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
Outline

➢ 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]**

  ![Thermal model image]

  **Statistical weather forecasting [2]**

  ![Statistical weather forecasting image]

  **Population pharmacokinetics modeling [3]**

  ![Population pharmacokinetics image]
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
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:

$$A_{[n \times p]} = U_{[n \times r]} S_{[r \times r]} V^T_{[r \times p]}$$

- $n$ – 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, $\gamma$:

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

- $V^*_{[p \times k]} = (V^T_{[k \times p]})^{-1}$
- $k \ll r$

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:

\[
\{A^*_{n \times p}\}_{sim} = \{\gamma_{n \times k}\}_{sim} \{V^*_{T} \}_{sim}
\]

\[
\{A^*_{m \times p}\}_{exp} = \{\gamma_{m \times k}\}_{exp} \{V^*_{T} \}_{sim}
\]
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:

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

➢ Sandia’s Thermally Activated Battery Simulator (TABS)

º Suite of thermal battery simulations for full battery and single cell configurations
º Characterize thermal, electrochemical, and species behavior over time
º 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

º Model parameters: 23
º Observations: time-dependent voltage traces
Experimental data: 5 boundary conditions

<table>
<thead>
<tr>
<th>Case</th>
<th>Thermal BC</th>
<th>Electrical BC</th>
<th># of Exp. Traces</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 °C Data</td>
<td>Constant, 400 °C</td>
<td>Pulsed</td>
<td>3</td>
</tr>
<tr>
<td>475 °C Data</td>
<td>Constant, 475 °C</td>
<td>Pulsed</td>
<td>4</td>
</tr>
<tr>
<td>550 °C Data</td>
<td>Constant, 550 °C</td>
<td>Pulsed</td>
<td>3</td>
</tr>
<tr>
<td>1 A/cm² Data</td>
<td>Constant, 525 °C</td>
<td>Constant, 1 A/cm²</td>
<td>1</td>
</tr>
<tr>
<td>2 A/cm² Data</td>
<td>Constant, 525 °C</td>
<td>Constant, 2 A/cm²</td>
<td>1</td>
</tr>
</tbody>
</table>

Mean experimental traces:
Simulation data

➢ 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 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

Latent Variables
- exp
- 1
- 2
- 5
- 10

Error (V)

Time
Calibration in latent response space

Latent variable retained per data set:

<table>
<thead>
<tr>
<th>Case</th>
<th>Thermal BC</th>
<th>Electrical BC</th>
<th># of Exp. Traces</th>
<th># of Latent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 °C Data</td>
<td>Constant, 400 °C</td>
<td>Pulsed</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>475 °C Data</td>
<td>Constant, 475 °C</td>
<td>Pulsed</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>550 °C Data</td>
<td>Constant, 550 °C</td>
<td>Pulsed</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>1 A/cm² Data</td>
<td>Constant, 525 °C</td>
<td>Constant, 1 A/cm²</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2 A/cm² Data</td>
<td>Constant, 525 °C</td>
<td>Constant, 2 A/cm²</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

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)
Observations

➢ 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
Conclusions

➢ 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/FeS$_2$ 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.
References


Thank you for your attention!

Questions?
BACKUP
475 °C Dataset
550 °C Dataset
1 A/cm² Dataset
2 A/cm² Dataset
Latent Variables per Dataset: 400 °C Dataset

Simulation Reconstruction Error (Dataset 1)

Reconstructed Simulation Spatio-Temporal (Dataset 1)

For comparison, MSE with mean: 0.0080364

Experiment Reconstruction Error (Dataset 1)

Reconstructed Experimental Spatio-Temporal (Dataset 1)

For comparison, MSE with mean: 4.9945e-05
Latent Variables per Dataset: 475 °C Dataset

Simulation Reconstruction Error (Dataset 2)

Experiment Reconstruction Error (Dataset 2)

Reconstructed Simulation Spatio-Temporal (Dataset 2)

Reconstructed Experimental Spatio-Temporal (Dataset 2)
Latent Variables per Dataset: 550 °C Dataset

Simulation Reconstruction Error (Dataset 3)

For comparison, MSE with mean: 0.010623

Experiment Reconstruction Error (Dataset 3)

For comparison, MSE with mean: 0.00025902
Latent Variables per Dataset: 1 A/cm² Dataset

Simulation Reconstruction Error
(Dataset 4)

For comparison, MSE with mean: 0.016441

Reconstructed Simulation Spatio-Temporal (Dataset 4)

Experiment Reconstruction Error
(Dataset 4)

For comparison, MSE with mean: 0.0012676

Reconstructed Experimental Spatio-Temporal (Dataset 4)
Latent Variables per Dataset: 2 A/cm² Dataset

Simulation Reconstruction Error (Dataset 5)

For comparison, MSE with mean: 0.025917

Reconstructed Simulation Spatio-Temporal (Dataset 5)

For comparison, MSE with mean: 0.001121