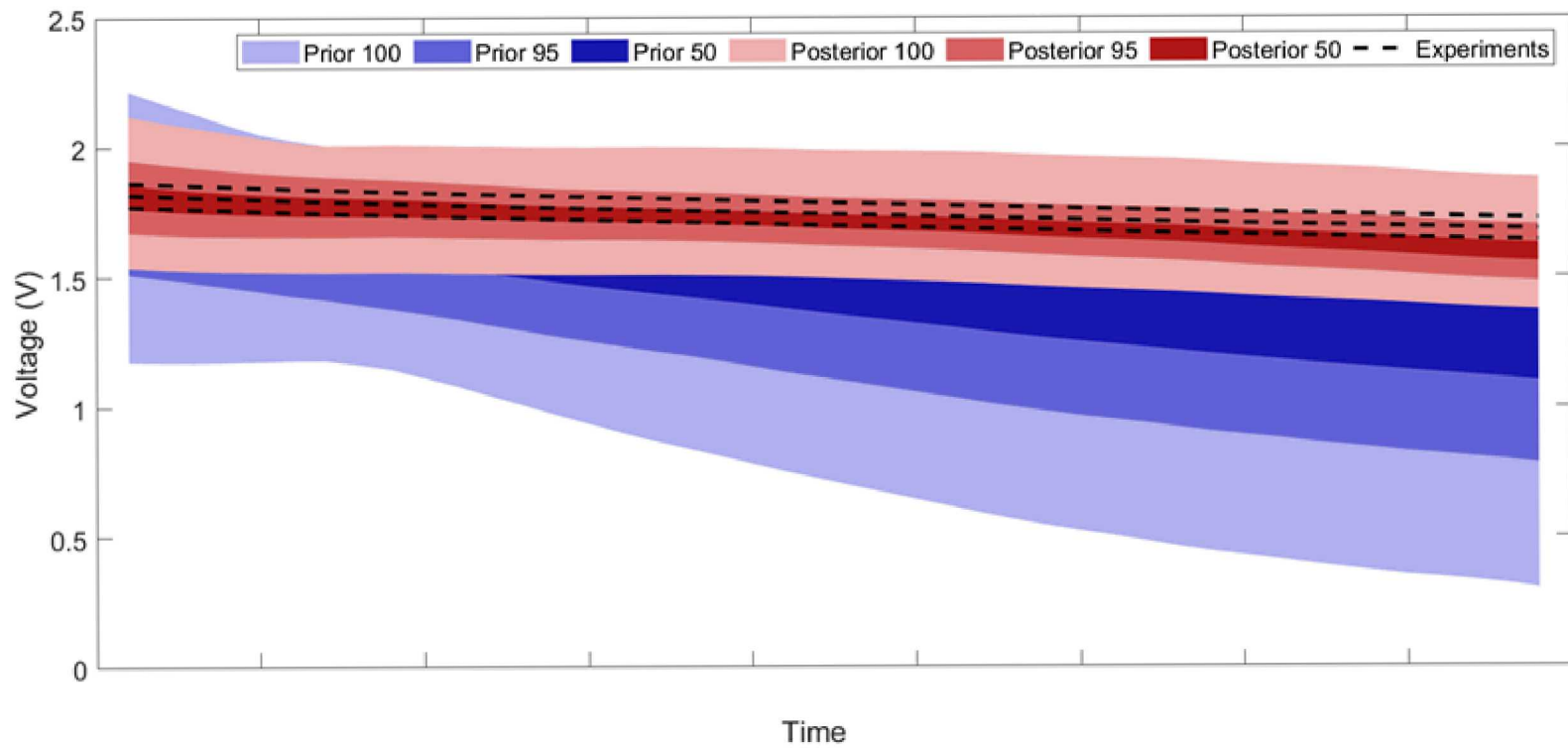
BACKUP



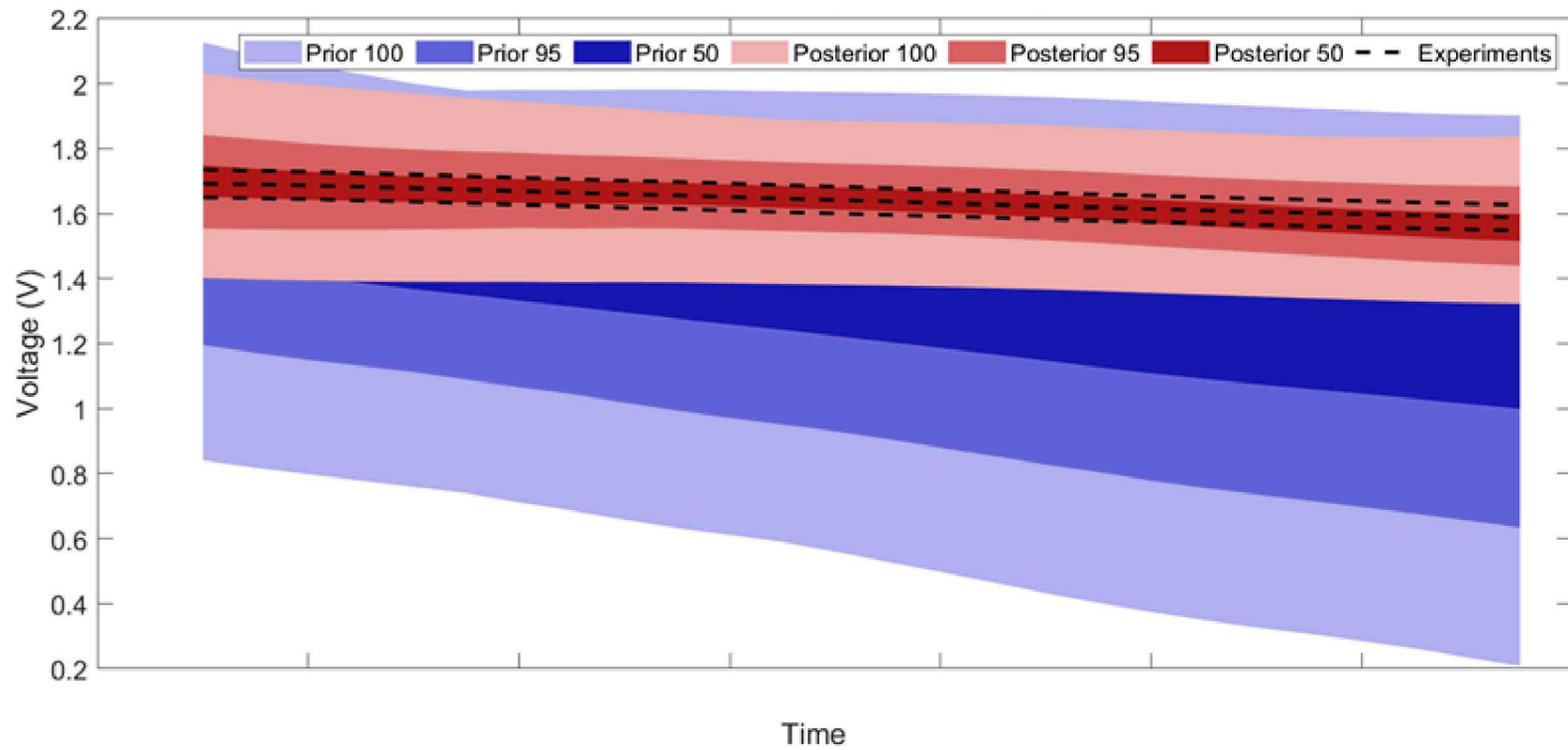


550 °C Dataset

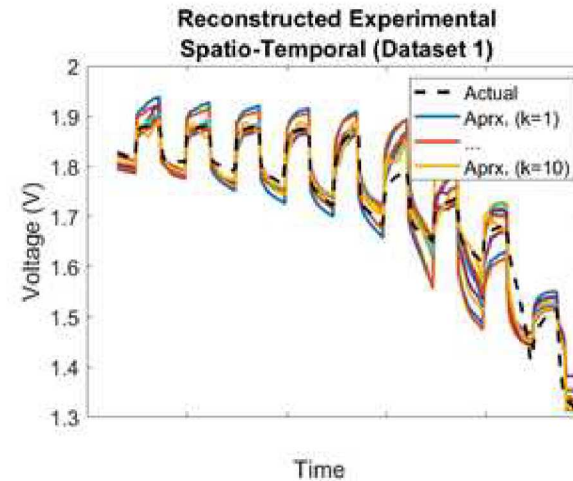
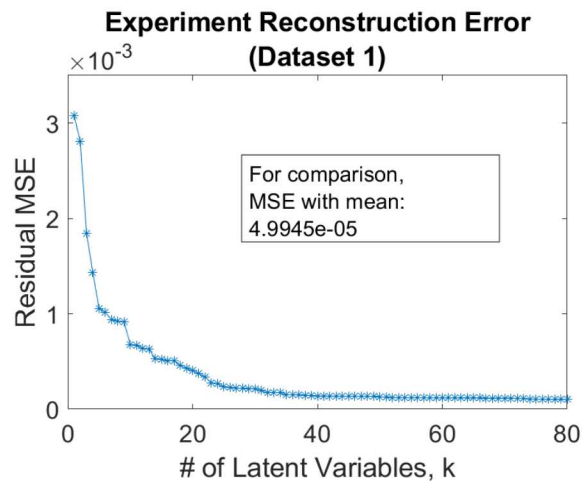
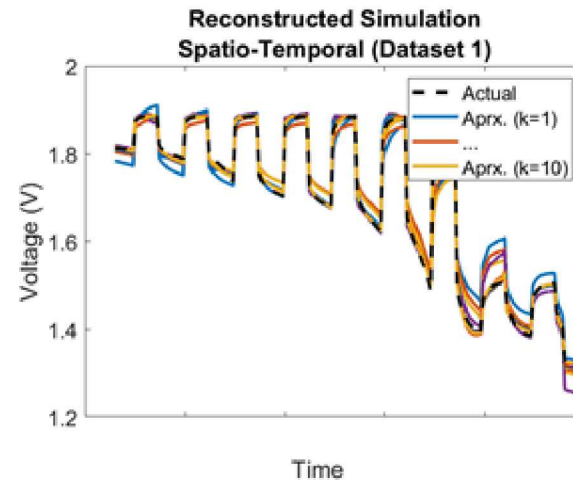
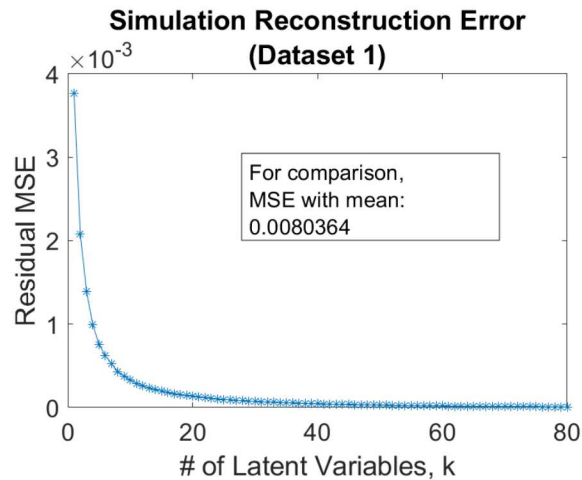




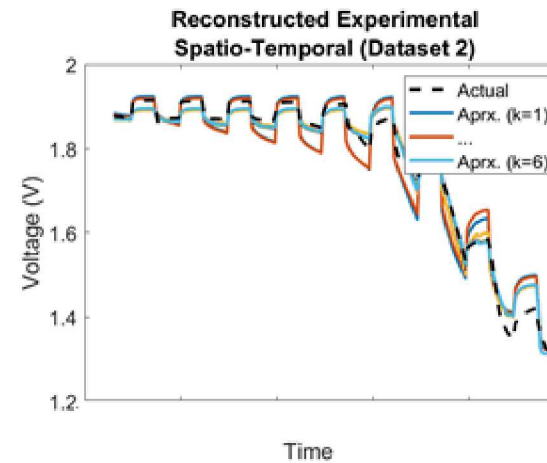
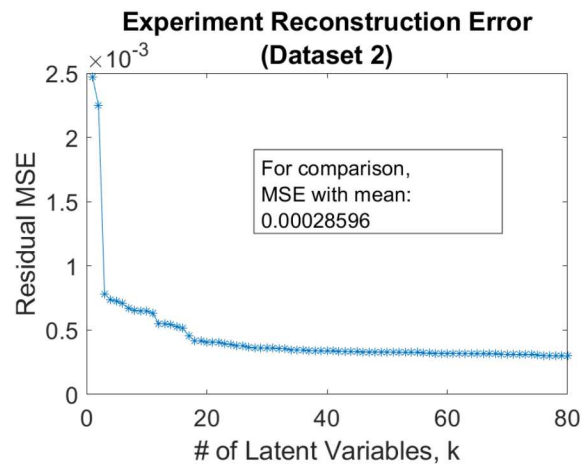
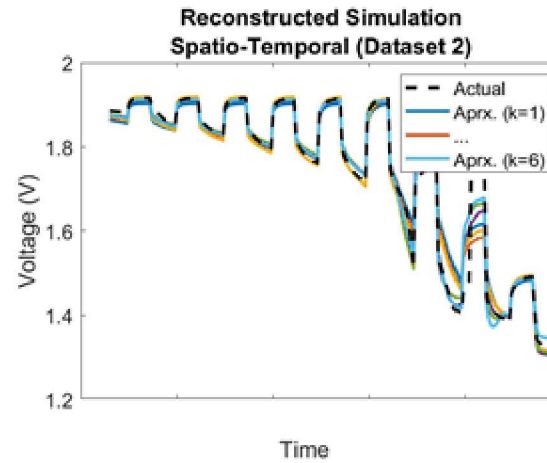
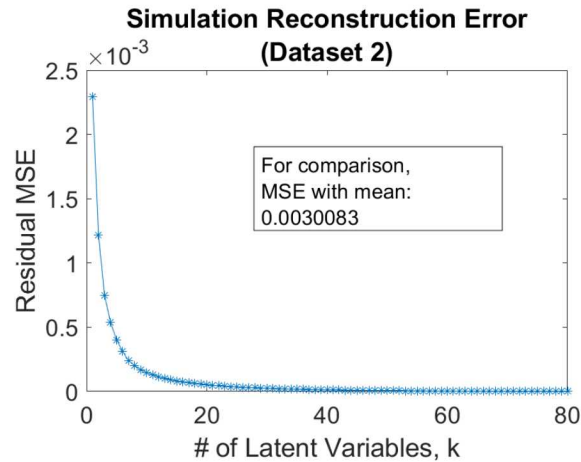
2 A/cm² Dataset



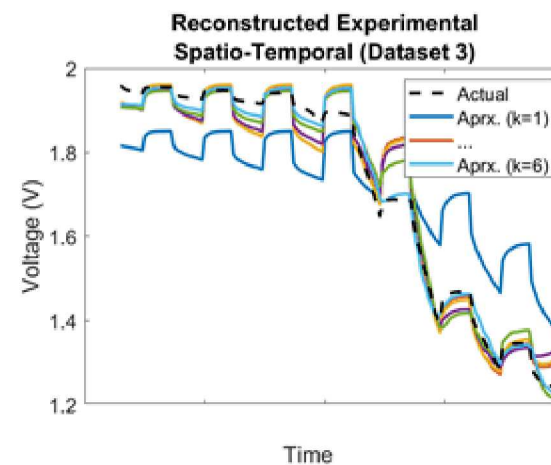
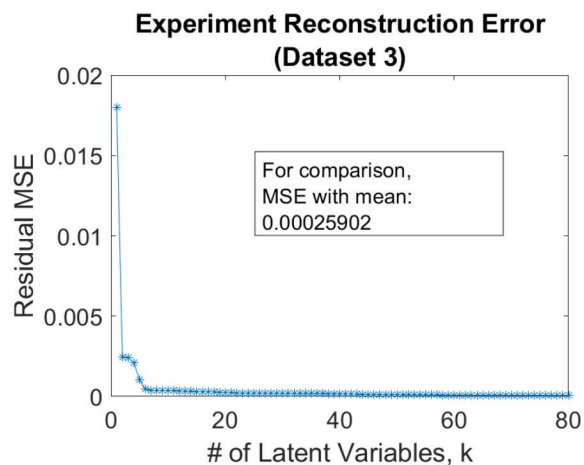
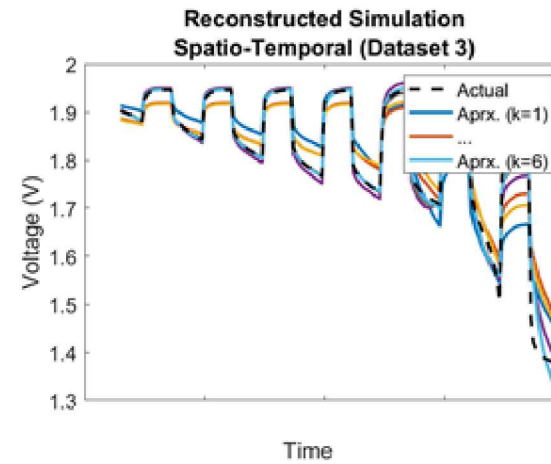
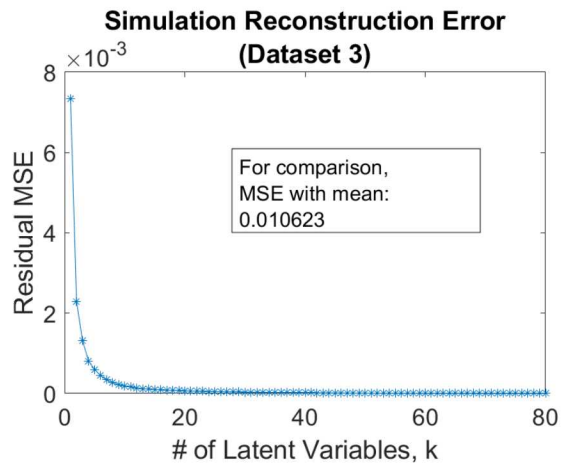
Latent Variables per Dataset: 400 °C Dataset



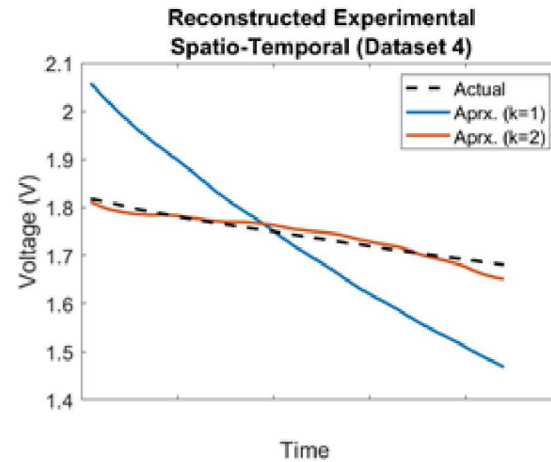
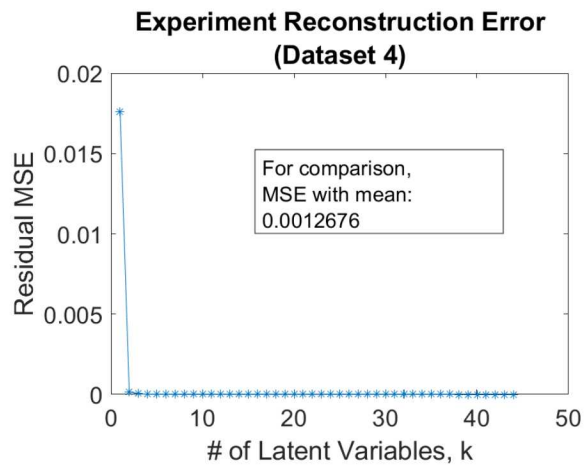
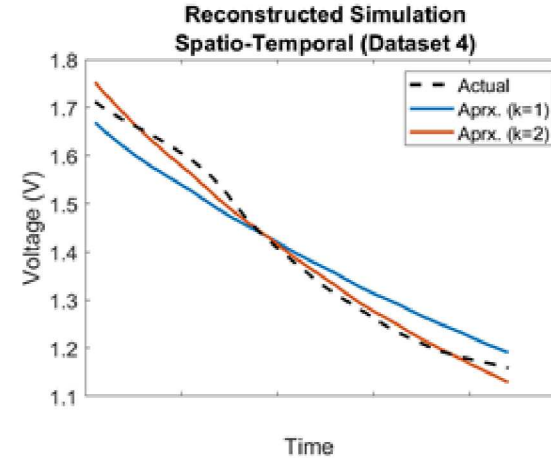
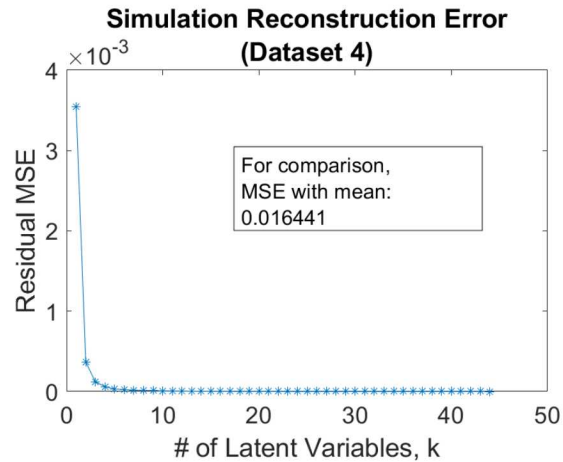
Latent Variables per Dataset: 475 °C Dataset



Latent Variables per Dataset: 550 °C Dataset



Latent Variables per Dataset: 1 A/cm² Dataset



Latent Variables per Dataset: 2 A/cm² Dataset

