Surrogate modeling and analysis toolset

Justin Weber, William Fullmer, Aytekin Gel
Research and Innovation Center (RIC)
• Application and framework for graphical programing through the use of nodes and connections
• Underlying library for the optimization/UQ work.
• Integrates with the MFiX GUI
Surrogate modeling and analysis toolset

Design of Experiments → Model evaluation → Response Surface Construction → Optimization
Design of Experiments | Variables

Add variables

Select variable parameters

Fuzzy search of Parameters and MFiX Keywords

Automatic population of Categorical Variables

Inside MFiX = MFiX Aware
Design of Experiments | Methods

Method
- Factorial
- Covary
- Montecarlo
- Latin hypercube
- Central composite
- Sobol
- Hammersly
- Halton

Options
- Import, Build, Export

Samples
- 2D plot of samples

Change variables
Response Surface | Samples

- Samples + Response
- Or-
- Read CSV
### Response Surface | Models

**Select models to fit**

- Gaussian process
- Polynomial
- Multilayer perceptron
- Support vector machine
- Decision tree
- Random forest
- Nearest
- Linear
- Cubic ($d \leq 2$)
- Radial basis function

**Model parameters**

- **Max terms**: 100
- **Max degree**: 1
- **Penalty**: 3.0

**Points to remove for cross validation**

- Cross validation points: 10

**Error metrics**

<table>
<thead>
<tr>
<th>fit</th>
<th>Model</th>
<th>MSE</th>
<th>$R^2$</th>
<th>$L_{\inf}$</th>
<th>$L_1$</th>
<th>$L_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>radial basis function</td>
<td>MARS</td>
<td>0.678</td>
<td>0.892</td>
<td>0.161</td>
<td>0.065</td>
<td>0.0813</td>
</tr>
<tr>
<td>cubic</td>
<td></td>
<td>0.956</td>
<td>0.848</td>
<td>0.209</td>
<td>0.087</td>
<td>0.0966</td>
</tr>
<tr>
<td>linear</td>
<td></td>
<td>1.07</td>
<td>0.83</td>
<td>0.643</td>
<td>0.0405</td>
<td>0.102</td>
</tr>
<tr>
<td>random forest</td>
<td></td>
<td>0.433</td>
<td>0.931</td>
<td>0.13</td>
<td>0.052</td>
<td>0.065</td>
</tr>
<tr>
<td>support vector machine</td>
<td></td>
<td>0.372</td>
<td>0.941</td>
<td>0.155</td>
<td>0.0482</td>
<td>0.0603</td>
</tr>
<tr>
<td>decision tree</td>
<td></td>
<td>0.433</td>
<td>0.931</td>
<td>0.13</td>
<td>0.052</td>
<td>0.065</td>
</tr>
<tr>
<td>nearest</td>
<td></td>
<td>0.468</td>
<td>0.926</td>
<td>0.131</td>
<td>0.0538</td>
<td>0.0676</td>
</tr>
<tr>
<td>multilayer perceptron</td>
<td></td>
<td>0.505</td>
<td>0.92</td>
<td>0.171</td>
<td>0.0594</td>
<td>0.0701</td>
</tr>
</tbody>
</table>

**Fit model**

- Output model: radial basis function
Response Surface | Models

1D model test

polynomial regressors

watch the edge!
Select model
Select plot

Save/manipulate plot
Select plot
Select model
Save/manipulate plot
### Optimization

<table>
<thead>
<tr>
<th>Operation</th>
<th>Attempts</th>
<th>Global Method</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimize</td>
<td>10</td>
<td>differential evolution</td>
<td>basin hopping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nelder-Mead</td>
<td>Powell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG</td>
<td>BFGS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-BFGS-B</td>
<td>TNC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COBYLA</td>
<td>SLSQP</td>
</tr>
</tbody>
</table>

Results of optimization attempts

**Optimal Point**

Schwefel function $f(420.99, 420.99) = 0$
Schwefel function $f(420.99, 420.99) = 0$
Sensitivity Analysis

<table>
<thead>
<tr>
<th>Options</th>
<th>Plot</th>
<th>Total</th>
<th>First Order</th>
<th>Second Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>sobol analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samples</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resamples</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-3.14</td>
</tr>
<tr>
<td>b</td>
<td>-3.14</td>
</tr>
<tr>
<td>c</td>
<td>-3.14</td>
</tr>
</tbody>
</table>

**SALib**

- Sobol
- Method of Morris
- Fourier Amplitude
- Delta Moment-independent
- Random balance
- Fourier Amplitude
Forward Propagation | Variables

Samples

Variables

Variable options
Forward Propagation | P-Box

- The probability that the value will be 2 or less is between 62.4 % and 76.5 %.
- Given the prescribed input uncertainties with 95 % probability, the quantity of interest will be between 4.32 and 5.61.

Calculate Propagation

Export bounds to file
Examples
We have an underperforming cyclone on 50 kWth Chemical Looping Reactor

- Increase efficiency
- Maintain or lower pressure drop
Base cyclone

Pressure outlet: 101.32 kPa
Gas + Solids

Solids loss

Gas
0.02 kg/s
Solids
0.08 kg/s

HDPE
Diameter: 871 µm
Density: 860 kg/m³

~1 particles/parcel
Design of experiments

<table>
<thead>
<tr>
<th>Variable</th>
<th>min (m)</th>
<th>max (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{\text{BARREL}}$</td>
<td>0.04</td>
<td>0.1</td>
</tr>
<tr>
<td>$R_{\text{vortex}}$</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>$H_{\text{vortex}}$</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>$H_{\text{inlet}}$</td>
<td>0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>$W_{\text{inlet}}$</td>
<td>0.015</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- genetically optimized Latin hypercube
- 100 samples (2x recommended)
- $L_2$-discrepancy measure of 0.00295
Model creation

Models created using Nodeworks and MFIX
Using Nodeworks and MFiX, Dispatch all models to the queue
Run the models!

• All models ran simultaneously
• Took 21 minutes to 7 hours per model
• Cell count varied from 40,320 to 169,764
• Three models failed (6%), due to bad mesh
Quantity of interest

\[ \text{Mass} + \Delta P = \text{QOI} \]
Surrogate model: Gaussian Process

Alpha: noise level or smoothing of the data

10% hold out CV
Sobol Sensitivity

Using SALib
Optimization

Using differential evolution
- 11 times lower pressure drop
- 2.3 times lower mass loss

<table>
<thead>
<tr>
<th>Variable</th>
<th>Original (m)</th>
<th>Optimal (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{barrel}}$</td>
<td>0.06</td>
<td>0.096</td>
</tr>
<tr>
<td>$r_{\text{vortex}}$</td>
<td>0.015</td>
<td>0.026</td>
</tr>
<tr>
<td>$h_{\text{vortex}}$</td>
<td>0.4</td>
<td>0.373</td>
</tr>
<tr>
<td>$h_{\text{inlet}}$</td>
<td>0.08</td>
<td>0.12</td>
</tr>
<tr>
<td>$w_{\text{inlet}}$</td>
<td>0.02</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Edge of design space
Putting it all together
Examples
Quantify mixing as the rate of decay of the **Alike Neighbor Fraction (ANF)**

\[
\text{ANF} = \text{fraction of particles within } 2.5r_p\text{-radius of a given particle with the same color (averaged over all particles)}
\]

thanks: Steven Dahl, Casey Q. LaMarche & Christine M. Hrenya
Problem definition

Model: MFiX-DEM
Rotation induced by angular gravity
Geometry considered fixed/known
Seven model parameters considered as unknown quantities
• Six of which are taken from measurements of real particles

Model uncertainties considered:

<table>
<thead>
<tr>
<th>DEM Model Parameter</th>
<th>Units</th>
<th>DOE Input Variable</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f ) (rpm)</td>
<td>( x_1 )</td>
<td>28.8</td>
<td>31.2</td>
<td></td>
</tr>
<tr>
<td>( d_p ) (cm)</td>
<td>( x_2 )</td>
<td>0.26</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>( \rho_p ) (g/cm(^3))</td>
<td>( x_3 )</td>
<td>2.22</td>
<td>2.92</td>
<td></td>
</tr>
<tr>
<td>( e_{pp} )</td>
<td>( x_4 )</td>
<td>0.92</td>
<td>0.9999</td>
<td></td>
</tr>
<tr>
<td>( e_{pw} )</td>
<td>( x_5 )</td>
<td>0.58</td>
<td>0.9999</td>
<td></td>
</tr>
<tr>
<td>( \mu_{pp} )</td>
<td>( x_6 )</td>
<td>0.1</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>( \mu_{pw} )</td>
<td>( x_7 )</td>
<td>0.02</td>
<td>0.42</td>
<td></td>
</tr>
</tbody>
</table>
DOE → Simulations → Surrogate

Design of Experiments
- Latin Hypercube
- Genetic optimization
- 7-D space
- 345 samples (overkill?)

Response Surface Model
Gaussian Process
- RBF kernel

Cross-validation Error

8/20/2019
Forward Propagation (of input uncertainties)

Hybrid/nested sampling approach of Roy & Obekampf
10 epistemic samples, each with 100 aleatory samples

Examples of surrogate model propagated p-boxes
Forward Propagation (of input uncertainties)

*What if...* we decide the p-box is too coarse for our use purpose and we need to increase the number of samples?

- Direct/full model: *expensive*
- Surrogate model: (once constructed) *cheap*
Examples
Example | Biomass Gasifier

Control Variables:
- \( x_1 = \) biomass mass flow rate
- \( x_2 = \) inlet gas mass flow rate
- \( x_3 = \) inlet gas steam mass fraction

System Response/QoI:
- \( y_1 = H_2/CO \) molar ratio of product syngas (time-averaged from 25 to 30s)

Objective Function:
- \( \min \frac{x_3 x_2}{x_1 | y_1 = 2} \), minimize the amount of steam required to produce a syngas with a 2:1 hydrogen to carbon monoxide ratio

thanks: Yupeng Xu, Mehrdad Shahnam & William Rogers
DOEs and Results

Sub-set of preliminary, scoping DOE
Secondary, refined DOE

Nodeworks
- Latin Hypercube
- Genetic Optimization
- Composite DOE not LH

Results for the QoI, H₂/CO

Region of interest
Q: How do we get a continuous surface of $y_1 = 2$
A: Construct a (4-D) response surface surrogate model and extract the (3-D) iso-surface characterizing $y_1 = 2$
Surrogate Modeling

Cross-Validation for the QoI, H₂/CO

- Best surrogate models:
  - Radial basis function (RBF) with smoothing parameter of 1e-14
  - Gaussian process (GP) with RBF kernel and noise parameter of 1e-14

Full Model Error

Selection: GP (more consistent)
Optimize: RSM == 2

Optimal condition

Iso-surface

\[
x_1 \quad x_2 \quad x_3 \quad y_1 \\
\#
\#
\#
\# 2.000.. \\
\#
\#
\#
\# 1.999..
\#
\#
\#
\# 1.999..
\#
\#
\#
\# 2.000..
\#
\#
\#
\# ...
\]

Another surrogate?
Nah, the GP is cheap.
Just iterate many times and interpolate.

Color: \( f = x_3 \cdot x_2 / x_1 \)
Validation of (surrogate) Optimum

\[ x_1 = 0.086 \text{ (g/s)}, \quad x_2 = 0.054 \text{ (g/s)}, \quad x_3 = 4.8 \times 10^{-4}, \quad \hat{y}_1 = 2, \quad y_1 = 2.2 \]

(within expected error from cross-validation test)