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

Jim Oreluk
Arun Hegde
Wenyu Li
Andrew Packard
Michael Frenklach

National Nuclear Security Administration

CARBON CAPTURE
MULTIDISCIPLINARY
SIMULATION CENTER

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

primekinetics.org

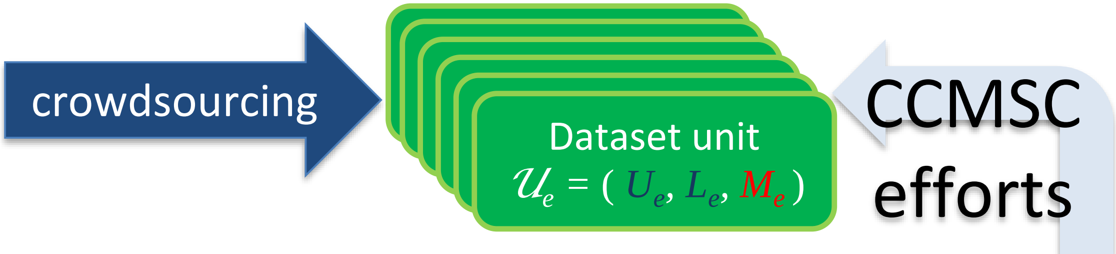
- What is PrIMe?
  - Data Warehouse – repository of experimental records
  - Applications – aid in development of predictive models

- 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



- 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_2O, Gas Mixture)

Experiment Types: *Devolatilization, Char oxidation*

In collaboration with Salvatore Iavaron and Alessandro Parente, Université Libre de Bruxelles

leveraging existing cloud infrastructure and data models


```
<?xml version="1.0"?>
<experiment xsi:schemaLocation="http://purl.org/NET/prime/
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  <copyright>©primekinetics.org 2017</copyright>
  <bibliographyLink primeID="b00019060" preferredKey="Leis
- <apparatus>
    <kind>drop-tube furnace</kind>
  - <property name="length" label="Length" units="m">
      <value>4</value>
    </property>
  - <property name="length" label="D_inner" units="mm">
      <value>150</value>
    </property>
  </apparatus>
- <commonProperties>
  + <property name="initial composition">
  + <property name="temperature" label="T_furnace" units
  + <property name="pressure" label="P" units="atm">
  </commonProperties>
- <dataGroup label="Residence time vs Weight Loss" id="dg
    <property description="Residence Time" name="resider
    <property description="Fraction of Total Weight Loss" r
  - <dataPoint id="dp1">
      <x1>0</x1>
      <x2>0</x2>
    </dataPoint>
  - <dataPoint id="dp2">
      <x1>232</x1>
      <x2>0.144</x2>
    </dataPoint>
  - <dataPoint id="dp3">
```

CCMSC Coal Database

primekinetics.org

github.com/oreluk/coalDB

CCMSC Coal Database

Options

Select	Coal Name	Coal Rank	% O2	% H2O	Gas Mixture	Temp [K]	Properties	Ref
☐	Ashland char	Coal	5.5	8.6	O2 / H2O / N2	1231	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Ashland char	Coal	5.7	11.6	O2 / H2O / N2	1498	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Ashland char	Coal	5.4	14.1	O2 / H2O / N2	1713	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Coal-2 sub coal, Powder River Basin c...	Coal / Subbit...	4	-	O2 / N2	1673	Residence Time, Fraction of Total Weight Loss	Leiser 2003
☐	Ashland char	Coal	5.78	13.85	O2 / H2O / N2	1713	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Ashland char	Coal	5.32	8.45	O2 / H2O / N2	1451	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Rietspruit	Bituminous	0	18.1	H2O / N2	1673	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	6.2	9	O2 / H2O / N2	1223	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	5.4	11.9	O2 / H2O / N2	1478	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	5.5	14.1	O2 / H2O / N2	1733	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	5.81	9.3	O2 / H2O / N2	1473	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Douglas Premium 1	Bituminous	0	-	N2	1473	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	4.4	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	5.5	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Coal-2 sub coal, Powder River Basin c...	Coal / Subbit...	4	-	O2 / N2	1473	Residence Time, Fraction of Total Weight Loss	Leiser 2003
☐	Douglas Premium 1 char	Coal / Bitumi...	8.3	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	11.7	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	6	-	O2 / N2	1473	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	3.4	-	O2 / N2	1673	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1	Bituminous	0	-	N2	1473	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1 char	Coal / Bitumi...	4.1	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1 char	Coal / Bitumi...	6.1	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1 char	Coal / Bitumi...	8	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998

Data Groups Found: 2710

Plot Data

Filter by Coal Name

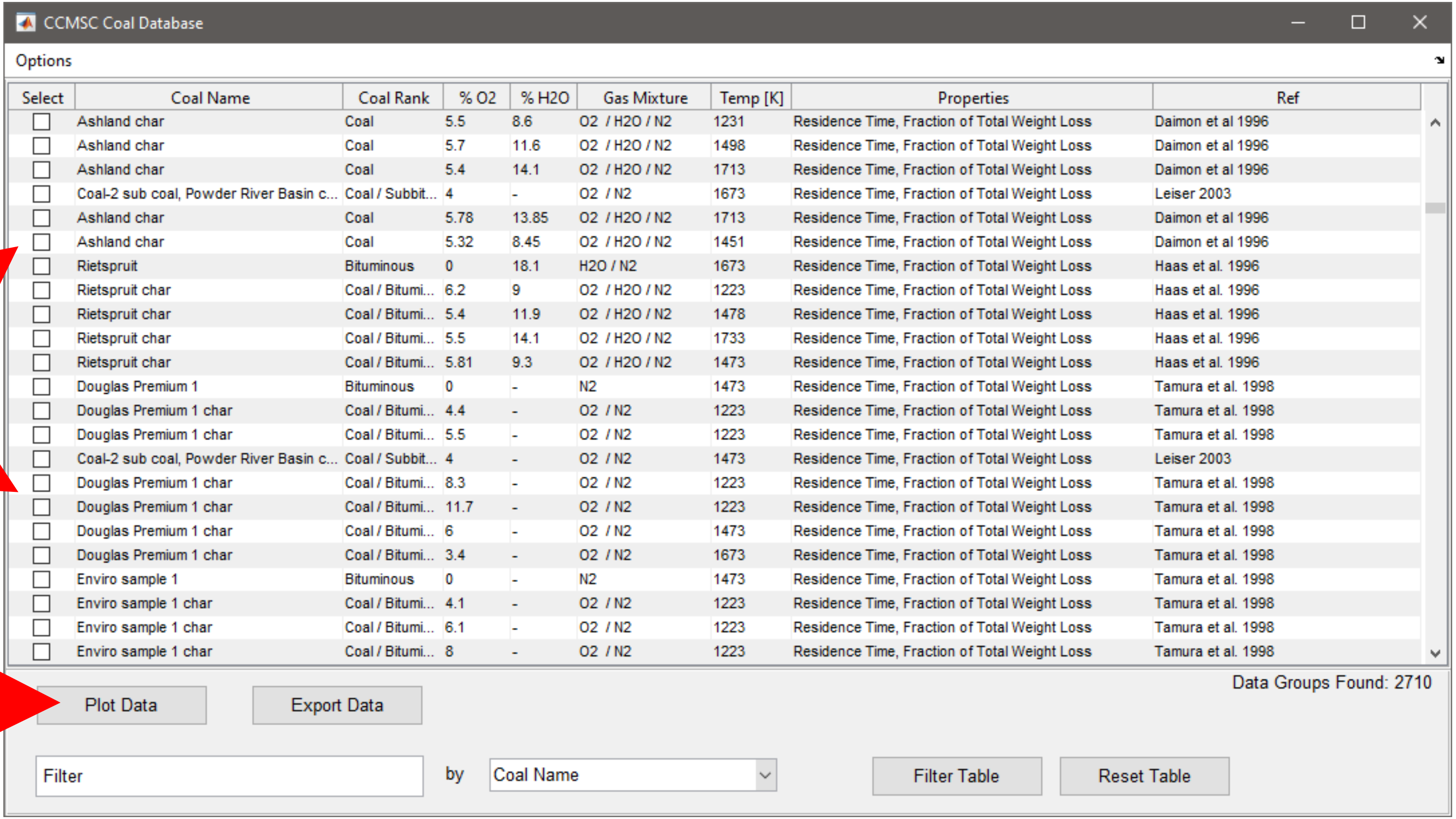
Filter Table

Reset Table

CCMSC Coal Database

Select Experiments

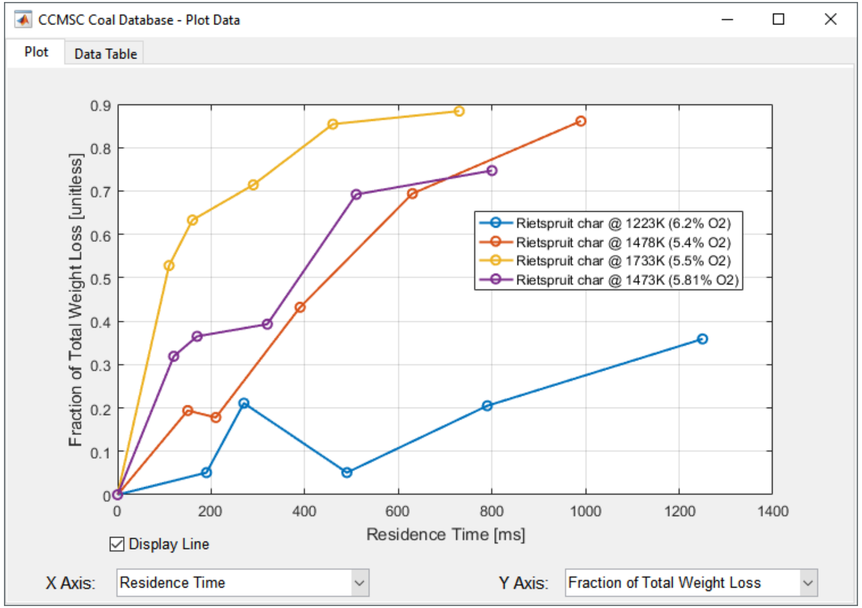
Plot & Export Data



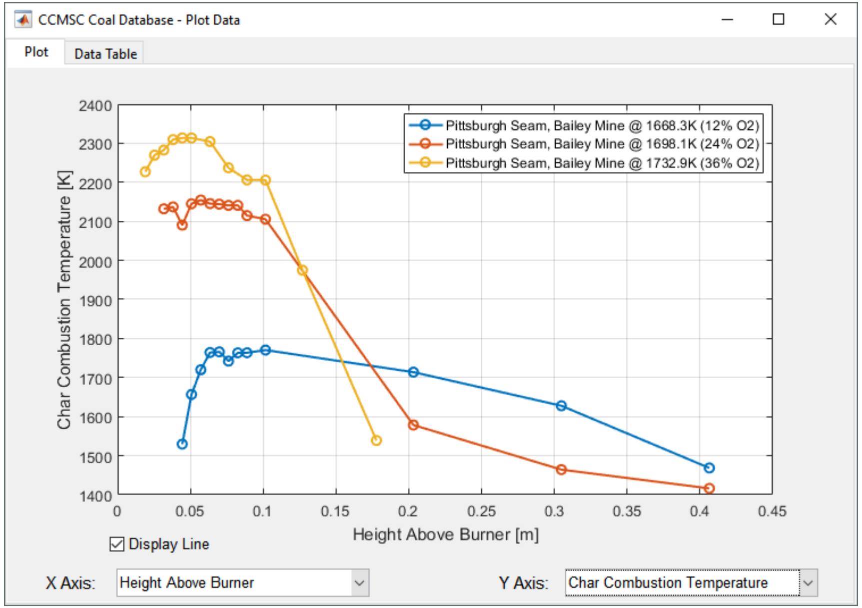
Select	Coal Name	Coal Rank	% O2	% H2O	Gas Mixture	Temp [K]	Properties	Ref
☐	Ashland char	Coal	5.5	8.6	O2 / H2O / N2	1231	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Ashland char	Coal	5.7	11.6	O2 / H2O / N2	1498	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Ashland char	Coal	5.4	14.1	O2 / H2O / N2	1713	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Coal-2 sub coal, Powder River Basin c...	Coal / Subbit...	4	-	O2 / N2	1673	Residence Time, Fraction of Total Weight Loss	Leiser 2003
☐	Ashland char	Coal	5.78	13.85	O2 / H2O / N2	1713	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Ashland char	Coal	5.32	8.45	O2 / H2O / N2	1451	Residence Time, Fraction of Total Weight Loss	Daimon et al 1996
☐	Rietspruit	Bituminous	0	18.1	H2O / N2	1673	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	6.2	9	O2 / H2O / N2	1223	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	5.4	11.9	O2 / H2O / N2	1478	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	5.5	14.1	O2 / H2O / N2	1733	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Rietspruit char	Coal / Bitumi...	5.81	9.3	O2 / H2O / N2	1473	Residence Time, Fraction of Total Weight Loss	Haas et al. 1996
☐	Douglas Premium 1	Bituminous	0	-	N2	1473	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	4.4	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	5.5	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Coal-2 sub coal, Powder River Basin c...	Coal / Subbit...	4	-	O2 / N2	1473	Residence Time, Fraction of Total Weight Loss	Leiser 2003
☐	Douglas Premium 1 char	Coal / Bitumi...	8.3	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	11.7	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	6	-	O2 / N2	1473	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Douglas Premium 1 char	Coal / Bitumi...	3.4	-	O2 / N2	1673	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1	Bituminous	0	-	N2	1473	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1 char	Coal / Bitumi...	4.1	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1 char	Coal / Bitumi...	6.1	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998
☐	Enviro sample 1 char	Coal / Bitumi...	8	-	O2 / N2	1223	Residence Time, Fraction of Total Weight Loss	Tamura et al. 1998

CCMSC Coal Database

Fraction of Weight Loss



Char Temperature



Char Oxidation Example

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

Features:

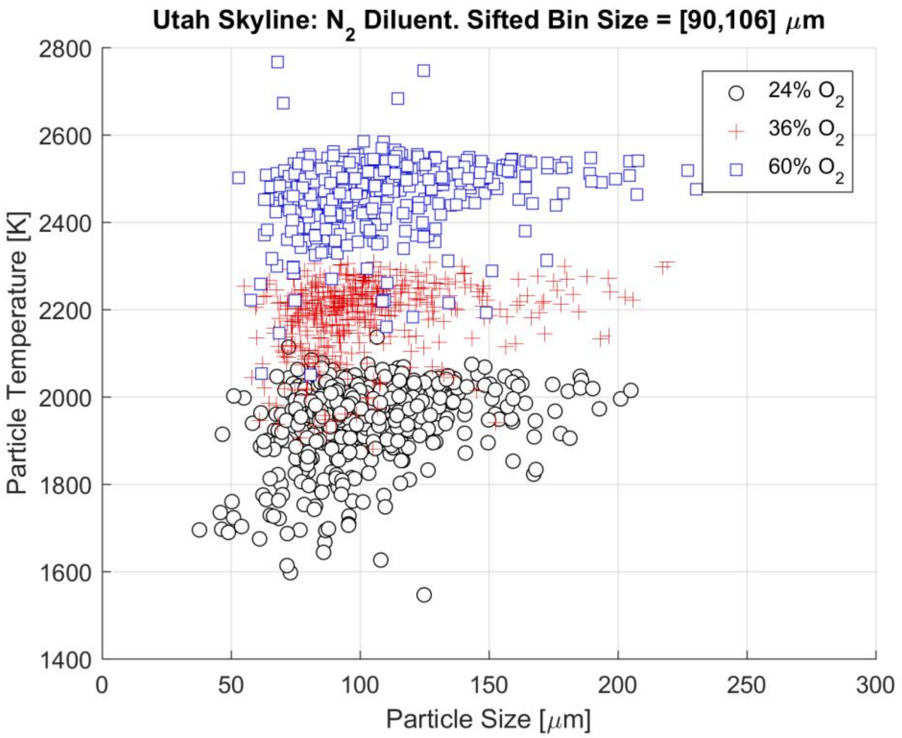
CO_2 or N_2 diluent

Initial Particle Diameter: $53 - 150\mu m$

O_2: $24 - 60\%$

H_2O: $10 - 16\%$

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

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

Feeley, Seiler, Packard, Frenklach. *J. Phy. Chem. A*, 2004, 108, 9573-9583.

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

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

Feeley, Seiler, Packard, Frenklach. *J. Phy. Chem. A*, 2004, 108, 9573-9583.

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

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

Feeley, Seiler, Packard, Frenklach. *J. Phy. Chem. A*, 2004, 108, 9573-9583.

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

Bound-to-Bound Data Collaboration (B2BDC)

Scalar consistency measure

$\gamma \geq 0 :$ Consistent Dataset

$\gamma < 0 :$ Inconsistent Dataset

$$C_{\text{Dataset}} = \max_{\gamma, x \in \mathcal{H}} \quad \gamma$$

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

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

Feeley, Seiler, Packard, Frenklach. *J. Phy. Chem. A*, 2004, 108, 9573-9583.

Bound-to-Bound Data Collaboration (B2BDC)

Scalar consistency measure

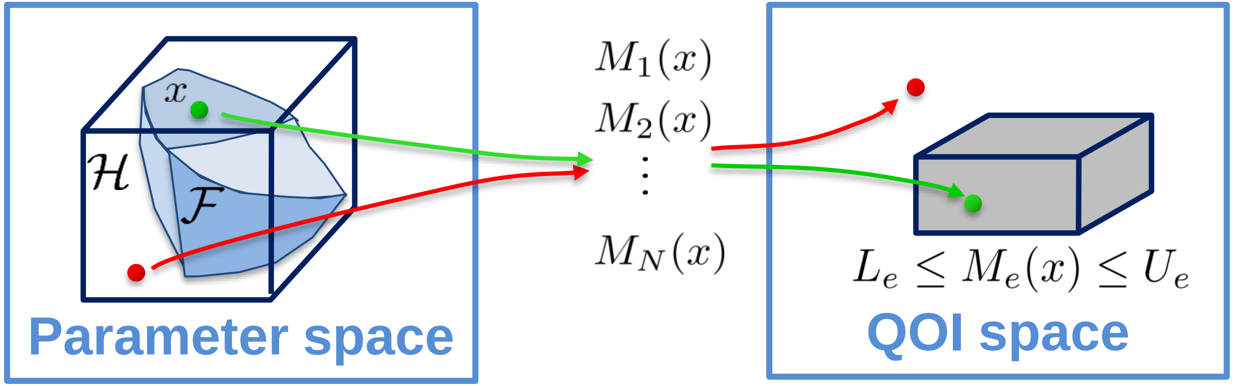
$\gamma \geq 0:$ Consistent Dataset

$\gamma < 0:$ Inconsistent Dataset

$$C_{\text{Dataset}} = \max_{\gamma, x \in \mathcal{H}} \quad \gamma$$

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

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



B2BDC Model Validation Workflow

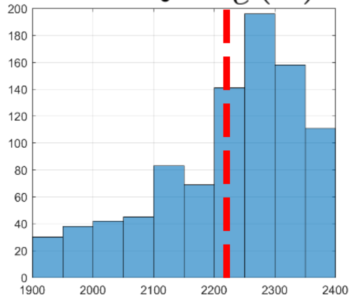
Uncertain Parameters

$x \in \mathcal{H}$

Scenario Parameters, $x_{s,e}$

Char Oxidation Model
(Instrument + Physics)

$\mathrm{QOI}_e(x)$



Particle Temperature

CCMSC Coal Database

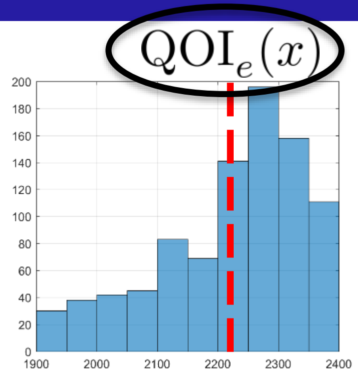
B2BDC Model Validation Workflow

Uncertain Parameters

$x \in \mathcal{H}$

Scenario Parameters, $x_{s,e}$

Char Oxidation Model
(Instrument + Physics)



Particle Temperature

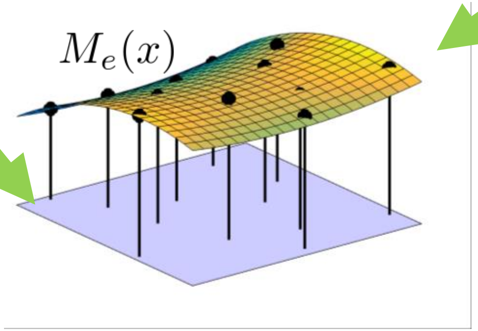
CCMSC Coal Database

B2BDC Model Validation Workflow

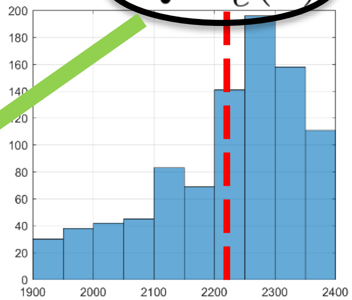
Uncertain Parameters $x \in \mathcal{H}$

Scenario Parameters, $x_{s,e}$

Char Oxidation Model (Instrument + Physics)



$\mathrm{QOI}_e(x)$



Particle Temperature

CCMSC Coal Database

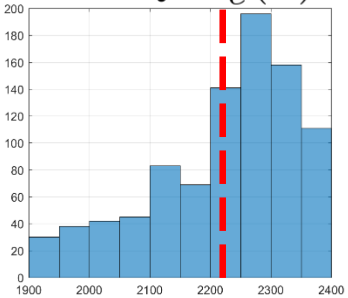
B2BDC Model Validation Workflow

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

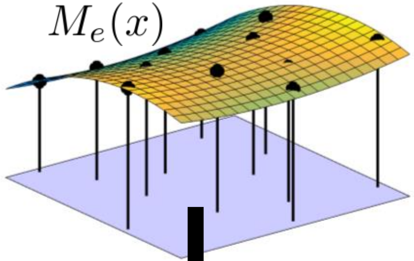
Scenario Parameters, $x_{s,e}$

Char Oxidation Model (Instrument + Physics)

$\mathrm{QOI}_e(x)$



Particle Temperature



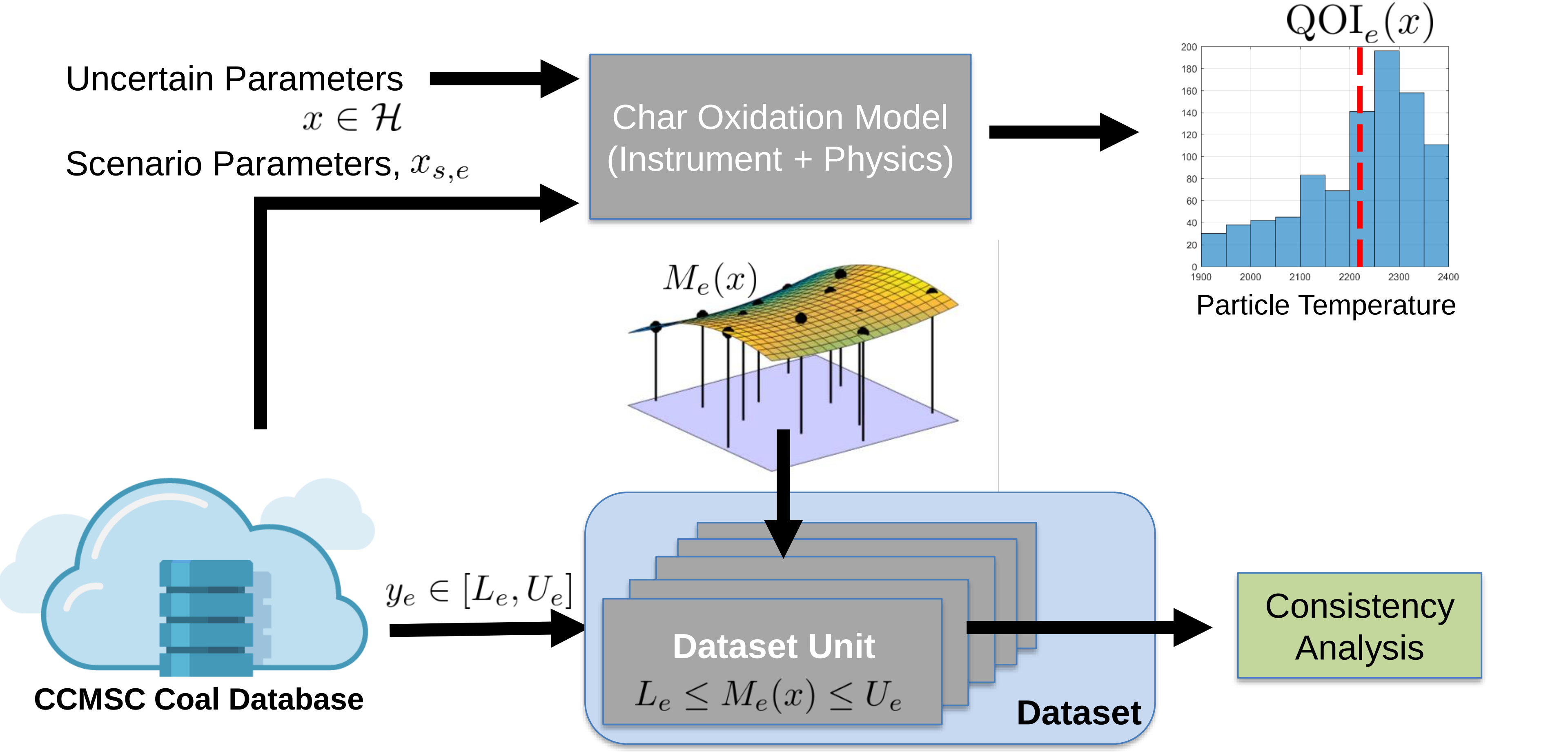
CCMSC Coal Database

$y_e \in [L_e, U_e]$

Dataset Unit

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

B2BDC Model Validation Workflow



Validation through consistency

Model Form

$$\left.\begin{aligned} C_{(s)} + O_2 &\longrightarrow CO_2 \\ 2\,C_{(s)} + O_2 &\longrightarrow 2\,CO \end{aligned}\right\} \quad r = \mathbf{A}\exp\left(\frac{-\mathbf{Ea}}{RT_p}\right)C_{\mathrm{oxid}}$$

Transport

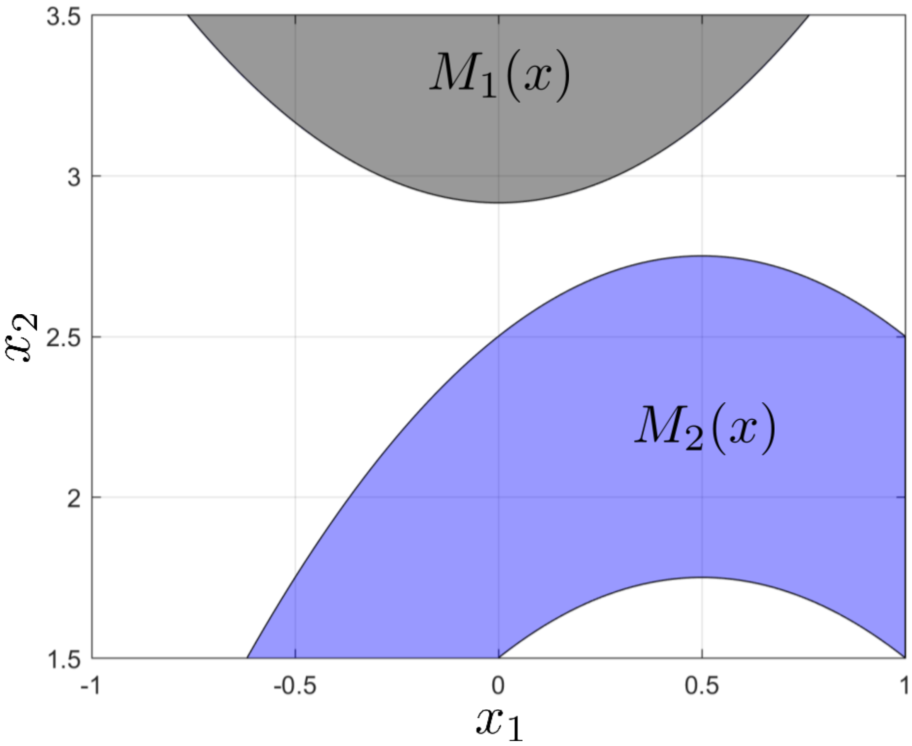
- Diffusion of oxidizer to particle surface
- Diffusion of products from particle surface

Scalar consistency measure:

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

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.



Iavarone et al. *Energy Procedia*, 120 (2017): 500-507.

Validation through consistency

Model Form

$$\left.\begin{array}{lll} C_{(s)} + O_2 & \longrightarrow & CO_2 \\ 2\,C_{(s)} + O_2 & \longrightarrow & 2\,CO \end{array}\right\} \quad r = \mathbf{A}\exp\left(\frac{-\mathbf{Ea}}{RT_p}\right) C_{\text{oxid}}$$

Transport

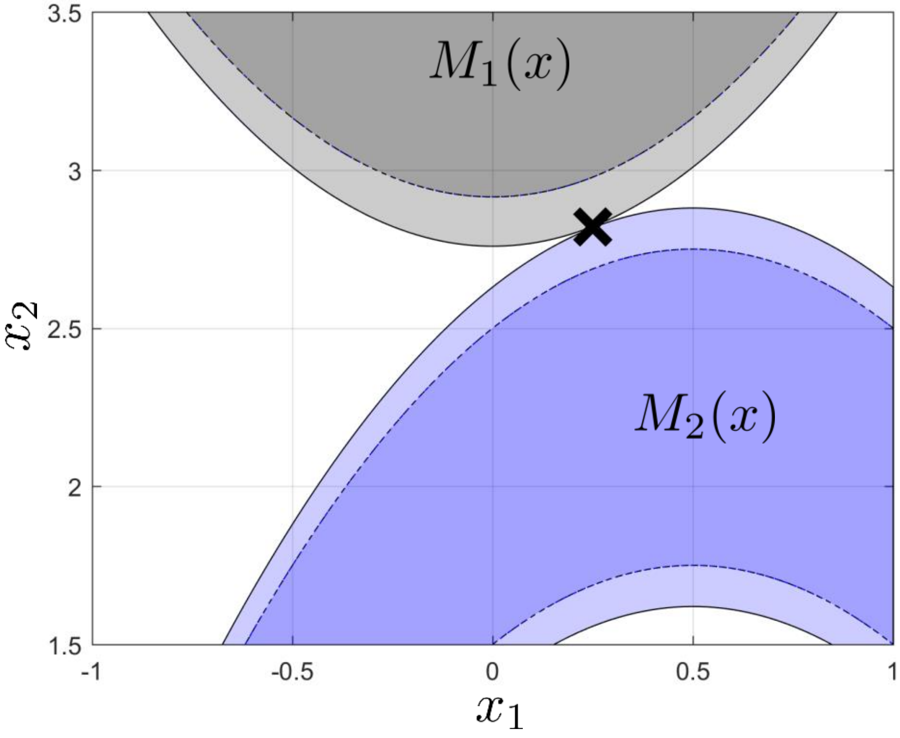
- Diffusion of oxidizer to particle surface
- Diffusion of products from particle surface

Scalar consistency measure:

$$\mathrm{C}_{\mathrm{Dataset}} = [-0.26, -0.19]$$

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

$$\left.\begin{array}{l} C_{(s)} + O_2 \longrightarrow CO_2 \\ 2\,C_{(s)} + O_2 \longrightarrow 2\,CO \end{array}\right\} \quad \{A_{O_2}, E_{O_2}\}$$

$$C_{(s)} + CO_2 \longrightarrow 2\,CO \quad \{A_{CO_2}, E_{CO_2}\}$$

$$C_{(s)} + H_2O \longrightarrow H_2 + CO \quad \{A_{H_2O}, E_{H_2O}\}$$

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

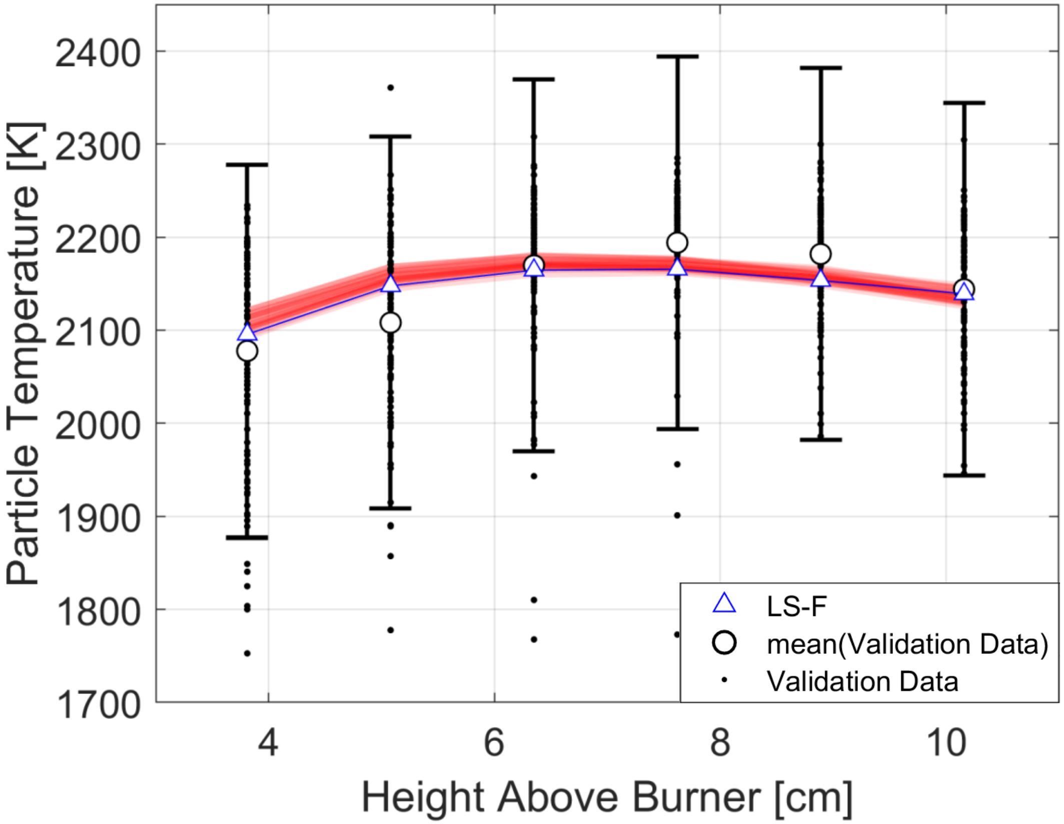
Scalar consistency measure:

$$C_{\text{Dataset}} = [0.06, 0.32]$$

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

Validation through consistency

$N_2 : 54\%, H_2O : 10\%, O_2 : 36\%, d_{\text{bin}} = 82.5\mu\text{m}$



$$x_{\text{LS-F}} = \underset{x \in \mathcal{F}}{\text{argmin}} \, ||M_e(x) - y_e||_2$$

y_e : 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.

National Nuclear Security Administration

CARBON CAPTURE MULTIDISCIPLINARY SIMULATION CENTER