

Predictive analytics with an advanced Bayesian modeling framework

**Jesper Kristensen*, Isaac Asher, You Ling
Kevin Ryan, Arun Subramaniyan, Liping Wang**

**General Electric Co. - Global Research
Niskayuna, New York**

***Corresponding author: jesper.kristensen@ge.com**

ABSTRACT

One of the main limitations in predictive analytics is the acquisition cost of engineering data due to slow-running computer code or expensive experiments. Also, data is often multi-dimensional and highly non-linear in nature, causing problems for standard statistical predictive models. Once data is collected and models are built, many applications require accurate and scalable uncertainty quantification (UQ) solutions to enable robust design and reliable decision making. A modeling framework that addresses these common problems is presented here. The advanced Bayesian modeling framework called “GE Bayesian Hybrid Modeling” (GEBHM) combines simulation and experimental data sources using machine learning techniques and Bayesian statistics to perform UQ, provides detailed sensitivity reports for design engineers, visualizes the problem and its solution succinctly, and aids in developing next-step plans for decision makers. GEBHM works well for cases with sparse data situations but also handles large-data problems. Fundamentally, GEBHM relies on Markov Chain Monte Carlo techniques for accurately learning the Bayesian model parameters and employs state-of-the-art techniques for self-validation. The GEBHM framework applied to an example engineering design problem requiring FE and/or CFD data is presented. It is demonstrated how the surrogate model is built, how uncertainty sources are quantified and propagated, how the impact of design decisions on performance is quantified, and how the engineering design is optimized. It is shown that the framework is capable of handling problems with limited data, accurately capture the uncertainty, and provide a high level of detail and accuracy for the designers and decision makers.

ABOUT THE AUTHORS

Jesper Kristensen is a researcher at GE’s Global Research center headquartered in Niskayuna, New York. He works in the Probabilistics Lab.

Isaac Asher works as a Lead Engineer in the Probabilistics Lab.

You Ling is a researcher in the Probabilistics Lab.

Kevin Ryan is a researcher in the Probabilistics Lab.

Arun Subramaniyan is an R&D manager for the Structures Lab.

Liping Wang is an R&D manager for the Probabilistics lab.

Predictive analytics with an advanced Bayesian modeling framework

Jesper Kristensen*, Isaac Asher, You Ling
Kevin Ryan, Arun Subramaniyan, Liping Wang

General Electric, Global Research Center
Niskayuna, New York

*Corresponding author: jesper.kristensen@ge.com

INTRODUCTION

In this paper, we present a Bayesian Hybrid Modeling (GEBHM) framework developed originally by Kennedy and O'Hagan (KOH) (Kennedy and O'Hagan, 2001). Researchers at Los Alamos National Lab (LANL) implemented the framework into a code which the Probabilistics Lab at General Electric Co.'s (GEs) Global Research Center (GRC) adopted and significantly enhanced. We call it the GEBHM approach and it follows Kennedy and O'Hagan's (KOHs) Bayesian calibration technique of using Gaussian Processes (GPs) to model the simulation output and discrepancy between the simulations and experiments (Kumar, Subramaniyan, and Wang, 2012; Subramaniyan et al., 2011). The GP model can then predict the uncertainty in the output parameters due to the uncertainty in the inputs. We illustrate this framework in Figure 1. Meta-models are created and the uncertainty is quantified simultaneously by construction through a Markov Chain Monte Carlo (MCMC) approach (Hastings, 1970; Higdon, Lee, and Holloman, 2003). GEBHM is an enhanced version of the LANL implementation and we have applied it to a wide variety of benchmark and industrial problems.

The BHM framework allows the combination of multi-fidelity and sparse data sources. Figure 1 shows a case where high-fidelity experimental data is known as well as lower-fidelity simulation results. In performing any probabilistic analysis, validation is surely important. Validation of simulations often involves a comparison with experimental data. This is at the heart of BHM. Furthermore, ideally, if we had a statistically large number of simulations and experiments, it is straightforward to use traditional statistical methods for validation. However, it is quite rare to have sufficiently large number of simulations and experiments in real engineering applications. Modern system level engineering models are typically complex and are computationally very expensive. Even with the fastest computers available today, it is impractical to perform a statistically large number of simulation runs. Through multiple different modules, the BHM framework lets the design engineer verify, validate, and confirm a multitude of properties related to the data quality, the model quality, and the physics being modeled. We demonstrate a handful of these in later sections. Finally, to address how we should expect BHM to handle sparse data, the technology fundamentally relies on GP meta-modeling which handles sparse-data scenarios inherently well (MacKay, 1999). This is through the Bayesian nature of the approach through which expert knowledge can readily be encoded in the priors to reflect our belief about the problem at hand.

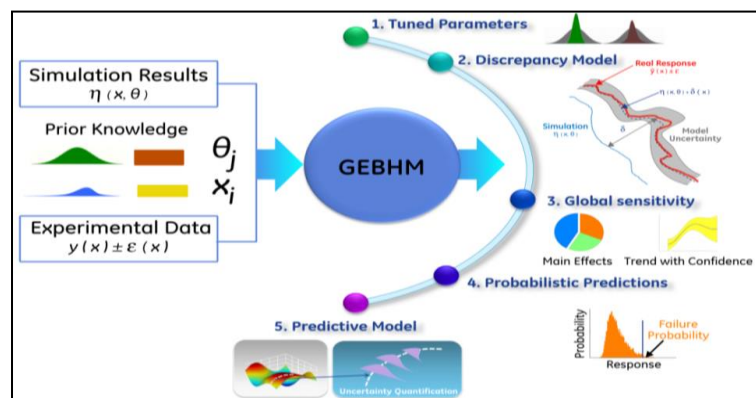


Figure 1. GE's Bayesian Hybrid Modeling (GEBHM) Framework Allowing for Data-Fusion

METHODS DEVELOPMENT AT GE

Since KOH first introduced the BHM algorithm in 2001, much research work has been done by government, academia and industry but many issues and technical challenges remain, particularly when applying the algorithm to solve real engineering problems. The key challenges include, but are not limited to:

- Curse of dimensionality (hundreds or thousands of parameters)
- Model inadequacy (identifiability) and characterization
- Transient data with multiple outputs
- Lack of data and extrapolation
- Multiple sets of experimental data
- Model validation metrics
- Measurement errors and noise
- Multi-physics & multi-fidelity models.

As previously mentioned, GRC has made extensive enhancements to the KOH framework which address the above technical issues and challenges. MCMC analysis is the largest bottleneck to fast execution of BHM. Computations quickly becomes intractable as the number of calibration or model-tuning parameters grow larger than ten. A few techniques including parallel MCMC and adaptive convergence of the Markov chain were developed at GRC to this aspect of the algorithm. These enhancements made the BHM algorithm applicable in solving complex engineering problems with large number of parameters (200+) at GE. Besides enhancements to the MCMC aspect of the algorithm, significant improvements were made to enable correct treatment of transient data with multiple outputs, forward uncertainty quantification, model validation metrics, flexibility and visualization of the BHM calibration process and outcomes. This effort led to efficient implementation of BHM on a wide variety of both steady state and transient problems. This in-house version of BHM, GEBHM, has been successfully applied to several engineering problems over the years.

Solutions Provided by GEBHM

GEBHM can be used for model calibration when both simulation and experiment data are available. It can also be used for building metamodels (surrogates) if only simulation or experiment data is available. Some of the key elements that sets BHM apart from other approaches include, but are not limited to:

- An explicitly formulated discrepancy $\delta(\cdot)$
- The building of GP emulators for the simulation- and discrepancy models during the calibration and updating process
- The model discrepancy formulation helps prevent over-tuned physics models

GEBHM is a strict superset of BHM and includes the ability to build meta-models (surrogate models) if only simulation or experimental data is available. The GEBHM methodology builds a flexible predictive model suitable for complex nonlinear and multi-dimensional processes. Care has been taken in the implementation to avoid over- and underfitting. The Bayesian approach enables reliable results even when only sparse data are available. BHM models have proven to be powerful as the basis of a variety of analyses including failure probability, reliability, uncertainty quantification/propagation, and robust design. A key advantage over other meta-model techniques is accurate predictive uncertainties. A fast, global sensitivity analysis module provides insight into the process being modelled and aids further model improvement. In circumstances where data is available at multiple fidelities, a data fusion methodology is employed to reduce the amount of high-fidelity data required. The result is high-fidelity predictive accuracy at low-fidelity cost. The high-fidelity model consists of tuned predictions based on the low-fidelity data along with a spatially varying discrepancy term.

In this work, we chose to demonstrate the meta-modeling aspect of GEBHM, but refer the interested reader to, e.g., the work in (Higdon et al., 2004; Kumar, Subramaniyan, and Wang, 2012) for more details on calibration.

THEORETICAL FRAMEWORK

The GEBHM technology was developed to combine data from numerical simulations and experimental sources (multi-fidelity) to build hybrid data-fused Bayesian models of potentially multi-dimensional responses. For more details on

the BHM framework please consult (Higdon et al., 2004; Higdon et al., 2008; Higdon, Nakhleh, and Williams, 2008). It can equally be used to combine low- and high-fidelity simulation models. As described by KOH, the experimental observations of outputs y can be expressed as:

$$y(x_i) \pm \epsilon(x_i) = \eta(x_i, \theta) + \delta(x_i), \text{ for } i = 1..n,$$

where n is the number of experimental observations, θ are the true values of the model/tuning parameters, δ is the discrepancy between the calibrated simulator η and the experimental observation y , and ϵ are the well-characterized observation errors (an optional input to the Bayesian framework).

As proposed by KOH, and as described by Higdon et al. (Higdon et al., 2004), the two sources of data are combined by simultaneously estimating the parameters in the η model as well as the δ model, which are described as GP models.

The parameters of the η and δ Gaussian process models are called hyperparameters and are computed using probabilistic techniques such as MCMC. We use the Metropolis-Hastings (Hastings, 1970) algorithm with univariate proposal distributions for the MCMC posterior updates. The initial values of the covariance matrices are updated with current realizations of the hyperparameters at every MCMC step and realizations from the posterior distributions of the model parameters are produced. This ensures that one does not over-fit the high-fidelity data since the MCMC process tries to find the most probable values for the model parameters that fit the data accurately. One-off values for the model parameters are automatically discarded.

Variance-Based Global Sensitivity Analysis

For design of large complex systems in which GEBHM is frequently employed, it is essential to understand the relative importance of system input variables and their interactions with respect to the desired system outputs. Having the knowledge of relative sensitivity of input design parameters or uncertain random variables can help the engineers in several ways including accelerating through the design of experiment matrices or recognizing a set of critical design or uncertain variables that need more attention to name a few. Primarily, sensitivity analysis is performed either by local or global methods, the latter being more comprehensive and informative. Variance based global sensitivity methods are getting increasingly popular and will be the topic of the following discussion based on the paper by Sudret et al. (Sudret, 2008):

Let $y(x) = g(x_1, x_2, \dots, x_n)$, where $x_i \in [0, 1]$. Sobol decomposition of $y(x)$ is written as:

$$y(x) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{i,j}(x_i, x_j) + \dots + f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) \quad (2)$$

The functions f are defined as:

$$f_0 = \text{Mean of } y(x) = \int g(x) dx \quad (3)$$

$f_i(x_i)$ = Main effect functions = $\int g(x) dx_{\sim i} - f_0$ = Integrate $f(x)$ with all inputs except x_i and subtract the mean. The result will be a function of only x_i .

$$f_{i,j}(x_i, x_j) = \text{Two way Interaction effect} = \int g(x) dx_{\sim \{i,j\}} - f_i(x_i) - f_j(x_j) - f_0 \quad (4)$$

To get interaction effect functions, integrate $g(x)$ with all inputs except the inputs x_i and x_j and subtract the main effect functions of x_i and x_j and the mean of $g(x)$. Higher order interaction effects can be written in a similar fashion. Using the above main and interaction effect functions we can compute the Sobol indices. We let D denote the variance of the true function $g(x)$.

$$D = \text{Var}[g(x)] = \int g^2(x) - f_0^2 \quad (5)$$

By integrating the squared of Eq. (1) and using the orthogonality property, one gets:

$$D = \sum_{i=1}^n D_i + \sum_{1 \leq i < j \leq n} D_{i,j} + D_{1,2,\dots,n} \quad (6)$$

$$D_i = \int f_i^2(x_i) dx_i \quad (7)$$

$$D_{i_1, i_2, \dots, i_s} = \int f_{i_1, i_2, \dots, i_s}^2(x_{i_1}, x_{i_2}, \dots, x_{i_s}) dx_{i_1} dx_{i_2} \dots dx_{i_s} \quad (8)$$

Sobol Indices (S_i) can then be computed as:

$$S_i = \frac{D_i}{D} \quad (9)$$

$$S_{i,j} = \frac{D_{i,j}}{D}$$

Main effect Sobol indices and interaction effect Sobol indices sum to 1:

$$\sum_{i=1}^n S_i + \sum_{1 \leq i < j \leq n} S_{i,j} + S_{1,2,\dots,n} = 1$$

Hence, each Sobol index is a sensitivity measure. It describes how much of the total variance is due to the uncertainties in a set of input parameters. The first order indices S_i give the influence of each parameter taken alone and the higher order indices account for possible mixed influence of multiple parameters.

Total effect of an input x_i is defined as the sum of all partial sensitivities involving the input x_i . For example, total sensitivity of input x_1 can be written as:

$$S_{T1} = S_1 + S_{1,2} + S_{1,3} + \dots + S_{1,n} + S_{1,2,3} + S_{1,2,4} + \dots + S_{1,2,n} + \dots + S_{1,2,\dots,n}$$

$$S_{T_i} = 1 - S_{\sim i} = 1 - \text{sum of all indices not involving input } x_i$$

Sobol indices are good descriptors of the model sensitivity to its input parameters, since they are void of assumptions regarding any monotonicity or linearity in the problem.

The GEBHM global sensitivity analysis module computes the Sobol indices for any GEBHM model. The model form may be integrated analytically, avoiding the use of sampling-based methods for computing Sobol indices. The result is a relatively fast sensitivity analysis even for models with many input parameters.

APPLICATION: OPTIMIZATION OF A GENERIC TWO-PHASE FLOW SYSTEM

As an advanced application of the GEBHM framework, we consider a gas-liquid two-phase flow system (Miki, 2014). Two-phase flows are encountered in applications such as turbo-machinery, gas-turbines, ram-jet and scram-jets, automotive engines and aircraft engines. We seek to understand the effect of operational parameters, geometric variations, and their interactions on two-phase flow behavior since this is essential for future combustion system designs. Two-phase flows such as jets and sheets can be significantly non-linear with respect to the geometry and flow parameters. The goal of our application is to apply a probabilistic optimization to minimize the number of simulations or experiments that are needed to obtain an optimal two-phase flow configuration. We start with a generic two-phase flow configuration and end up with a final design where four geometrical design parameters are optimized. This procedure consists mainly of four processes, see Figure 2. Due to time constraints, all data is from numerical simulations. We stress that, given the GEBHM framework, the procedure presented here can also use a subset of experimental data as well with great benefits to the final predictions.

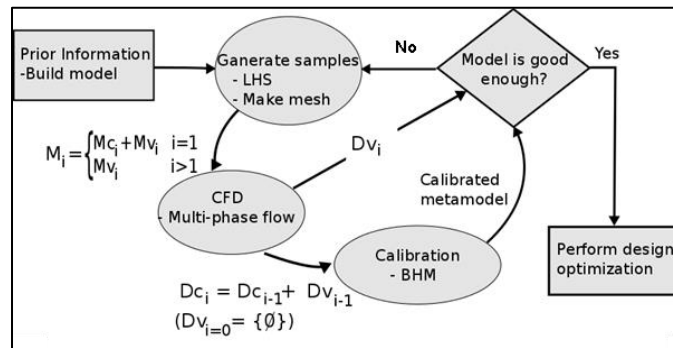


Figure 2. Two-Phase Flow Optimization Procedure

Figure 3 shows the two-dimensional axisymmetric cross section of the generic two-phase flow configuration considered. It consists of a single stream of air (inflow1) and two co-axial streams of liquid (inflows 2 and 3). The

small separation between air and liquid results in an unstable shear layer and promotes the breakup of the liquid. Domain A is the separation between the liquid streams and is the primary region for design optimization. To quantify and vary the separation shape, introduce two random variables θ_1 and R_1 , defined as L_{A2}/L_{A1} . L_{A1} is calculated by $H/\tan(\theta_1)$ (see bottom left of Figure 3). When $R_1 \geq 1$, there is no vertical cut near D. Consider now inflow 3. The bottom portion of the circuit, denoted Domain B, acts as a guide to the liquid stream. It can be critical for the two-phase flow characteristics. Introduce further two random geometrical variables, namely θ_2 and R_2 . Here, R_2 is the ratio of L_{B2} to L_{B1} . Point P is the location where slope F-G reaches the height of the of the upper liquid separation. R_2 is enforced to be less than unity so that the wall does not interfere with the flow path.

The Reynolds number based on the inflow 1 height and velocity is around 40,000 making it reasonable to assume a turbulent inflow profile. At the outflow, impose a 0 Pa gauge pressure as the pressure-boundary and allow the formation of a recirculation zone.

Design of Experiments (Sampling)

An optimized Latin hypercube design (Viana, Venter, and Balabanov, 2010) is used to generate the initial input design space. Find the bounds of the variables listed in Table 1. A total of 19 DOE points in 4 dimensions were generated with corresponding geometries shown in Figure 4. The corresponding geometries are shown in Fig 4. Computational Fluid Dynamics (CFD) was simulated at each DOE point with the Volume-of-Fluid (VOF) model available in ANSYS-Fluent Version 14.5 The data obtained consists of the liquid location distributions (LLD), the liquid size distributions (LSD), and the mass flow rate distributions (MFRD) at a location downstream.

Table 1. Variable Bounds

| | Lower limit | Upper limit |
|-------------------|-------------|-------------|
| R_1 | 0 | 1.2 |
| R_2 | 0 | 1 |
| θ_1 (deg.) | 0 | 90 |
| θ_2 (deg.) | 0 | θ_1 |

At this point, we can mention an additional benefit of GEBHM, namely its ability to accurately predict responses while given that input variables are correlated or confounded. In addition to the BHM model, global sensitivity analysis and independent validation was performed to ensure the validity of the model predictions.

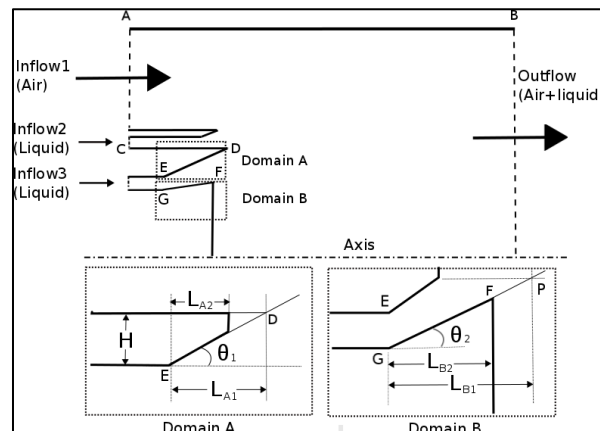


Figure 3. Computational Domain with emphasis on the region of study (A and B)

The geometric design is optimized via the following quantities of interest (QoIs): (1) LSD (2) LLD, and (3) MFRD. The median values and their 95% variation are used as optimization objectives for (1) and (2), whereas the variance of the mass flow rate is used for (3). The cumulative density function (cdf) of each of the QoIs was discretized at the following fractional probabilities: [0.1, 2.3, 15.9, 50, 84.1, 97.7, 99.9], see Figure 5.

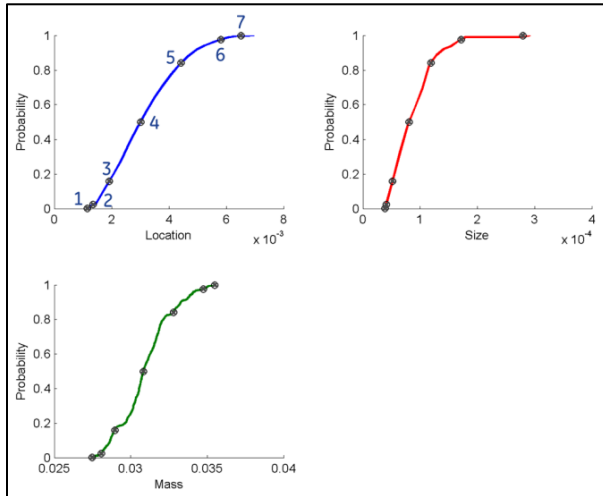


Figure 5. Example cdf's of the LSD, LLD, and MFRD

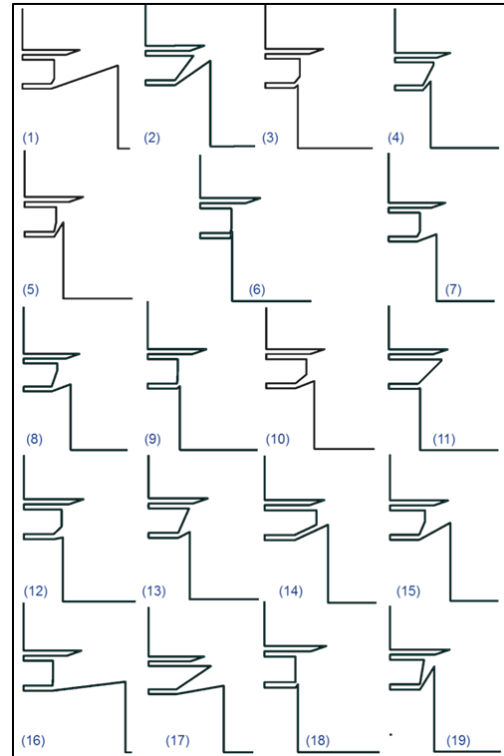


Figure 4. Corresponding Geometries of DOE Points

GEBHM was built for the cdfs of location, size and mass flow rate. The Bayesian framework captured the response of all three cdfs as a function of the four geometric parameters as shown below:

$$\begin{bmatrix} \{C_{loc}^i\} \\ \{C_{size}^i\} \\ \{C_m^i\} \end{bmatrix} = f(R_1, R_2, \theta_1, \theta_2)$$

where C_{loc}^i is the cdf location at a probability of p_i with $i=1,2\dots7$, C_{size}^i is the size at a probability of p_i and C_m^i is the mass flow rate at a probability of p_i . The model predictions are compared to the data for the location cdf's in Figure 6. The other cdf's are captured equally as well. From now on, only results for the location distribution are presented.

Next, an effort was created to identify the most significant geometric parameters and their interactions driving the location distribution. Variance-based global sensitivity analysis was performed within GEBHM. The parameters that affect the median of the location distribution are shown in Figure 7 in decreasing order of importance.

The contribution of the variance in location distribution median due to the interaction between R_2 and θ_2 is shown in Figure 8. The DOE data points are shown as white dots. Two regions with minimal uncertainty are identified (blue). GEBHM provides these uncertainty maps as a useful guide in determining the next set of simulations and/or experiments. The maximum uncertainty is low enough that the current models are deemed satisfactory to perform geometric optimization of the two-phase flow problem.

Optimization

Given that a satisfactory BHM model built in the previous sections, a next logical step is to employ it in an optimization setting to find the optimal geometry of the flow system. To this end, Particle Swarm Optimization (PSO) was used for global optimization (Eberhart and Kennedy, 1995; Bratton and Kennedy, 2007). PSO has been applied successfully to non-linear optimization problems. PSO has been shown to achieve near global optimum solutions like those computed using evolutionary algorithms like Genetic Algorithms (Golberg, 1989) at much faster convergence rates. PSO also requires fewer parameters that need adjusting.

The Bayesian models were used to optimize the geometry for several different objectives. The objectives for the various cases are listed in Table 5. The multiple objective cases 1 through 3 were the design optimization runs and the other cases were used for identifying the maximum achievable improvement for each of the objectives individually while disregarding the other objectives.

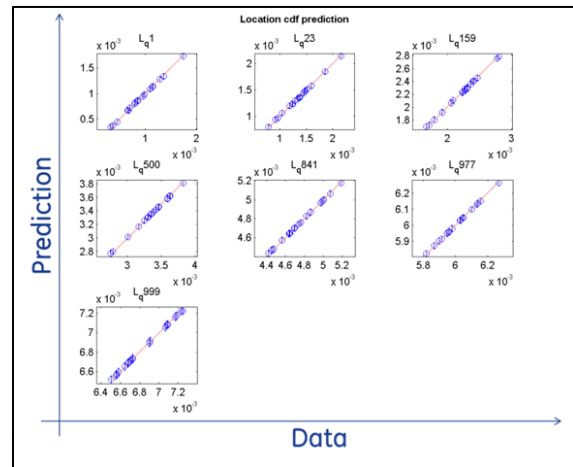


Figure 6. Comparison of Location cdf Predictions and Data

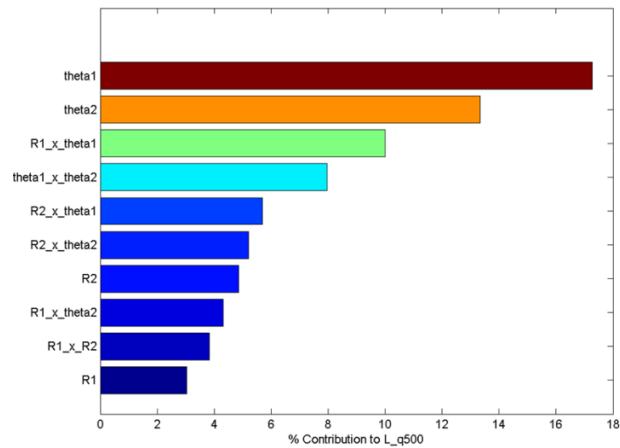


Figure 7. Global Sensitivity of Median Location to Input Parameters

Table 5. Optimization Cases

| Case | Objective |
|------|---|
| 1 | Maximize median location, minimize location 2 sigma, minimize Mass 2 sigma, minimize median size and minimize size 2 sigma |
| 2 | Maximize median location, minimize location 2 sigma, minimize Mass 2 sigma, fix median size at 75 and minimize size 2 sigma |
| 3 | Case 2 with higher weight for Mass 2 sigma |
| 4 | Maximize median location |
| 5 | Minimize Location 2 sigma |
| 6 | Minimize Mass 2 sigma |
| 7 | Minimize Size 2 sigma |
| 8 | Fix Size median to 75 and minimize Size 2 sigma |
| 9 | Fix Size median at 75 ± 3 and minimize Size 2 sigma |
| 10 | Fix Size median at 75 ± 1 and minimize Size 2 sigma |

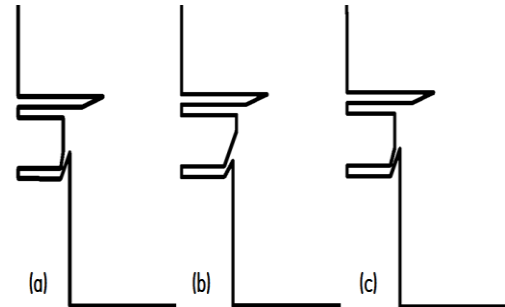


Figure 8. Optimized Geometries for (a) Case 1, (b) Case 2, and (c) Case 3

The optimum geometries resulting from the main optimization tasks, cases 1-3, are shown in Figure 8. The VOF for each of the optimized geometries are shown in Figure 9. Subtle differences in the geometry produce large variations in the response as seen by the instantaneous VOF plots (column 3). Once the optimal geometries were identified, they were simulated using CFD again to validate the BHM model predictions. In Figure 10, we show the comparison of the location distribution as predicted by BHM and the one produced from CFD. GEBHM accurately predicted the cdf's from the CFD run, further verifying the overall approach. Thus, we show that fast predictions from the trained GEBHM model can replace expensive CFD computations.

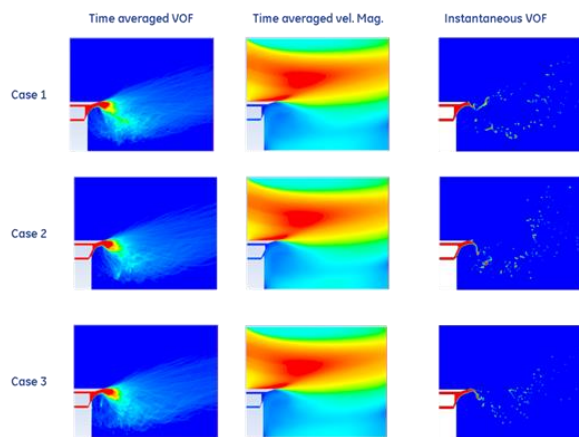


Figure 9. VOF for Optimized Geometries

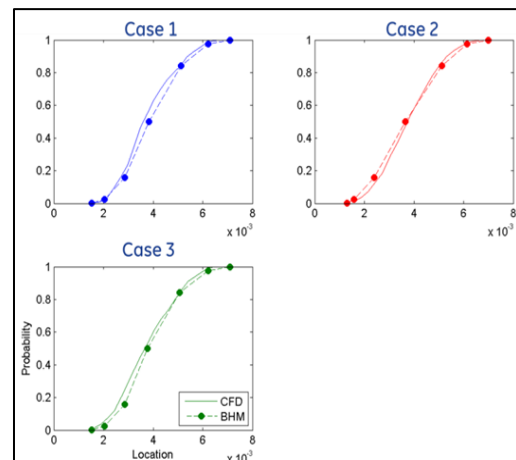


Figure 10. Comparison of predicted BHM location distributions to CFD

CONCLUSIONS

GE's BHM (GEBHM) technology was presented in detail through overview discussions, the theoretical framework was presented, and finally via a computationally expensive Computational Fluid Dynamics (CFD) application of optimizing the geometry of a two-phase flow system. Accurate Bayesian models for predicting the distribution of the liquid phase location, size, and mass flow rates were built at minimal computational cost. Global sensitivity analysis was performed with the BHM models. This helped to identify the most significant parameters and rank their relative significance. The BHM framework was used to optimize the geometry under various objectives when coupled with a global optimization technique. The visualization module of BHM helped guide the design engineer towards the optimum solution and provided further insights into the problem at hand. BHM could accurately predict the full distribution of droplet location, size, and mass flow rate accurately despite the sparse data available. The mass flow rates were also predicted accurately for cases with unimodal distributions. GEBHM was thus shown to reduce the computational cost of design process as well as accurately model nonlinear responses with sparse datasets

REFERENCES

- Kennedy, M. and O'Hagan, A. (2001). "Bayesian calibration of computer models", *Journal of the Royal Statistical Society (Series B)*, Vol. 68, 2001, pp 425–464.
- N. Kumar, A. K. Subramaniyan and L. Wang (2012). "Improving high-dimensional physics models through Bayesian calibration with uncertain data", 2012, GT2012-69058, ASME TurboExpo 2012, June 11-15, 2012, Copenhagen, Denmark.
- Subramaniyan, A. K., Wang, L., Beeson, D., Nelson, J., Berg, R., Ceyress, R. (2011). "A Comparative Study On Accuracy And Efficiency Of Metamodels For Large Industrial Datasets", *Proceedings of ASME Turbo Expo 2011*, Vancouver, Canada, June 6-10, 2011.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57: 97–109.
- Higdon, D. M., Lee, H. and Holloman, C. (2003). "Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems", *Bayesian Statistics 7*, edited by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, *Proceedings of the Seventh Valencia International Meeting*, Oxford University Press, pp. 181–197.
- MacKay, D. J. C. (1999). "Introduction to Gaussian Processes", Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, England, UK, <http://www.inference.phy.cam.ac.uk/mackay/gpB.pdf>
- Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J. and Ryne, R. D. (2004). "Combining field observations and simulations for calibration and prediction", *SIAM Journal of Scientific Computing*, Vol. 26, No.2, 2004, pp 448–466.
- Higdon, D., Gattiker, J., Williams, B. J., Rightley, M. (2008). "Computer Model Calibration using High Dimensional Output", *Journal of the American Statistical Association*, Vol. 103, No. 482, June 1, 2008, pp 570-583.
- Higdon, D., Nakhleh, C. and Williams, B. (2008). "A Bayesian Calibration Approach to the Thermal Problem", *Computer Methods in Applied Mechanics and Engineering* 197, 2431-2441.
- Miki, K., Subramaniyan, A., Pai, M., & Balasubramanyam, P. (2014, June). Probabilistic Optimization of Two-Phase Flow Using Bayesian Models. In *ASME Turbo Expo 2014: Turbine Technical Conference and Exposition* (pp. V02BT45A018-V02BT45A018). American Society of Mechanical Engineers.
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7), 964-979.
- Golberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. *Addion wesley*, 1989, 102.
- Viana, F. A., Venter, G., & Balabanov, V. (2010). An algorithm for fast optimal Latin hypercube design of experiments. *International journal for numerical methods in engineering*, 82(2), 135-156.
- Eberhart, R. C. and Kennedy, J. (1995). "A new optimizer using particle swarm theory". *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan. pp. 39-43.
- Bratton, D.; Kennedy, J. (2007). "Defining a Standard for Particle Swarm Optimization," *Swarm Intelligence Symposium*, 2007. SIS 2007. IEEE, vol., no., pp.120,127, 1-5 April 2007.