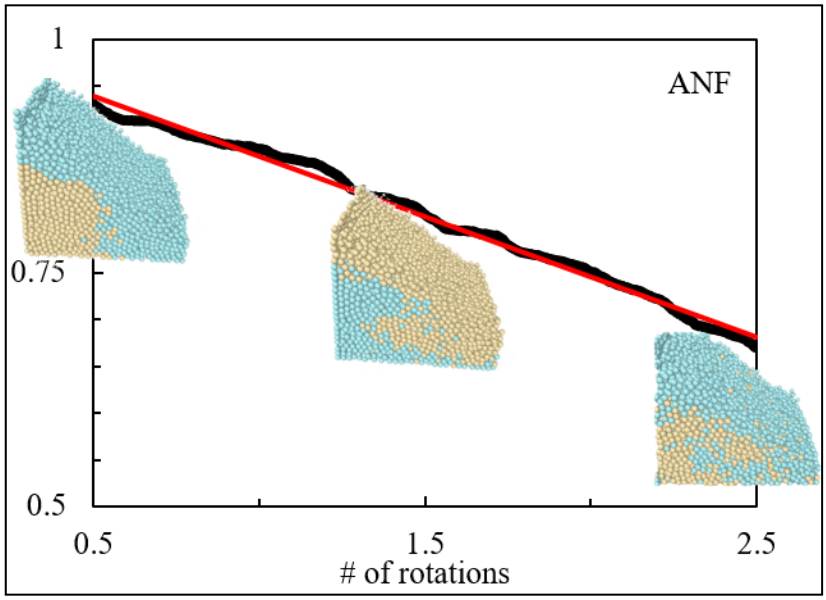
NATIONAL ENERGY TECHNOLOGY LABORATORY




# Surrogate Modeling Approach to Uncertainty Quantification for a DEM Model of a Rotating Cubic Tumbler

21 May 2019

U.S. DEPARTMENT OF ENERGY | NATIONAL ENERGY TECHNOLOGY LABORATORY

**Office of Fossil Energy**

NETL-TRS-5-2019

# Disclaimer

This work was funded by the Department of Energy, National Energy Technology Laboratory, an agency of the United States Government, through a support contract with Leidos Research Support Team (LRST). 

**Cover Illustration:** Decrease in the alike neighbor fraction (ANF) from 0.5 to 2.5 rotations of the baseline simulation, red line shows linear fit, insets show snapshots (at end of tumbler rotation axis) at approximately 0.5, 1.5, and 2.5 rotations.

**Suggested Citation:** Fullmer, W. D.; Dahl, S.; Weber, J. *Surrogate Modeling Approach to Uncertainty Quantification for a DEM Model of a Rotating Cubic Tumbler*; NETL-TRS-5-2019; NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2019, p 24. DOI: 10.18141/1514272.

**An electronic version of this report can be found at: https://netl.doe.gov/TRS**

# Surrogate Modeling Approach to Uncertainty Quantification for a DEM Model of a Rotating Cubic Tumbler

**William D. Fullmer[1], Steven Dahl[2], Justin Weber[3]**

**[1] U.S. Department of Energy, National Energy Technology Laboratory, Leidos Research Support Team, 3610 Collins Ferry Road, Morgantown, WV 26507**

**[2] Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80309**

**[3] U.S. Department of Energy, National Energy Technology Laboratory, 3610 Collins Ferry Road, Morgantown, WV 26507**

---

**NETL-TRS-5-2019**

21 May 2019

NETL Contacts:

William A. Rogers, Principal Investigator

Jonathan Lekse, Technical Portfolio Lead

Bryan Morreale, Executive Director, Research & Innovation Center

This page intentionally left blank

# Table of Contents

# List of Figures



# List of Tables

# Acronyms, Abbreviations, and Symbols

| Term | Description |
|---|---|
| ANF | Alike neighbor fraction |
| CDF | Cumulative distribution function |
| CFD | Computational fluid dynamics |
| DEM | Discrete element method |
| DOE | Design of experiments |
| GP | Gaussian process |
| KDE | Kernel density estimator |
| LH | Latin Hypercube |
| LSD | Linear-spring dashpot |
| MFiX | Multiphase Flow with Interphase eXchanges |
| NETL | National Energy Technology Laboratory |
| RSM | Response surface model |
| SRQoI | System response quantity of interest |
| UQ | Uncertainty quantification |
| VV | Verification and validation |
| VVUQ | Verification, validation and uncertainty quantification |

# Acknowledgments

This work was performed in support of the U.S. Department of Energy's Fossil Energy Crosscutting Technology Research Program. The research was executed through the National Energy Technology Laboratory's (NETL) Research and Innovation Center's Advanced Reaction Systems Field Work Proposal. Research performed by Leidos Research Support Team staff was conducted under the RSS contract 89243318CFE000003.

The authors would like to thank Casey Q. LaMarche for taking the particle input uncertainty data, Peiyuan Liu for providing the modified Multiphase Flow with Interphase eXchanges (MFiX) source code files necessary to simulate the tumbler, and Christine M. Hrenya for initiating the research in Phase-II of the MFiX-DEM Enhancement for Industrially-Relevant Flows project.

## EXECUTIVE SUMMARY

This report studies the forward propagation of input uncertainty on a discrete element method (DEM) model of a square, granular rotating tumbler. The National Energy Technology Laboratory's (NETL) Multiphase Flow with Interphase eXchanges (MFiX) computational fluid dynamics (CFD) code is used for numerical solution. Comparisons are made to the original work of Dahl et al. (2019) who used direct propagation, i.e., the full MFiX-DEM model itself is used directly for forward propagation. Here, the forward propagation is performed on a surrogate model which was constructed with another open source NETL software, the Nodeworks graphical programming library.

In this report, Nodeworks' Surrogate Modeling and Analysis nodes were used to:

- Devise and assess a seven-dimensional (7D) input parameter, space-filling optimal Latin Hypercube (LH) Design of Experiments (DOE)
- Create and optimize a Gaussian process (GP) surrogate model by minimizing cross-validation error
- Compute probability boxes (p-box) describing the uncertainty in the system response quantity of interest (SRQoI) for the prescribed model input uncertainties
- Performed a global sensitivity analysis by reporting first and second order Sobol' indices

The primary conclusions drawn from this work include:

- P-boxes computed from the surrogate model are in good qualitative agreement with the p-box evaluated directly from the full model by Dahl et al. (2019). (Note that quantitative comparison is complicated due to the stochastic nature of the sampling.)
- It is simple and computationally efficient to increase the number of samples and observe statistical convergence of the p-box.
- A global sensitivity analysis of the surrogate model by first- and second-order Sobol' indices computed from 100,000 samples is in good agreement (order of importance) with the local sensitivity analysis originally performed on the full model (Dahl et al., 2019).

# 1. INTRODUCTION

Uncertainty quantification (UQ), part of the broader field of verification and validation (V&V or VVUQ) in computational science and engineering, focuses on determining the accuracy of numerical predictions (Roache, 1998; Oberkampf and Roy, 2008). While VVUQ practices have been adopted in many single-phase flow computational fluid dynamics (CFD) communities, e.g., aerospace and automotive industries, they have been lagging in multiphase flow CFD communities, namely in particle technology and fluidization. However, some foundational VVUQ work with multiphase gas-solids particulate flow models have been spearheaded by the National Energy Technology Laboratory (NETL) (Gel et al., 2013a, 2013b; Gel et al., 2014; Choudhary et al., 2016; Gel et al., 2016; Shahnam et al., 2016; Gel et al., 2017; Syamlal et al., 2017; Bakshi et al., 2018; Fullmer and Musser, 2018; Gel et al., 2018).

The lag in adopting VVUQ practices is largely due to the increased complexity of multiphase flows compared to traditional single-phase flows. The increased complexity of multiphase flow phenomena requires more complex models for their prediction leading to an increased number of uncertain model parameters. These models are often more computationally expensive, which places a greater restriction on the number of simulations which may be carried out. A recent work by Dahl et al. (2019) studied methods to reduce the computational overhead associated with forward propagation of input uncertainties in a discrete element method (DEM) model of a rotating tumbler. Specifically, Dahl et al. (2019) considered: 1) full model evaluation requiring 1,000 simulations; 2) using a local sensitivity analysis to determine and eliminate unimportant input parameters; and 3) using a local sensitivity analysis to estimate the two most extreme cases.

Another approach to reducing computational overhead associated with VVUQ and optimization studies is by the design and construction of a data fitted surrogate, i.e., a response surface model (RSM). The graphical programming library Nodeworks, recently developed at and released by NETL, contains a specialized set of tools, called nodes, to reduce the entry barrier to surrogate modeling and analysis. The toolset was specifically designed for (although not limited to) simulations using Multiphase Flow with Interphase eXchanges (MFiX), NETL's open source, multiphase computational fluid dynamics (CFD) code. This work uses the surrogate modeling and analysis nodes within Nodeworks to: 1) create a statistical design of experiments (DOE); 2) construct a data fitted surrogate model; 3) perform a global sensitivity analysis; and 4) forward propagate input uncertainties to output uncertainty for the square tumbler case of Dahl et al. (2019).

The remainder of this work is outlined as follows. In Section 2, the MFiX-DEM numerical method is briefly overviewed. A baseline model without uncertainties is defined and simulated in Section 3 and analyzed to identify a suitable quantity of interest. The model is then simulated at many conditions specified by a statistical DOE which is used to construct a CFD-data-fitted surrogate model in Section 4. The surrogate model is then used for input uncertainty forward propagation and a global sensitivity analysis in Sections 5 and 6, respectively. The report closes with an overview of the principal conclusions in Section 7.

## 2. MODEL

This work utilizes a pure DEM model of granular flow in which the interstitial gas-phase is neglected. The position, velocity and angular momentum of each particle is calculated via Newton's laws of motion:

$$\frac{d\boldsymbol{X}_i}{dt} = \boldsymbol{V}_i \,, \tag{1}$$

$$m_i \frac{d\boldsymbol{V}_i}{dt} = m\boldsymbol{g} + \sum_{j=1}^{N_i^{(c)}} \boldsymbol{F}_{ji} \,, \tag{2}$$

and

$$I_i \frac{d\boldsymbol{\omega}_i}{dt} = \sum_{j=1}^{N_i^{(c)}} \boldsymbol{T}_{ji} \,, \tag{3}$$

where $\boldsymbol{X}_i$, $\boldsymbol{V}_i$, and $\boldsymbol{\omega}_i$ are the position, translational velocity, and angular velocity of the $i^{\text{th}}$ particle, respectively. All cases considered in this work are monodisperse so that the particle diameter, $d_p$, and density, $\rho_p$, are constants in a given simulation, though the values will change from simulation to simulation in the UQ study. Consequently, the particle mass, $m_i = m = \pi \rho_p (d_p)^3/6$, and moment of inertia $I_i = I = m(d_p)^2/10$, are also constants. Gravity, $\boldsymbol{g}$, is the only body force acting on the system. The gravity vector is defined as,

$$g_x = -|\boldsymbol{g}| \sin\frac{2\pi t'}{T}, \quad g_y = -|\boldsymbol{g}| \cos\frac{2\pi t'}{T}, \quad g_z = 0 \,, \tag{4}$$

where $|\boldsymbol{g}| = 980.665$ cm/s$^2$, $t' = \max(t - t_{settle}, 0)$, and $T = 1/f$ is the period of rotation. The sinusoidal body force mimics the granular dynamics of a rotating tumbler without requiring moving boundaries (which is not fully supported in MFiX at this time). However, it should be noted that this setup is not entirely physically consistent with an actual (moving wall) tumbler. The simulation time $t'$ is delayed by a constant offset which allows a prescribed uniform, random particle configuration to settle into random packed state at the bottom of the tumbler before rotation begins. The initial settling delay is set at $t_{settle} = 1$ s for all simulations.

The two remaining variables yet to be specified in Equations (1)–(3) are the forces, $\boldsymbol{F}_{ji}$, and torques, $\boldsymbol{T}_{ji}$, acting on the $i^{\text{th}}$ particle due to the $j^{\text{th}}$ neighboring particle or wall in contact with the $i^{\text{th}}$ particle, where $N_i^{(c)}$ is the number of particles and walls contacting the $i^{\text{th}}$ particle which may be zero. MFiX-DEM uses the so-called "soft-sphere" approach to DEM (Pöschel and Schwager, 2005), where collisions are resolved in time which cause particles to "deform" (mathematically), imparting forces and torques on one another. The linear-spring dashpot (LSD) model of Cundall and Strack (1979) is applied here. Details of the LSD model as implemented in MFiX can be found in Garg et al. (2012a, 2012b).

## 3. BASELINE CONDITION

The DEM model described previously is now used to simulate the self-diffusion of a monodisperse granular "mixture" in a rotating, square tumbler. A few system parameters are assumed to be known exactly. Namely, the size of tumbler, $L_x = L_y = L_z = L = 8.942$ cm, and the mass of the particles in the tumbler, $M = 518.1$ g, are assumed to have no associated uncertainties. Additionally, the normal spring stiffness, $k_n = 10^6$ dyne/cm, an LSD model parameter, is assumed to be sufficiently large that the results of all simulations are insensitive to further increases in its value.

The remaining model parameters are all considered uncertain input variables: $f$ (or equivalently $T$), $d_p$, $\rho_p$, the particle-particle restitution coefficient $e_{pp}$, the particle-wall restitution coefficient, $e_{pw}$, the particle-particle friction coefficient $\mu_{pp}$, and the particle-wall friction coefficient $\mu_{pw}$. The distribution ranges and distributions of these input parameters will be discussed further in Sections 4 and 5 , respectively. Here, the base values (typically the median or mean) for the seven uncertain parameters are applied which make up the baseline condition: $f = 30$ rpm (= 0.5 Hz), $d_p = 0.310134$ cm, $\rho_p = 2.513$ g/cm$^3$, $e_{pp} = 0.969565$, $e_{pw}$, = 0.954934, $\mu_{pp} = 0.273370$, and $\mu_{pw} = 0.251369$. It should be noted that this exercise is a hypothetical problem—there is no (known) existing tumbler in this configuration with associated experimental data. The particle properties, however, are real and the characterization of all uncertainties, except $f$, were measured in the laboratory (Dahl et al., 2019).

At baseline conditions, the number of particles in the system are $N_p = M/m = 13{,}200$ and the solids concentration of the system is approximately $\phi = M/\rho_p L^3 \approx 0.29$. The solids concentration is slightly too high to generate the initial particle configuration in a random array. Therefore, the particles are initialized in a uniformly spaced primitive cubic lattice with a slight gap between each particle and a small, thermal velocity. The particles reorganize slightly from the initial thermal speed and collapse into a static, packed bed during the 1 s settling time before rotation begins.

The system response quantity of interest (SRQoI) in this case is the rate of self-diffusion or mixing of the particles amongst themselves. There are several ways in which the self-diffusion coefficient can be calculated (Garzó and Montanero, 2004). Here, the straightforward approach of Dahl et al. (2019) is followed in which the first half of the seeded particles are "colored" by one value (e.g. 1) and the second half are colored by another value (e.g. 2) as shown graphically in Figure 1. Then, the number of nearby particles $N_i^{(\text{nearby})}$, defined as particles with centroids separated by no more than 2.5 radii, surrounding the $i^{\text{th}}$ particle are isolated. If $N_i^{(\text{nearby})} \geq 4$, the fraction of $N_i^{(\text{nearby})}$ which have the same color as the $i^{\text{th}}$ particle itself is taken as the alike neighbor fraction (ANF) for the $i^{\text{th}}$ particle. This procedure is repeated for all particles with at least four nearby neighbors and averaged into a global ANF.

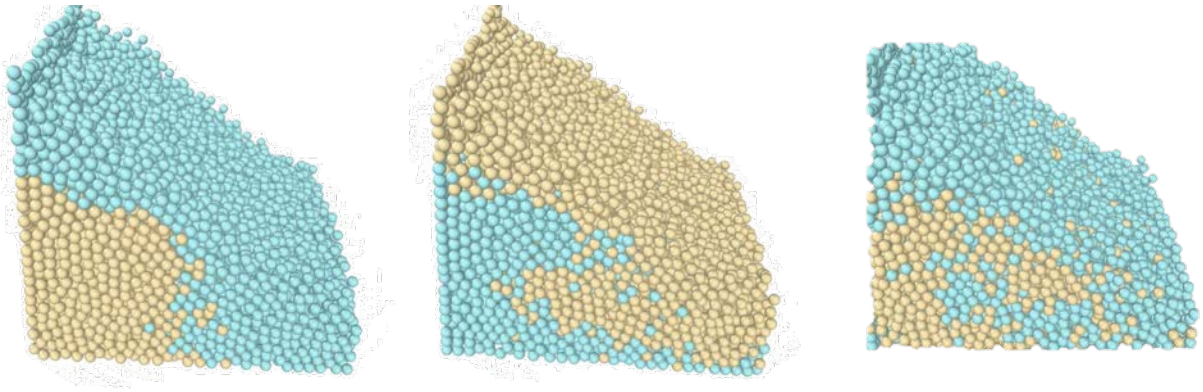


**Figure 1: State of the baseline tumbler after (approximately) 1/2 (left), 3/2 (center) and 5/2 (right) rotations.**

Initially the ANF is near 1 (not 1 exactly due to the interface), meaning that almost all particles are surrounded by particles of the same color. After the settling period ($t_{settle}$), rotation begins and the particles begin mixing. As the particles mix, the ANF decreases in value, eventually approaching the fully mixed state where ANF ~ 0.5. The self-diffusion process is shown qualitatively in Figure 1. The mixing rate can be quantified by calculating the slope of the ANF. Specifically, between approximately one half and two and a half "turns" of the tumbler, the ANF decays linearly on average which can be readily fit with a first-order regression, as shown in Figure 2. The magnitude of slope of this linear fit is taken as the mixing rate, which, for a single realization of the baseline condition, is approximately 0.129. This value is in good agreement with the original study (Dahl et al., 2019) which considered three replicates of the base case (three different stochastic realizations of the same initial configuration) finding baseline mixing rates of 0.129, 0.129, and 0.128.

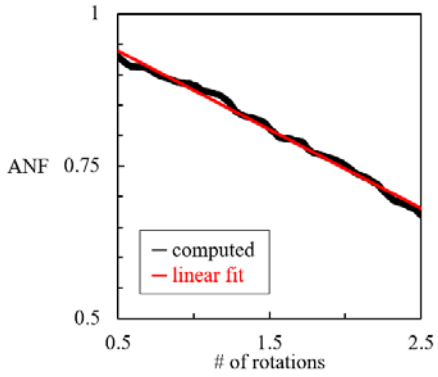



**Figure 2: Decay of the ANF from 1/2 to 5/2 "turns" (i.e. rotations of the gravity vector) with the linear regression superimposed.**

# 4. SURROGATE MODELING

## 4.1 DESIGN OF EXPERIMENTS

The first step in constructing a data-fitted surrogate model is to collect the data, in this case by running DEM simulations, at carefully selected sample locations within the parameter space. This sampling is often referred to as a (statistical) design of experiments (DOE). Here, the DOE state-space is defined by the seven uncertain input parameters: $x_1 = f$, $x_2 = d_p$, $x_3 = \rho_p$, $x_4 = e_{pp}$, $x_5 = e_{pw}$, $x_6 = \mu_{pp}$, and $x_7 = \mu_{pw}$. The max and min of the seven parameters used in the DOE is provided in Table 1. The input parameter ranges are selected to extend over the smallest and largest measured value of each parameter by 10% on each side so that when the surrogate model is constructed, the model accurately captures the "edges" of the measured values. Two exceptions to the extended DOE coverage are the restitution coefficients; both upper limits are set at 0.9999 so as to not exceed the physically realistic upper limit of one.

**Table 1: Input Parameter Ranges for DOE**

| DEM Model Parameter | Units | DOE Input Variable | Min | Max |
|---|---|---|---|---|
| $f$ | (rpm) | $x_1$ | 28.8 | 31.2 |
| $d_p$ | (cm) | $x_2$ | 0.26 | 0.35 |
| $\rho_p$ | (g/cm$^3$) | $x_3$ | 2.22 | 2.92 |
| $e_{pp}$ | - | $x_4$ | 0.92 | 0.9999 |
| $e_{pw}$ | - | $x_5$ | 0.58 | 0.9999 |
| $\mu_{pp}$ | - | $x_6$ | 0.1 | 0.45 |
| $\mu_{pw}$ | - | $x_7$ | 0.02 | 0.42 |

Nodeworks' Design of Experiments node from the Surrogate Modeling and Analysis node collection is used to create the DOE. The selected sampling method is Latin Hypercube (LH). An initially randomized LH sample is optimized to maximize the design's space filling property by swapping variable values between samples. The seed for the initial LH is 6212624. The DOE node's genetic algorithm is used for optimization with 100 iterations. Following a recommendation of Oberkampf and Roy (2010) for the number of required samples for LH designs: $7^3+2 = 345$ samples are generated. The resulting lower triangular scatter plot matrix of the genetically optimized LH samples shown in Figure 3.

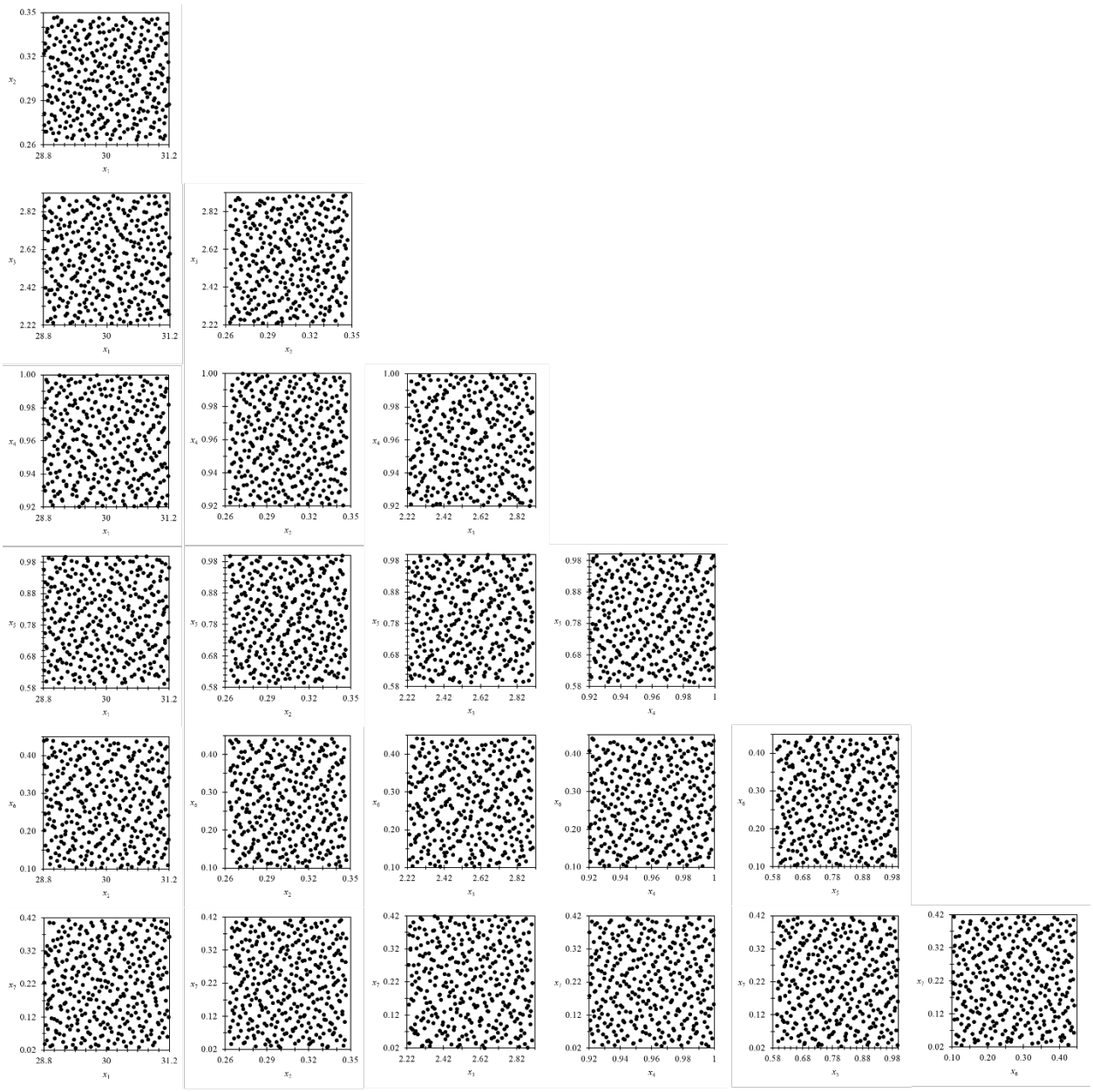


**Figure 3: Lower triangular scatter plot matrix of the input parameter DOE.**

After construction of the DOE, a simulation campaign is carried out. The DEM model input variable information contained in the 345 samples is transformed into 345 distinct simulations. It is worth noting that since the uncertain input parameters are related to the particles themselves (and the particle initial state is provided as an input, namely through a `particle_input.dat` file), the sample information is propagated to several streams. First, the particle diameter, density and number of particles (computed for each sample) are entered into a small `Pgen.in` file. Next, a particle generator, here a small standalone FORTRAN code, is executed which takes `Pgen.in` as input and creates a sample-specific `particle_input.dat` file, which is the MFiX code input file describing the initial particle configuration. Secondly, all seven values are entered in a

sample-specific `mfix.dat` file. Then, the 345 MFiX simulations are launched. Each simulation is carried out for $t'$ = 7 s, i.e., 6 s of tumbler rotation time, sufficient to extend 2.5 rotations for all conditions. All simulations were run on NETL's Joule2 supercomputer in serial on the shared queue. Wall clock times took between 1.7 and 11.2 hours. Once all simulations are completed, the resulting output data files are post-processed using the same[1] MATLAB script as in the original study by Dahl et al. (2019). Only one SRQoI is considered: $y_1$ = mixing rate, the magnitude of the slope in the ANF decay calculated from a linear regression between 0.5 to 2.5 rotations, as shown in Figure 2. The SRQoI is computed for all 345 simulations which is agglomerated in Figure 4. When isolated to the effects of single inputs on the SRQoI, only x3 in Figure 4 appears to show a strong dependence. A sensitivity analysis may be able to pare down the seven input parameters (see Section 6), however, since all seven have been included in the original DOE, there is little increased computational effort in including all parameters in the surrogate model, constructed next in Section 4.2.

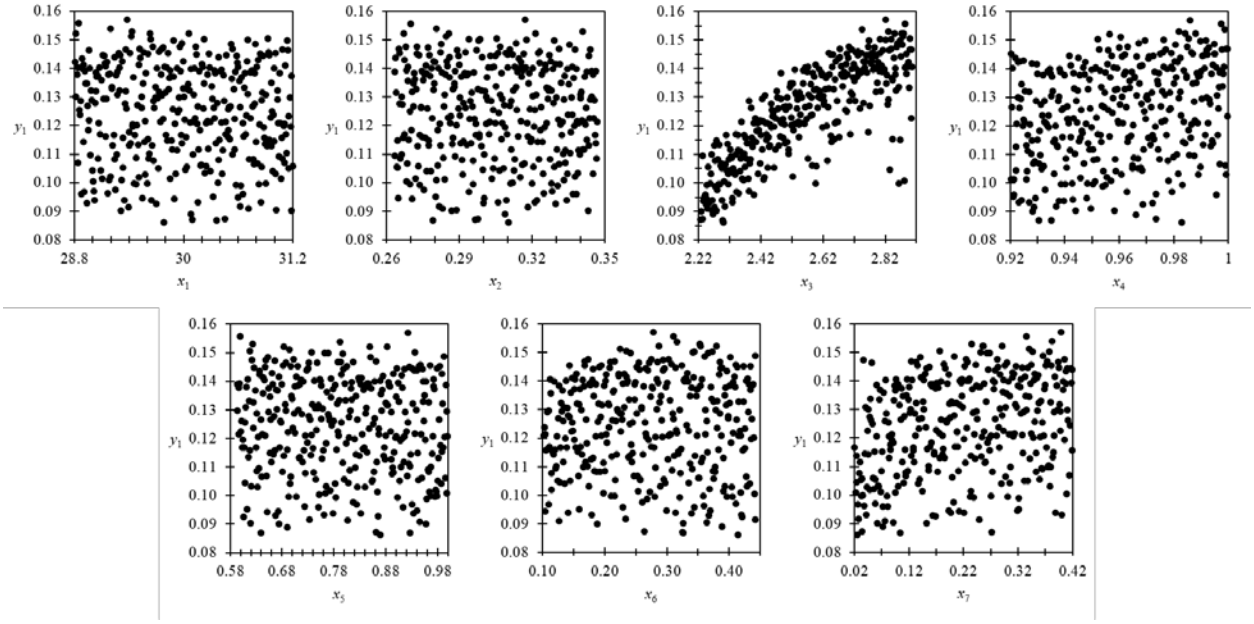



**Figure 4. Scatter plot of the SRQoI, $y_1$, versus the seven inputs $x_1 - x_5$.**

[1] To minimize post-processing computational time, the original script was streamlined by removing all calculations not related to the ANF and restricting calculation of the ANF to the time period between 0.5 and 2.5 rotations used for extrapolation.

## 4.2 CROSS-VALIDATION

The DOE input, $x_1 - x_5$, and processed response, $y_1$, are imported into the Nodeworks' Response Surface Model (RSM) node for construction of a DEM-data-fitted surrogate model. Although there are several surrogate modeling options available in the RSM node, focus is limited here to the Gaussian Process (GP) model with the `1*RBF` kernel, made available via the scikit-learn toolkit (Scikit-learn, 2019; Pedregosa et al., 2011). This GP model has been found to be reliable and robust in recent surrogate modeling studies of CFD-data also using Nodeworks (Xu et al., 2019; Weber et al., 2019). The GP automatically fits a variety of hyper-parameters by using the `fmin_l_bfgs_b` optimizer with ten restarts. The only remaining tuning variable is `alpha`, a noise parameter added to the diagonal of the kernel matrix. The value of `alpha` essentially determines how tight the surrogate fits to the data.

In this tumbler case, some sensitivity in the SRQoI, $y_1$, to the initial particle randomization is known to exist (Dahl et al., 2019). There are several ways to deal with this dependence, here the simplest option is taken: it is assumed that the response contains inherent noise because each $y_1$ only represents a single sample from a distribution of possible results (resulting from the stochastic initial condition). In other words, the values of $y_1$ themselves contain some unknown amount of uncertainty which is assumed to be properly accounted for when the GP noise level, `alpha`, minimizes the error of the surrogate. The error in the surrogate model is assessed by cross-validation, i.e., some data is held out from the fitting and then used to assess the discrepancy in the surrogate at the withheld locations. Otherwise, the "error" is minimized when `alpha` = 0 and the regression becomes a pure interpolator. However, such an error would not assess the accuracy of the model when interrogated at points which are not exactly one of the 345 DOE sample points.

A cross-validation study to determine the optimal `alpha` is conducted with a 20% holdout. The samples being withheld are randomized. Therefore, each test is repeated 20 times to compute 95% confidence intervals. Figure 5 shows the change in the cross-validation error with `alpha`. The error in Figure 5 is assessed as an $L_2$-norm,

$$L_2\text{-norm} = \sqrt{\frac{\sum_{i=1}^{N}[f(\boldsymbol{x}_i) - y_i]^2}{\sum_{i=1}^{N} y_i^2}}, \tag{5}$$

where $f(\boldsymbol{x}_i)$ corresponds to the GP model evaluation at the input variable vector corresponding to output $y_i$, and N is the subset corresponding to the 20% holdout points (of the full 345 samples). At low values of `alpha`, the GP model asymptotes to the noise level observed of a pure interpolator when the GP guarantees that $f(\boldsymbol{x}_i) = y_i$. However, exact evaluation only holds for the 80% of the samples used to construct the surrogate. The 20% holdout points experience an ($L_2$-norm) error of approximately 10%. Over a relatively narrow range, `alpha` $\approx 10^{-6}$ to $10^{-2}$, the error is reduced when the surrogate model noise (roughly) matches the noise inherent in the $y_1$. The optimal value for `alpha` is approximately $10^{-4}$, which is selected for the remainder for this study. Above `alpha` $\approx 10^{-2}$, the GP quickly becomes overwhelmed by noise and saturates to zero causing the error to approach unity. It is briefly noted that the optimal value of `alpha` $\approx 10^{-4}$ is somewhat larger than found in similar, previous studies (Xu et al., 2019; Weber et al., 2019), which is perhaps a consequence of the non-negligible impact of the stochastic initial condition on the SRQoI.

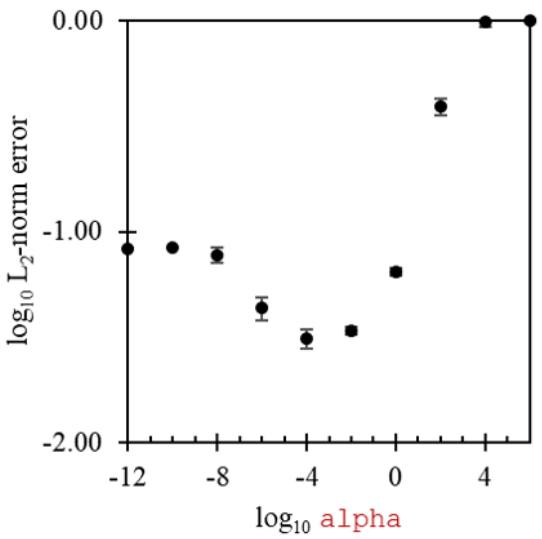



**Figure 5: Cross-validation error analysis of the GP surrogate model as a function of the noise parameter.**

## 4.3 FINAL MODEL

With the determined optimal value of `alpha` ≈ $10^{-4}$, cross-validation is deactivated, i.e., 0% holdout and the GP model is refit to all 345 points. The parity and error plots of the final GP surrogate model are provided in Figure 6. The mean square error of the fit is $5.6\times10^{-6}$. The R-squared value of the regression is 0.98. The $L_1$, $L_2$, and $L_\infty$ error norms are 1.4%, 1.9%, and 5.8%, respectively. The GP model is visualized in a reduced three-dimensional (3D) space in Figure 7. Note the remaining parameters, $x_1$, $x_2$, $x_4 - x_6$, are fixed at their mean values in Figure 7.

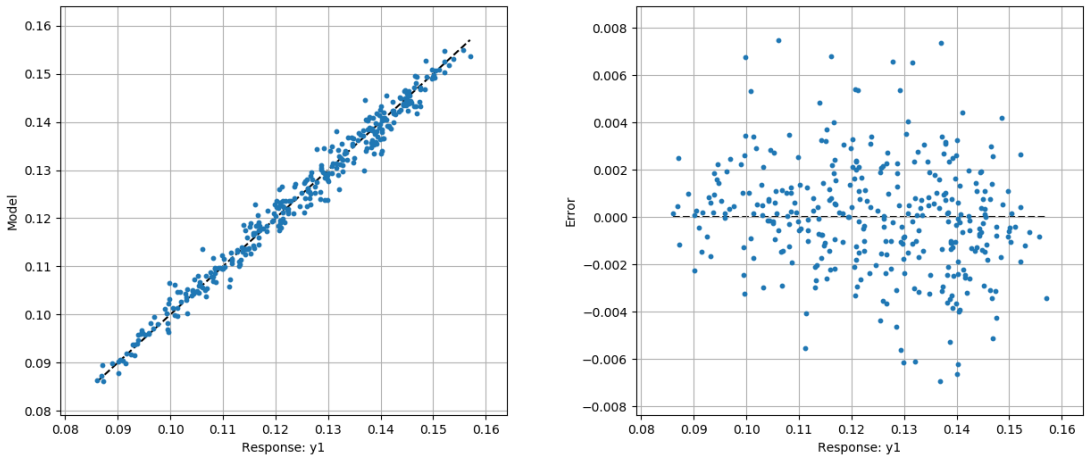



**Figure 6: Parity (left) and error (right) plots of GP surrogate model compared to full model (MFiX-DEM) evaluation without holdout.**

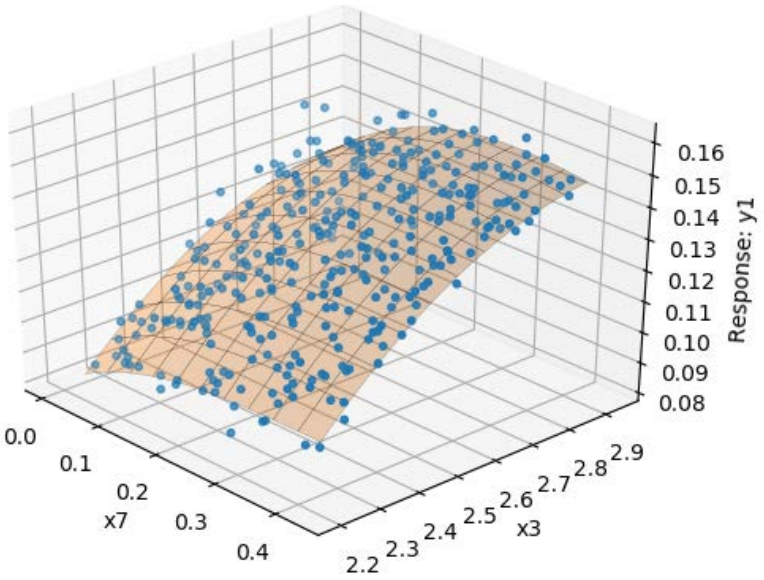



**Figure 7: Selected GP surrogate model visualized in $x_3$-$x_7$ space.**

## 5. FORWARD PROPAGATION

Having constructed a well-fit surrogate, the primary UQ objective of this study can be carried out: quantifying how model input uncertainty translates to uncertainty in the model output, here the $y_1$ SRQoI specifically. In the VVUQ framework, this study is often called *forward propagation*. Like Dahl et al. (2019), the comprehensive VVUQ framework of Roy and Oberkampf (2011) is followed. The first (UQ) step is identifying and characterizing all input uncertainties—with specific emphasis on differentiating aleatory uncertainties (characterized by a likelihood distribution) from epistemic uncertainties (characterized by a uniform interval).

While the uncertainties were identified prior to DOE sampling, they were only characterized based on the range of their measured values. Here, construction of probability distributions for aleatory uncertainties are necessary for sampling. Five of the seven input parameters with experimental measurements are the particle diameter and restitution and friction coefficients, $x_2$, $x_4 - x_7$. The measured values are used with MATLAB's kernel density estimator (KDE). The input space of each distribution is taken to be the same width as that used in the DOE, i.e., extending the measured data by 10% except for physical limits of $0 \leq x_4\text{-}x_7 \leq 1$. While this results in some non-negligible probabilities at the boundaries of the distributions, it ensures that the surrogate model will be sampled within its support region. Each range is then sampled by 1,000 equally spaced points for KDE. The resulting estimated distributions are provided in Figure 8 compared to the underlying data (circles) used for KDE.

The two remaining input parameters, $x_1$ and $x_3$, are considered epistemic. Because this is not a physical experiment, the rotation rate is not a measurable value and is taken to have an interval of 28–31 rpm. The particle density is a measurable quantity, however, taking single particle density measurements is challenging. Hence, batch (many particle) measurements for particle density are often reported, such is the case here. Therefore, $x_7$ is taken as an epistemic uncertainty with a range of 2.2835–2.85 g/cm$^3$.

As with the preceding steps, the forward propagation analysis is carried out in the Forward Propagation node of Nodeworks' Surrogate Modeling and Analysis node-set. The epistemic uncertainty ranges are input directly into the node and the KDE distributions of Figure 8 are imported as user-specified aleatory distributions. Nodeworks' Forward Propagation node follows the framework of Roy and Oberkampf (2011), where the sampling is separated for combined epistemic-aleatory uncertainty. First, one of the $N_e$ epistemic samples is drawn. The epistemic values are fixed, and $N_a$ aleatory samples are drawn. This produces a single cumulative distribution function (CDF) of the response, $y_1$. The procedure is then repeated for all $N_e$ samples, each time re-drawing all $N_a$ samples. The output is a collection of $N_e$-CDFs, each with $N_a$ steps. Because each of the epistemic samples are equally likely, no probable distinction can be drawn between them. Instead, the minimum and maximum values at given probability define the probability box, or p-box, of the uncertain response. This can be thought of as the bounding CDFs, however, the p-box may be defined by more than just two CDFs.

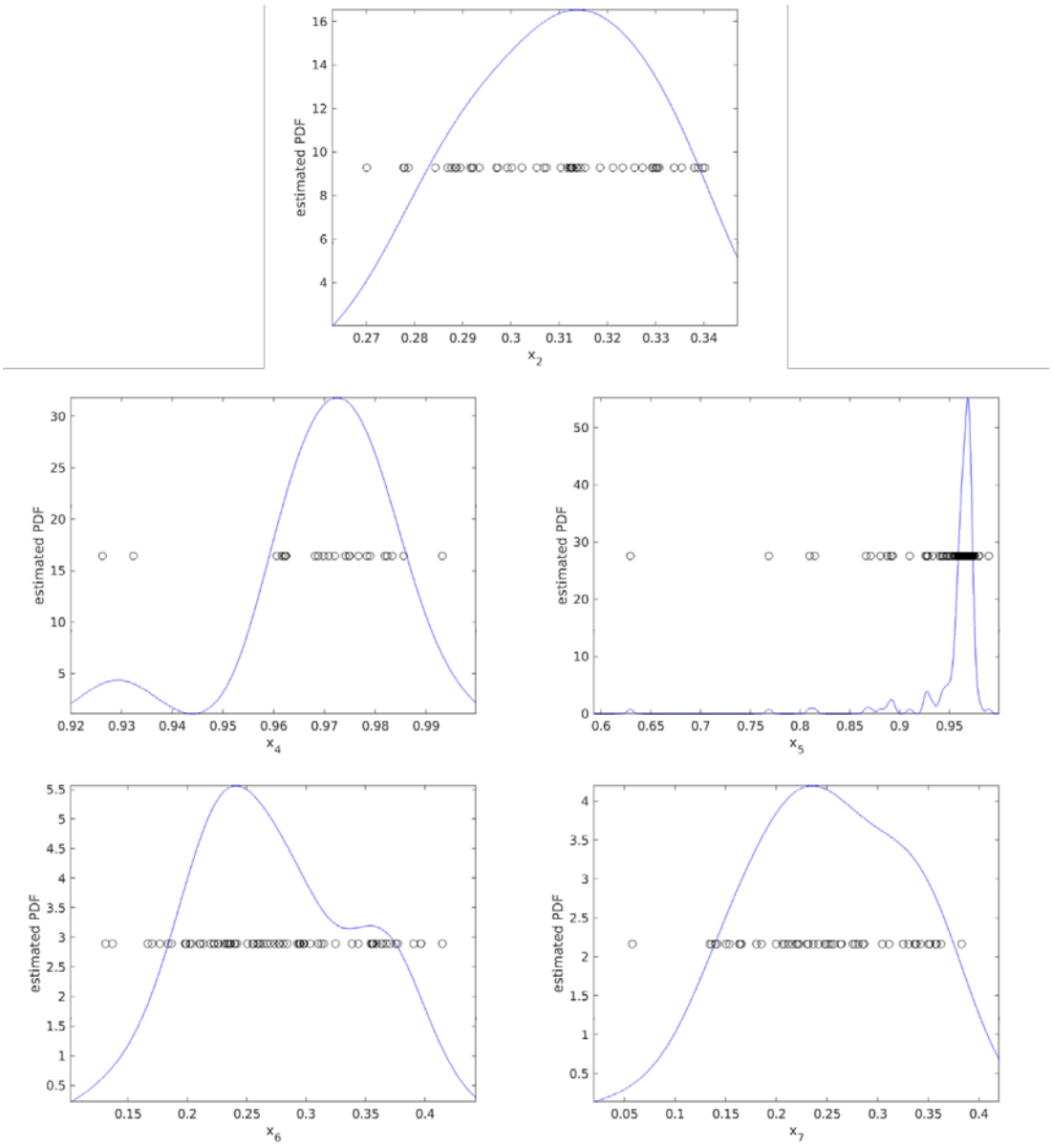



**Figure 8: Estimated probability distributions for the five aleatory uncertainties, black circles represent measured data.**

For comparison to the original results (Dahl et al., 2019), forward propagation is first considered with $N_e$ = 10 and $N_a$ = 100. LH epistemic sampling is selected. Two example p-boxes with these settings are shown in Figure 9 surrounding the original result of Dahl et al. (2019) which was generated via direct model evaluation by which the DEM model itself was evaluated at the 10×100 points necessary to construct the p-box. The full-model and surrogate model p-boxes compare favorably. Re-evaluating the forward propagation (which is nearly instantaneous for the surrogate) can produce p-boxes which may look more or less qualitatively similar, depending on the sampling.

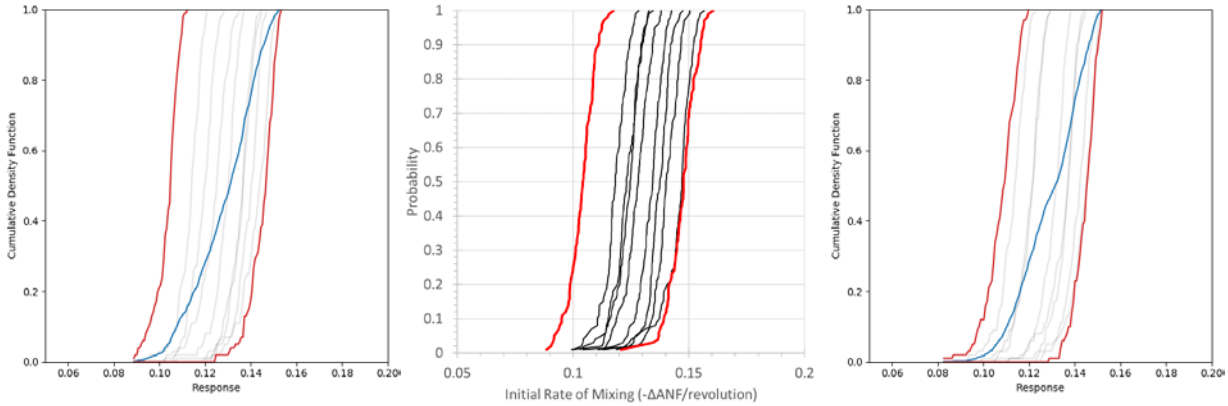



**Figure 9: Example p-boxes from surrogate model forward propagation (left, right) inclosing original full-model evaluation (center) of Dahl et al. (2019).**

While there was not a substantial computational savings in constructing the surrogate model over the full model evaluation (i.e., 345 vs. 1,000 simulations), an advantageous feature of the surrogate approach is that the dependence of the p-box on the sampling (either the method or the number of samples) can be easily changed. Figure 10 shows how the p-box changes with increasing number of samples keeping a factor of ten in the $N_a/N_e$ ratio. It is noted that this case converges nicely—specifically the tails of the p-box are not significantly enhanced with continued sampling—due to the truncated ranges of the KDE distributions.

Finally, it is noted that this (hypothetical) study was focused on this forward propagation step. If this study were part of a larger VVUQ study aimed at evaluation of model form uncertainty (by comparison to experimental data), numerical uncertainty, largely due to discretization, must be evaluated and included (buffering the p-box) for both surrogate and full (direct) mode forward propagation. Additionally, the surrogate model generated p-box would need to include the uncertainty (error) resulting from the p-box itself.

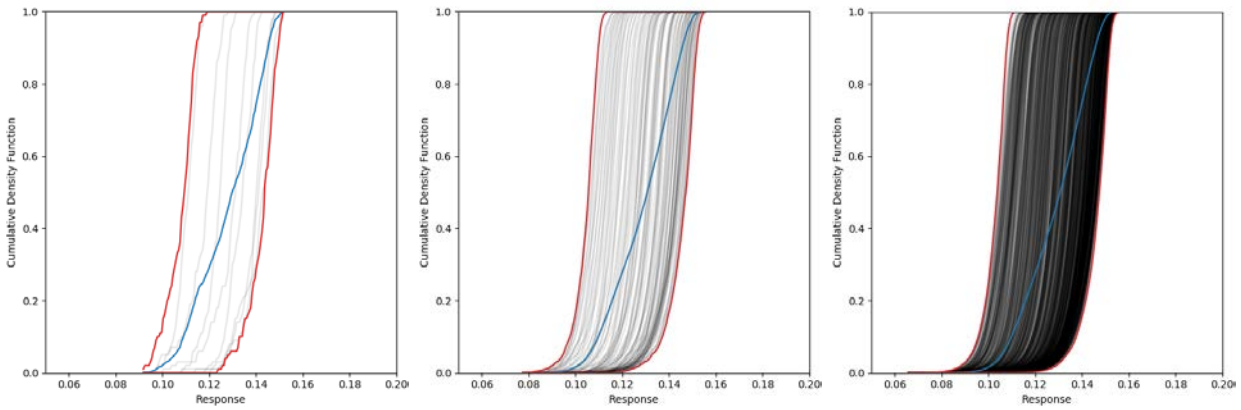



**Figure 10: Evolution of p-box with increased sampling: $N_e$ = 10, $N_a$ = 100 (left), $N_e$ = 100, $N_a$ = 1,000 (center), $N_e$ = 1,000, $N_a$ = 10,000 (right).**

## 6. SENSITIVITY ANALYSIS

While the primary focus of this work was the forward propagation (Section 5), the surrogate model can also be used for further analysis. In the original study by Dahl et al. (2019), the full model evaluated p-box was compared to reduced (less computationally expensive) forward propagations by beginning with a sensitivity analysis and eliminating unimportant inputs from the UQ study. Using a local sensitivity analysis with replication, Dahl et al. (2019) found that $x_3$, $x_7$, $x_4$ and $x_6$, in that order, had the largest impact on the SRQoI. The local sensitivity analysis is checked by a global sensitivity analysis on the surrogate model. For the analysis the RSM node containing the GP model is connected to Nodeworks' Sensitivity Analysis node, also from the Surrogate Modeling and Analysis toolkit. Sobol' indices are computed using 100,000 samples with 1,000 resamples. The results in Figure 11 are in good agreement, at least qualitatively, with the local sensitivity analysis. Ranking by total indices, the most important parameters are $x_3$ ($0.737 \pm 0.005$), $x_7$ ($0.131 \pm 0.002$), $x_4$ ($0.101 \pm 0.001$) and $x_5$ ($0.035 \pm 0.000$). While the strong dependence of the results on $x_3$ may have been obvious from the scatter plot of Figure 4, the agreement, at least in the order of importance, between the local, full-model sensitivity analysis and the global, surrogate model sensitivity analysis is encouraging.

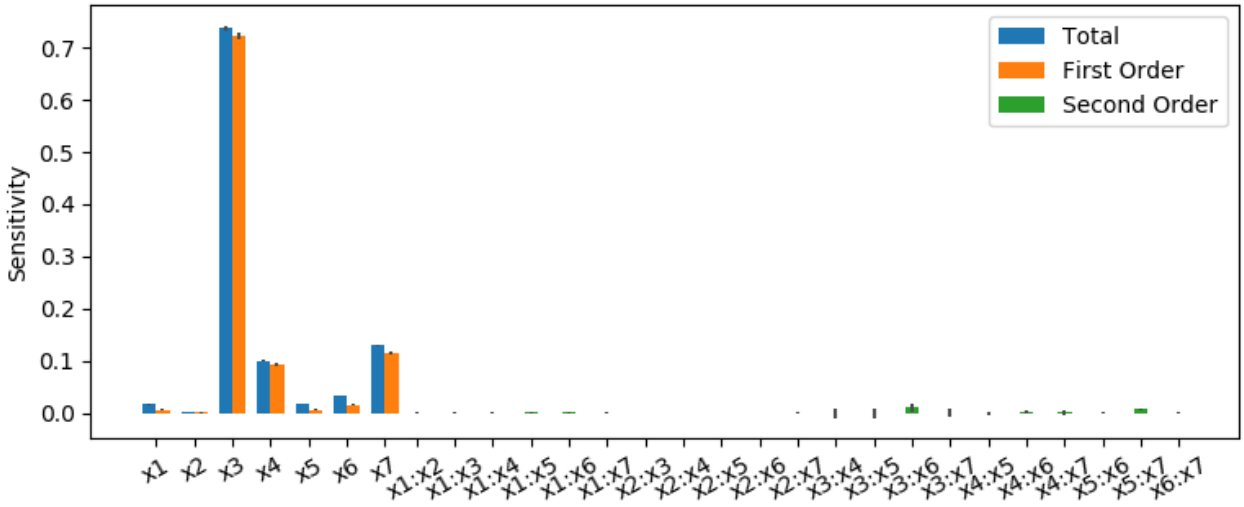



**Figure 11: Global Sobol' sensitivity indices computed from 100,000 samples of the GP surrogate model.**

## 7. CONCLUSIONS AND FUTURE WORK

This work followed a recent study by Dahl et al. (2019) on the forward propagation of seven input uncertainties through an MFiX-DEM model of a square, rotating tumbler. In the original work, the uncertainties were propagated directly through the MFiX-DEM model. Here, Nodeworks was used extensively to drive a companion effort in which a (numerical) data-fitted surrogate model was constructed and used for the forward propagation. It was found that the resulting probability bands or p-boxes computed from the surrogate model were in good qualitative agreement with the original p-box evaluated directly from the full model by Dahl et al. (2019). (Note that quantitative comparison is complicated due to the stochastic nature of the sampling.) A global sensitivity analysis was also carried out on the surrogate model. First- and second-order Sobol' indices computed from 100,000 samples were used to rank the importance of the uncertain inputs. The ranking (order of importance) was is in good agreement with the local sensitivity analysis originally performed on the full model (Dahl et al., 2019).

The most apparent benefit of the surrogate modeling approach to forward propagation is the substantially increased flexibility over the direct propagation method. For instance, if the p-box was judged to be too coarse after the evaluation of a 10x100 (epistemic-by-aleatory) sampling, it is straightforward and computationally inexpensive to arbitrarily increase the sampling until the desired p-box resolution is reached, as demonstrated in Section 5. Further, for sensitivity analysis, the sampling and even this method itself can be easily adjusted when used with a surrogate model.

The surrogate model approach employed here did not substantially reduce the overall computational overhead: 1,000 simulations in the original direct propagation compared to 345 simulations (plus the effort of constructing the surrogate). The LS sampling criterion followed here was recommended in the context of forward propagation (Oberkampf and Roy, 2010); however, not necessarily for the construction of surrogate models. Other guidelines for surrogate construction are more favorable. For instance, Loeppky et al. (2009) recommend as few as 10 samples per dimension may be sufficient. This would reduce the number of samples used in this work from 345 to 70, however this would come at a cost of increased error in the surrogate itself. Future work will be aimed at evaluating the connection between the number of samples, the dimensionality of the problem, and the corresponding (surrogate) model error for multiphase CFD data should be considered.

## 8. REFERENCES


Bakshi, A.; Shahnam, M.; Gel, A.; Li, T.; Altantzis, C.; Rogers, W.; Ghoniem, A. F. Comprehensive multivariate sensitivity analysis of CFD-DEM simulations: Critical model parameters and their impact on fluidization hydrodynamics. *Powder Technology* **2018,** *338*, 519–537.

Choudhary, A.; Roy, C. J.; Dietiker, C. F.; Shahnam, M.; Garg, R.; Musser, J. Code verification for multiphase flows using the method of manufactured solutions. *International Journal of Multiphase Flow* **2016,** *80*, 150–163.

Cundall, P. A.; Strack, O. D. L. A discrete numerical model for granular assemblies. *Geotechnique* **1979,** *29*, 47–65.

Dahl, S.; Liu, P.; LaMarche, C. Q.; Hrenya, C. M. Comparison of State-of-the-Art Uncertainty Quantification (UQ) Method and Reduced UQ Method as Applied to a DEM Model of Particle Mixing in a Small-Scale Tumbler; MFiX-DEM Enhancement for Industrially-Relevant Flows, Task 8 Milestone Report, 2019.

Fullmer, W. D.; Musser, J. CFD-DEM solution verification: Fixed-bed studies. *Powder Technology* **2018,** *339*, 760–764.

Garg, R.; Galvin, J.; Li, T.; Pannala, S. *Documentation of open-source MFIX-DEM software for gas-solids flows*; Technical Report; National Energy Technology Laboratory, Morgantown, WV, 2012a.

Garg, R.; Galvin, J.; Li, T.; Pannala, S. Open-source MFIX-DEM software for gas–solids flows: Part I—Verification studies. *Powder Technology* **2012b,** *220*, 122–137.

Garzó, V.; Montanero, J. M. Diffusion of impurities in a granular gas. *Physical Review E* **2004,** *69*, 021301.

Gel, A.; Chaudhari, K.; Turton, R.; Nicoletti, P. Application of uncertainty quantification methods for coal devolatilization kinetics in gasifier modeling. *Powder Technology* **2014,** *265*, 66–75.

Gel, A.; Garg, R.; Tong, C.; Shahnam, M.; Guenther, C. Applying uncertainty quantification to multiphase flow computational fluid dynamics. *Powder Technology* **2013b,** *242*, 27–39.

Gel, A.; Li, T.; Gopalan, B.; Shahnam, M.; Syamlal, M. Validation and uncertainty quantification of a multiphase computational fluid dynamics model. *Industrial & Engineering Chemistry Research* **2013a,** *52*, 11424–11435.

Gel, A.; Shahnam, M.; Musser, J.; Subramaniyan, A. K.; Dietiker, J. F. Nonintrusive uncertainty quantification of computational fluid dynamics simulations of a bench-scale fluidized-bed gasifier. *Industrial & Engineering Chemistry Research* **2016a,** *55*, 12477–12490.

Gel, A.; Shahnam, M.; Subramaniyan, A. K. Quantifying uncertainty of a reacting multiphase flow in a bench-scale fluidized bed gasifier: A Bayesian approach. *PowderTechnology* **2017,** *311*, 484–495.

Gel, A.; Vaidheeswaran, A.; Musser, J.; Tong, C. H. Toward the Development of a Verification, Validation, and Uncertainty Quantification Framework for Granular and Multiphase Flows—Part 1: Screening Study and Sensitivity Analysis. *Journal of Verification, Validation and Uncertainty Quantification* **2018,** *3*, 031001.

https://mfix.netl.doe.gov/documentation/dem_doc_2012-1.pdf

Loeppky, J. L.; Sacks, J.; Welch W.J. Choosing the sample size of a computer experiment: A practical guide. *Technometrics* **2009,** *51*, 366–376.

Oberkampf, W. L.; Roy, C. J. *Verification and Validation in Scientific Computing*; Cambridge University Press: Cambridge, UK, 2010.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **2011,** *12*, 2825–2830.

Pöschel, T.; Schwager, T. *Computational Granular Dynamics: Models and Algorithms*: Springer Science & Business Media, 2005.

Roache, P. J. *Verification and Validation in Computational Science and Engineering*; Hermosa Publisheres: Albuquerque, NM, 1998.

Roy, C. J.; Oberkampf, W. L. A comprehensive framework for verification, validation and uncertainty quantificaiton in scientific computing. *Computer Methods in Applied Mechanics and Engineering* **2011,** *200*, 2131–2144.

Scikit-learn, https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html, accessed (February 27, 2019).

Shahnam, M.; Gel, A.; Dietiker, J.-F.; Subramaniyan, A. K.; Musser, J. The effect of grid resolution and reaction models in simulation of a fluidized bed gasifier through nonintrusive uncertainty quantification techniques. *Journal of Verification, Validation and Uncertainty Quantification* ***2016,*** *1*, 041004.

Syamlal, M.; Celik, I. B.; Benyahia, S. Quantifying the uncertainty introduced by discretization and time-averaging in two-fluid model predictions. *AIChE Journal* ***2017,*** *63*, 5343–5360.

Weber, J.; Musser, J.; Gel, A.; Fullmer, W. D. Optimization of a cyclone using multiphase flow computational fluid dynamics. Proceedings of the ASME-JSME-KSME 2019 Joint Fluids Engineering Conference, AJKFLUIDS2019-5182, San Francisco, CA, July 28–Aug 1, 2019.

Xu, Y.; Shahnam, M.; Fullmer, W. D.; Rogers, W. A. *CFD-Driven Optimization of a Bench-Scale Fluidized Bed Biomass Gasifier using MFiX-TFM and Nodeworks-OT*; NETL-TRS-3-2019; NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2019; p 28. DOI: 10.18141/1506664.

**Brian Anderson**
Director
National Energy Technology Laboratory
U.S. Department of Energy

**Bryan Morreale**
Executive Director
Research & Innovation Center
National Energy Technology Laboratory
U.S. Department of Energy

**John Wimer**
Associate Director
Strategic Planning
Science & Technology Strategic Plans
& Programs
National Energy Technology Laboratory
U.S. Department of Energy