UNCERTAINTY ANALYSIS IN MULTIPHASE FLOW PREDICTIONS IN PRESENCE OF SOLIDS

Wei Dai[a], Selen Cremaschi[a, *], Hariprasad J. Subramani[b], Haijing Gao[b]
[a] Chemical Engineering Department, Auburn University, Auburn, AL 36849, USA
[b] Chevron Energy Technology Company, Houston, TX, USA

Abstract

The transport of solids in multiphase flows is common practice in energy industries due to the unavoidable extraction of solids from oil and gas bearing reservoirs. For safe and efficient operation and design of pipelines, reliable estimates of erosion rates are required. Prediction of erosion rates in multiphase flow is a complex problem due to the lack of accurate models for predicting particle movements in the flow and their impact velocities to the wall. The erosion-rate calculations also depend on the accuracy of the flow regime predictions in the pipeline. Our preliminary comparisons of existing model predictions to experimental data revealed that the predictions may differ by several orders of magnitude for some operating conditions. The goal of this paper is to introduce a framework for estimating expected erosion rates and for generating the corresponding confidence intervals of these estimates. The inputs are a model predicting erosion rates and a database containing erosion-rate measurements at various operating conditions. The framework combines a novel data clustering approach with non-parametric regression analysis that uses Gaussian Process Modeling (GPM). The results reveal that the proposed data clustering approach significantly reduces the confidence intervals of the expected erosion rates.

Keywords

Uncertainty analysis, Data clustering, Gaussian process modeling, Multiphase flow, Erosion.

Introduction

The solid transport and management system (STMS) is an important component of production systems for energy industry because of the unavoidable extraction of solids from oil and gas bearing reservoirs either onshore or offshore sites. In STMS, one of the integral design and operational decisions, is the maximum fluid flow rate. If the amount and velocity of the solids traveling with the fluids (oil, water and gas) in the transportation lines are too high, they might cause erosion in the pipelines resulting in facility integrity issues. The erosion risks can result in high operational costs for repairing pipelines and equipment, especially for offshore deep-water production due to the limited access to fields.

The erosion process, especially in conduits with multiphase flows, is a complex phenomenon. It depends on many factors including fluid characteristic, solid characteristics, the construction material properties, and the geometry of the flow lines. Given this complexity, most of the modeling work in this area focuses on developing empirical or semi-empirical models. Although all of these models have been compared to limited sets of experimental data, it has been reported that their predictions for the same input set can vary up to two orders of magnitude (Vieira, 2014). Our preliminary comparisons of model predictions to experimental data revealed that the predictions may be up to 10 orders of magnitude higher (resulting in considerable overdesign) for some input sets whereas up to one order of

[*] selen-cremaschi@auburn.edu

magnitude lower (causing facility integrity issues) for others.

There are many sources of uncertainty in the models used to predict erosion rates. These sources include the model inputs, assumptions, and equations, the data used to develop the model, and the computational methods used to solve the governing equation sets. Furthermore, the models are routinely extrapolated beyond their capabilities as they are originally developed and validated using data collected from bench-scale or at best pilot-scale experiments. Few of the existing erosion models explicitly address the prediction uncertainty or discuss the model's performance when extrapolated.

Mazumder (2004) developed a mechanistic model to predict erosion rates in single- and multi-phase flows, and propagated the input uncertainties. The study revealed that the uncertainties in sand sizes, liquid rates and gas rates are respectively 21%, 6% and 4% of the total erosion-rate prediction uncertainties, and they may yield 20% to 70% uncertainty in erosion-rate predictions depending on input conditions. However, the impact of model-form uncertainty was not considered. Zhang et al. (2007) extrapolated the model developed by Oka et al. (2005) to predict erosion rates caused by fine particles at low flowrates although the original model was developed for relatively large particles traveling at high flowrates. They concluded that the model's predictions closely matched the experimentally observed erosion rates in the extrapolated regions, but the uncertainty of these predictions was not addressed.

For quantifying model uncertainty, recent studies apply statistical analysis and machine learning approaches. Roy and Oberkampf (2011) gives a comprehensive overview of sources of uncertainty in scientific computing, and introduces a procedure for including estimates of numerical error and model-form uncertainty. Lin et al. (2012) compiled several available uncertainty quantification (UQ) approaches to a report. The approaches were forward uncertainty propagation, sensitivity analysis, response surface methods, and dimensional reduction. They also included a list of available tools and software packages that embeds some or all of these approaches. In a recent paper, Thacker et al. (2015) studied two interpolation approaches to quantify the uncertainties of the concentration of oil at the sea surface resulting from a certain spill site in the Gulf of Mexico. They approximated the response with a flexible *n*-degree polynomial and with a Gaussian process. Both approaches provided reliable estimates of the output uncertainty, and the authors concluded that the Gaussian process was less likely to exhibit erratic behavior in uncertainty estimates when extrapolated.

This paper introduces a framework to quantify prediction uncertainty and its confidence interval for erosion models under a wide range of input conditions especially focusing on regions where experimental data is scarce or not available. A comprehensive database, consisting of measurement approaches, flow regimes and experimental set-up details, is assembled from open literature. The data are clustered using a unified similarity metric that is shown to perform well with datasets containing both numerical and categorical attributes (Cheung and Jia, 2013). For each cluster, a non-parametric model based on Gaussian processes is trained to estimate the model discrepancy - the difference between the model predictions and experimental observations. The methodology is applied to an erosion-rate prediction model that is commonly used by oil and gas industry. The results reveal that the confidence intervals of the erosion-rate predictions are reduced by 41% using the clustering approach compared to the ones obtained by dividing the data based on flow-regime only (Dai and Cremaschi, 2015).

Experimental Database and Data Preprocessing

We have collected approximately seven hundred experimental data points in single or multiphase flow with detailed operating conditions from open literature (The reference list is available upon request from the corresponding author). The independent variables are geometry and diameter of the pipe, hardness of the pipe material, particle size and rate, densities and viscosities of the liquid and gas, and liquid and gas flow rates. The dependent variable is the measured erosion rate (experimental ER) in mils/lb. The database also includes the approach used to measure erosion rate, flow orientation and particle impingement angle, if provided by the experimenters.

The data is preprocessed to consolidate erosion-rate measurement discrepancies. Training a Gaussian process model is prone to ill-conditioning as the distances between its training points decrease. These training points result in linearly-dependent equations and cause instability in the model predictions (Giunta et al., 2006). The following steps avoid that the training points overlap with each other: (1) For experiments at identical operating conditions, the average of the erosion rate measurements is used. (2) For data points with same operating conditions except flow orientation, the erosion rate from vertical orientation is selected because the erosion model used in this paper was developed for predicting maximum erosion rate under vertical orientation. (3) For data points taken at the same operating condition except their particle flow-rates, the one with the highest measured erosion rate is kept. (4) For experiments conducted at the same conditions except their particle impingement angles, the one with higher erosion rate is selected. After the preprocessing, the database contained 585 linearly-independent data points.

Overview of the Erosion-Rate Prediction Model

The model predicts maximum erosion rate given system geometry and materials, flow conditions, and particle properties (McLaury and Shirazi, 1999). It calculates the maximum erosion by defining how a hypothetical representative particle will impinge the target material. The abrasion caused by this particle is defined by thickness loss in the target specimen, and is calculated using the momentum of impingement. Given flow conditions, particle and pipe properties, the model first calculates the characteristic impact velocity. The erosion ratio, which is defined as the ratio of measured target material mass loss to the mass of all particles in the carrier fluid, is calculated using a power law correlation of the characteristic impact velocity. The maximum erosion rate (predicted ER), which is defined as the target specimen thickness loss per particle weight, is calculated using the erosion ratio and accounts for pipe geometry, size and material; fluid properties (density and viscosity); and sand sharpness, density and rate via empirical constants.

The characteristic particle impact velocity depends on the flow regime in multiphase flows. The multiphase flow mixtures exhibit different flow patterns in the conduit, i.e., different flow regimes, depending on the relative ratios of liquid and gas amounts, their densities and viscosities. The erosion model can predict the erosion rate for mist (Mi), annular (An), churn (Ch), slug (Sl), bubbly (Bu) and dispersed bubble (DB) flows.

Methodology

Data Clustering

The data clustering approach uses the object-cluster similarity metric (OCIL) developed by Cheung and Jia (2013). The clustering problem of N data points, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\mathrm{N}}\}$, with mixed attributes into k different clusters, denoted with $C_1, C_2, \ldots, C_k$, can be formulated as

$$\mathbf{Q}^* = \arg\max_{\mathbf{Q}} F(\mathbf{Q}) = \arg\max_{\mathbf{Q}} \left[\sum_{j=1}^{k}\sum_{i=1}^{N} q_{ij}\, s(\mathbf{x}_i, C_j)\right] \tag{1}$$

In Eq. (1), $s(\mathbf{x}_i, C_j)$ is the OCIL between data point $\mathbf{x}_i$ and cluster C_j, and $\mathbf{Q} = (q_{ij})$ is an $N \times k$ partition matrix satisfying $\sum_{j=1}^{k} q_{ij} = 1$ and $0 < \sum_{i=1}^{N} q_{ij} < N$ where $\mathrm{q_{ij}} \in \{0,1\}$, $i = 1,2, \ldots, N$, and $j = 1,2, \ldots, k$.

The OCIL is calculated as a combination of the similarity measures obtained for categorical, $\mathbf{x}_i^{\mathrm{c}}$, and numerical, $\mathbf{x}_i^{u}$, attributes. The similarity between $\mathbf{x}_i^{\mathrm{c}}$ and C_j, $s(\mathbf{x}_i^{\mathrm{c}}, C_j)$ is defined as

$$s(\mathbf{x}_i^{\mathrm{c}}, C_j) = \sum_{r=1}^{d_c} w_r\, s(\mathrm{x}_{ir}^{\mathrm{c}}, C_j) \tag{2}$$

where the weight factor, w_r, accounts for possible unequal importance of each attribute, and d_c is the number of categorical attributes. The similarity between a categorical attribute value x_{ir}^{c} and cluster C_j, $i \in \{1,2, \ldots, N\}$, $j \in \{1,2, \ldots, k\}$, $r \in \{1,2, \ldots, d_c\}$, is defined as

$$s(\mathrm{x}_{ir}^{\mathrm{c}}, C_j) = \frac{\sigma_{A_r = x_{ir}^{c}}(C_j)}{\sigma_{A_r \neq NULL}(C_j)} \tag{3}$$

where $\sigma_{A_r = x_{ir}^{c}}(C_j)$ is the number of data points that have the value x_{ir}^{c} for attribute A_r in cluster C_j, NULL refers to the empty set, and $\sigma_{A_r \neq NULL}(C_j)$ is the number of data points that have the attribute A_r whose value is not equal to NULL in cluster C_j. The value of $s(\mathrm{x}_{ir}^{\mathrm{c}}, C_j)$ becomes one when all the data points have the value $\mathrm{x}_{ir}^{\mathrm{c}}$ for attribute A_r in cluster C_j, and becomes zero when none of the data points has the value $\mathrm{x}_{ir}^{\mathrm{c}}$ for attribute A_r in cluster C_j.

The similarity measure between numerical attribute $\mathbf{x}_i^{u}$ and cluster C_j, $i \in \{1,2, \ldots, N\}, j \in \{1,2, \ldots, k\}$ is given by

$$s(\mathbf{x}_i^{u}, C_j) = \frac{\exp(-0.5\mathrm{Dis}(\mathbf{x}_i^{u}, \mathbf{c}_{\mathrm{j}}))}{\sum_{t=1}^{k} \exp(-0.5\mathrm{Dis}(\mathbf{x}_i^{u}, \mathbf{c}_{\mathrm{t}}))} \tag{4}$$

where $\mathbf{c}_{\mathrm{j}}$ is the cluster center (i.e., a vector of average values of each numerical attribute in cluster C_j), and $\mathrm{Dis}(\cdot)$ stands for the Euclidean distance. The value of $s(\mathbf{x}_i^{\mathrm{u}}, C_j)$ is within the interval [0,1] (Eq. (4)).

The OCIL between $\mathbf{x}_i$ and cluster C_j can be obtained as the average of the similarity measures calculated based on each feature,

$$s(\mathbf{x}_i, C_j) = \frac{d_c}{d_f} s(\mathbf{x}_i^{\mathrm{c}}, C_j) + \frac{1}{d_f}\, s(\mathbf{x}_i^{u}, C_j) \tag{5}$$

where d_f denotes the total number of attributes. Because the numerical attributes are often treated as a vector and handled together in clustering analysis, $d_f = d_c + 1$.

We use the iterative clustering algorithm of Cheung and Jia (2013). The optimal $\mathbf{Q}^* = \{q_{ij}^*\}$ in Eq. (1) is defined by

$$q_{ij}^* = \begin{cases} 1 & \text{if } s(\mathbf{x}_i, C_j) \geq s(\mathbf{x}_i, C_r) \quad \forall 1 \leq r \leq k \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Each data point $\mathbf{x}_i$ is assigned to the cluster with the largest OCIL among the k clusters for that data point (Eq. (6)). The number of clusters is determined by a penalization mechanism that gradually eliminates redundant clusters. The algorithm assumes that the number of clusters, k, is initialized to a value greater than the true value (i.e., $k \geq k^*$), and assigns a weight to each cluster. This weight measures the importance of each cluster to the whole clustering structure. After a data point is assigned to the winning cluster, which has the largest OCIL for that data point, the weight of the winning cluster is increased and the weight of the cluster with the second largest OCIL is decreased as a penalty. As the algorithm proceeds, the clusters with very low weights are assigned fewer data points, and hence, may eventually be eliminated. The algorithm terminates when the maximum learning epoch is reached.

In general, the algorithm is executed multiple times, because the data points are randomly assigned to clusters at initialization. In this work, we propose a special initialization scheme based on the concept that cluster centers are surrounded by more data points with a greater density and positioned relatively far away from other centers (Rodriguez and Laio, 2014). The local density and the distance between high density points are calculated to determine the number of clusters, cluster centers, and to make the initial data point assignments to clusters. A parameter, cutoff distance, controls the average number of neighboring data points that is used for density calculations. The number of clusters obtained and cluster centers change when the cutoff distance is changed. As a rule of thumb, Rodriguez and Laio (2014) recommend selecting a cut-off distance such that the average number of neighboring data points is around 1-2% of the total number of data points.

For the erosion data set, eleven different cluster sets are obtained by changing cut-off distance so that the average number of neighboring data points increased from 1% to 2% with an interval of 0.1%. The corresponding operating conditions of each cluster center is compiled, and the overlapping centers are removed. The Euclidean distances between the remaining cluster centers are calculated, and the cluster centers that are deemed very close to each other based on these distances are consolidated to a single cluster center. After the consolidation step, the remaining unique cluster centers are used to determine the number of clusters, and the cluster centers to initialize the clustering algorithm.

Gaussian Process Modeling for Uncertainty Analysis

According to general model uncertainty quantification formulation (Kennedy and O'Hagan, 2001), the experimental response, $\mathrm{y^e}$ can be expressed as $\mathrm{y^e = y^m + \delta' + \varepsilon}$, where $\mathrm{y^m}$ is the model response, δ' is the model bias, and ε is the experimental uncertainty. In this work, the experimental response is the measured erosion rate, and the model response is the predicted erosion rate by the model. The experimental uncertainty is assumed to follow a zero-mean normal distribution. The model bias includes the uncertainties associated with estimated model parameters, the numerical errors and the model form discrepancies, and is expressed as a Gaussian random process (Jiang et al., 2013). Because the experimental uncertainty is assumed to follow a zero-mean normal distribution, the model bias and the experimental uncertainty can be combined into one term, which we will refer as the model discrepancy (δ).

The Gaussian process, $\mathcal{GP}(m,k)$, is a natural generalization of the Gaussian distribution (Rasmussen and Williams, 2006). Let x denote a point in multidimensional space, then $m(x)$ is the mean function of the $\mathcal{GP}(m,k)$, and $k(x,x')$ is the covariance function of the $\mathcal{GP}(m,k)$, representing the spatial covariance between any two points (x and x') at the process. There are different mean and covariance function forms that can be used for constructing the Gaussian process model – GPM. The mean and covariance functions' hyper-parameters are determined via the maximum-likelihood estimation using the available data, and this procedure is referred to as training the GPM. Once trained, the GPM can be used to predict $m(y)$ and its variance, where y is a point in the multidimensional space that was not in the training data set. The variance is calculated using the covariance function.

A constant mean function and a neural network covariance function are determined to be appropriate for estimating model discrepancy for the dataset and model used in this study through trial-and-error. Both categorical and numerical attributes, and the corresponding actual model discrepancies are used to train the GPM. The actual model discrepancy is defined as the difference between experimental erosion rates and the corresponding erosion rate predictions of the model. The experimental database covers a wide range of input conditions resulting in significantly different erosion rates, and, at times, up to five orders of magnitude differences in actual model discrepancies. To minimize the impact of scaling issues on the GPM training, both numerical attributes and the actual model discrepancies are normalized to the range [0.1, 1].

The normalized data set is divided into two subsets: (1) a training set (3/4th of the data points), and (2) a test set (1/4th of the data points). The data in the training set is used to calculate the maximum-likelihood estimators of the GPM hyper-parameters, and the test set is used to assess the performance of the trained GPM. A four-fold cross validation is used to generate GPM predictions for all data points in the dataset. The process is repeated for 30 times, and the results are averaged to minimize the impact of local solutions on GPM predictions. A Matlab® based toolbox, Gaussian Process for Machine Learning - GPML (Rasmussen and Williams, 2006), is used to train the GPM. The same toolbox is used to calculate the expected mean ($\hat{\delta}$) and variance ($\hat{\sigma}^2$) at the test data points. The expected mean gives the model discrepancy prediction, and the confidence interval of the prediction at α significance levels is calculated using the estimated variance ($\hat{\sigma}$).

Assessing the Quality of GPM Predictions

Area metric (AM) is a measure that quantifies how well a model estimates the experimental observation of a physical variable (Ferson et al., 2008). It is defined as the disagreement area between the estimated variable ($\hat{y}^e$) and experimental observation ($\mathrm{y^e}$). In Figure 1, total area of the shaded region gives the AM for an input condition.

In Figure 1, the experimental measurement is given as a single value whereas the model estimate is expressed as a probability distribution. A smaller AM indicates a better predictive capability of the model for that input condition. An overall AM can be calculated by summing the AMs for a set of input conditions. The predictive capabilities of different models can be assessed using the overall AM.

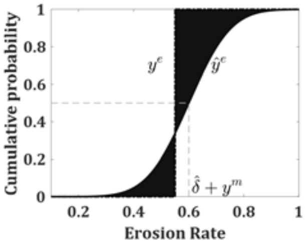

Figure 1. Definition of area metric

Implementation of the Overall Methodology

The steps for the uncertainty analysis is summarized in Figure 2. After the preprocessing of database, density-peak based initialization scheme is employed to find the cluster centers for initialization, and the clustering algorithm using OCIL is applied to assign all the data points to the proper clusters. Each of the clustered data sets is separated as training set and test set. A GPM is trained using the training set, and used to estimate the uncertainty in the test set. For each cluster, the uncertainty in erosion-rate predictions is obtained after four-fold cross validation. Finally, the AM for each data point is calculated, and the overall AM is obtained. The overall algorithm is implemented in MATLAB R2016a and executed on a 2.30GHz Intel Xeon E5 PC with 32GB memory running Windows 10 Enterprise.

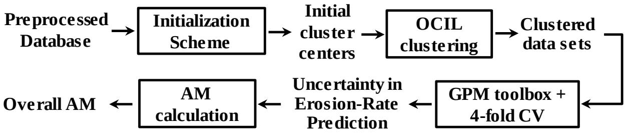

Figure 2. Flowchart for the uncertainty analysis

Results and Discussion

The numerical attributes are pipe material hardness and size, particle diameter, liquid viscosity and velocity, gas density and velocity. The categorical attributes are pipe geometry, flow regime and flow orientation.

Impact of Proposed Initialization Scheme

Using the proposed cut-off values, a total of 11 cluster sets were obtained. Three of these sets contained seven, nine and eleven elements respectively, two of them contained eight elements and six of them contained ten elements, which yielded 103 cluster centers. Of these cluster centers, 17 of them were identical (appeared in several sets) resulting in 27 distinct cluster centers. The Euclidean distances between these 27 cluster centers recommended further consolidation leaving a total of 12 unique cluster centers for initialization. The OCIL clustering algorithm constructed the final cluster set, which contained 12 clusters. A GPM is trained for each of the clusters, and the overall AM is calculated to be 14.26.

In order to assess the performance of the proposed initialization scheme, the OCIL clustering algorithm is run 50 times using randomly generated 12 initial cluster centers. The GPMs are trained for each cluster, and the overall AM is calculated for each cluster set. The cluster set with the minimum overall AM value is selected. This cluster set contained eight clusters with an overall AM of 13.99.

Both approaches yielded clusters with similar operating conditions, where data are separated based on material hardness, pipe diameters and particle sizes. The proposed initialization scheme grouped data points in horizontal orientation into three clusters, while the random initialization grouped data points in horizontal orientation into one cluster. The overall AMs are comparable. However, the proposed initialization scheme considerably reduces the computational time (30 CPUs) compared to the random one that requires multiple initializations (4170 CPUs).

Table 1 gives the cluster centers of pipe and particle diameters (D, d_p), and the number of data points, flow regime and average AM for each cluster. Nine clusters have around or more than 30 data points, which are enough to develop highly non-linear relationships using GPMs (Sankararaman et al., 2011). For the remaining three clusters, the variance of the GPM may decrease with additional data points, however, there is not a correlation between the variance and the number of data points. Clusters six and seven contain data mostly collected from slug flow and have the highest average AMs. The data points with higher AM values are from conditions of very small particle sizes (20 µm) and relatively denser liquid viscosity (40 cp) where the erosion model used tends to under predict. The particle movement and impingement under this conditions may not follow the model assumptions for larger particles (>50 µm).

Table 1. Cluster centers and relevant data

	D (in)	d_p (µm)	No.	Flow regime	AM_{avg}
1	1	150	12	Gas, An	5.4×10^{-3}
2	1	250	20	Gas, Mi, An, Ch	5.6×10^{-3}
3	2	150	95	Mi, An	7.4×10^{-3}
4	2	300	83	Mi, An, Ch, Sl	4.2×10^{-2}
5	2	350	28	Gas, Mi	6.3×10^{-3}
6	3	20	15	Sl	1.7×10^{-1}
7	3	300	85	Sl, liquid	1.3×10^{-1}
8	3	300	71	Gas, Mi, An, Sl	3.6×10^{-3}
9	3	300	32	Mi, An (horizontal)	1.6×10^{-3}
10	4	150	37	An, Ch (horizontal)	3.4×10^{-3}
11	4	150	78	Gas, An, Sl	4.9×10^{-3}
12	4	300	29	Mi, An (horizontal)	4.0×10^{-3}

Figure 3 shows the average measured (y_{avg}^e) and predicted (y_{avg}^m) erosion rates, average (δ_{avg}) and standard deviations (δ_{std}) of actual model discrepancies, average predicted erosion rates using GPM ($\hat{y}_{avg}^e$) and the prediction uncertainties (represented by the standard deviation: $\hat{\sigma}_{avg}$) for each cluster. Figure 3 suggests that, on average, the predicted erosion rates using GPM (5th set) are closer to measured values (1st set) than model predictions (2nd set) for all clusters except for the 1st cluster, which is developed with limited number of data points. It can also be observed that the model tends to under-predict erosion rates of the 2nd, 5th, 6th and 8th clusters. Due to the large standard deviation of δ (see δ_{std} in Figure 3), the predicted erosion rates using GPM also have large standard deviations ($\widehat{\sigma}$).

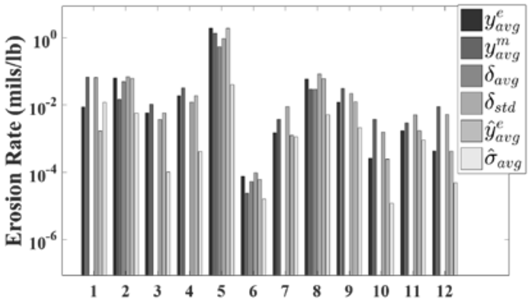

Figure 3. Average erosion rate predictions using GPM with OCIL clustering

Conclusions and Future Directions

This paper applied data mining methods and Gaussian process modeling to estimate the model prediction uncertainties for erosion. A clustering approach for datasets with mixed attributes is adopted for identification of data with similar characteristics. A new initialization scheme is proposed for the clustering approach. This scheme is shown to reduce the computational burden for clustering due to random initialization. For each cluster, model discrepancy predicted by GPM is added to the prediction from the model to remove model's bias. In a previous work, we clustered the data based on the predicted flow regimes, and trained a GPM for each cluster (Dai and Cremaschi, 2015). There were six clusters: gas, mist, annular, churn, slug and liquid. The overall AM using clustering based on flow regime only was 23.03. The proposed clustering approach in this paper reduced the overall AM by 41% compared to flow regime based clustering. The results recommend clustering the data using the OCIL metric and using the proposed initialization scheme for obtaining the best predictive capability for erosion rates most efficiently. This framework can also be applied to other models and used as a guide for future model development and experimental designs.

Acknowledgments

This work was supported by NSF CAREER Award #1623417 and Chevron Energy Technology Company. Authors would like to thank Drs. Shirazi and McLaury (University of Tulsa) for discussion on erosion modeling and providing some of the experimental data.

References

Cheung, Y. M., Jia, H. (2013). Categorical-and-numerical-attribute data clustering based on a unified similarity metric without knowing cluster number. *Pattern Recognition*, *46*, 2228-2238.

Dai, W., Cremaschi, S. (2015). Quantifying Model Uncertainty in Scarce Data Regions – A Case Study of Particle Erosion in Pipelines. *12th Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*, Copenhagen, Denmark.

Ferson, S., Oberkampf, W. L., and Ginzburg, L. (2008). Model Validation and Predictive Capability for the Thermal Challenge Problem, *Computer Methods in Applied Mechanics and Engineering*, Vol. 197, No. 29-32, pp 2408-2430.

Giunta, A. A., McFarland, J. M., Swiler, L. P., Elder, M. S. (2006). The promise and peril of uncertainty quantification using response surface approximations, *Structure and Infrastructure Engineering, Vol. 2, Nos. 3-4*, 175-189

Jiang, Z., Chen, W., Fu, Y. and Yang, R. (2013). Reliability-Based Design Optimization with Model Bias and Data Uncertainty, *SAE International*, 10.4271.

Kennedy, M. C., O'Hagan, A. (2001). Bayesian calibration of computer models, *Journal of the Royal Statistical Society Series B-Statistical Methodology, 63*, 425-450

Lin, G., Engel, D. W. and Esliner, P. W. (2012). Survey and Evaluate Uncertainty Quantification Methodologies. *PNNL-20914 Report*, Richland

Mazumder, Q. H. (2004). Development and Validation of a Mechanistic Model to Predict Erosion in Single-Phase and Multiphase Flow, PhD. Dissertation, Department of Mechanical Engineering, The University of Tulsa, Tulsa, Oklahoma, USA

McLaury, B.S. and Shirazi, S.A. (1999). Generalization of API RP 14E for Erosive Service in Multiphase Production, *Society of Petroleum Engineer*, 423-432

Oka, Y. I., Okamura, K., Yoshida, T. (2005). Practical estimation of erosion damage caused by solid particle impact. Part 1: effects of impact parameters on a predictive equation, *Wear 259*, 95–101.

Rasmussen, C. E., Williams, C. K. (2006). Gaussian Process for Machine Learning, *The MIT Press*.

Roy, C. J., Oberkampf, W. L. (2011). A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, *Comput. Methods Appl. Mech. Engrg., 200*: 2131-2144

Sankararaman, S., Ling, Y., Shantz, C., Mahadevan, S. (2011). Uncertainty quantification in fatigue crack growth prognosis. International Journal of Prognostics and Health Management, Vol. 2 (1) 001, pages: 15.

Thacker, W. C., Iskandarani M., Goncalves, R. C., Srinivasan, A. and Knio, O. M. (2015). Pragmatic aspects of uncertainty propagation: A conceptual review, *Ocean Modelling, 95*: 25-36

Vieira, R. (2014). Sand Erosion Model Improvement for Elbows in Gas Production, Multiphase Annular and Low-Liquid Flow. PhD. Dissertation, Department of Mechanical Engineering, *University of Tulsa*, Tulsa, Oklahoma, USA

Zhang, Y., Reuterfors, E. P., McLaury, B. S., Shirazi, S. A., Rybicki, E. F. (2007). Comparison of Computed and Measured Particle Velocities and Erosion in Water and Air Flows. *Wear 263*, 330-338