

PAPER • OPEN ACCESS

Hybrid analysis and modeling for next generation of digital twins

To cite this article: Suraj Pawar *et al* 2021 *J. Phys.: Conf. Ser.* **2018** 012031

View the [article online](#) for updates and enhancements.

You may also like

- [Seismic isolation of Advanced LIGO: Review of strategy, instrumentation and performance](#)
F Matichard, B Lantz, R Mittleman et al.
- [HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization](#)
Zara Abdurashidova, James E. Aguirre, Paul Alexander et al.
- [Input variable selection for data-driven models of Coriolis flowmeters for two-phase flow measurement](#)
Lijuan Wang, Yong Yan, Xue Wang et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Extended abstract submission deadline: April 22, 2022

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD



Submit your abstract



Hybrid analysis and modeling for next generation of digital twins

Suraj Pawar¹, Shady E. Ahmed¹, Omer San¹, Adil Rasheed²

¹School of Mechanical & Aerospace Engineering, Oklahoma State University, Stillwater, OK 74078, USA.

²Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7465 Trondheim, Norway.

E-mail: supawar@okstate.edu, shady.ahmed@okstate.edu, osan@okstate.edu, adil.rasheed@ntnu.no

Abstract. The physics-based modeling has been the workhorse for many decades in many scientific and engineering applications ranging from wind power, weather forecasting, and aircraft design. Recently, data-driven models are increasingly becoming popular in many branches of science and engineering due to their non-intrusive nature (i.e., they are equation-free) and online learning capability. Despite the robust performance of data-driven models, they are faced with challenges of poor generalizability and difficulty in interpretation. These challenges have encouraged the integration of physics-based models with data-driven models, herein denoted hybrid analysis and modeling (HAM). We propose two different frameworks under the HAM paradigm for applications relevant to wind energy in order to bring the physical realism within emerging digital twin technologies. The physics-guided machine learning (PGML) framework reduces the uncertainty of neural network predictions by embedding physics-based features from a simplified model at intermediate layers and its performance is demonstrated for the aerodynamic force prediction task. Our results show that the proposed PGML framework achieves approximately 75% reduction in uncertainty for smaller angle of attacks. The interface learning (IL) framework illustrates how different solvers can be coupled to produce a multi-fidelity model and is successfully applied for the Boussinesq equations that govern a broad class of transport processes. The IL approach paves the way for seamless integration of multi-scale, multi-physics and multi-fidelity models (M^3 models).

1. Introduction

Some of the major challenges in realizing the potential of wind energy to meet the global electricity demand are the need for a deeper understanding of the physics of the atmospheric flow, science, and engineering of these large dynamic rotating machines and synergistic optimization and control of fleets of wind farms within the electricity grid [1]. The decades of research and development in fluid dynamics, systems engineering, manufacturing processes, material discovery can now be complemented with unprecedented amounts of data generated from in-situ measurements, lab experiments, and numerical simulations to tackle these challenges. The combination of physics-based and data-driven models is increasing in every branch of science leading to the hybrid analysis and modeling (HAM) approach for many scientific applications.

The HAM paradigm can be applied to a variety of tasks related to wind research, such as digital twinning for the optimization and real-time control of wind farms. A digital twin is



defined as a virtual replica of a physical system enabled through data collected from sensors and simulations in real-time to solve problems such as control, optimization, monitoring, and improved decision-making [2]. The advancement in multiphysics solver, computational infrastructure, big data, and artificial intelligence has allowed us to make a digital replica of large physical systems such as wind farms [3]. The digital twin lets us examine *what if* scenarios, evaluate the system's response, and select the corresponding mitigation strategies. For example, the digital twin of wind farms can be useful for different purposes such as optimal control of wind turbines to achieve the maximum performance, accurate forecast of the power production, wind farm optimization [4], and improved decision-making. The data collected from IoT sensors can provide insights into when the equipment maintenance is anticipated and help to plan in advance. Figure 1 illustrates the typical digital twin framework for a wind farm where the HAM becomes the key enabler for modeling physics of the system.

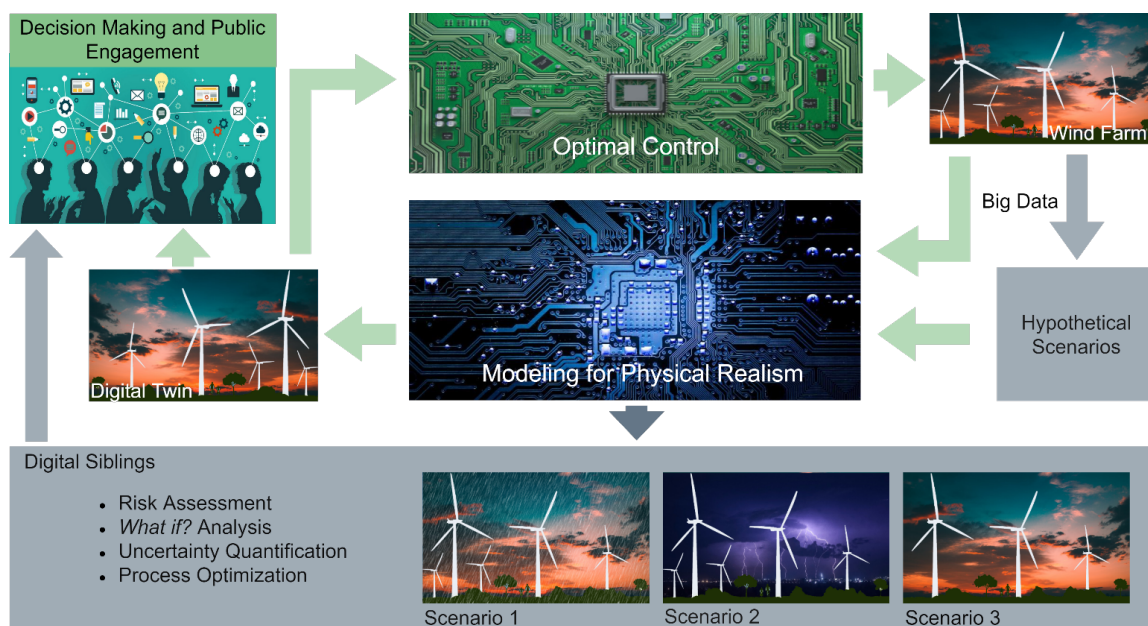


Figure 1: Digital twin of the wind farm can allow us to evaluate different scenarios using hybrid models. The big data collected from IoT sensors can be continuously assimilated to correct hybrid models and improve the state estimation.

The success of the digital twins depends upon the type of approach that we employ for modeling the system. The ability of the HAM to combine the generalizability of physics-based approaches and the automatic pattern-identification feature of data-driven approaches makes it an attractive choice to model virtual replica of physical systems in digital twins. For example, there are several turbine wake models that capture the flow in the wake of a single turbine and these models predict sufficiently accurate aerodynamics of a single turbine [5, 6]. However, these models are not enough to capture the flows in wind farm due to wake superposition, complex terrains, deep array effects, and neglected physics [7, 8, 9]. The observational data collected from a variety of sensors can be assumed to comprise of these complex flow interactions and the manifestation of all physical processes. Therefore, the hybrid model will aid in the robustness of the digital twin rather than a pure physics-based or pure data-driven approach.

In this work, we introduce two different frameworks under the umbrella of HAM. The first framework is called physics-guided machine learning (PGML) where information from simplified physics-based models is incorporated within neural network architectures to improve

the generalizability of data-driven models [10]. The second framework is our interface learning approach that deals with coupling different solvers and mathematical descriptions using statistical inference tools [11, 12, 13]. One of the potential applications of this framework can be to specify the physically accurate inlet boundary condition for the wind farm simulation. For example, it is very common to use a coarse-grid solver to resolve the atmospheric boundary layer flow and employ the fine-grid solver to resolve the flow field around wind turbines in the wind farm. For such problems, it is very important to exchange the information between two solvers that leads to physically consistent boundary conditions, and statistical tools like deep learning can be exploited for this.

Concisely, bringing physical realism in digital twins will need

- new modelling approaches that are accurate and certain, generalizable, computationally efficient, and trustworthy,
- seamless integration of multi-scale, multi-physics and multi-fidelity models (M^3 models).

To this end, we propose two HAM approaches that address the above needs. While the first approach PGML provides a mechanism to guide the learning of a machine learning algorithm using simplified physics, the second approach IL enables coupling of M^3 models.

2. Physics-guided machine learning

A wide variety of problems in wind energy such as aerodynamic performance prediction [14, 15] for optimal control, design optimization, uncertainty quantification requires the prediction of the quantity of interest in real-time. The prediction of flow around an airfoil is a high-dimensional, and nonlinear problem that can be solved using high-fidelity methods like computational fluid dynamics (CFD). However, these methods are computationally infeasible for real-time prediction. On one hand, in certain flow regimes, the simplified methods like panel codes come with a non-negligible difference between the actual dynamics and approximate models for real-world problems. On the other hand, the full-order CFD simulations are computationally demanding, thus limiting their use in many inverse modeling methodologies that require a model run to be performed in each iteration. To overcome these challenges, combining CFD models with machine learning to build a non-intrusive surrogate model is gaining widespread popularity [16, 17]. One of the main challenges with these non-intrusive models is their predictive capabilities for unseen data and their interpretation. Even though there are methods to predict the uncertainty estimate in the prediction of machine learning models [18, 19], the generalizability of non-intrusive models is not on par with physics-based models. Meanwhile, the simplified models like the Blasius boundary layer model in fluid mechanics are highly generalizable across different conditions. Therefore, it is important to incorporate the domain knowledge into learning, and to this end, we exploit the relevant physics-based features from panel methods through the PGML framework to enhance the generalizability of data-driven surrogate model.

We now introduce different components of the PGML framework depicted in Figure 2(a). In supervised machine learning, the input vector $\mathbf{x} \in \mathbb{R}^m$ is fed to the machine learning model (for example, the neural network in our case), and the mapping from the input vector to the output vector $\mathbf{y} \in \mathbb{R}^n$ is learned through training. The neural network is trained to learn the function F_θ , parameterized by θ , that includes the weights and biases of each neuron. The parameters of the neural network are optimized using the backpropagation algorithm to minimize the cost function. Usually, for the regression problems, the cost function is the mean squared error between true and predicted output, i.e., $C(\mathbf{x}, \theta) = \|\mathbf{y} - F_\theta(\mathbf{x})\|_2$. The features extracted from simplified physics-based models are embedded into hidden layers along with latent variable. Therefore, in the PGML framework, the neural network is augmented with the output of the simplified physics-based model. This is in contrast to admitting physics-based features at the

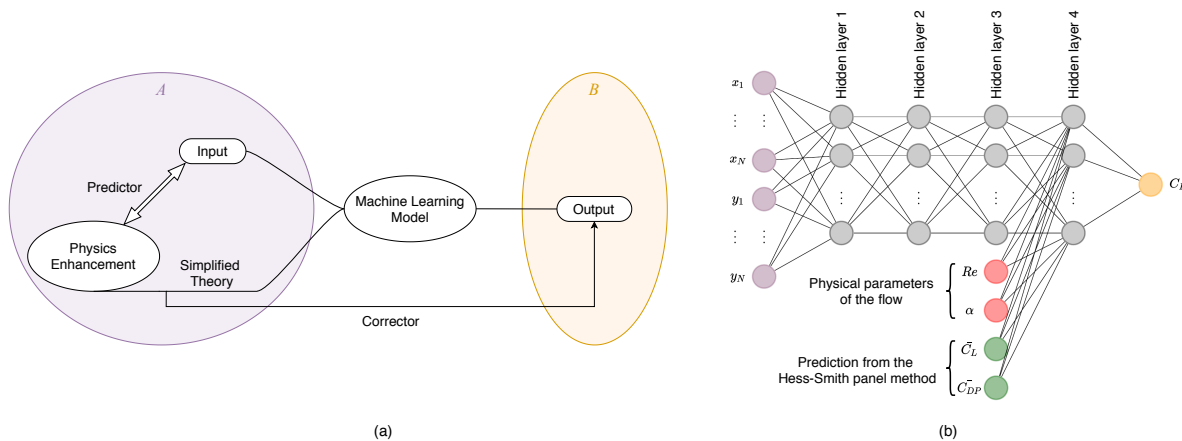


Figure 2: Physics guided machine learning (PGML) framework to train a learning engine between processes A and B : (a) a conceptual PGML framework which shows a predictor-corrector approach to incorporate physics into machine learning models through embedding simplified theories directly into neural network models, and (b) the representative neural network architecture of the PGML framework used in this study for aerodynamic forces prediction task.

input layer in conventional neural network architectures, which might lead to underestimation of the effect of such physical information, especially for high dimensional systems. For example, imagine stacking a few parameters to an input vector of a dimension of $O(10^3 - 10^6)$. It is highly probable that the learning algorithm overlooks the effect of such prior knowledge in the minimization algorithm, and a modification of the cost function becomes necessary. In the PGML, a proper latent space is first identified and the information from the physics-based model aids the neural network in constraining its output to a manifold of physically realizable models.

The training data for the neural network is generated using a series of numerical simulations performed in XFOIL [20]. We highlight here that the the neural network can also be trained using the data gathered from CFD simulations or wind-tunnel experiments. The lift coefficient data were obtained for different Reynolds numbers Re between 1×10^6 and 4×10^6 and several angles of attacks α in the range of -20 to $+20$. A total of 168 sets of two-dimensional airfoil geometry were generated from NACA 4-digit, NACA210, NACA220, and NACA250 series for training the neural network. Each airfoil is represented by 201 points. The maximum thickness of all airfoils in the training dataset was between 6% to 18% of the chord length. We use the NACA23012 and NACA23024 airfoil geometry as the test dataset to evaluate the predictive capability of the trained neural network. The simplified model used to generate the physics-based feature corresponds to the Hess-Smith panel method [21] based on potential flow theory. We note here that our testing airfoils are selected not only from a different NACA230 series (i.e., not used in the training dataset), but also the maximum thickness of 24% is well beyond the thickness ratio limit included in the training dataset.

The adopted neural network architecture has four hidden layers with 20 neurons in each hidden layer. The physical parameters, i.e., the Reynolds number and the angle of attack are concatenated at the third hidden layer along with the latent variables at that layer. In the PGML model, we augment the latent variables at the third layer with the lift coefficient and the pressure drag coefficient predicted by the panel method along with physical parameters of the flow (i.e., the Reynolds number and angle of attack). Therefore, the third layer of the neural network in the PGML framework has 24 latent variables. The representative neural network for the PGML framework to predict the aerodynamic forces on an airfoil is displayed in Figure 2(b). We utilize the deep ensembles method to predict the epistemic uncertainty [22, 23]. In this mehtod, the weights and biases of each model are initialized using the Glorot

uniform initializer and different random seed numbers are used to ensure that different values of weights and biases are assigned for each model. The inference from ensemble of all these models indicates the model uncertainty estimate of the predicted lift coefficient. Figure 3 shows the actual and predicted lift coefficient for the NACA23012 and NACA23024 airfoil geometry. The reference *True* performance is obtained by XFOIL. The ML corresponds to a simple feed-forward neural network that uses the airfoil x and y coordinates as the input features, and the physical parameters of the flow are concatenated at the third hidden layer along with the latent variables at that layer.

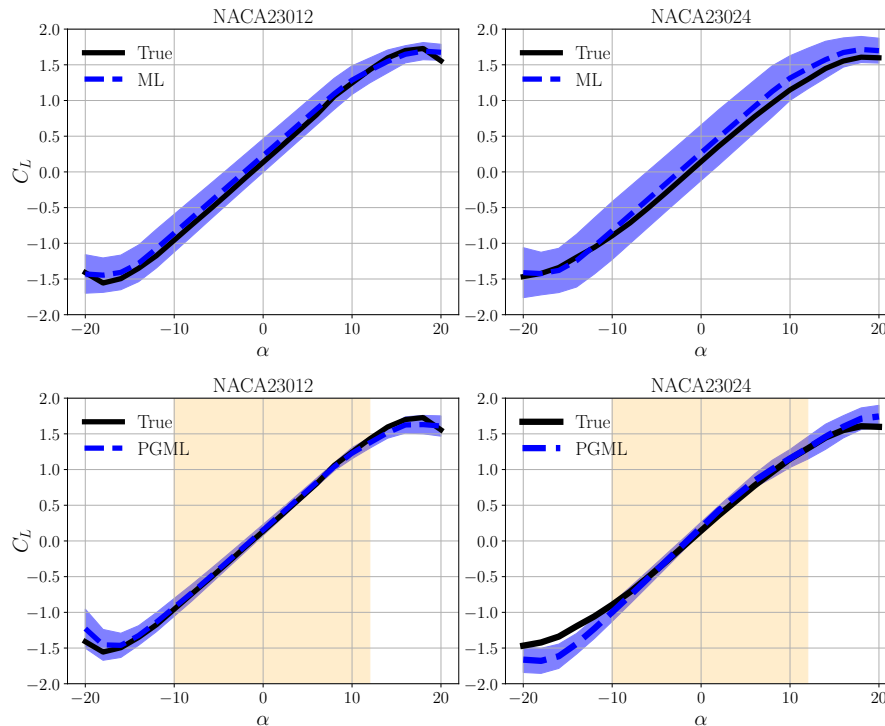


Figure 3: Actual versus predicted lift coefficient (C_L) for NACA23012 and NACA23024 airfoils at $Re = 3 \times 10^6$ using ML and PGML framework. The dashed blue curve represent the average of the predicted lift coefficient by all data-driven models (i.e., testing runs with different initialization seeds). The uncertainty band represent the two standard deviation in the prediction from ensemble of neural networks.

As shown in Figure 3, we can see that the uncertainty in the prediction of the lift coefficient by the PGML model is considerably less than the ML model for both NACA23012 and NACA23024 airfoils. The proposed PGML framework provides significantly more accurate predictions with uncertainty reduced approximately by 75% especially for the angle of attacks between -10 and +12 degrees. This further illustrates the viability of the proposed PGML framework, since the physics embedding considered here employs constant source panels and a single vortex to approximate the potential flow around the airfoil. We can also notice that the uncertainty is higher for the angle of attacks outside the range of -10 to +12 degrees. This finding is not surprising as the Hess-Smith panel method is a proven method for analysis of inviscid flow over airfoil for the smaller angle of attacks regime. The maximum thickness of an airfoil included in the training dataset is 18% of the chord length. Therefore, the uncertainty in the prediction of the lift coefficient by the ML model is higher for the NACA23024 airfoil compared to the NACA23012 airfoil. These results clearly show the potential of the PGML framework for building trustworthy models that can enable the digital twin of physical systems.

3. Interface learning

The second framework we are introducing under the umbrella of HAM is the interface learning (IL). Multi-scale, multi-physics and multi-fidelity models (M^3 models) are the main beneficiaries from the IL methodology. Many complex systems relevant to scientific and engineering applications include multiple spatiotemporal scales and comprise a multifidelity problem sharing an interface between various formulations or heterogeneous computational entities. We refer the readers to our previous discussion about the potential of IL approaches, with demonstrations on truncated domains [12], and micro-macro scale solvers coupling [11]. In the present study, we are interested in situations where part of the domain, physics, or scales are characterized by repeating coherent structures, and can thus be represented by a reduced order model (ROM) for computational speed-up. In the meantime, a high-fidelity full order model (FOM) is dedicated for the rest of domain/dynamics for accuracy requirements. However, both solvers are coupled and information should be communicated and matched at their *interface*. To this end, we present a robust HAM approach combining a physics-based FOM and a data-driven ROM to form the building blocks of an integrated approach among mixed fidelity descriptions toward predictive digital twin technologies. This is depicted in Figure 4(a), where HAM approaches are introduced at the interface (marked in orange) to efficiently couple the ROM solvers (in blue) with the FOM ones (in red).

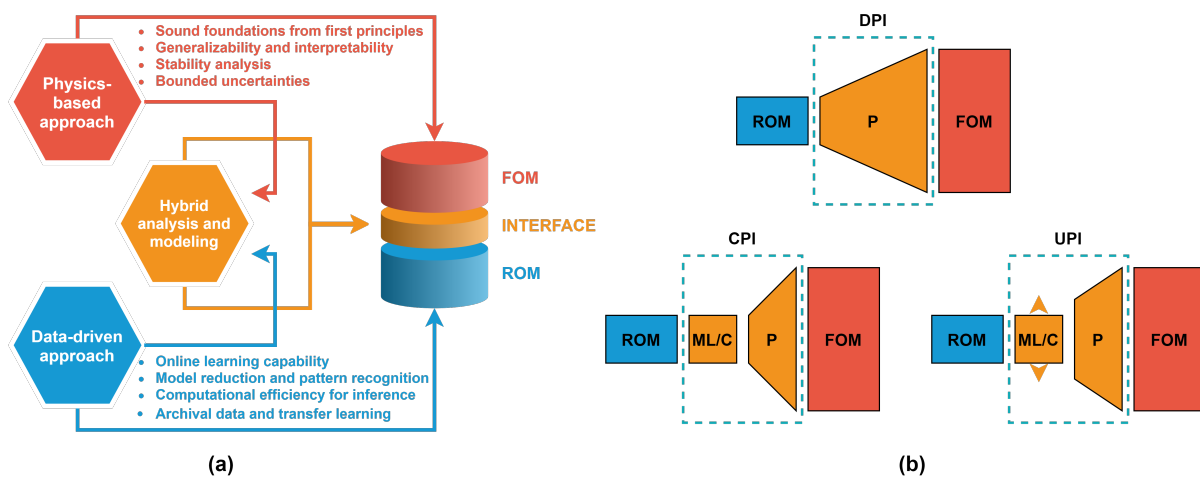


Figure 4: Hybrid analysis and modeling (HAM) as a key enabler for ROM-FOM coupling problems toward predictive digital twins: (a) an overview, and (b) proposed ROM-FOM coupling frameworks (Section 3.1).

3.1. ROM-FOM coupling framework

In order to demonstrate the ROM-FOM coupling framework, we consider a coupled system as follows,

$$\frac{\partial u}{\partial t} = f_1(u; \mu_1) + g_1(u, v; \mu_1, \mu_2), \quad (1)$$

$$\frac{\partial v}{\partial t} = f_2(v; \mu_2) + g_2(u, v; \mu_1, \mu_2), \quad (2)$$

where u and v are the coupled variables and g_1 and g_2 define this coupling, while μ_1 and μ_2 denote the set of system's parameters. We highlight that the coupled variables might represent the state variables at different regions of the domain (e.g., multi-component systems), different physics (e.g., fluid-structure interactions) and/or different scales within the same domain (e.g., multiscale systems). We suppose that the dynamics of u can be approximated by a ROM while a FOM resolves v and both solvers need to communicate information to satisfy the coupling.

3.1.1. Reduced order model Introducing a spatial discretization to Eq. (1), it can be rewritten in a semi-discrete continuous-time as follows,

$$\frac{d\mathbf{u}}{dt} = \mathcal{F}(\mathbf{u}, \mathbf{v}; \boldsymbol{\mu}) = \mathcal{L}_1\mathbf{u} + \mathcal{L}_2\mathbf{v} + \mathcal{N}(\mathbf{u}, \mathbf{v}), \quad (3)$$

where the boldfaced symbols represent the arrangement of discretized variables in a column vector, $\boldsymbol{\mu}$ defines the system's parameters, and \mathcal{F} is a deterministic operator with linear and nonlinear components \mathcal{L} , and \mathcal{N} , respectively. We exploit the advances in ROM techniques to build surrogate models to economically resolve portions of domain and/or physics. The ROM solution can thus be used to infer the flow conditions at the interface so that a FOM solver can be efficiently employed for the regions of interest. The standard Galerkin ansatz is applied for the dynamics of \mathbf{u} as $\mathbf{u}(t) \approx \Phi\boldsymbol{\alpha}(t)$, where the columns of matrix $\Phi = [\phi_1, \phi_2, \dots, \phi_r]$ form the orthonormal bases of a reduced subspace, and $\boldsymbol{\alpha}$ defines their amplitudes. Proper orthogonal decomposition (POD) is one popular technique to systematically construct Φ such that the solution manifold preserves as much variance as possible when projected onto the subspace spanned by Φ [24]. By substituting this approximation into Eq. (3), performing the inner product with Φ , and making use of the quadratic nonlinearity in most fluid flow systems, we get the following,

$$\frac{d\boldsymbol{\alpha}}{dt} = \mathfrak{L}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \mathfrak{N}\boldsymbol{\alpha} + \mathfrak{C}, \quad (4)$$

where \mathfrak{L} and \mathfrak{N} signify the model coefficients while \mathfrak{C} defines the contribution of \mathbf{v} into the ROM of \mathbf{u} . At the interface, we introduce a long short-term memory network to bridge the low-fidelity descriptions to high-fidelity models in various forms of interfacial error correction or prolongation. An array of interface modeling paradigms is sketched in Figure 4(b) and summarized as follows (see [13] for more details),

- (i) *DPI: Direct Prolongation Interface.* The DPI approach provides an estimate of the flow variables at the interface from the ROM solution. Indeed, this prolongation map naturally results from the Galerkin ansatz, without any interference from the ML side. However, it is known that truncated Galerkin ROMs might yield erroneous and even unstable predictions for complex systems [25]. Therefore, the solution from the DPI approach is potentially inaccurate, and a correction needs to be introduced.
- (ii) *CPI: Correction followed by Prolongation Interface.* The CPI methodology works by introducing the correction in the latent subspace and addresses the deviation in modal coefficients predicted from solving the Galerkin ROM, known as closure error. Specifically, the LSTM for CPI takes the values of modal coefficients acquired from integrating Eq. (4) and predicts the discrepancy between these values and their optimal values. We highlight here that the size of the input and output vectors is $O(r)$, independent of the FOM resolution, which offers a potential flexibility dealing with 2D and 3D problems.
- (iii) *UPI: Uplifted Prolongation Interface.* Although the CPI methodology cures the closure error and provides a stabilized solution, it does not address the projection error. Unless a large number of modes are resolved, the projection error can be significant. An uplifting ROM has been proposed [25], where both closure and projection errors are taken care of. In addition to the closure modeling, the ROM subspace is expanded to recover some of the smaller scales missing in the initial subspace as $\mathbf{u} \approx \Phi\boldsymbol{\alpha} + \Psi\boldsymbol{\beta}$, where Ψ forms orthonormal basis for a q -dimensional subspace complementing that spanned by Φ and with $\boldsymbol{\beta}$ being the corresponding amplitudes. Indeed, we define Φ as the first r POD modes and Ψ as the following q modes, as they are orthogonal by construction. We highlight that the Galerkin ROM equations only solve for $\boldsymbol{\alpha}$ to keep the computational cost as low as possible. Therefore, a complementary model for $\boldsymbol{\beta}$ has to be constructed so that the

uplifting approach can be employed. A mapping from the first r modal coefficients to the next q modes is assumed to exist and we exploit the LSTM learning capabilities to infer this map from data. In particular, the UPI architecture is trained to read the Galerkin ROM prediction for the first r modal coefficients as input, and return the true coefficients of the first $r + q$ modes, defined as projection of the FOM data onto the first $r + q$ POD modes. The latter represents the optimal solution that can be obtained using the subspace spanned by the first $r + q$ basis. Thus, it provides a *closure* correction for the first r modes and a *superresolution* effect for the next q modes, simultaneously in a single network.

3.2. Demonstration using Boussinesq equations

The dimensionless form of the 2D incompressible Boussinesq equations can be represented by the following two coupled transport equations in vorticity-streamfunction formulation,

$$\frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{\text{Re}} \nabla^2 \omega + \text{Ri} \frac{\partial \theta}{\partial x}, \quad (5)$$

$$\frac{\partial \theta}{\partial t} + J(\theta, \psi) = \frac{1}{\text{RePr}} \nabla^2 \theta, \quad (6)$$

where ω , ψ and θ denote the dimensionless vorticity, streamfunction and temperature fields, respectively. We utilize the 2D Boussinesq equation to illustrate the ROM-FOM coupling in multi-fidelity environments. In particular, we suppose that we are more interested in the temperature field predictions. Thus, we dedicate a FOM solver for Eq. (6). However, the solution of this equation requires evaluating the streamfunction field at each time step. The kinematic relationship between vorticity and streamfunction is given by the Poisson equation (i.e., $\nabla^2 \psi = -\omega$), the solution of which consumes significant amount of time and computational resources and is considered the bottleneck for most incompressible flow solvers. Therefore, we consider a ROM solver for the vorticity dynamics and FOM solver for the temperature field.

For demonstration, we explore the lock-exchange problem, defined by a fluid in a rectangular domain $(x, y) \in [0, 8] \times [0, 1]$ with a vertical barrier dividing the domain at $x = 4$, keeping the temperature of the left half at 1.5 and temperature of the right half at 1. Initially, the flow is at rest, with uniform temperatures at the right and left regions, and the barrier is removed at $t = 0$. Reynolds number of $\text{Re} = 10^4$, Richardson number of $\text{Ri} = 4$, and Prandtl number of $\text{Pr} = 1$ are set. For FOM simulations, a Cartesian grid of 4096×512 with a timestep of $\Delta t = 5 \times 10^{-4}$ are used to capture the dynamical evolution up to $t = 8$.

A two-layer LSTM with 20 hidden units in each LSTM cell constitutes our ML architecture. We store 800 time snapshots for POD basis construction and we retain $r = 8$ modes for the Galerkin ROM solver. We also utilize the dataset of the stored 800 snapshots for LSTM training and validation. During the testing phase, the trained neural networks are deployed to enforce the ROM-FOM coupling at each timestep. Figure 5 shows the predictions of the temperature field at final time (i.e., $t = 8$) computed from DPI, CPI, and UPI approaches compared to the FOM field. We emphasize that the ROM-FOM coupling results correspond to the solution of the vorticity equation with a ROM solver, which feeds the FOM solver with streamfunction to solve the temperature equation only as opposed to the FOM results which comes from the solution of both equations using a fully FOM simulation. Although the CPI results are better than those of DPI, we can observe that the fine details of the flow field are not accurately captured. On the other hand, the implementation of the UPI approach with $r = 8$ and $q = 8$ recovers an increased amount of the fine flow structures that are not well-represented by the first $r = 8$ modes. We note that the CPI and UPI have similar computational speed-ups as they both involve the same steps which are as follows; (1) solution of Galerkin ROM with r modes for the vorticity and streamfunction fields, (2) deployment of LSTM prediction with similar architecture and complexity, and (3) solution of FOM equation for temperature fields. Nonetheless, DPI is slightly faster because it does not require any neural network predictions.

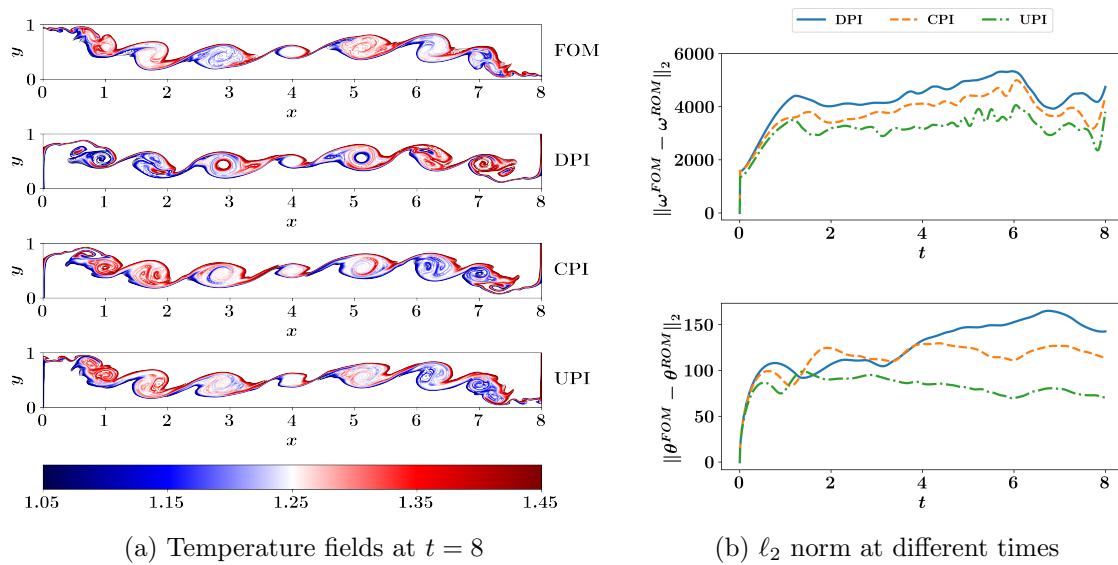


Figure 5: Final temperature fields as obtained from different ROM-FOM coupling approaches at final time $t = 8$, compared to the FOM solution (a), and the ℓ_2 norm of the ROM-FOM solutions at different times (b).

4. Concluding remarks

The data-driven methods are increasingly being applied in many branches of science and engineering due to their success in automatic pattern-identification using the data collected from sensors and numerical simulations. Even though they offer an alternative to physics-based modeling derived from phenomenological arguments, they are usually black box in nature and lack generalizability. The hybrid analysis and modeling (HAM) is a newly emerging paradigm that combines physics-based and data-driven modeling to deliver robust and generalizable models that can enable the digital twins of large scale physical systems.

The major contributions of this work towards HAM paradigm are

- A novel deep neural network architecture that makes it possible to inject physics during the training process. This resulted in a significant reduction of uncertainty.
- An interface learning technique that makes seamless coupling of multi fidelity models possible.

The physics-guided machine learning (PGML) framework enhances the generalizability of the neural network based surrogate model and its performance is illustrated for real time aerodynamic performance prediction task. The interface learning (IL) framework allows integration of two different models seamlessly to build multi-fidelity models that are orders of magnitude faster than full order physics-based models. Although PGML approach was demonstrated to work with deep neural network, it can easily be extended to other machine learning algorithms. Likewise the interface learning technique that was demonstrated to couple multi-fidelity models can be used for coupling multiscale and multiphysics models.

Acknowledgement

The authors acknowledge the financial support from the Department of Engineering Norwegian Research Council, the industrial partners of OPWIND: Operational Control for Wind Power Plants (Grant No.: 268044/E20), the U.S. DOE Early Career Research Program (Award Number DE-SC0019290), and the National Science Foundation under Award Number DMS-2012255.

References

- [1] Paul Veers, Katherine Dykes, Eric Lantz, Stephan Barth, Carlo L Bottasso, Ola Carlson, Andrew Clifton, Johny Green, Peter Green, Hannele Holttinen, et al. Grand challenges in the science of wind energy. *Science*, 366(6464):eaau2027, 2019.
- [2] Adil Rasheed, Omer San, and Trond Kvamsdal. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, 8:21980–22012, 2020.
- [3] Digital Wind Farm. <https://www.ge.com/renewableenergy/wind-energy/onshore-wind/digital-wind-farm>. Accessed: 2021-01-01.
- [4] Garrett E Barter, Amy Robertson, and Walter Musial. A systems engineering vision for floating offshore wind cost optimization. *Renewable Energy Focus*, 34:1–16, 2020.
- [5] Majid Bastankhah and Fernando Porté-Agel. A new analytical model for wind-turbine wakes. *Renewable Energy*, 70:116–123, 2014.
- [6] Carl R Shapiro, Dennice F Gayme, and Charles Meneveau. Modelling yawed wind turbine wakes: a lifting line approach. *Journal of Fluid Mechanics*, 841, 2018.
- [7] Charles Meneveau. Big wind power: Seven questions for turbulence research. *Journal of Turbulence*, 20(1):2–20, 2019.
- [8] Evangelos S Politis, John Prospathopoulos, D Cabezon, Kurt S Hansen, PK Chaviaropoulos, and Rebecca J Barthelme. Modeling wake effects in large wind farms in complex terrain: the problem, the methods and the issues. *Wind Energy*, 15(1):161–182, 2012.
- [9] Wolfgang Schlez and Anja Neubert. New developments in large wind farm modelling. In *European Wind Energy Conference*, page 8, 2009.
- [10] Suraj Pawar, Omer San, Burak Aksoylu, Adil Rasheed, and Trond Kvamsdal. Physics guided machine learning using simplified theories. *Physics of Fluids*, 33(1):011701, 2021.
- [11] Suraj Pawar, Shady E Ahmed, and Omer San. Interface learning in fluid dynamics: Statistical inference of closures within micro–macro-coupling models. *Physics of Fluids*, 32(9):091704, 2020.
- [12] Shady E Ahmed, Omer San, Kursat Kara, Rami Younis, and Adil Rasheed. Interface learning of multiphysics and multiscale systems. *Physical Review E*, 102(5):053304, 2020.
- [13] Shady E Ahmed, Omer San, Kursat Kara, Rami Younis, and Adil Rasheed. Multifidelity computing for coupling full and reduced order models. *PLoS ONE*, 16(2):e0246092, 2021.
- [14] Ge-Cheng Zha, Bruce F Carroll, Craig D Paxton, Clark A Conley, and Adam Wells. High-performance airfoil using coflow jet flow control. *AIAA Journal*, 45(8):2087–2090, 2007.
- [15] Patrick LeGresley and Juan Alonso. Airfoil design optimization using reduced order models based on proper orthogonal decomposition. In *Fluids 2000 conference and exhibit*, page 2545, 2000.
- [16] Yao Zhang, Woong Je Sung, and Dimitri N Mavris. Application of convolutional neural network to predict airfoil lift coefficient. In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1903, 2018.
- [17] Saakaar Bhatnagar, Yaser Afshar, Shaowu Pan, Karthik Duraisamy, and Shailendra Kaushik. Prediction of aerodynamic flow fields using convolutional neural networks. *Computational Mechanics*, 64(2):525–545, 2019.
- [18] Yin hao Zhu, Nicholas Zabarar, Phaedon-Stelios Koutsourelakis, and Paris Perdikaris. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *Journal of Computational Physics*, 394:56–81, 2019.
- [19] Romit Maulik, Kai Fukami, Nesar Ramachandra, Koji Fukagata, and Kunihiko Taira. Probabilistic neural networks for fluid flow surrogate modeling and data recovery. *Physical Review Fluids*, 5(10):104401, 2020.
- [20] Mark Drela. XFOIL: An analysis and design system for low Reynolds number airfoils. In *Low Reynolds number aerodynamics*, pages 1–12. Springer, 1989.
- [21] John L Hess. Panel methods in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 22(1):255–274, 1990.
- [22] Robert Tibshirani. A comparison of some error estimates for neural network models. *Neural Computation*, 8(1):152–163, 1996.
- [23] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems*, pages 6402–6413, 2017.
- [24] Philip Holmes, John L Lumley, Gahl Berkooz, and Clarence W Rowley. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge University Press, New York, 2012.
- [25] Shady E Ahmed, Omer San, Adil Rasheed, and Traian Iliescu. A long short-term memory embedding for hybrid uplifted reduced order models. *Physica D: Nonlinear Phenomena*, page 132471, 2020.