Providing structure to experimental data:
A large scale heterogeneous database for collaborative model validation

Jim Oreluk
Arun Hegde
Wenyu Li
Andrew Packard
Michael Frenklach
Overview

• Introduction

• Giving structure to experimental data
  • PrIMe Data Warehouse

• New PrIMe application
  • front-end application to the CCMSC coal database (filter, visualization, and export data)

• Bound-to-Bound Data Collaboration workflow for model validation

• Summary
Introduction

• Predictive modeling starts with validation

• Experimental data stored in various file formats
  – CSV, Excel, tab delimited, ASCII, etc.
  – No standard

• Each record requires specialized knowledge of how the data was stored
  – Can be an incomplete record of experiment with missing information

• We would like automated access to data
  – Without structure, query requests are quickly intractable across a diverse collection of data

• Efficiently discover validation data to incorporate in the model validation process
Providing Structure to Experimental Data

- What is PrIMe?
  - Data Warehouse – repository of experimental records
  - Applications – aid in development of predictive models
- Transformation of information into a usable form
- PrIMe’s data models use XML schemas to provide structure
  - Contains complete information of an experiment
  - Experimental data is stored in XML or HDF5 files
- Storage of raw experimental data and derived properties
  - Ability for instrumentation modeling
CCMSC Coal Database for V/UQ

Dataset unit $U_c = (U_e, L_e, M_e)$

- International Flame Research Foundation, Livorno, Italy
- Sandia National Laboratory, Livermore, CA

269 Solid Fuels & Blends
- Fossil, Biomass, Sludge, Waste, Char

2710 Data Groups collected from 1016 Records
- Varying Conditions (Temperatures, %O$_2$, %H$_2$O, Gas Mixture)
- Experiment Types: Devolatilization, Char oxidation

In collaboration with Salvatore Iavaron and Alessandro Parente, Université Libre de Bruxelles
### CCMSC Coal Database

#### primekinetics.org  
github.com/oreluk/coalDB

<table>
<thead>
<tr>
<th>Coal Name</th>
<th>Coal Rank</th>
<th>% O₂</th>
<th>% H₂O</th>
<th>Gas Mixture</th>
<th>Temp [K]</th>
<th>Properties</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashland char</td>
<td>Coal</td>
<td>5.5</td>
<td>8.6</td>
<td>O₂ / H₂O / N₂</td>
<td>1231</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Ashland char</td>
<td>Coal</td>
<td>5.7</td>
<td>11.6</td>
<td>O₂ / H₂O / N₂</td>
<td>1496</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Ashland char</td>
<td>Coal</td>
<td>5.4</td>
<td>14.1</td>
<td>O₂ / H₂O / N₂</td>
<td>1713</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Coal-2 sub coal, Powder River Basin c.</td>
<td>Coal / Subbit.</td>
<td>4</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1673</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Liser 2003</td>
</tr>
<tr>
<td>Ashland char</td>
<td>Coal</td>
<td>5.78</td>
<td>13.85</td>
<td>O₂ / H₂O / N₂</td>
<td>1713</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Ashland char</td>
<td>Coal</td>
<td>5.32</td>
<td>8.45</td>
<td>O₂ / H₂O / N₂</td>
<td>1451</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Retorutor</td>
<td>Bituminous</td>
<td>0</td>
<td>18.1</td>
<td>H₂O / N₂</td>
<td>1673</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Hoas et al 1996</td>
</tr>
<tr>
<td>Retorutor</td>
<td>Coal / Bituminous</td>
<td>6.2</td>
<td>9</td>
<td>O₂ / H₂O / N₂</td>
<td>1223</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Retorutor</td>
<td>Coal / Bituminous</td>
<td>5.4</td>
<td>11.9</td>
<td>O₂ / H₂O / N₂</td>
<td>1478</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Retorutor</td>
<td>Coal / Bituminous</td>
<td>5.5</td>
<td>14.1</td>
<td>O₂ / H₂O / N₂</td>
<td>1733</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Retorutor</td>
<td>Coal / Bituminous</td>
<td>5.81</td>
<td>9.3</td>
<td>O₂ / H₂O / N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Dalmon et al 1998</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Bituminous</td>
<td>0</td>
<td>-</td>
<td>N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Coal / Bituminous</td>
<td>4.4</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1223</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Coal / Bituminous</td>
<td>5.5</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1223</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Coal-2 sub coal, Powder River Basin c.</td>
<td>Coal / Subbit.</td>
<td>4</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Liser 2003</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Coal / Bituminous</td>
<td>8.3</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Coal / Bituminous</td>
<td>11.7</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Coal / Bituminous</td>
<td>6</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Douglas Premium 1</td>
<td>Coal / Bituminous</td>
<td>3.4</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1673</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Enviro sample 1</td>
<td>Bituminous</td>
<td>0</td>
<td>-</td>
<td>N₂</td>
<td>1473</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Enviro sample 1</td>
<td>Coal / Bituminous</td>
<td>4.1</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1223</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Enviro sample 1</td>
<td>Coal / Bituminous</td>
<td>0.1</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1223</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
<tr>
<td>Enviro sample 1</td>
<td>Coal / Bituminous</td>
<td>8</td>
<td>-</td>
<td>O₂ / N₂</td>
<td>1223</td>
<td>Residence Time, Fraction of Total Weight Loss</td>
<td>Tamura et al 1996</td>
</tr>
</tbody>
</table>

**Data Groups Found:** 2710
CCMSC Coal Database

Select Experiments

Plot & Export Data

Select experiments based on Coal Name.

Plot Data

Export Data

Filter by Coal Name.
Char Oxidation Example

Experimental Data of Utah Skyline coal from Sandia’s Laminar Entrained Flow Reactor

Features:
- CO₂ or N₂ diluent
- Initial Particle Diameter: 53 – 150 μm
- O₂: 24 – 60%
- H₂O: 10 – 16%

Validation data at 399 different gas conditions & heights above burner
Bound-to-Bound Data Collaboration (B2BDC)

Uncertain parameters: \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[ \mathcal{H} = \{ x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, \ i = 1, \ldots, n \} \]

Bounds on QOI model: \( L_e \leq M_e(x) \leq U_e, \ \text{for} \ e = 1, \ldots, N \)

**Dataset:** \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[ L_e \leq M_e(x) \leq U_e, \ \text{for} \ e = 1, \ldots, N \]

Bound-to-Bound Data Collaboration (B2BDC)

Uncertain parameters: \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[
\mathcal{H} = \{ x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, \; i = 1, \ldots, n \}
\]

Bounds on QOI model: \( L_e \leq M_e(x) \leq U_e, \; \text{for} \; e = 1, \ldots, N \)

\[Dataset:\]

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

\( L_e \leq M_e(x) \leq U_e, \; \text{for} \; e = 1, \ldots, N \)

Bound-to-Bound Data Collaboration (B2BDC)

Uncertain parameters: \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[ \mathcal{H} = \{ x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, \ i = 1, \ldots, n \} \]

Bounds on QOI model: \( L_e \leq M_e(x) \leq U_e \), for \( e = 1, \ldots, N \)

Dataset: \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[ L_e \leq M_e(x) \leq U_e, \ \text{for} \ e = 1, \ldots, N \]

Bound-to-Bound Data Collaboration (B2BDC)

Uncertain parameters: \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[ \mathcal{H} = \{ x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, \ i = 1, \ldots, n \} \]

Bounds on QOI model: \( L_e \leq M_e(x) \leq U_e \), for \( e = 1, \ldots, N \)

**Dataset:** \( x \in \mathcal{H} \subseteq \mathbb{R}^n \)

\[ L_e \leq M_e(x) \leq U_e, \ \text{for} \ e = 1, \ldots, N \]

Feasible set: \( \mathcal{F} = \bigcap_{e=1}^{N} \{ x \in \mathcal{H} : L_e \leq M_e(x) \leq U_e \} \)
Scalar consistency measure

\[ \gamma \geq 0 : \quad \text{Consistent Dataset} \]

\[ \gamma < 0 : \quad \text{Inconsistent Dataset} \]

\[ C_{\text{Dataset}} = \max_{\gamma, x \in H} \gamma \]

s.t. \[ L_e + \frac{(U_e - L_e)}{2} \gamma \leq M_e(x) \leq U_e - \frac{(U_e - L_e)}{2} \gamma \]

for \( e = 1, \ldots, N. \)

Bound-to-Bound Data Collaboration (B2BDC)

Scalar consistency measure

\[ \gamma \geq 0 : \ \text{Consistent Dataset} \]

\[ \gamma < 0 : \ \text{Inconsistent Dataset} \]

\[
C_{\text{Dataset}} = \max_{\gamma, x \in H} \gamma
\]

\[
\text{s.t.} \quad L_e + \frac{(U_e - L_e)}{2} \gamma \leq M_e(x) \leq U_e - \frac{(U_e - L_e)}{2} \gamma
\]

for \( e = 1, \ldots, N \).
B2BDC Model Validation Workflow

Uncertain Parameters $x \in \mathcal{H}$
Scenario Parameters, $x_{s,e}$

Char Oxidation Model (Instrument + Physics)

CCMSC Coal Database

Response $[\text{Image of Particle Temp distribution $&$ highlight QOI}]

$QOÎ_e(x)$

Particle Temperature
B2BDC Model Validation Workflow

Uncertain Parameters $x \in \mathcal{H}$

Scenario Parameters, $x_{s,e}$

Char Oxidation Model (Instrument + Physics)

CCMSC Coal Database

Response $\text{QOI}_e(x)$

Particle Temperature
B2BDC Model Validation Workflow

Uncertain Parameters, $x \in \mathcal{H}$

Scenario Parameters, $x_{sc}$

Char Oxidation Model (Instrument + Physics)

$M_e(x)\rightarrow QOI_e(x)$

Particle Temperature

CCMSC Coal Database
B2BDC Model Validation Workflow

Uncertain Parameters $x \in \mathcal{H}$

Scenario Parameters, $x_{s,e}$

Char Oxidation Model (Instrument + Physics)

Dataset Unit $L_e \leq M_e(x) \leq U_e$

CCMSC Coal Database

$y_e \in [L_e, U_e]$
B2BDC Model Validation Workflow

Uncertain Parameters $x \in \mathcal{H}$

Scenario Parameters, $x_{s,e}$

Char Oxidation Model (Instrument + Physics)

CCMSC Coal Database

$y_e \in [L_e, U_e]$  

$L_e \leq M_e(x) \leq U_e$

Dataset Unit  

Dataset

Consistency Analysis

$QOI_{e}(x)$

Particle Temperature
Validation through consistency

Model Form

\[
\begin{align*}
C_{(s)} + O_2 & \rightarrow CO_2 \\
2C_{(s)} + O_2 & \rightarrow 2CO
\end{align*}
\]

\[ r = A \exp \left( -\frac{E\alpha}{RT_p} \right) C_{\text{oxid}} \]

Transport

- Diffusion of oxidizer to particle surface
- Diffusion of products from particle surface

Scalar consistency measure:

\[
C_{\text{Dataset}} = [-0.26, -0.19]
\]

If all constraints are expanded by at least 26% the inconsistency can be resolved.

If all constraints are expanded by no more than 19% the inconsistency cannot be resolved.

Validation through consistency

**Model Form**

\[
\begin{align*}
C_{(s)} + O_2 &\rightarrow CO_2 \\
2C_{(s)} + O_2 &\rightarrow 2CO
\end{align*}
\]

\[r = A \exp\left(\frac{-E_a}{RT_p}\right)C_{oxid}\]

**Transport**
- Diffusion of oxidizer to particle surface
- Diffusion of products from particle surface

**Scalar consistency measure:**

\[C_{\text{Dataset}} = [-0.26, -0.19]\]

If **all** constraints are expanded by at least 26% the inconsistency can be resolved.

If **all** constraints are expanded by no more than 19% the inconsistency **cannot** be resolved.
Validation through consistency

**Model Form**

**Uncertain Kinetic Parameters**

\[
\begin{align*}
C_{(s)} + O_2 & \rightarrow CO_2 \quad \{A_{O_2}, E_{O_2}\} \\
2C_{(s)} + O_2 & \rightarrow 2CO \quad \{A_{CO_2}, E_{CO_2}\} \\
C_{(s)} + CO_2 & \rightarrow 2CO \quad \{A_{H_2O}, E_{H_2O}\}
\end{align*}
\]

**Transport**

- Diffusion of oxidizer to particle surface
- Diffusion of products from particle surface
- Diffusion of oxidizer through coal particle
  - coal particle is a porous medium with internal surface area

**Scalar consistency measure:**

\[
\mathcal{F} = \bigcap_{e=1}^{N} \{x \in \mathcal{H} : L_e \leq M_e(x) \leq U_e\}
\]

\[
C_{\text{Dataset}} = [0.06, 0.32]
\]
Validation through consistency

$N_2 : 54\%, \ H_2 \ O : 10\%, \ O_2 : 36\%, \ d_{\text{bin}} = 82.5\mu m$

$$x_{\text{LS-F}} = \arg\min_{x \in \mathcal{F}} \| M_e(x) - y_e \|_2$$

$y_e : \text{mean of the validation data}$
Summary

• Developed new data models for coal data

• Easy filtering through a diverse collection of experimental data

• B2BDC based test-bed for exploring parameter and model form uncertainty
  – With a consistent dataset we can do prediction of posterior QOI or parameter bounds, and sample the feasible set for correlations between parameters and QOIs
Acknowledgements

This material is based upon work supported by the Department of Energy, National Nuclear Security Administration, under Award Number(s) DE-NA0002375. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.