# When models and data disagree: sparse resolutions to inconsistent datasets

**Arun Hegde**
Wenyu Li
Jim Oreluk
Andrew Packard
Michael Frenklach

This project is supported by the U.S. Department of Energy, National Nuclear Security Administration, under Award Number DE-NA0002375

# Overview



- Bound-to-Bound Data Collaboration (B2BDC)
  - models + data = *dataset (model-data system)*
- Dataset Consistency – agreement between models and data
  - scalar consistency measure
  - **vector consistency measure**
- Dataset examples
  - GRI-Mech 3.0
  - DLR-SynG
- Summary

# Bound-to-Bound Data Collaboration



UQ as constrained optimization: **parameters** constrained by **models** and **data**

## Models

"True" QOI models

$$f_e : \mathbb{R}^n \to \mathbb{R}$$

Surrogate QOI models

$$M_e(x) \approx f_e(x)$$

Fitting error

$$|M_e(x) - f_e(x)| \leq \epsilon_e$$

$$e = 1, \ldots, N$$

+

## Data

- Prior knowledge on uncertain parameters

$$x \in \mathcal{H} \subseteq \mathbb{R}^n$$

- QOI measurements (w/ uncertainty)

$$L_e \leq y_e \leq U_e$$

$$e = 1, \ldots, N$$

=

## Dataset

$$x \in \mathcal{H} \subseteq \mathbb{R}^n$$

$$L_e - \epsilon_e \leq M_e(x) \leq U_e + \epsilon_e$$

$$\text{for } e = 1, \ldots, N$$

$$\mathcal{F} \subseteq \mathcal{H}$$

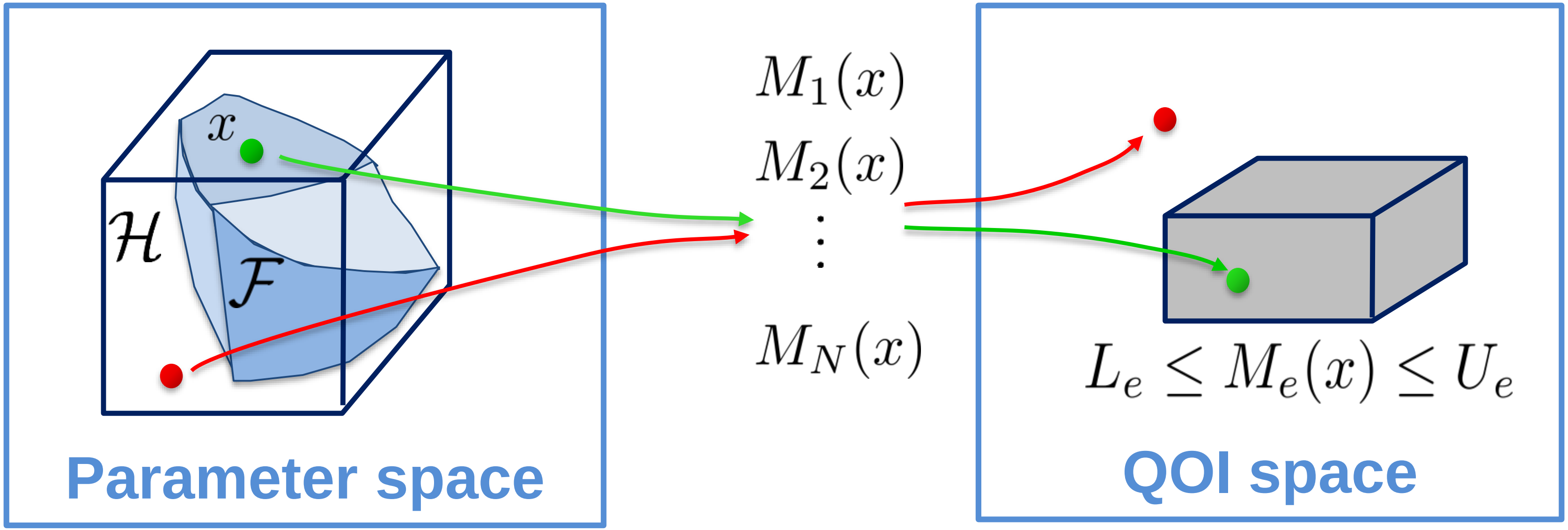
**Feasible set** – parameters for which the models and data agree.

Prediction bounds the range of a model $M_P(x)$ subject to the feasible set

$$\left( \begin{array}{ll} \min & M_P(x) \\ \text{s.t.} & x \in \mathcal{F} \end{array} \;\middle|\; \begin{array}{ll} \max & M_P(x) \\ \text{s.t.} & x \in \mathcal{F} \end{array} \right)$$

# Consistency as Model Validation

- A dataset is **consistent** if it is feasible
  - Parameters exist for which model predictions match the experiments




- Consistency analysis provides measures of validation

# Quantifying Consistency

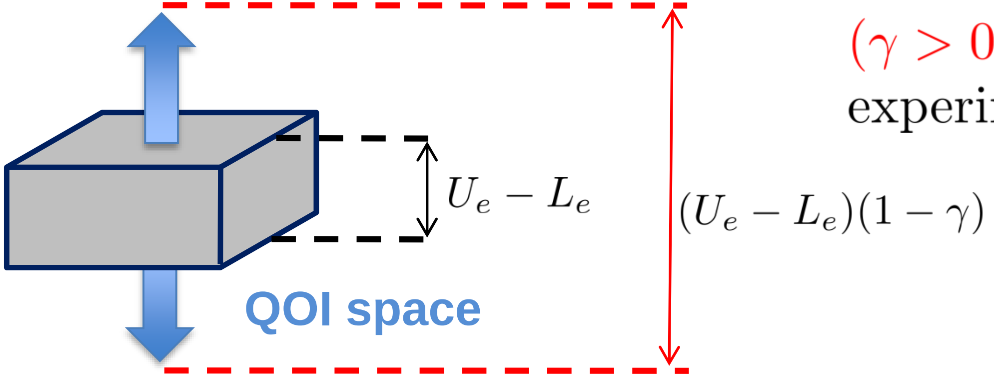
**<u>Q:</u>** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**<u>A:</u>** Compute the *scalar consistency measure* (**SCM**)

## Scalar Consistency Measure (SCM)*

$$
\begin{aligned}
\max_{x,\gamma} \quad & \gamma \\
\text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \leq M_e(x) \leq U_e - \frac{(U_e - L_e)}{2}\gamma \\
& x \in \mathcal{H}, \gamma \in \mathbb{R} \\
& \text{for } e = 1, \ldots, N
\end{aligned}
$$

The SCM produces a symmetric tightening $(\gamma > 0)$ or stretching $(\gamma < 0)$ of all experimental bounds.




*Feeley, R.; Seiler, P.; Packard, A.; Frenklach., M.; *J. Phys. Chem. A.* 2004, *108*, 9573.

# Quantifying Consistency



**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

## Scalar Consistency Measure (SCM)*

$$
\begin{aligned}
\max_{x,\gamma} \quad & \gamma \\
\text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \le M_e(x) \le U_e - \frac{(U_e - L_e)}{2}\gamma \\
& x \in \mathcal{H}, \gamma \in \mathbb{R} \\
& \text{for } e = 1, \ldots, N
\end{aligned}
$$

If consistent, go to prediction


*Feeley, R.; Seiler, P.; Packard, A.; Frenklach., M.; *J. Phys. Chem. A.* 2004, *108*, 9573.

# Quantifying Consistency



**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

## Scalar Consistency Measure (SCM)*

$$
\begin{aligned}
\max_{x,\gamma} \quad & \gamma \\
\text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \le M_e(x) \le U_e - \frac{(U_e - L_e)}{2}\gamma \\
& x \in \mathcal{H}, \gamma \in \mathbb{R} \\
& \text{for } e = 1, \ldots, N
\end{aligned}
$$

If **inconsistent**, … ???


*Feeley, R.; Seiler, P.; Packard, A.; Frenklach., M.; *J. Phys. Chem. A.* 2004, *108*, 9573.

# Quantifying Consistency



**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

**Scalar Consistency Measure (SCM)***

$$
\begin{aligned}
\max_{x,\gamma} \quad & \gamma \\
\text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \le M_e(x) \le U_e - \frac{(U_e - L_e)}{2}\gamma \\
& x \in \mathcal{H}, \gamma \in \mathbb{R} \\
& \text{for } e = 1, \ldots, N
\end{aligned}
$$

- Inconsistency → models and data disagree
- Follow-up questions:
  - **What** are the sources of inconsistency?
  - **Where** do we begin to look?

*Feeley, R.; Seiler, P.; Packard, A.; Frenklach., M.; *J. Phys. Chem. A.* 2004, *108*, 9573.

# Quantifying Consistency

**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

## Scalar Consistency Measure (SCM)

$$\begin{aligned} \max_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \leq M_e(x) \leq U_e - \frac{(U_e - L_e)}{2}\gamma \\ & x \in \mathcal{H}, \gamma \in \mathbb{R} \\ & \text{for } e = 1, \ldots, N \end{aligned}$$

Lagrange multipliers from dual form

Sensitivities

Local: $\lambda_j \approx \frac{\partial(\text{SCM})}{\partial(\text{bound } j)}$

Global: $\Delta(\text{SCM}) \leq \lambda^T \Delta(\text{bounds})$

Feeley, R.; Seiler, P.; Packard, A.; Frenklach., M.; *J. Phys. Chem. A.* 2004, *108*, 9573.

# Quantifying Consistency



**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

**Scalar Consistency Measure (SCM)**

$$\begin{aligned} \max_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \le M_e(x) \le U_e - \frac{(U_e - L_e)}{2}\gamma \\ & x \in \mathcal{H}, \gamma \in \mathbb{R} \\ & \text{for } e = 1, \dots, N \end{aligned}$$

New question: What is the **fewest** number of constraint relaxations required to render the dataset consistent?

# Vector Consistency



**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

## Scalar Consistency Measure (SCM)

$$\begin{aligned} \max_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \leq M_e(x) \leq U_e - \frac{(U_e - L_e)}{2}\gamma \\ & x \in \mathcal{H}, \gamma \in \mathbb{R} \\ & \text{for } e = 1, \ldots, N \end{aligned}$$

If inconsistent, compute the *vector consistency measure* (**VCM**)

- Offers detailed analysis of inconsistency by allowing independent relaxations.
- Can be used to flag constraints contributing to inconsistency

## Vector Consistency Measure (VCM)

$$\begin{aligned} \min_{x,\Delta} \quad & \|\Delta^L\|_1 + \|\Delta^U\|_1 \\ \text{s.t.} \quad & L_e - \Delta_e^L \leq M_e(x) \leq U_e + \Delta_e^U \\ & \Delta_e^L, \Delta_e^U \in \mathbb{R}, x \in \mathcal{H} \\ & \text{for } e = 1, \ldots, N \end{aligned}$$

# Vector Consistency



**Q:** Does there exist a parameter vector $x \in \mathcal{H}$ for which the models and data agree, within uncertainty?

**A:** Compute the *scalar consistency measure* (**SCM**)

## Scalar Consistency Measure (SCM)

$$\begin{aligned} \max_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & L_e + \frac{(U_e - L_e)}{2}\gamma \le M_e(x) \le U_e - \frac{(U_e - L_e)}{2}\gamma \\ & x \in \mathcal{H}, \gamma \in \mathbb{R} \\ & \text{for } e = 1, \ldots, N \end{aligned}$$

If inconsistent, compute the *vector consistency measure* (**VCM**)

- Offers detailed analysis of inconsistency by allowing independent relaxations.
- Can be used to flag constraints contributing to inconsistency

## Vector Consistency Measure (VCM)

$$\begin{aligned} \min_{x,\Delta} \quad & \boxed{\|\Delta^L\|_1 + \|\Delta^U\|_1} \\ \text{s.t.} \quad & L_e - \Delta_e^L \le M_e(x) \le U_e + \Delta_e^U \\ & \Delta_e^L, \Delta_e^U \in \mathbb{R}, x \in \mathcal{H} \\ & \text{for } e = 1, \ldots, N \end{aligned}$$

heuristic for fewest # of nonzeros (sparsity)

# Examples*

* Hegde, A.; Li, W.; Oreluk, J.; Packard, A.;  Frenklach, M., *SIAM/ASA J. Uncert. Quantif.,* 2018, 6(2), 429-456.

# Example 1: GRI-Mech 3.0



GRI-Mech 3.0 dataset (**77 QOIs, 102 uncertain parameters**) for natural gas combustion.

### Scalar Consistency

- **Procedure:** apply SCM, use sensitivities to flag problems.

### Vector Consistency

# Example 1: GRI-Mech 3.0



GRI-Mech 3.0 dataset (**77 QOIs, 102 uncertain parameters**) for natural gas combustion.

### Scalar Consistency

- **Procedure:** apply SCM, use sensitivities to flag problems.
- SCM < 0. Analyze ranked sensitivities

### Vector Consistency

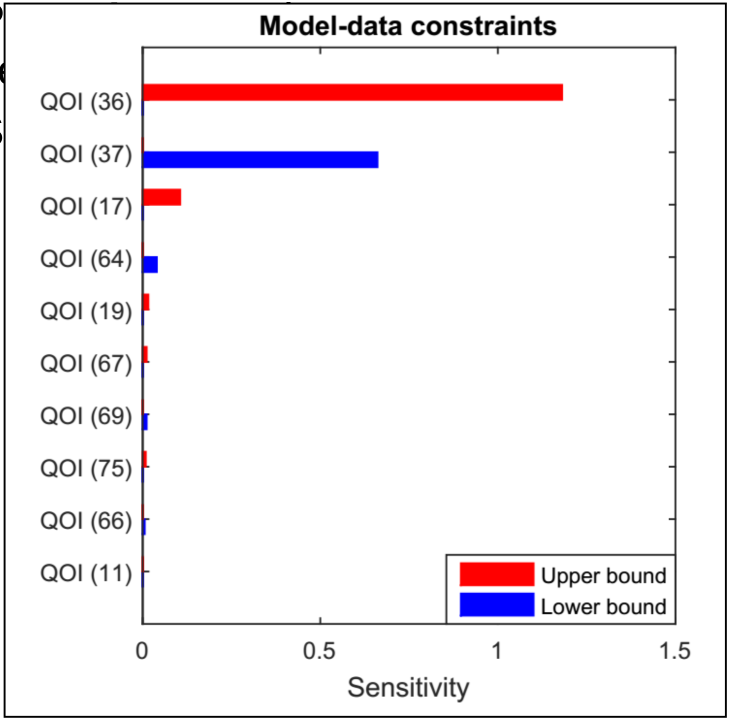
# Example 1: GRI-Mech 3.0



GRI-Mech 3.0 dataset (**77 QOIs, 102 uncertain parameters**) for natural gas combustion.

## Scalar Consistency

- **P** ...
  s...
- S...



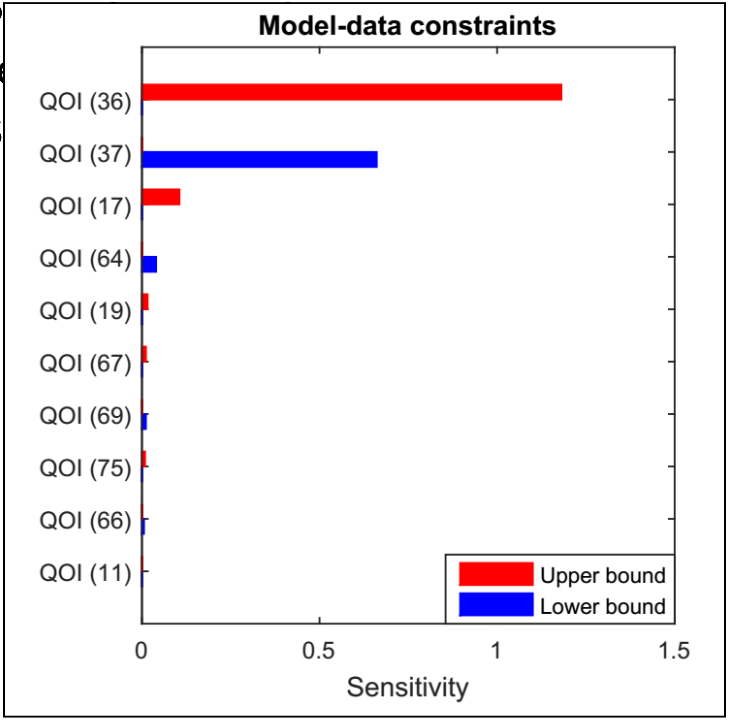

## Vector Consistency

# Example 1: GRI-Mech 3.0

GRI-Mech 3.0 dataset (**77 QOIs, 102 uncertain parameters**) for natural gas combustion.

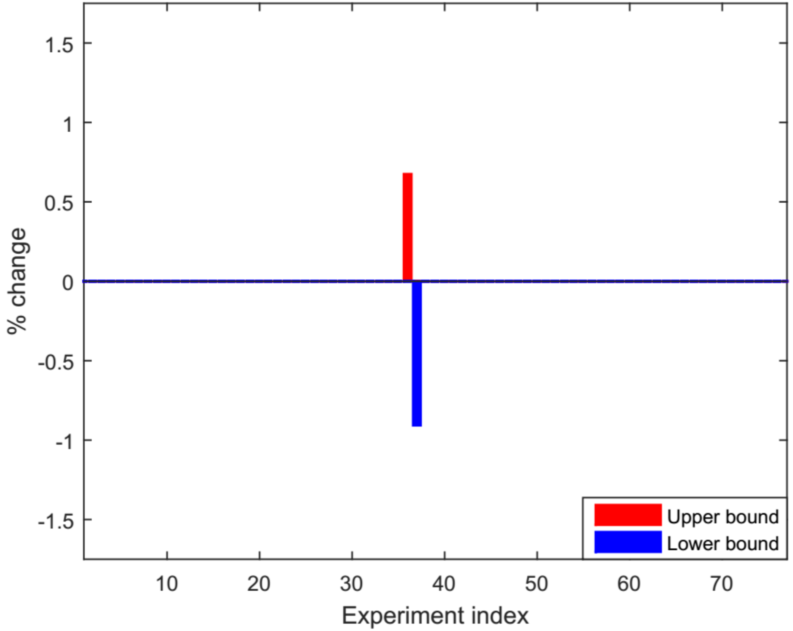
## Scalar Consistency

- **P** se
- S




## Vector Consistency

- Compute VCM.



# Example 1: GRI-Mech 3.0



GRI-Mech 3.0 dataset (**77 QOIs, 102 uncertain parameters**) for natural gas combustion.

### Scalar Consistency

- **Procedure:** apply SCM, use sensitivities to flag problems.
- SCM < 0. Analyze ranked sensitivities
- SCM > 0. Two QOIs removed, dataset consistent.

### Vector Consistency

- Compute VCM.
- Two QOIs relaxed (same as in SCM), dataset consistent.

# Example 1: GRI-Mech 3.0



GRI-Mech 3.0 dataset (**77 QOIs, 102 uncertain parameters**) for natural gas combustion.

### Scalar Consistency

- **Procedure:** apply SCM, use sensitivities to flag problems.

Rapid and interpretable resolution of inconsistency

### Vector Consistency

- Compute VCM.
- Two QOIs relaxed (same as in SCM),

Rapid and interpretable resolution of inconsistency

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR*.

**Scalar Consistency**

**Vector Consistency**

* Slavinskaya, N.; et al. *Energy & Fuels.* 2017, 31, pp 2274–2297

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

**Scalar Consistency**

- SCM < 0. Analyze ranked sensitivities.

**Vector Consistency**

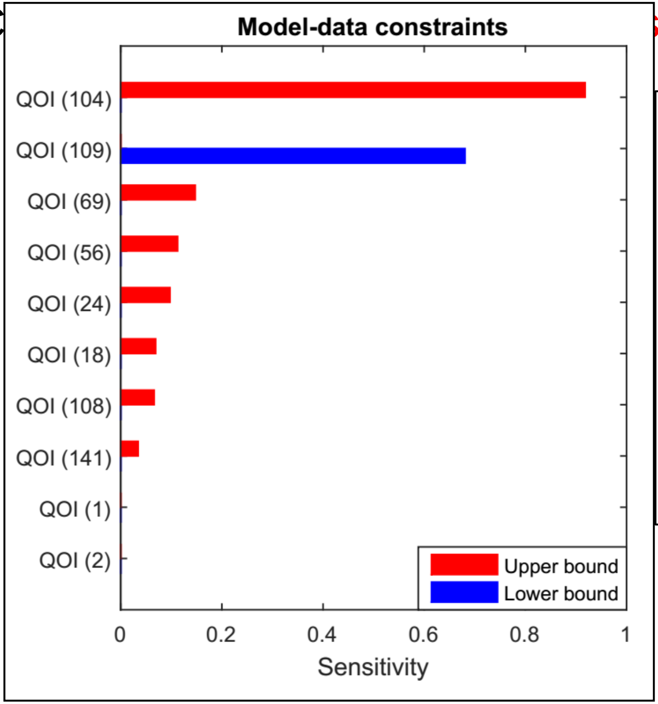
# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

## Scalar Consistency

- SC




## Vector Consistency

Set aside the top most sensitive QOI

Set aside the top two most sensitive QOIs

Set aside the top *n* most sensitive QOIs

Set aside the second most sensitive QOI (counter intuitive)

.

.

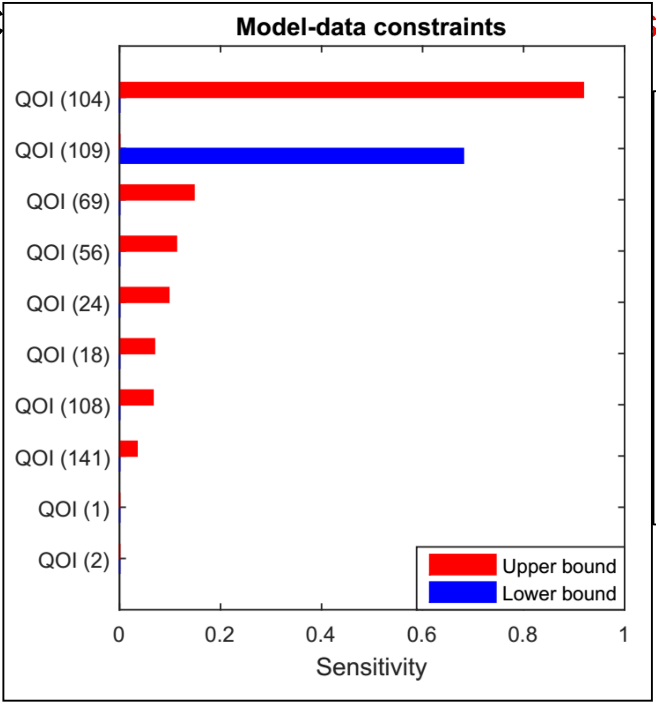
.

# Example 2: DLR-SynG

DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

## Scalar Consistency

- SC




## Vector Consistency

**Set aside the top most sensitive QOI**

~~Set aside the top two most sensitive QOIs~~

~~Set aside the top *n* most sensitive QOIs~~

~~Set aside the second most sensitive QOI (counter intuitive)~~

.

.

.

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

## Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.

## Vector Consistency

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

## Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.

## Vector Consistency

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

### Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

### Vector Consistency

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

## Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

## Vector Consistency

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

### Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

- This strategy results in the removal of 73 QOIs.

### Vector Consistency

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

**Scalar Consistency**

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

- This strategy results in the removal of 73 QOIs.
- Another strategy results in 56 QOIs removed.

**Vector Consistency**

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

### Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

- This strategy results in the removal of 73 QOIs.
- Another strategy results in 56 QOIs removed.

### Vector Consistency

- Compute VCM.

# Example 2: DLR-SynG

DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

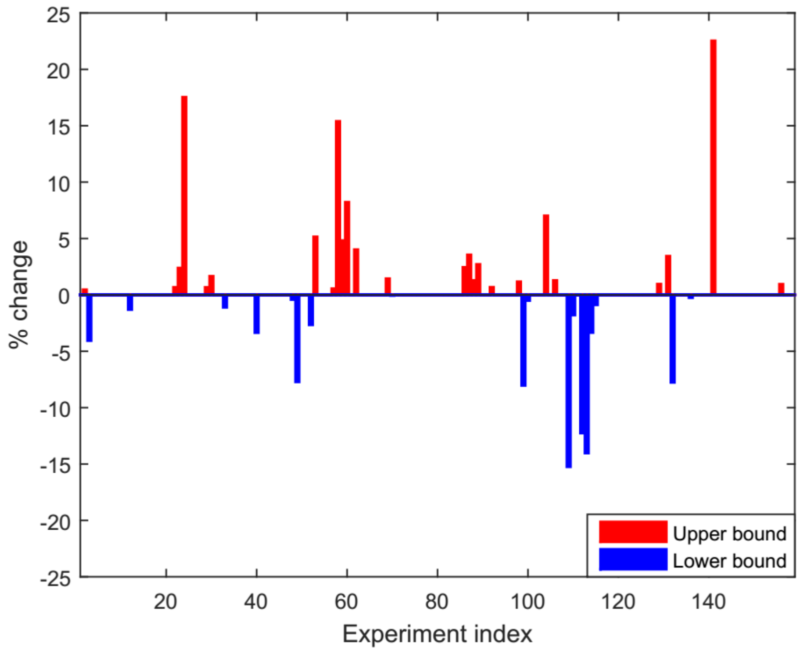
**Scalar Consistency**

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

- This strategy results in the removal of 73 QOIs.
- Another strategy results in 56 QOIs removed.

Implementing 43 relaxations results in a consistent dataset



# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

### Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

- This strategy results in the removal of 73 QOIs.
- Another strategy results in 56 QOIs removed.

### Vector Consistency

- Compute VCM.
  - Recommends 43 relaxations (18 to lower bounds, 25 to upper bounds)

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

### Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI # 109.

Repeat until consistent

- This strategy results in the removal of 73 QOIs.
- Another strategy results in 56 QOIs removed.

### Vector Consistency

- Compute VCM.
  - Recommends 43 relaxations (18 to lower bounds, 25 to upper bounds)

Example of what we termed **massive inconsistency**

# Example 2: DLR-SynG



DLR-SynG dataset (**159 QOIs, 55 uncertain parameters**) for syngas combustion developed at DLR.

### Scalar Consistency

- SCM < 0. Analyze ranked sensitivities.
  - Remove QOI #104 from dataset.
- SCM < 0. Analyze ranked sensitivities

Indirect and inefficient resolution of inconsistency

73 QOIs.

- Another strategy results in 56 QOIs removed.

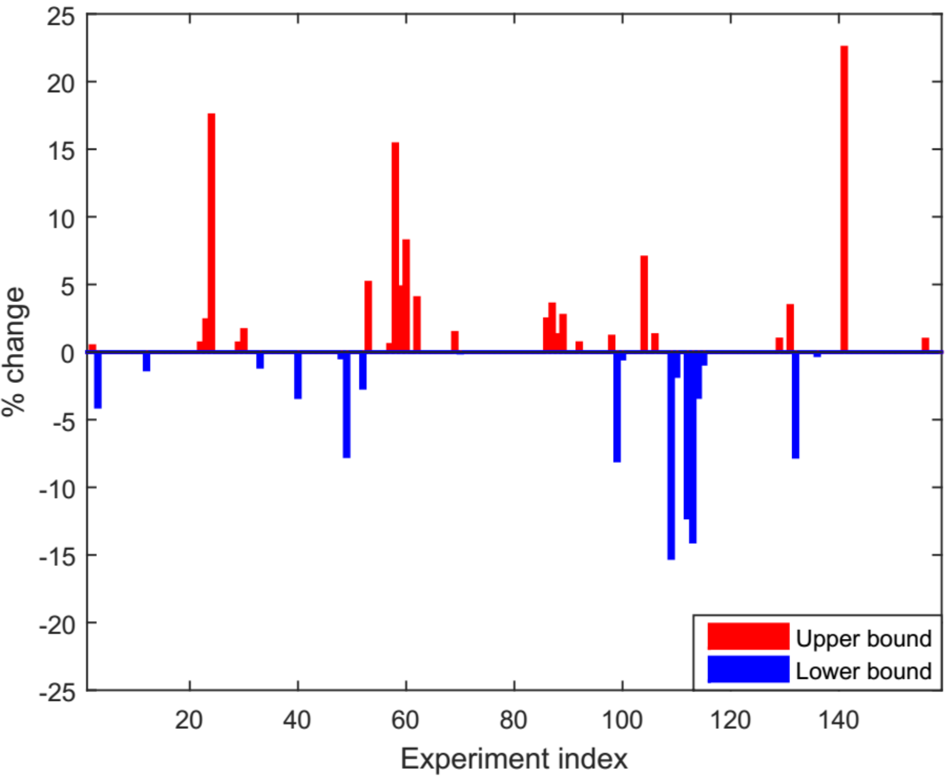
### Vector Consistency

- Compute VCM.
  - Recommends 43 relaxations (18 to lower bounds, 25 to upper bounds)

Direct, one-shot resolution of inconsistency

# Including weights in the VCM






What if we are unwilling to change certain experimental bounds?

# Including weights in the VCM

- **Goal:** To allow domain expert knowledge and opinions enter VCM as weights.

- **Motivation:** If a <u>dataset</u> is inconsistent, one should be less willing to relax model-data constraints they trust and more willing to relax constraints that are less reliable. The same goes for parameter bounds.

**Weighted VCM**

$$\begin{aligned} \min_{x,\Delta^L,\Delta^U,\delta^l,\delta^u} \quad & \|\Delta^L\|_1 + \|\Delta^U\|_1 + \|\delta^l\|_1 + \|\delta^u\|_1 \\ \text{s.t.} \quad & L_e - W_e^L \Delta_e^L \leq M_e(x) \leq U_e + W_e^U \Delta_e^U \qquad \text{for } e = 1, \dots, N \\ & l_i - w_i^l \delta_i^l \leq x_i \leq u_i + w_i^u \delta_i^u \qquad \text{for } i = 1, \dots, n \end{aligned}$$

- Small weight - less willing to change bound.
- Large weight - more willing to change bound.

# Including weights in the VCM



- **Goal:** To allow domain expert knowledge and opinions enter VCM as weights.

- **Motivation:** If a dataset is inconsistent, one should be less willing to relax model-data constraints they trust and more willing to relax constraints that are less reliable. The same goes for parameter bounds.

**Weighted VCM**

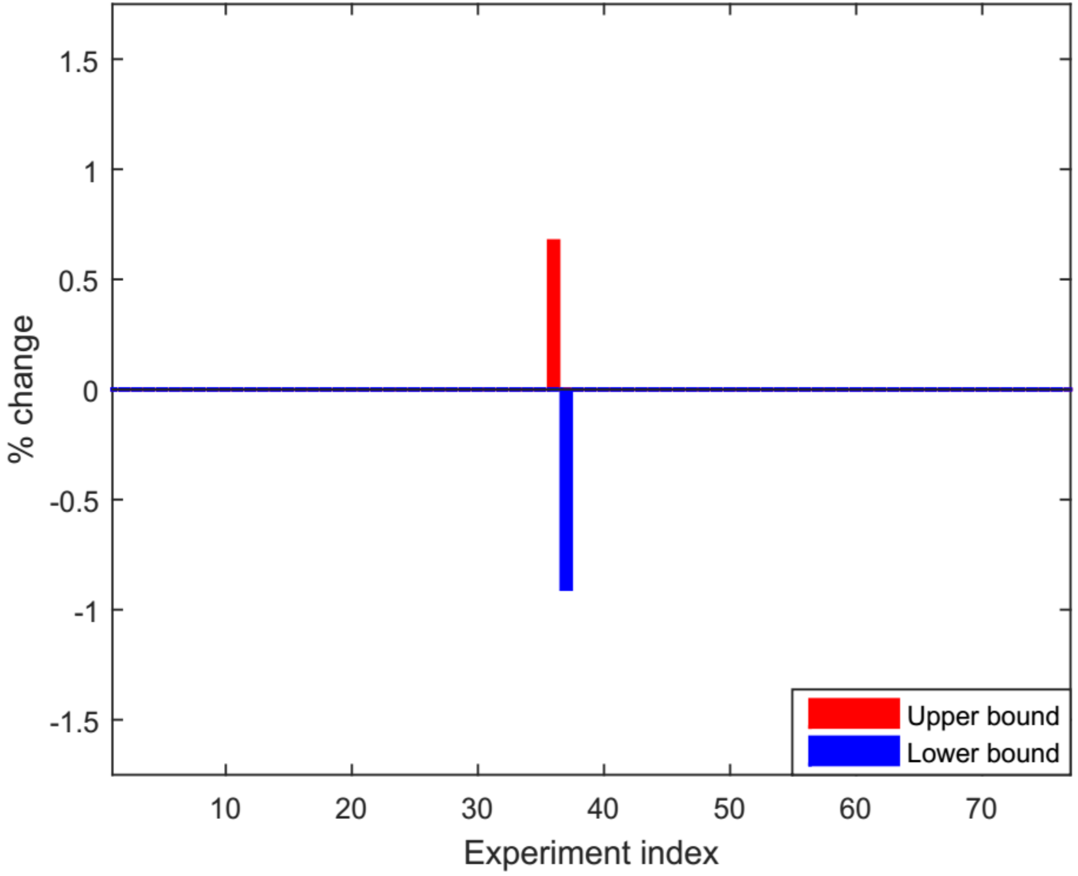
$$\begin{aligned}
\min_{x,\Delta^L,\Delta^U,\delta^l,\delta^u} \quad & \|\Delta^L\|_1 + \|\Delta^U\|_1 + \|\delta^l\|_1 + \|\delta^u\|_1 \\
\text{s.t.} \quad & L_e - W_e^L \Delta_e^L \leq M_e(x) \leq U_e + W_e^U \Delta_e^U \qquad \text{for } e = 1, \ldots, N \\
& l_i - w_i^l \delta_i^l \leq x_i \leq u_i + w_i^u \delta_i^u \qquad \text{for } i = 1, \ldots, n
\end{aligned}$$

$$W_e^{L/U} = U_e - L_e$$

With these weights, DLR-SynG can be made consistent by adjusting 37 QOIs.

# Weights and GRI-Mech 3.0

- Single application of VCM identifies two experimental bounds.

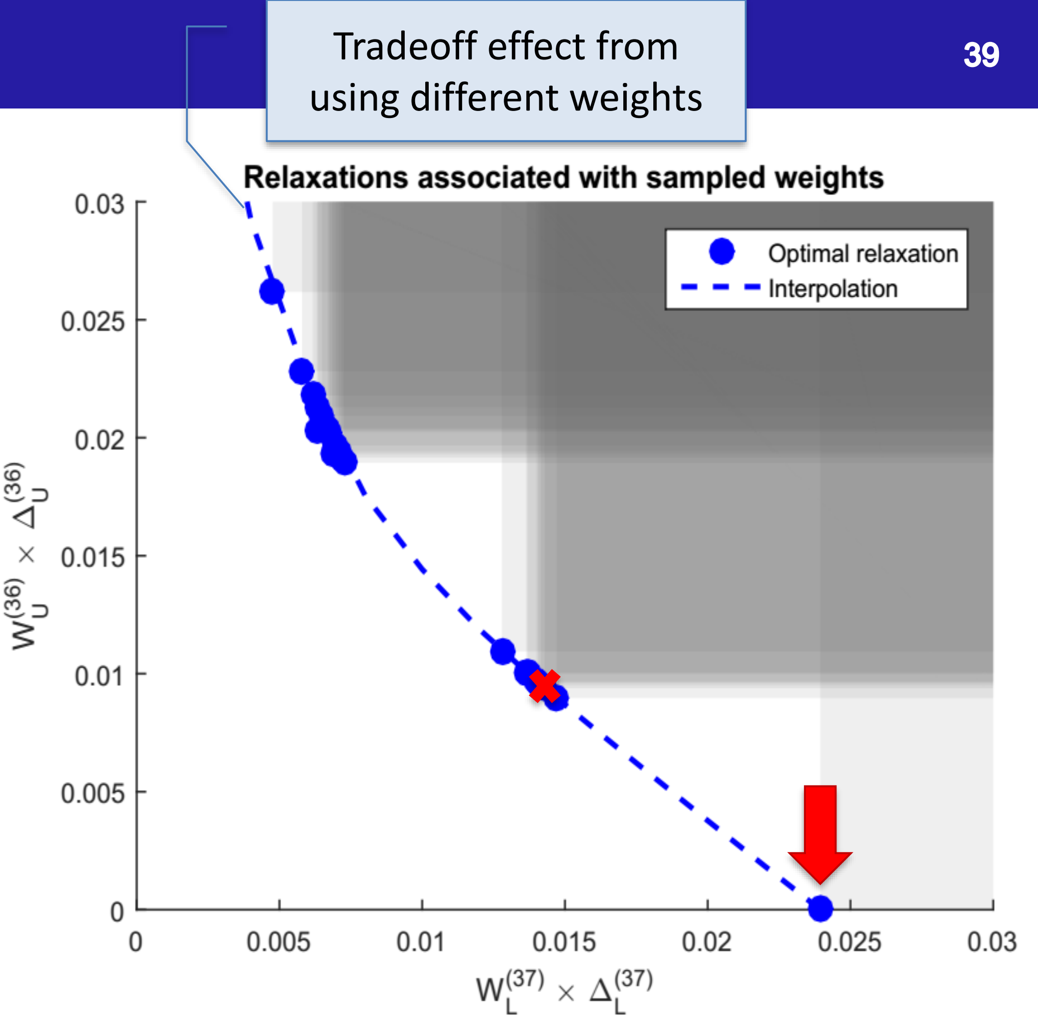


# Weights and GRI-Mech 3.0



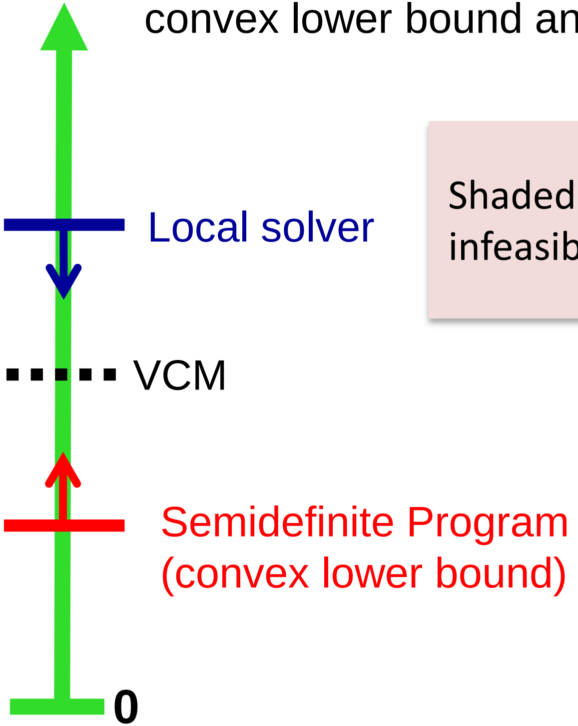
- Single application of VCM identifies two experimental bounds.

- Weights applied to only the previous two bounds.



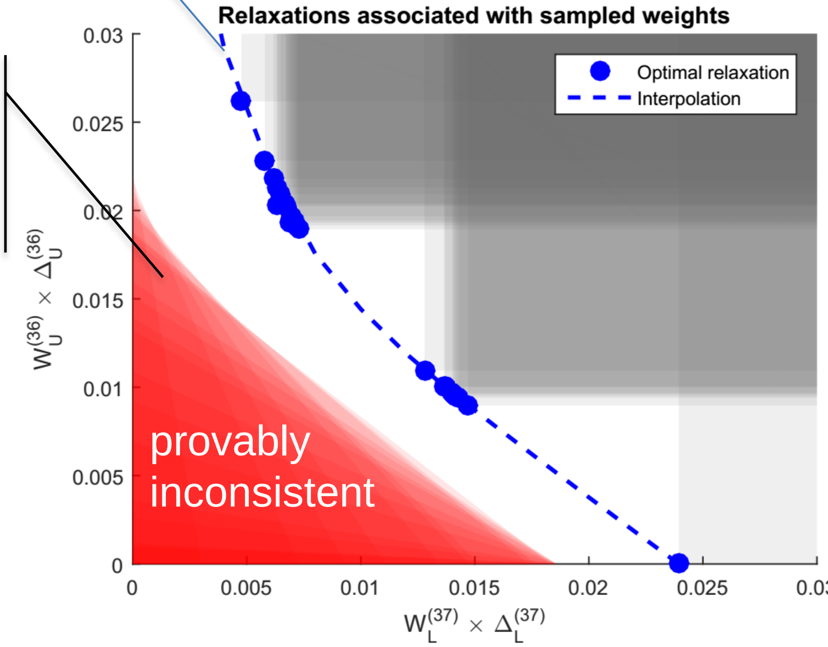
# Computing the VCM

For **quadratic surrogate models**, we can approximate the true solution (NP-hard) using convex lower bound and local optima.




Shaded red region certified infeasible by SDP.

Tradeoff effect from using different weights




Example: GRI-Mech 3.0 dataset – VCM relaxations with varying weights, relaxations allowed to two constraints

# Summary



- **Consistency measures** are tools to accomplish **validation**
  - **Scalar Consistency Measure (SCM)** – are we consistent?
  - **Vector Consistency Measure (VCM)** – diagnose inconsistency

- **VCM** particularly efficient for resolving **massive inconsistency**

- **Weighted VCM** allows inclusion of expert opinions

- Application: GRI-Mech 3.0, DLR-SynG

# Acknowledgements



This work is supported as a part of the CCMSC at the University of Utah, funded through PSAAP by the National Nuclear Security Administration, under Award Number DE-NA0002375.

NNSA
National Nuclear Security Administration

CARBON CAPTURE
MULTIDISCIPLINARY
SIMULATION CENTER

# Questions?