INL/CON-19-56281-Revision-0

DEMONSTRATION OF A DATA-DRIVEN PHYSICS-BASED APPROACH FOR COMPUTATIONALLY EFFICIENT CFD PREDICTION OF TWO-PHASE BUBBLY FLOW

Han Bao, jinyong feng, Nam Dinh, Hongbin Zhang

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/CON-19-56281-Revision-0

DEMONSTRATION OF A DATA-DRIVEN PHYSICS-BASED APPROACH FOR COMPUTATIONALLY EFFICIENT CFD PREDICTION OF TWO-PHASE BUBBLY FLOW

Han Bao, jinyong feng, Nam Dinh, Hongbin Zhang

March 2020

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

DEMONSTRATION OF A DATA-DRIVEN PHYSICS-BASED APPROACH FOR COMPUTATIONALLY EFFICIENT CFD PREDICTION OF TWO-PHASE BUBBLY FLOW

Han Bao
Idaho National Laboratory
P.O. Box 1625, MS 3860, Idaho Falls, ID 83415
han.bao@inl.gov

Jinyong Feng
Department of Nuclear Science and Engineering
Massachusetts Institute of Technology
Cambridge, 02139, Massachusetts, USA
fengjinyong2008@gmail.com

Nam Dinh
Department of Nuclear Engineering
North Carolina State University
ntdinh@ncsu.edu

Hongbin Zhang
Idaho National Laboratory
P.O. Box 1625, MS 3870, Idaho Falls, ID 83415
hongbin.zhang@inl.gov

ABSTRACT

To realize efficient computational fluid dynamics (CFD) prediction of two-phase flow, a multi-scale physics-guided data-driven approach, Feature Similarity Measurement (FSM) technique was developed for error estimation in two-phase flow simulation using coarse-mesh CFD, to achieve a comparable accuracy as fine-mesh simulations with fast-running feature. By defining physics-guided parameters and variable gradients as physical features, FSM has the capability to capture the underlying local patterns in the coarse-mesh CFD simulation. Massive low-fidelity data and respective high-fidelity data are used to explore the underlying information relevant to the main simulation errors and the effects of phenomenological scaling. By learning from previous simulation data, a surrogate model using deep feedforward neural network (DFNN) can be developed and trained to estimate the simulation error of coarse-mesh CFD. In a demonstration case of two-phase bubbly flow, the DFNN model well captured and corrected the unphysical "peaks" in the velocity and void fraction profiles near the wall in the coarse-mesh configuration, even for extrapolative predictions. The research documented supports the feasibility of the physics-guided deep learning methods for coarse mesh CFD simulations which has a potential for the efficient industrial design.

KEYWORDS
Data-driven, physics-based, two-phase bubbly flow, coarse-mesh CFD, physical feature

1. INTRODUCTION

Owing to the advancement of high-performance computing and computational methods, modeling and numerical simulations have become instrumental in the design, analysis and licensing of nuclear power plants. Compared to system codes using lumped-parameter models, computational fluid dynamics (CFD) methods have been widely used for solving transport equations of fluid mechanics by using local instantaneous formulations with finer mesh sizes, where small-scale flow features could be captured. While CFD has the potential to accurately predict the flow behavior and reduce the need for dedicated reactor-operational experiments, it suffers from two key challenges for the system analysis of NPP behaviors, namely high computational costs and limited understanding on error sources of simulation.

The main limitation of applying CFD methods to practical industrial applications is the computational cost. Since discretizing the temporal and spatial space on a much smaller scale, CFD simulations require many more cells than a system thermal-hydraulic simulation. One of the most representative examples is direct numerical simulation (DNS) method. As a first principle-based method, DNS directly solves the Navier-Stokes equations without any closure models, thus making it serve as high fidelity benchmark data, especially in the single-phase study. In the state-of-the-art two-phase DNS simulation [1] in order to resolve each individual bubble and turbulent eddies down to the smallest turbulent length scale, i.e., Kolmogorov scale, it requires 1.10 billion cells and ~730,000 core-hours to simulate the reactor subchannel with hydraulics Reynolds number of 80,000 whereas the hydraulics Reynolds number under reactor operational conditions is ~500,000. To bypass the computational cost of the fully resolved high Reynolds number case, researchers either conducted separate effect studies with well-controlled flow conditions [2–4] to develop individual closures, or adopt computationally efficient Reynolds-averaging Navier-Stokes method.

The last limitation is the limited understanding on error sources of CFD simulations. There are two major error sources in the CFD applications: physical model error and mesh-induced numerical error. Physical model error arises from physical assumptions and mathematical approximations in the form of the model equations (misrepresentation of the underlying physical system) and errors in assigning values for model parameters or calibration coefficients. Mesh-induced numerical error comes from the solution discretization in space and time, and the approximation used in over-cell integration of the variables that are non-uniformly distributed within the cell. While CFD simulations are advantageous for its high fidelity and resolution, these benefits have largely not been realized in reactor safety applications due to the lack of established and agreed upon approaches for assessing those error sources in CFD simulations. Two mainstream research directions of error estimation in CFD simulations are numerical error estimation, which are translated as the mesh error or discretization scheme error [5, 6], and model error estimation where the model error estimations are usually performed by varying the model coefficients [7–10] to justify the solution range. In two-phase flow CFD simulation, two main error sources are both tightly connected with local mesh sizes, which makes it difficult to analyze them separately. Furthermore, some of the applied models in CFD codes are not scalable for extrapolative predictions because of the lack of validation data during the development phase. The uncertainty stemming from using these models increases outside of their applicable ranges. Mesh configuration is crucial to the accurate prediction of the multiphase flow phenomena using CFD approaches. The calculation of flow variables gradients depends on the mesh resolution between two adjacent cells which are directly relevant to certain physical models.

To deal with these difficulties, this paper applied a data-driven approach, Feature Similarity Measurement (FSM), for error estimation in two-phase flow simulation using coarse-mesh CFD, to achieve a comparable accuracy as fine-mesh simulations with the featured fast-running capability. The approach was preliminarily developed by the authors before where results demonstrated good predictions on single phase flow [11] and mixed convection [12, 13] using GOTHIC as low-fidelity simulation tool. As a coarse-mesh thermal-hydraulic simulation tool, GOTHIC has been widely used for reactor safety analysis [14]. The performance of this proposed approach is also evaluated considering the interpolative and extrapolative predictions.

Computational cost for system-level thermal hydraulic modeling and simulation could be reduced by using coarse-mesh CFD, meanwhile, this data-driven approach treats model error and mesh-induced numerical error together by taking their tight connection with local mesh size into consideration. The deep learning model trained in FSM approach treats the physical correlations, coarse mesh sizes and numerical solvers as an integrated model, which can be considered as a surrogate of governing equations and closure correlations of coarse-mesh CFD.

In this paper, the data-driven approach is further developed and demonstrated on a two-phase bubbly flow case study. Technical background of the proposed approach described in Section 2 provides a guidance about how to explore local physics and bridge the global scale gap. Section 3 illustrates the workflow of applying FSM to realize computationally efficient CFD simulation, which is demonstrated using a bubbly flow case study in Section 4. Key findings are summarized in Section 5.

2. TECHNICAL BACKGROUND

There are two "ideal" approaches to explore and predict behaviors in prototype-scale applications: (1) prototype-scale experiment, which presumably duplicates a prototype-scale phenomenon existing in the real applications, (2) DNS modeling where the local information is solved accurately with very fine mesh. However, prototype-scale experiments are hard to build while many prototype-scale tests are required in the real applications, and DNS is computationally expensive to deal with the system-level predictions. Traditional CFD approaches, such as RANS, are treated as reduced-order models because their predictions of prototype-scale processes are made using models developed based on scaled experiments. However, these reduced-order CFD approaches are still not practical for wide use in system-level NPP analysis because of high computational cost compared with systems codes. The development of FSM enables CFD approaches with capability to realize the computationally efficient prediction by exploring local physical features instead of global characteristics.

Over the past few decades, many concepts of nuclear reactor have been proposed with different components, geometries, and powers. The respective global characteristics might be out of the domain of previous designs or simulations, which brings large uncertainty into the application. The relevant thermal-hydraulic experiments with a wide range of scale and structure must be designed and implemented for code validation and licensing of new reactor designs. The extrapolation of global characteristics, e.g., dimension, geometry/structure, initial conditions/boundary conditions (ICs/BCs), or powers, may limit the applicability of the previously developed models or experiments. However, local physics such as the interaction between liquid, vapor and heat structure may not change significantly: quantities of interest (temperature, velocities and vapor fraction) remain approximately the same although their different values lead to different flow patterns [13]. This makes it possible that some well-defined local physical features in the local cells are similar even if the global characteristics vary significantly.

According to the similarity or coverage of global characteristics and local physical features, four different physics coverage conditions (PCCs) are classified: global interpolation through local interpolation (GILI), global interpolation through local extrapolation (GILE), global extrapolation through local interpolation (GELI) and global extrapolation through local extrapolation (GELE). The GELI condition refers to the situation where the global characteristics of target case are identified as an extrapolation of existing cases, but the local physical features of target case are similar to or mostly covered by the ones of existing cases. These well-defined local physical features are supposed to represent the specific underlying local physics. The interpolation or similarity of local physics between the target case and existing cases depends on the identification of physical features, data quality and quantity. The local similarity in GELI condition makes it feasible to derive benefits from the existing data to estimate the target case. Instead of endlessly evaluating the applicable ranges of models and scaling uncertainty in extrapolative predictions, exploring the similarity of local physics provides a potential to bridge the scale gap in global extrapolations.

Although some advanced codes have been widely used in system-level NPP analysis, the V&V of these codes still suffer from the lack of validation data. The application domain for new rector designs are always not met by the validation domain defined based on global characteristics, as shown in the left part of Figure 1. A major fraction of the validation domain belongs to GILI condition, the grounded physics coverage condition for code/model V&V, where the existing data or models has the capability to estimate the target case due to both the global and local similarities. For the GILE condition, even if global physical condition of the target case is covered by existing cases, data from existing cases is not able to predict the target case since local physics are different. For instance, the models developed from experiments of laminar flow or turbulent flow are not applicable for transition prediction, although the global Reynolds number is covered. In contrast, the GELI condition has a potential to be added into the validation domain once the similarity of local physical features can be well defined and determined. From the perspective of data analysis, the previous validation domain defined by global characteristics may be expanded if it is separated into several new validation domains re-classified by local physical features, as illustrated in the right part of Figure 1. Focusing on GELI condition, FSM has a potential to provide insights on the designs of experiments and numerical tests to enlarge the validation domain to reach the required application domain.

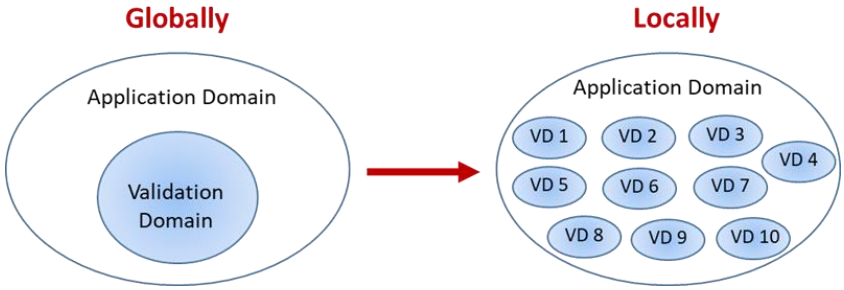


Figure 1. Exploring Similarity of Local Physical Features: A Way to Enlarge Validation Domain.

3. DESCRIPTION OF FEATURE SIMILARITY MEASUREMENT

The proposed data-driven approach, Feature Similarity Measurement (FSM), was developed to identify the local physical features, measure the data similarity of defined physical features between training data and target data, and investigate the relationship between physical feature similarity and accuracy of machine learning prediction in GELI condition. There are some basic requirements to apply FSM on CFD modeling and simulation: (1) coarse-mesh CFD simulation can generate reasonable results which capture the basic behaviors of targeted phenomena within an acceptable uncertainty; (2) simulation error is mainly impacted by model error and mesh-induced numerical error where mesh size is one of key model parameters that makes two main error sources tightly connected. (3) training data is qualified (i.e., performance of physical feature) and sufficiently efficient (i.e., size of relevant data for training) for machine learning to learn from and find the underlying patterns of identified local physical features. The basic idea of FSM is to develop a surrogate model to identify the relationship between simulation error and specific local physical features, as shown in Figure 2. These local physical features are identified based on the system information (e.g., IC/BC, geometry, structure), closure models that contain the information of phenomena of interest and relevant to model error, and local mesh sizes that affect the model error and mesh-induced numerical error. Since the values of physical features are not only determined by mesh sizes but also other physical parameters, the gaps between simulations with different mesh sizes are reduced, which makes it possible to use this well-trained surrogate model to predict the extrapolation of local mesh sizes and use fine-mesh simulation to inform coarse-mesh simulation.

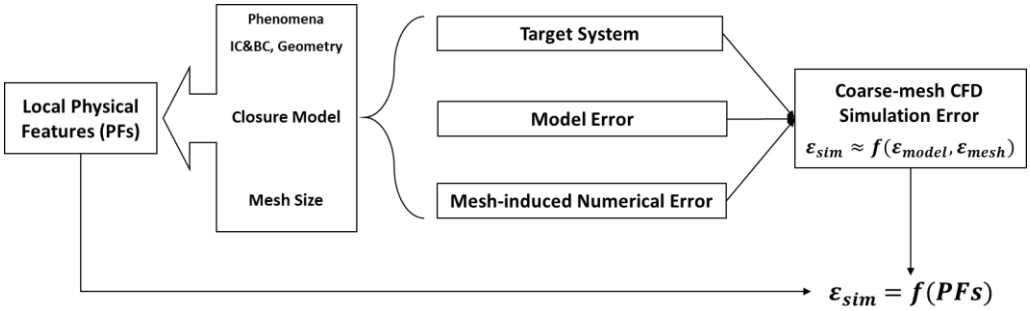


Figure 2. Basic idea of FSM: surrogate modeling of local simulation error and physical features.

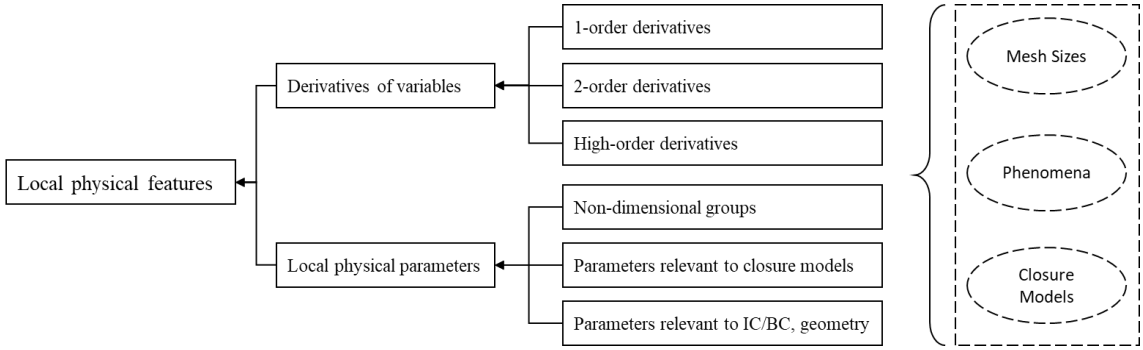


Figure 3. Classification of Local Physical Features.

FSM includes two basic parts: establishing and enhancing local similarity. The local similarity is established by defining a group of local physical features and constructing data warehouse. Local physical features are defined based on the phenomena, ICs/BCs and geometry of target system, relevant closure models, and local mesh sizes ensuring local physics to be well represented. The local physical features are classified into two types: derivatives of variables that indicate regional information and local physical parameters as shown in Figure 3. The derivatives of variables include 1-order, 2-order and high-order derivatives of variables calculated by central-difference formulas, which not only contain the information in local cell but also the information that represents regional physical patterns and connection with adjacent cells. It is analogous to the identification of a person, not only his/her personal information such as height and weight is important, but also his/her connections with other people should be considered. More detailed regional information may be involved if higher-order derivatives are added into the local physical feature group. All available high-fidelity data that is relevant to the phenomena involved in the target case should be collected and processed to build the data warehouse, which includes experimental observation, DNS data, and validated high-resolution numerical results. The definition of "high-fidelity" depends on the requirements of simulation accuracy for the target case. According to the physical conditions of limited high-fidelity data, low-fidelity data can be collected or generated using CFD codes with coarse mesh sizes and closure models.

To ensure the target case is similar enough to training data as a GELI condition, some means can be applied to enhance the established local similarity. One way to enhance local similarity is to select the "closest" training data points for the prediction of target case. It aims to answer the question: which kind of data in the data warehouse should be used as the training database? Training database should be sufficient to capture the local physics in the target case, and efficient to avoid huge computational cost on data training and processing. Therefore, another question needs to be answered: how to determine the similarity of target data and training data? One of the claims is that if target data is more like training data, machine learning

prediction error on the target case is smaller. Considering these target-similar data points distribute in different cases in training data warehouse, the optimal training database should be constructed by choosing these target-similar data points in one case instead of choosing the entire case.

4. DEMONSTRATION OF FEATURE SIMILARITY MEASUREMENT ON BUBBLY FLOW

4.1. Problem Statement

In this paper, a case study based on two-phase bubbly flow was performed to evaluate the predictive capability of FSM in two-phase flow coarse-mesh CFD simulation which adopts the Eulerian-Eulerian two-fluid model as discussed in the Appendix. It is upward bubbly pipe flow experiments for water at atmospheric pressure and temperature of 10 ^{0}C without phase change. 42 reference experimental datasets with different injection rates and void fractions are from Liu and Bankoff [15, 16] and used for the validation of high-fidelity simulation data. Physical experiment exhibit uncertainties in the measurement of velocity and void fraction and quantitative analysis of those uncertainties will be performed in future work. Table I lists the global injection conditions of the 42 experimental cases.

Table I. Summary of the flow conditions of experimental cases

Set	Liquid rate u_l (m/s)	Vapor rate u_g (m/s)	Void fraction α	Set	Liquid rate u_l (m/s)	Vapor rate u_g (m/s)	Void fraction α
1	0.376	0.027	0.0407	**22**	0.974	0.027	0.0204
2	0.376	0.067	0.1167	**23**	0.974	0.067	0.0514
3	0.376	0.112	0.1843	**24**	0.974	0.112	0.0791
4	0.376	0.18	0.2449	**25**	0.974	0.18	0.1242
5	0.376	0.23	0.3079	**26**	0.974	0.23	0.1512
6	0.376	0.293	0.3657	**27**	0.974	0.293	0.1869
7	0.376	0.347	0.4168	**28**	0.974	0.347	0.2108
8	0.535	0.027	0.0312	**29**	1.087	0.027	0.0176
9	0.535	0.067	0.0877	**30**	1.087	0.067	0.0473
10	0.535	0.112	0.1406	**31**	1.087	0.112	0.0737
11	0.535	0.18	0.2016	**32**	1.087	0.18	0.1096
12	0.535	0.23	0.2344	**33**	1.087	0.23	0.1497
13	0.535	0.293	0.3102	**34**	1.087	0.293	0.1777
14	0.535	0.347	0.3398	**35**	1.087	0.347	0.1976
15	0.753	0.027	0.0235	**36**	1.391	0.027	0.0148
16	0.753	0.067	0.0622	**37**	1.391	0.067	0.0387
17	0.753	0.112	0.1091	**38**	1.391	0.112	0.0581
18	0.753	0.18	0.1554	**39**	1.391	0.18	0.0964
19	0.753	0.23	0.1816	**40**	1.391	0.23	0.1176
20	0.753	0.293	0.2381	**41**	1.391	0.293	0.1504
21	0.753	0.347	0.2692	**42**	1.391	0.347	0.1724

Both high-fidelity and low-fidelity simulations were performed using the commercial CFD package, STAR-CCM+12.06, with the following two-phase interfacial forces closures and turbulence models: drag force

model from Tomiyama [17], lift correction from Shaver and Podowski [18] with a base coefficient of 0.025, turbulent dispersion force from Burns [19] and the standard $k-\varepsilon$ turbulence model. The set of these models which is referred as the Bubbly and Moderate Void Fraction (BAMF) model has been tested and validated for 12 cases from the Liu and Bankoff experimental datasets, which provided reasonable predictions for mean flow profiles of void fraction and phase velocities [20]. In this paper, all the experimental datasets are simulated with the BAMF model to gain sufficient database. To accelerate the simulation, as shown in Figure 4, only one-quarter of the domain is simulated, and symmetric boundary conditions are applied on the two side surfaces. For each of the 42 experimental cases, four sets of mesh configurations are selected and the cross sectional view of the mesh configurations are shown in Figure 4, where the mesh configuration of the reference high-fidelity case is identical to the cases reported in [20] and the other three low-fidelity cases adopt coarse mesh resolutions to analyze the integration of mesh-induced numerical and physical model error. The numbers of cells from the wall to the pipe center are 10, 15, 20 and 25, thus having the total number of cells in the domain as 0.07, 0.18, 0.35, and 0.63 million. The application of FSM on bubbly flow coarse-mesh CFD simulation is described in this section. Of all the 42 cases, Case 35 can be considered as a global extrapolation of large u_g for other 41 cases, and Case 42 is a global extrapolation of both large u_g and u_l. Therefore, the goal of this case study is to predict the 10-cell coarse-mesh CFD simulation errors of Case 35 and 42 using part of other 41 cases for data training.

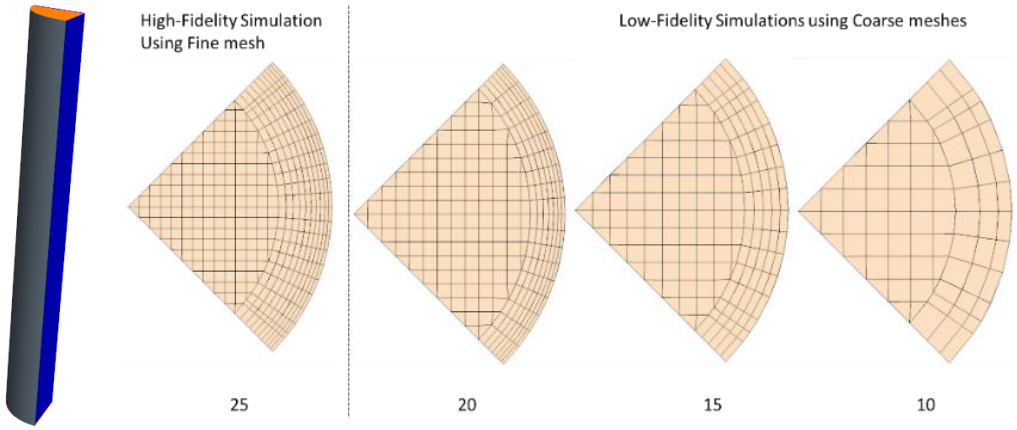


Figure 4. Cross-sectional view of the mesh configuration for high and low fidelity simulations.

4.2. Local Similarity Establishment

4.2.1. Identification of Local Physical Features

According to the involved phenomena, applied closure models, IC/BC and geometry information, 27 potential local physical features are defined as shown in Table II. 16 of all are 1-order and 2-order derivatives of variables, which include liquid and vapor velocity (u_l and u_g), void fraction (α), pressure (P), liquid and vapor kinetic energy (k_l and k_g), liquid and vapor turbulence dissipation rate (ε_l and ε_g). A set of non-dimensional parameters which are adopted and customized to characterize the flow features is included as well. The principle is to include as many relevant non-dimensional parameters as possible while the importance of each parameter will be justified by the machine learning algorithms. Defined as the ratio between the inertial forces to viscous forces, in this work, Reynolds number is customized to have three different expressions representing local flow features. Three different local Reynolds number Re_Δ, Re_b, and Re_y respectively use local mesh size (Δ), pre-set bubble size (D_b) and wall distance (y) as characteristic

lengths, which take the effects of mesh, closure model and geometry into consideration. Weber number characterizes the relative importance of the fluid's inertia compared to its surface tension. Turbulent intensity (I_l and I_g) provides a measurement of the flow fluctuations versus the mean flow velocity. R_l and R_g are defined as the ratio between turbulent length scale, i.e., $\frac{k^{\frac{3}{2}}}{\varepsilon}$, and the bubble diameter, D_b. Non-dimensional wall distance, R_b, is included as well since the velocity and void fraction distributions crucially depend on the wall distance. r_l represents the ratio of liquid turbulent eddy viscosity and liquid molecular viscosity, which is supposed to become important for the high Reynolds number regime. R_μ represents the ratio between the gas and liquid eddy viscosity which characterizes the magnitude of modeled turbulence level for liquid and gas.

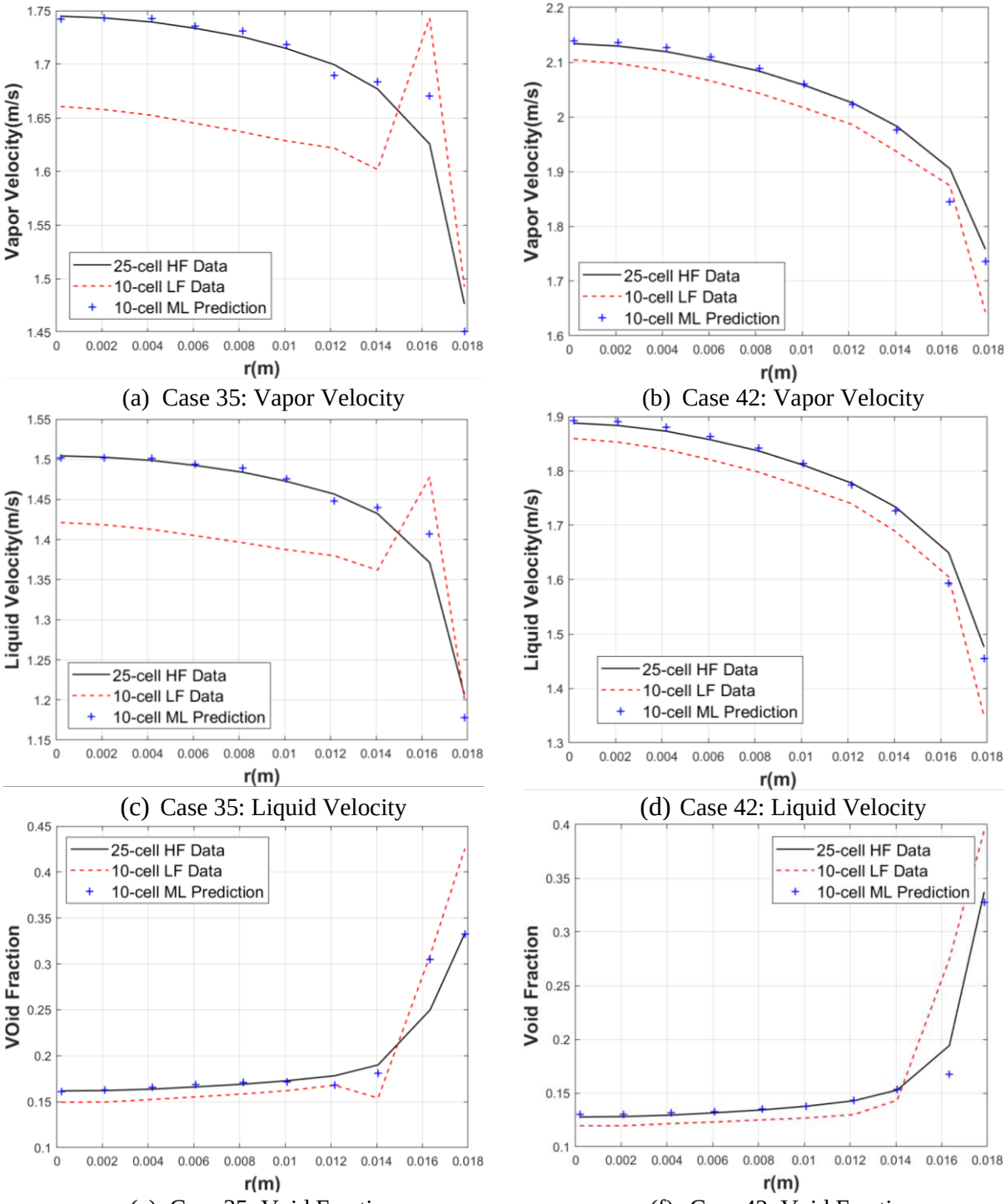
Table II. Identification of physical feature for bubbly flow CFD simulation

<table>
<tr><th colspan="4">Derivatives of variable</th><th colspan="4">Local physical parameters</th></tr>
<tr><th colspan="2">1-order derivatives</th><th colspan="2">2-order derivatives</th><th colspan="2">Non-dimensional groups</th><th colspan="2">Parameters relevant to closure models, IC/BC, geometry</th></tr>
<tr><td>$\frac{du_l}{dx}$</td><td>$\frac{du_g}{dx}$</td><td>$\frac{d^2u_l}{dx^2}$</td><td>$\frac{d^2u_g}{dx^2}$</td><td>$Re_\Delta = \frac{\rho_l \Delta \cdot \Delta u}{\mu_l}$</td><td>$I_l = \frac{k_l}{{u_l}^2}$</td><td>$R_l = \frac{{k_l}^{\frac{3}{2}}}{\varepsilon_l D_b}$</td><td>$R_\mu = \frac{\mu_g^t}{\mu_L^t}$</td></tr>
<tr><td>$\frac{d\alpha}{dx}$</td><td>$\frac{dP}{dx}$</td><td>$\frac{d^2\alpha}{dx^2}$</td><td>$\frac{d^2P}{dx^2}$</td><td>$Re_b = \frac{\rho_l D_b \Delta u}{\mu_l}$</td><td>$I_g = \frac{k_g}{{u_g}^2}$</td><td>$R_g = \frac{{k_g}^{\frac{3}{2}}}{\varepsilon_g D_b}$</td><td>$r_l = \frac{\mu_l^t}{\mu_l}$</td></tr>
<tr><td>$\frac{dk_l}{dx}$</td><td>$\frac{dk_g}{dx}$</td><td>$\frac{d^2k_l}{dx^2}$</td><td>$\frac{d^2k_g}{dx^2}$</td><td colspan="2" rowspan="2">$We = \frac{\rho D_b \Delta u^2}{\sigma}$</td><td rowspan="2">$Re_y = \frac{\rho_l y \Delta u}{\mu_l}$</td><td rowspan="2">$R_b = \frac{D_b}{\Delta}$</td></tr>
<tr><td>$\frac{d\varepsilon_l}{dx}$</td><td>$\frac{d\varepsilon_g}{dx}$</td><td>$\frac{d^2\varepsilon_l}{dx^2}$</td><td>$\frac{d^2\varepsilon_g}{dx^2}$</td></tr>
</table>

4.2.2. Construction of Data Warehouse for Deep Learning Training and Prediction

Using the BAMF model, high-fidelity (10, 15, 20-cell) and low-fidelity data (25-cell) were generated by STAR-CCM+12.06 with fine mesh and three different coarse meshes respectively. The point-to-point method is applied to calculate the simulation errors of local Quantities of Interest (QoIs, u_l, u_g, α) in this case study. With 27 inputs and 3 outputs, a deep feedforward neural network (DFNN) containing 3 hidden layers and 20 neurons in each hidden layer (i.e., 20-20-20) is applied for data training and simulation error prediction. The DFNN predicted results have smaller NRMSEs than the original low-fidelity simulation results for all the QoIs. After being corrected by the predicted simulation errors from well-trained DFNN model, the accuracy of coarse-mesh simulation results is improved. Figure 5 shows the comparisons between original low-fidelity simulation results and modified results based on the DFNN simulation error prediction for Case 35 and 42. The original coarse-mesh low-fidelity results using 10-cell configuration show different patterns with the high-fidelity data, there are unphysical "peaks" locating at the ninth point. For the coarse mesh (10-cell) case, the unphysical peaks are attributed to the overpredicted high void fraction on the near wall cell. After the training, the DFNN model well captures the pattern of these unexpected "peaks" and provides an appropriate correction to match the high-fidelity results for Case 35. The capability of capturing reginal patterns results from identifying 1-order and 2-order derivatives of QoIs as the physical features, because not only the characteristics at this point are captured, but also the connections of this point with its neighboring points. The results indicate that these defined local physical

features can represent local physics and provide sufficiently accurate prediction on simulation error of QoIs. FSM represents good predictive capability on estimating the local simulation error even for the extrapolation of global physics.



(a) Case 35: Vapor Velocity

(b) Case 42: Vapor Velocity

(c) Case 35: Liquid Velocity

(d) Case 42: Liquid Velocity

(e) Case 35: Void Fraction

(f) Case 42: Void Fraction

Figure 5. Comparisons between Original 10-cell Low-fidelity (LF) CFD Simulation Results and DFNN Predictions with 25-cell High-fidelity (HF) CFD Simulation Results for Case 35 and 42.

4.3. Local Similarity Enhancement

To ensure the quantity of training database is sufficient, 100 data points with small values of data distance are selected for each target data point. Therefore, there are 1000 data points used to build the training database of a 10-cell coarse-mesh configuration. Provided that the number (Q) of target data points is much smaller than the number (M) of data points in the data warehouse, the target-similar data points can be determined by measuring their data distance with each target data point. After standardizing values of physical features into [-1, 1], Euclidean distance is used as a metric to calculate the distance between single data points as expressed as below,

$$D_{m,i,q} = d\left(PF_m, PF_q\right) = \sqrt{\sum_{k=1}^{N_i}\left(x_{tr,k} - x_{ta,k}\right)^2} \tag{1}$$

where $D_{m,i,q}$ is the Euclidean distance between training data point PF_m and target data point PF_q for QoI_i, $(1 \le m \le M \; and \; 1 \le q \le Q)$. $\boldsymbol{N_i}$ is the number of optimal physical feature for each QoI, $x_{tr,k}$ and $x_{ta,k}$ are respectively the values of physical feature # k of PF_m and PF_q. Therefore, the optimal training database can be constructed by including these target-similar data points. Here Case 35 is used as the target case. The results in Table III show that the 10-cell and 15-cell data points have higher level of similarity than 20-cell data points since nearly all selected data points come from 10-cell and 15-cell simulations, even there are more data points from 20-cell simulations. More than half of the cases are involved in the training data selection, even the global conditions (e.g., u_g, u_l and α) of some cases are quite different from those of Case 35.

The data similarity between training data and target case can be measured using data similarity measure metrics. In this paper, a method called Kernel Density Estimation (KDE) is introduced to develop a measure metric. As a non-parametric way to estimate the probability density function, KDE assumes the data distribution can be approximated as a sum of multivariate Gaussians. A kernel distribution can be used if a parametric distribution cannot properly describe the data, or to avoid making assumptions about the distribution of the data. KDE can be considered as the probability that the data point ($\boldsymbol{q}$) locates in the distribution of training data ($\boldsymbol{p_i}, i = 1,2, \dots, N$). In this step, the similarity between training data ($\boldsymbol{p_i}, i = 1,2, \dots, N$) and target data ($\boldsymbol{q}_j, k = 1,2, \dots, M$) is expressed as the mean of KDEs below,

$$S_{KDE} = \frac{1}{M}\sum_{k=1}^{M} p_{KDE,k} = \sum_{k=1}^{M} \frac{1}{N \cdot h_1 h_2 \dots h_d} \sum_{i=1}^{N} \prod_{j=1}^{d} ker\left(\frac{q_{j,k} - p_{i,j}}{h_j}\right) \tag{2}$$

where d is the number of variables in $\boldsymbol{q}$ and $\boldsymbol{p_i}$. ker is the kernel smoothing function. h_j is the bandwidth for each variable. A multivariate kernel distribution is defined by a smoothing function (k) and a bandwidth matrix defined by $H = h_1, h_2, \dots, h_d$, which control the smoothness of the resulting density curve. Therefore, KDE method can be used to measure the data similarity by estimating the probability of a given point locating in a set of training data points.

A greater value of S_{KDE} means higher level of similarity. The KDE similarity values of Case 35 with its optimal training database also are listed in Table III. The distribution of selected data points from other 41 cases based on the data similarity with Case 35 is shown in Figure 6. Then the error prediction of local QoIs can be performed for Case 35 by using the optimal training database (Table III). By adding the DFNN

predicted simulation errors, the modified coarse-mesh CFD simulation results are compared with the high-fidelity results as shown in Figure 7. Informed by previous low-fidelity and high-fidelity simulation results, the simulation errors of a new extrapolative case using coarse-mesh CFD can be well predicted using the proposed data-driven approach (Feature Similarity Measurement). DFNN prediction errors for Case 35 with all data points for training and optimal training database are compared in Table IV.

Table III. Summary of optimal training database for 10-cell simulation error prediction

Target Case	Quantity of training data points (no repeat)				Data similarity (S_{KDE})	
	10-cell	15-cell	20-cell	Total	Before selection	After selection
Case 35	263	233	19	515	0.1792	0.2861

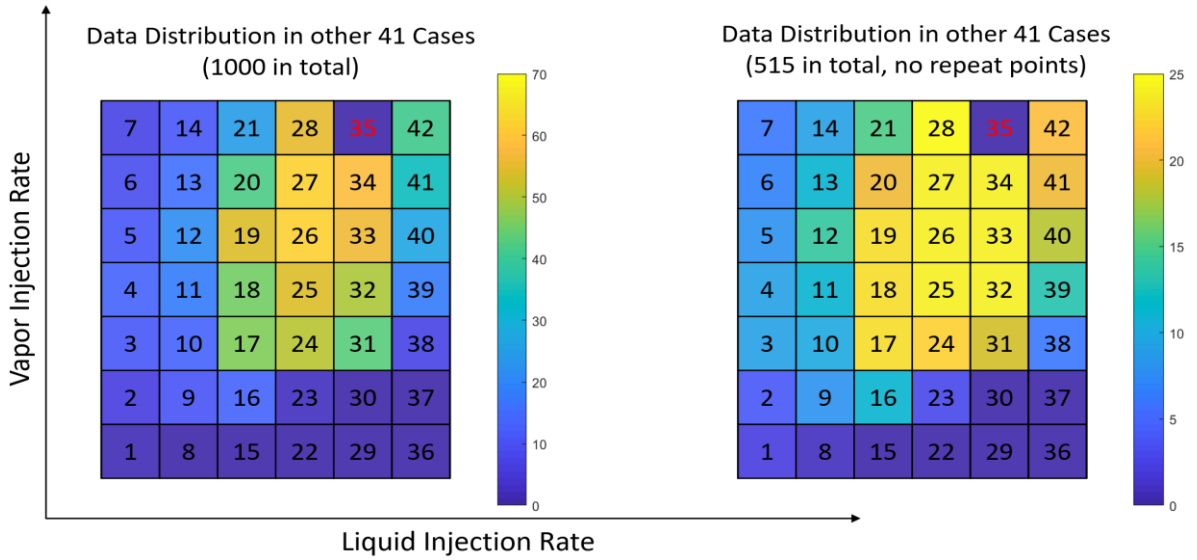
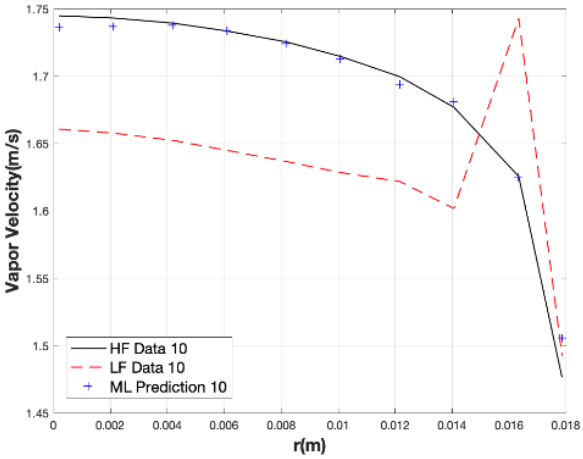


Figure 6. Distribution of Selected Data from Other 41 Cases based on Data Similarity with Case 35

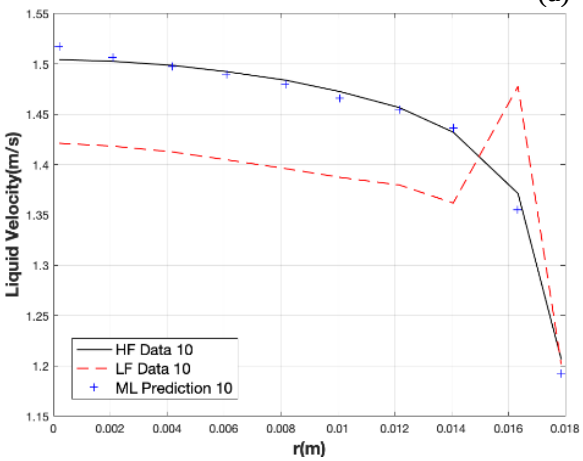
Table IV. DFNN predictions with optimal physical features and training data

Target case		$NRMSE_{u_g}$	$NRMSE_{u_l}$	$NRMSE_{\alpha}$	Global Condition
Case 35	DFNN prediction with all training data	0.0101	0.0105	0.0914	Extrapolation of Large u_g
	DFNN prediction with optimal training data	0.0057	0.0059	0.0158	
	Low-fidelity simulation	0.0499	0.0564	0.1934	

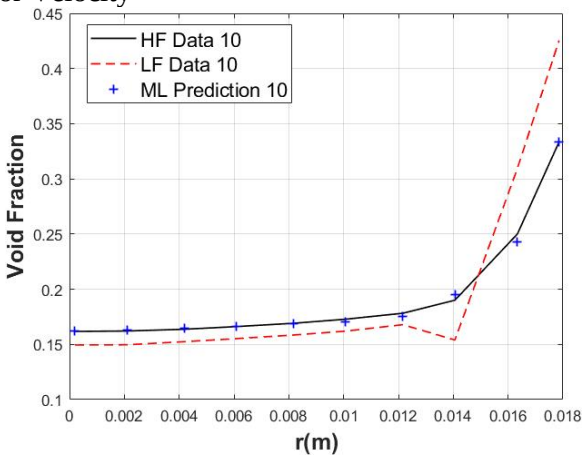
FSM was also applied for the predictions of Case 34 (global interpolation for other 41 cases) and Case 42 (global extrapolation of large u_g and u_l for other 41 cases), DFNN predictions are compared with original low-fidelity simulation results and high-fidelity simulation results in Figure 8. It shows that FSM performed well for both global interpolative and extrapolative conditions.



(a) Vapor Velocity



(b) Liquid Velocity

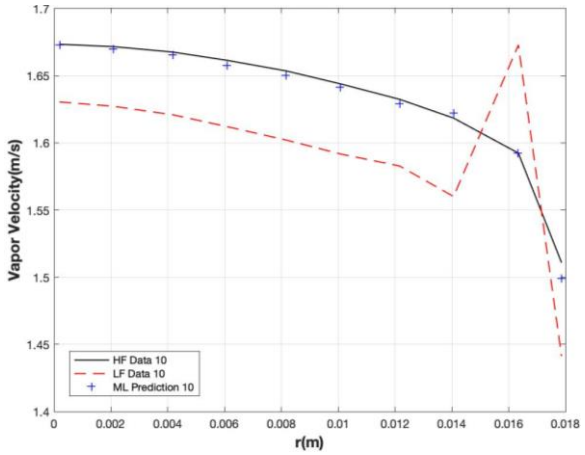


(c) Void Fraction

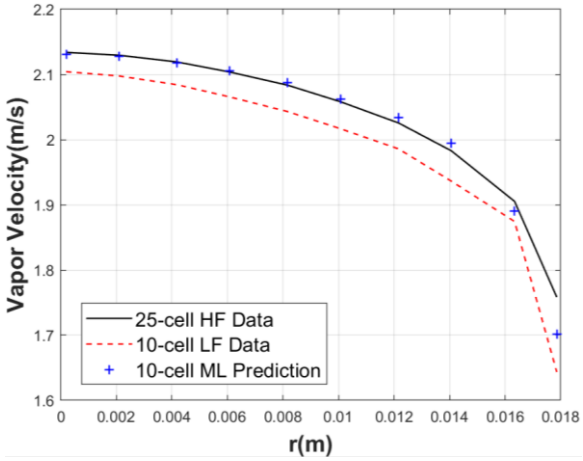
Figure 7. Comparisons of 25-cell High-fidelity (HF) Simulation, Original 10-cell Low-fidelity (LF) Simulation and DFNN Predictions with Optimal Training Database for Case 35.

5. CONCLUSIONS

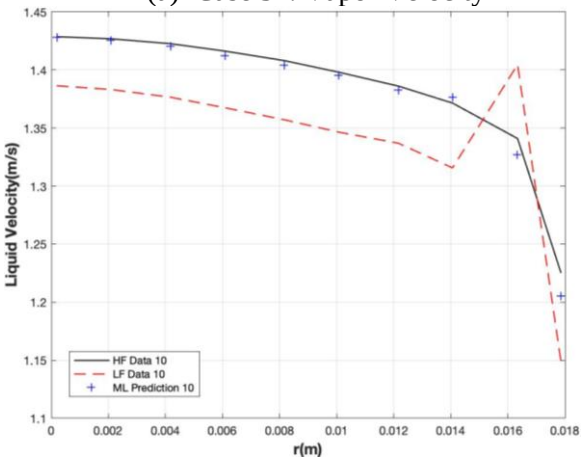
The results of the case study show that Feature Similarity Measurement (FSM) has the capability to estimate coarse-mesh CFD simulation errors. The CFD simulation with 10-cell configuration only has a total number of cells in the domain as 0.07 million while the 25-cell configuration validated as high-fidelity simulation has 0.63 million cells. Compared with the fine-mesh high-fidelity simulation, coarse-mesh simulation with FSM correction has a comparable accuracy and affordable computational cost. For the 10-cell coarse mesh CFD case, the total core-hours including simulation time and training cost are only ~6.6% of those of the high fidelity 25-cell configuration case. Even the unphysical "peaks" near the wall in the 10-cell configuration are well captured and corrected by the well-trained DFNN model. The capability of capturing reginal patterns results from identifying 1-order and 2-order derivatives of QoIs as the physical features, because not only the characteristics at this point are captured, but also the connections of this point with its neighboring points. Another part of physical features defined as non-dimensional local physical parameters enable FSM the capability for extrapolative predictions.



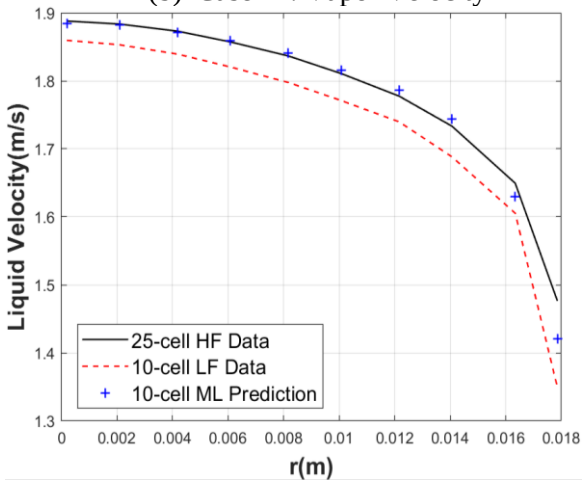
(a) Case 34: Vapor Velocity



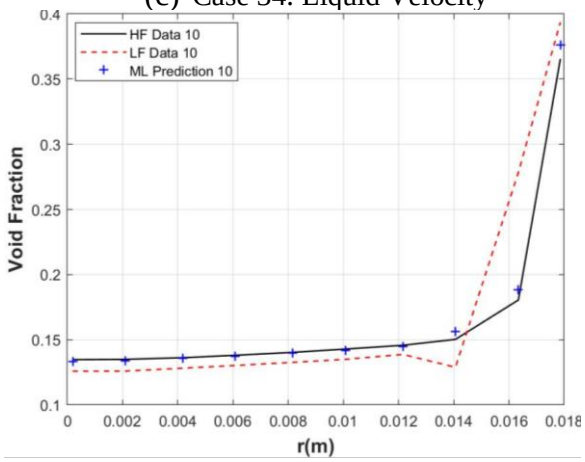
(b) Case 42: Vapor Velocity



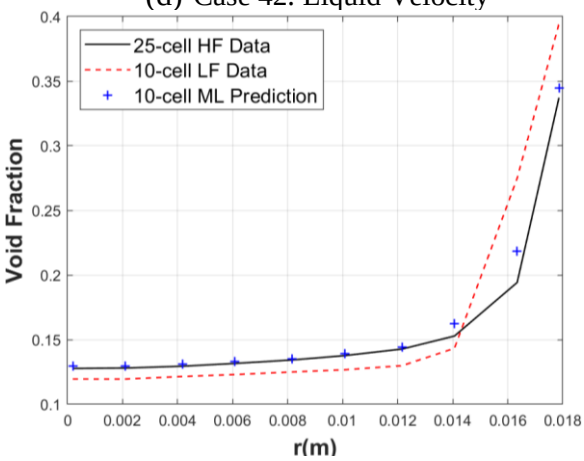
(c) Case 34: Liquid Velocity



(d) Case 42: Liquid Velocity



(e) Case 34: Void Fraction



(f) Case 42: Void Fraction

Figure 8. Comparisons of 25-cell High-fidelity (HF) Simulation, Original 10-cell Low-fidelity (LF) Simulation and DFNN Predictions with Optimal Training Database for Case 34 and Case 42.

ACKNOWLEDGMENTS

This work is supported by the U.S. Department of Energy, under Department of Energy Idaho Operations Office Contract DE-AC07-05ID14517. Accordingly, the U.S. Government retains a nonexclusive, royalty-

free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. The authors also would like to acknowledge the support from Baglietto CFD Research Group at MIT.

REFERENCES

1. J. Fang, J. J. Cambareri, C. S. Brown, J. Feng, A. Gouws, M. Li, I. A. Bolotnov, Direct numerical simulation of reactor two-phase flows enabled by high-performance computing. *Nucl. Eng. Des.* **330**, 409–419 (2018).
2. B. Bunner, G. Tryggvason, Effect of bubble deformation on the properties of bubbly flows. *J. Fluid Mech.* **495**, 77–118 (2003).
3. J. Feng, I. A. Bolotnov, Interfacial force study on a single bubble in laminar and turbulent flows. *Nucl. Eng. Des.* **313**, 345–360 (2017).
4. J. Feng, I. A. Bolotnov, Effect of the wall presence on the bubble interfacial forces in a shear flow field. *Int. J. Multiph. Flow*. **99**, 73–85 (2017).
5. L. Eça, M. Hoekstra, A procedure for the estimation of the numerical uncertainty of CFD calculations based on grid refinement studies. *J. Comput. Phys.* **262**, 104–130 (2014).
6. J. H. Ferziger, M. Peric, *Computational methods for fluid dynamics* (Springer Science & Business Media, 2012).
7. Edeling, Cinnella, Dwight, Bijl, Bayesian estimates of the parameter variability in the k-ϵ turbulence model. *J. Comput. Phys.* **258**, 73–94 (2014).
8. Y. Liu, X. Sun, N. T. Dinh, Validation and uncertainty quantification of multiphase-CFD solvers: A data-driven Bayesian framework supported by high-resolution experiments. *Nucl. Eng. Des.* **354**, 110200 (2019).
9. X. Wu, T. Kozlowski, H. Meidani, K. Shirvan, Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory. *Nucl. Eng. Des.* **335**, 339–355 (2018).
10. X. Wu, T. Kozlowski, H. Meidani, K. Shirvan, Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian Process, Part 2: Application to TRACE. *Nucl. Eng. Des.* **335**, 417–431 (2018).
11. H. Bao, N. Dinh, J. W. Lane, R. Youngblook, Study of Data-Driven Mesh-Model Optimization in System Thermal-Hydraulic Simulation. *Trans. Am. Nucl. Soc.*, Orlando, Florida, **118** (2018).
12. H. Bao, R. Youngblood, H. Zhang, N. Dinh, L. Lin, J. Lane, A Data-driven Approach to Scale Bridging in System Thermal-hydraulic Simulation. *Proceedings of NURETH-18*, Portland, Oregon (2019).
13. H. Bao, N. Dinh, J. W. Lane, R. Youngblood, A data-driven framework for error estimation and mesh-model optimization in system-level thermal-hydraulic simulation. *Nucl. Eng. Des.* **349**, 27–45 (2019).
14. H. Bao, H. Zhao, H. Zhang, L. Zou, P. Sharpe, N. Dinh, Safe reactor depressurization windows for BWR Mark I Station Blackout accident management strategy. *Ann. Nucl. Energy*. **114**, 518–529 (2018).
15. T. J. Liu, S. G. Bankoff, Structure of air-water bubbly flow in a vertical pipe—I. Liquid mean velocity and turbulence measurements. *Int. J. Heat Mass Transf.* **36**, 1049–1060 (1993).
16. T. J. Liu, S. G. Bankoff, Structure of air-water bubbly flow in a vertical pipe—II. Void fraction, bubble velocity and bubble size distribution. *Int. J. Heat Mass Transf.* **36**, 1061–1072 (1993).
17. A. Tomiyama, I. Kataoka, I. Zun, T. Sakaguchi, Drag coefficients of single bubbles under normal and micro gravity conditions. *JSME Int. J. Ser. B Fluids Therm. Eng.* **41**, 472–479 (1998).
18. D. R. Shaver, M. Z. Podowski, Modeling of interfacial forces for bubbly flows in subcooled boiling conditions. *Trans. of ANS Winter Meeting*, Washington D.C. **113**, 1368–1371 (2015).
19. A. D. Burns, T. Frank, I. Hamill, J.-M. Shi, The Favre averaged drag model for turbulent dispersion in Eulerian multi-phase flows. *5th Int. Conf. Multiph. flow, ICMF*. **4**, 1–17 (2004).
20. R. Sugrue, B. Magolan, N. Lubchenko, E. Baglietto, Assessment of a simplified set of momentum closure relations for low volume fraction regimes in STAR-CCM and OpenFOAM. *Ann. Nucl. Energy*. **110**, 79–87 (2017).