

MOX-Report No. 83/2024

Multi-fidelity reduced-order surrogate modelling

Conti, P.; Guo, M.; Manzoni, A.; Frangi, A.; Brunton, S. L.; Kutz, J.N.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

Multi-fidelity reduced-order surrogate modeling

Paolo Conti^a, Mengwu Guo^b, Andrea Manzoni^c, Attilio Frangi^a, Steven L. Brunton^d, J. Nathan Kutz^d

^a*Department of Civil Engineering, Politecnico di Milano, Milano, I-20133, Italy*

^b*Department of Applied Mathematics, University of Twente, Enschede, 7522NB, the Netherlands*

^c*MOX – Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133, Milano, Italy*

^d*Department of Mechanical Engineering, and ⁵Department of Applied Mathematics, University of Washington, Seattle, WA 98195, United States*

Abstract

High-fidelity numerical simulations of partial differential equations (PDEs) given a restricted computational budget can significantly limit the number of parameter configurations considered and/or time window evaluated. Multi-fidelity surrogate modeling aims to leverage less accurate, lower-fidelity models that are computationally inexpensive in order to enhance predictive accuracy when high-fidelity data are scarce. However, low-fidelity models, while often displaying the qualitative solution behavior, fail to accurately capture fine spatio-temporal and dynamic features of high-fidelity models. To address this shortcoming, we present a data-driven strategy that combines dimensionality reduction with multi-fidelity neural network surrogates. The key idea is to generate a spatial basis by applying proper orthogonal decomposition (POD) to high-fidelity solution snapshots, and approximate the dynamics of the reduced states — time-parameter-dependent expansion coefficients of the POD basis — using a multi-fidelity *long short-term memory* (LSTM) network. By mapping low-fidelity reduced states to their high-fidelity counterpart, the proposed reduced-order surrogate model enables the efficient recovery of full solution fields over time and parameter variations in a non-intrusive manner. The generality of this method is demonstrated by a collection of PDE problems where the low-fidelity model can be defined by coarser meshes and/or time stepping, as well as by misspecified physical features.

Keywords: Reduced-order modeling, multi-fidelity surrogate modeling, LSTM networks, proper orthogonal decomposition, parametrized PDEs

1. Introduction

Scientific computing has revolutionized science and engineering by enabling advancements that have transformed almost every field of application. While becoming an increasingly critical component of any real-world modeling, accurate and well-resolved simulations for multi-fidelity and multi-physics system also come at an elevated computational cost that can strain limited computational resources. Indeed, it is often challenging to generate many high-fidelity (HF) simulations from large-scale models with limited availability of computing power, thus imposing restrictions on how comprehensive, either in parametric studies or length of time evolution, such numerical simulations can be. In particular, computational costs can easily become prohibitive or intractable when parameterized, time-dependent systems of partial differential equations (PDEs) are solved with detailed full-order models (FOMs) in a multi-query context (i.e., at many instances of the input parameters characterizing the systems), such as in uncertainty quantification [1, 2], optimal control [3, 4], shape optimization [5], parameter estimation [6, 7], and model calibration [8, 9, 10]. In such cases, the construction of efficient surrogate models is of paramount importance in order to produce model proxies which can cheaply and accurately characterize the PDE system. By mapping low-fidelity reduced states to their high-fidelity counterpart, we demonstrate a reduced-order surrogate model paradigm that enables the efficient recovery of full solution fields over time and parameter variations in a non-intrusive manner.

Reduced-order models (ROMs) have been developed to construct low-dimensional representations of high-dimensional systems for a significant reduction in computational costs with controlled accuracy [11, 12, 13, 14, 15, 16, 17]. Among the available strategies, *intrusive* ROM techniques explicitly incorporate

full-order governing equations at the reduced level and often yield reliable and physically meaningful solutions. However, the requirement for full-order simulators has limited the flexibility, generality, and industrial relevance of these approaches. On the other hand, *non-intrusive* approaches learn reduced-order systems primarily from solution data, including from numerical or experimental data. Examples are dynamic mode decomposition [18, 19], reduced-order operator inference [20, 21, 22, 23], sparse identification of reduced latent dynamics [24, 25, 26, 27, 28, 29], manifold learning using deep auto-encoders [30, 31, 32, 33], data-driven approximation of time-integration schemes [34], and Gaussian processes for reduced representations [35, 36]. Other relevant, recent works on non-intrusive ROM include [37, 38, 39, 40]. These data-driven ROM techniques do not rely on direct operations on the full-order solvers, and are especially advantageous for applications with well-established, readily-executed legacy codes.

However, the applicability and reliability of these numerical methods can break down when the collection of HF data for model reduction is too computationally expensive, even in the offline stage of model training. In addition, it is increasingly common to encounter scenarios where a wealth of data sources are readily available, easily accessible, and/or cheaply computable, albeit not perfectly accurate. These low-fidelity (LF) data can be generated from coarse discretizations, linearization, simplified geometric or physical assumptions, or computationally efficient surrogate models. Despite limitations in accuracy, the LF data can represent a useful addition of information to the limited HF data used for model training. Thus, multi-fidelity (MF) methods aim to achieve an effective data fusion from various fidelity levels, and enable strong generalization performance of data-driven models in regions where HF data are scarce or even absent. A wide range of MF surrogate modeling techniques have been developed based on Gaussian processes [41, 42, 43] and neural networks (NNs) [44, 45, 46, 47, 48]. They have found recent applications in many areas of scientific computing, including uncertainty quantification, inference, and optimization [49, 50, 51, 52, 53, 54, 55]. Nevertheless, MF techniques often become impractical when approximating high-dimensional systems, thereby limiting their ability to directly approximate the full solution fields of PDEs. Fortunately, with the aid of dimensionality reduction, MF data fusion can be feasible for the representation of reduced states in a predominant low-dimensional latent space.

To combine the computational flexibility of non-intrusive ROMs and data efficiency of MF modeling, we present an MF method of reduced-order surrogate modeling, abbreviated as MF-POD, which integrates MF regression with dimensionality reduction via the proper orthogonal decomposition (POD). The core idea is to approximate the solution manifold by a reduced subspace spanned by a small number of spatial bases using the POD, and then employ MF regression to represent their time-parameter-dependent expansion coefficients. The essence of MF regression here is inferring HF POD coefficients from their LF counterpart, so as to approach HF accuracy at the computational cost of LF evaluations. To achieve this, a time-parameter-dependent mapping from the LF to HF POD coefficients is constructed by means of long-short term memory (LSTM) NNs [56, 57]. LSTM models have been shown to be effective in time series analysis, e.g., in speech recognition [58, 59], natural language processing [60], and computational mechanics [61, 62]. In particular, LSTM networks have proven to be successful in the detection of both long- and short-term temporal patterns and nonlinear correlations between datasets, with potential relevance in the construction of non-intrusive reduced order models [55, 63, 64, 65, 66]. A recent work [62] has also shown their success in MF surrogate modeling for simultaneous parametric generalization and temporal forecast. Thus, instead of relying on expensive HF full-order evaluations, MF-POD allows for efficient online approximation of solution fields over time and parameter variation by running fast LF simulations and then mapping their POD coefficients to the HF level. In this way, parameter regions with sparse (or even no) HF data coverage can be conveniently explored. From the time evolution of the LF model, the long-term HF forecast is enabled through LSTM models. A schematic representation of the method is represented in Fig. 1.

The major advantage of the proposed method lies in a guaranteed light-weight offline stage. The computational cost of MF data generation is reduced, as the need for HF samples (for both the POD and LSTM training) is limited and the computation of LF samples (for LSTM training) is extremely efficient. The POD reduction also ensures that the LSTM time series is modeled in a relatively low dimension. While leveraging the advantages of being non-intrusive, MF-POD overcomes the potential lack of physical consistency by incorporating physically meaningful LF data, hence enabling interpretability and reliability in generalization. We test MF-POD's performance in parametric generalization and temporal forecasting on a diverse set of PDE benchmark problems. Since the method has the potential to capture phenomena such as transients or the onset of instabilities, this work focuses on unsteady PDEs, including spiral wave propagation in a

parametrized reaction-diffusion system, vorticity approximation of advection-diffusion in shallow water, and velocity and pressure approximation of fluid flow past a cylinder in a channel. Compared to detailed HF simulations, LF data are generated by lower-quality discretizations over space and time, and/or with corrupted values of critical physical features.

This paper is structured as follows. In section 2, we introduce the *offline* training process of the proposed multi-fidelity reduced-order surrogate model, which consists of the POD reduced basis construction and the LF-to-HF mapping with LSTM network models, both based on bi-fidelity data at limited time-parameter locations; thereafter, using this trained model, we present the *online* procedure to infer HF solutions from LF evaluations over a wider range of parametric configurations and forward in time. Results for the aforementioned numerical tests are reported and discussed in sections 3 to 5, and conclusions are finally drawn in section 6. The source code of the proposed method is made available in the public repository `MultiFidelity_POD` [67].

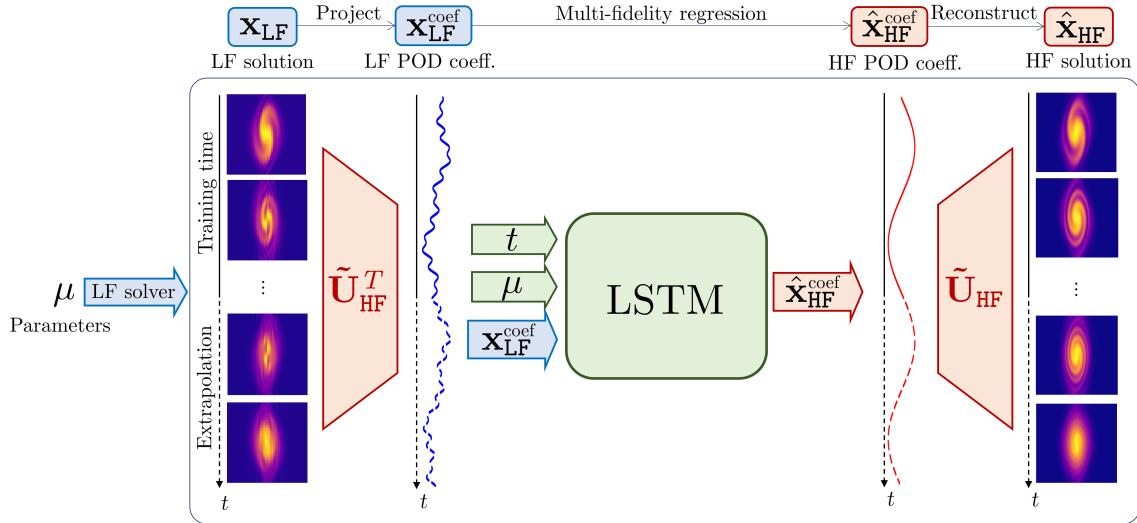


Figure 1: Using the proposed multi-fidelity reduced-order surrogate model to approximate time-parameter-dependent solution fields. Given a new parameter configuration μ , low-fidelity solutions \mathbf{x}_{LF} are evaluated over a desired time domain that can be much longer than the time window covered by training data. Low-fidelity POD coefficients $\mathbf{x}_{\text{LF}}^{\text{coef}}$ are computed via direct projection onto the reduced basis $\tilde{\mathbf{U}}_{\text{HF}}$, and their mapped to their high-fidelity counterpart $\hat{\mathbf{x}}_{\text{HF}}^{\text{coef}}$ through an LSTM neural network. Finally, the approximation of full solution fields is reconstructed as linear combinations of the reduced basis vectors.

2. MF-POD: offline/online framework

In this section, we introduce an algorithmic method for MF reduced-order surrogate modeling — MF-POD, which is decoupled into the stages of offline training and online testing.

2.1. Offline training

A schematic presentation of the offline flow is illustrated in Fig. 3.

- **Step 1: Generating MF training datasets**

HF and LF solution data for training are computed over a limited set of parametric configurations $\mathcal{P}_{\text{train}}$ by running the respective solvers. HF (resp. LF) snapshots are stacked in a matrix $\mathbf{X}_{\text{HF}} \in \mathbb{R}^{N_{\text{dof}}^{\text{HF}} \times N_{\mu} N_t^{\text{HF}}}$ (resp. $\mathbf{X}_{\text{LF}} \in \mathbb{R}^{N_{\text{dof}}^{\text{LF}} \times N_{\mu} N_t^{\text{LF}}}$), where $N_{\mu} = |\mathcal{P}_{\text{train}}|$, and $N_{\text{dof}}^{\text{HF}}$ (resp. $N_{\text{dof}}^{\text{LF}}$) is the number of spatial degrees of freedom and N_t^{HF} (resp. N_t^{LF}) the number of time instances for the HF (resp. LF) level. Note that solutions at different fidelity levels are evaluated at the same parameter values, but not necessarily at the same spatio-temporal locations, because the latter depends on the spatial mesh and time stepping of choice. This step demands the highest computational cost in the offline stage, as it requires querying the HF full-order model. Therefore, we consider a small number of parameter instances N_{μ} for the training data.

- **Step 2: Dimensionality reduction via POD**

A set of N_{POD} reduced basis vectors, collected in $\tilde{\mathbf{U}}_{\text{HF}} \in \mathbb{R}^{N_{\text{dof}}^{\text{HF}} \times N_{\text{POD}}}$, is extracted from the HF snapshot matrix \mathbf{X}_{HF} via the POD. Note that $\tilde{\mathbf{U}}_{\text{HF}}^T \tilde{\mathbf{U}}_{\text{HF}} = \mathbf{I}_{N_{\text{POD}}}$.

Proper Orthogonal Decomposition (POD): The POD takes advantage of the singular-value decomposition (SVD) to linearly extract principal components from high-dimensional data, and hence provide low-dimensional orthonormal basis that can be computed as follows:

1. Compute the SVD of \mathbf{X}_{HF} , such that

$$\mathbf{X}_{\text{HF}} = \mathbf{U}_{\text{HF}} \Sigma_{\text{HF}} \mathbf{V}_{\text{HF}}^T, \quad \text{with } \Sigma_{\text{HF}} = \text{diag}(\sigma_1, \dots, \sigma_{N_\mu N_t^{\text{HF}}}) \quad \text{and} \quad \mathbf{U}_{\text{HF}}^T \mathbf{U}_{\text{HF}} = \mathbf{I}, \quad \mathbf{V}_{\text{HF}}^T \mathbf{V}_{\text{HF}} = \mathbf{I}. \quad (1)$$

Here $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{N_\mu N_t^{\text{HF}}} \geq 0$.

2. Define N_{POD} as the minimum integer that satisfies

$$\frac{\sum_{i=1}^{N_{\text{POD}}} \sigma_i^2}{\sum_{i=1}^{N_\mu N_t^{\text{HF}}} \sigma_i^2} \geq 1 - \epsilon_{\text{POD}}^2, \quad (2)$$

where $\epsilon_{\text{POD}} > 0$ is a given tolerance that determines how much of the variance of the signal should be captured. As an alternative to providing a fixed tolerance ϵ_{POD} , one can select the dimension N_{POD} of the reduced basis according to the singular-value decay.

3. Form the POD basis $\tilde{\mathbf{U}}_{\text{HF}}$ as the first N_{POD} columns of \mathbf{U}_{HF} , i.e., $\tilde{\mathbf{U}}_{\text{HF}} = (\mathbf{U}_{\text{HF}})_{:,1:N_{\text{POD}}}$.

The POD truncation provides a low-rank approximation of \mathbf{X}_{HF} , i.e., $\mathbf{X}_{\text{HF}} = \mathbf{U}_{\text{HF}} \Sigma_{\text{HF}} \mathbf{V}_{\text{HF}}^T \approx \tilde{\mathbf{U}}_{\text{HF}} \tilde{\Sigma}_{\text{HF}} \tilde{\mathbf{V}}_{\text{HF}}^T$, where $\tilde{\mathbf{V}}_{\text{HF}}$ contains the first N_{POD} columns of \mathbf{V}_{HF} and $\tilde{\Sigma}_{\text{HF}}$ contains the first $N_{\text{POD}} \times N_{\text{POD}}$ block of Σ_{HF} .

- **Step 3: Computing POD coefficients via direct projection**

The time-parameter-dependent combination coefficients of the POD basis for both fidelity levels are obtained by projecting \mathbf{X}_{HF} and \mathbf{X}_{LF} onto the basis $\tilde{\mathbf{U}}_{\text{HF}}$, respectively:

$$\mathbf{X}_{\text{HF}}^{\text{coef}} = \tilde{\mathbf{U}}_{\text{HF}}^T \mathbf{X}_{\text{HF}}, \quad \mathbf{X}_{\text{LF}}^{\text{coef}} = \tilde{\mathbf{U}}_{\text{HF}}^T \mathcal{L}(\mathbf{X}_{\text{LF}}), \quad (3)$$

where $\mathbf{X}_{\text{HF}}^{\text{coef}}, \mathbf{X}_{\text{LF}}^{\text{coef}} \in \mathbb{R}^{N_{\text{POD}} \times N_\mu N_t^{\text{HF}}}$ collect respectively the POD coefficients of the HF and LF data. Here, \mathcal{L} represents an operator which lifts the LF solution vectors to the HF spatio-temporal resolution via interpolation. In particular, spatial interpolation is required when the LF model is on a different (coarser) mesh than the HF one, thus the LF solution vectors should be transformed to match with the HF resolution and hence enable a projection onto the POD modes $\tilde{\mathbf{U}}_{\text{HF}}$; temporal interpolation is necessary when LF and HF data are computed with different time discretizations, because compatible sequential data between the two fidelity levels are required by the NN model in the next step. In this work, we consider either linear interpolation (i.e., $\mathcal{L}(\mathbf{X}_{\text{LF}}) = \mathbf{P} \mathbf{X}_{\text{LF}} \mathbf{Q}^T$ with $\mathbf{P} \in \mathbb{R}^{N_{\text{dof}}^{\text{HF}} \times N_{\text{dof}}^{\text{LF}}}$ and $\mathbf{Q} \in \mathbb{R}^{N_\mu N_t^{\text{HF}} \times N_\mu N_t^{\text{LF}}}$ being the spatial and temporal interpolating matrices, respectively) or nearest-neighbor interpolation, both of which are computationally inexpensive. An alternative option for matching LF and HF solution vectors is manifold alignment [68, 69].

- **Step 4: Training LSTM surrogate model**

The goal of this step is to learn the mapping \mathbf{f} from the LF POD coefficients $\mathbf{x}_{\text{LF}}^{\text{coef}} \in \mathbb{R}^{N_{\text{POD}}}$ at time instance t and parameter μ to their HF correspondence $\mathbf{x}_{\text{HF}}^{\text{coef}} \in \mathbb{R}^{N_{\text{POD}}}$, written as

$$(t, \mu, \mathbf{x}_{\text{LF}}^{\text{coef}}(t, \mu)) \mapsto \mathbf{f}(t, \mu, \mathbf{x}_{\text{LF}}^{\text{coef}}(t, \mu)) = \mathbf{x}_{\text{HF}}^{\text{coef}}(t, \mu). \quad (4)$$

To approximate the mapping \mathbf{f} , we make use of a long short-term memory (LSTM) NN model denoted by $\mathbf{f}_{\text{NN}}(\cdot) = \mathbf{f}_{\text{NN}}(\cdot; \Theta_{\text{NN}})$, where Θ_{NN} are the network parameters. This NN model \mathbf{f}_{NN} is trained on the input-output pairs of POD coefficients $\{(\mathbf{x}_{\text{LF}}^{\text{coef}}(t_i, \mu_i), \mathbf{x}_{\text{HF}}^{\text{coef}}(t_i, \mu_i))\}_{i=1}^{N_\mu N_t^{\text{HF}}} \in \mathbb{R}^{N_{\text{POD}}}$ (the columns of the matrices $\mathbf{X}_{\text{LF}}^{\text{coef}}$ and $\mathbf{X}_{\text{HF}}^{\text{coef}}$, respectively, for each time-parameter instance (t_i, μ_i)). Using the Adam [70] algorithm, the LF to HF

mapping \mathbf{f}_{NN} is determined by minimizing the mean squared error loss function as follows:

$$\Theta_{\text{NN}} = \arg \min_{\Theta_{\text{NN}}} \frac{1}{N_{\mu} N_t^{\text{HF}}} \sum_{i=1}^{N_{\mu} N_t^{\text{HF}}} \left\| \mathbf{x}_{\text{HF}}^{\text{coef}}(t_i, \mu_i) - \mathbf{f}_{\text{NN}}(t_i, \mu_i, \mathbf{x}_{\text{LF}}^{\text{coef}}(t_i, \mu_i); \Theta_{\text{NN}}) \right\|_2^2. \quad (5)$$

Long Short-Term Memory (LSTM) NNs for MF regression: In this work, an LSTM network model is used for the time evolution of HF POD coefficients $\mathbf{x}_{\text{HF}}^{\text{coef}}$ by providing the corresponding sequence of LF coefficients $\mathbf{x}_{\text{LF}}^{\text{coef}}$ in the inputs together with the parameters μ and time t . By including time as an additional input along with LF coefficients $\mathbf{x}_{\text{LF}}^{\text{coef}}$ and parameters μ , the LSTM network features an explicit dependency on time that allows predictions at arbitrary time points. As shown in Fig. 2, an LSTM unit has two latent recurrent states \mathbf{c} and \mathbf{h} . At each time-step n , $(\mathbf{c}_n, \mathbf{h}_n)$ are recurrently updated from the combination of their last-step states $(\mathbf{c}_{n-1}, \mathbf{h}_{n-1})$ and the input values $\mathbf{x}_n = [t_n, \mu, \mathbf{x}_{\text{LF}}^{\text{coef}}(t_n, \mu)]$ through a three-fold *gate mechanism* with a *forget gate* \mathbf{G}_f , a *update gate* \mathbf{G}_u , and a *output gate* \mathbf{G}_o . The current state $\mathbf{h}_n \equiv \mathbf{f}_{\text{NN}}(t_n, \mu, \mathbf{x}_{\text{LF}}^{\text{coef}}(t_n, \mu))$ is produced as the unit's output \mathbf{y}_n to approximate $\mathbf{x}_{\text{HF}}^{\text{coef}}(t_n, \mu)$ (refer to (5)). The gate mechanism is the core of an LSTM unit, designed to allow a refined memory management by weighting the contribution of past and present information, and to overcome major drawbacks of recurrent networks such as exploding and vanishing gradients. The recurrent states are updated as follows:

$$\mathbf{c}_n = \mathbf{G}_f \circ \mathbf{c}_{n-1} + \mathbf{G}_u \circ \tilde{\mathbf{c}}_n, \quad \mathbf{h}_n = \mathbf{G}_o \circ \tanh \mathbf{c}_n \equiv \mathbf{y}_n,$$

in which the operation \circ denotes an element-wise product, $\tilde{\mathbf{c}}_n = \tanh(\mathbf{W}_c[\mathbf{h}_{n-1}, \mathbf{x}_n] + \mathbf{b}_c)$ represents the new candidate state to replace \mathbf{c}_{n-1} , and the gates $\{\mathbf{G}_i\}_{i \in \{f, g, u\}}$ are defined as $\mathbf{G}_i = \sigma(\mathbf{W}_i[\mathbf{h}_{n-1}, \mathbf{x}_n] + \mathbf{b}_i)$. Here, $\{\mathbf{W}_c, \mathbf{W}_f, \mathbf{W}_u, \mathbf{W}_o\}$ and $\{\mathbf{b}_c, \mathbf{b}_f, \mathbf{b}_u, \mathbf{b}_o\}$ are the trainable parameters – weights and biases – of the LSTM unit to be included in Θ_{NN} , and σ denotes the sigmoid activation function.

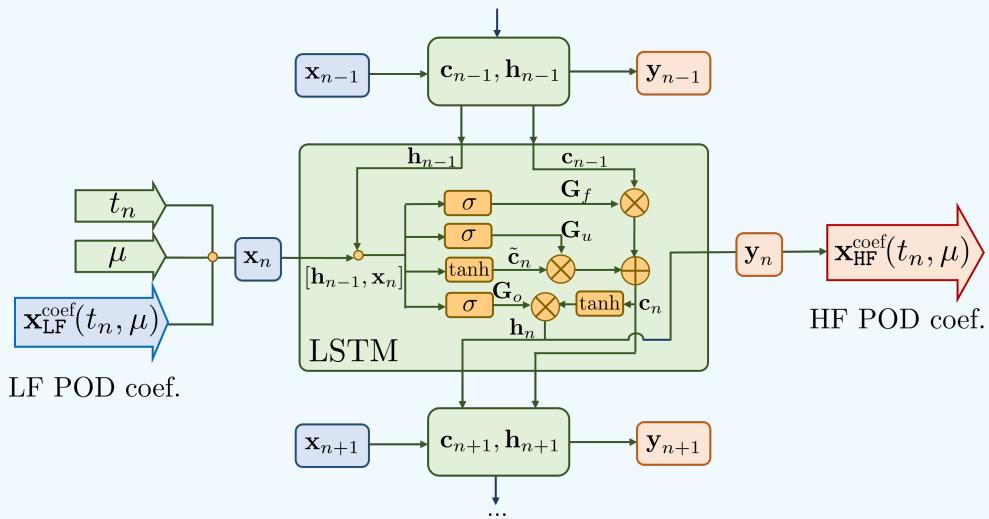


Figure 2: The visualization of an LSTM unit with its input-output setting [71, 62].

Since LSTM units are recurrent NNs that deal with sequential data, in practice, training data are grouped in batch subsequences with shape $n_{\text{batch}} \times K \times N_{\text{POD}}$, in which n_{batch} is the batch size and K is the length of batch subsequences. To determine the NN model's hyperparameters, including the network architecture (i.e., the numbers of layers and nodes in each layer) and the parameters associated to model training (e.g., the optimizer learning rate), we use a Bayesian optimization technique [72] implemented by the Python package Hyperopt [73].

2.2. Online testing

Once the MF-POD model is trained offline, it can be used online to efficiently evaluate PDE solutions for unseen parameter locations and longer time horizons, all at a very limited computational cost. As

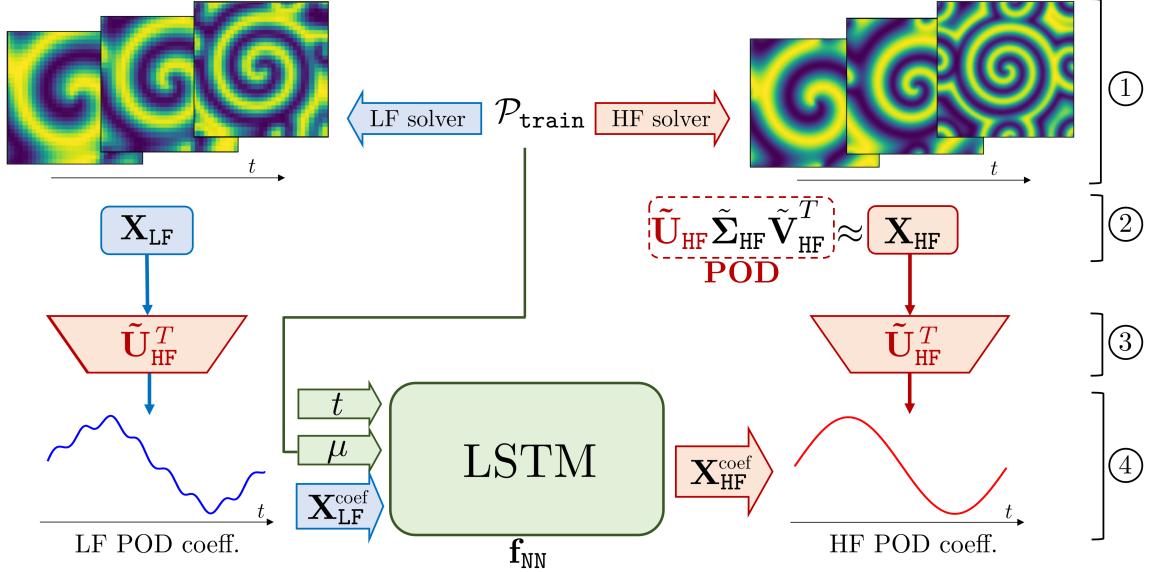


Figure 3: Offline training process of the proposed MF-POD strategy. The steps are ① generating multi-fidelity training datasets, ② dimensionality reduction via POD, ③ computing POD coefficients via direct projection, and ④ training LSTM surrogate model.

illustrated in Fig. 1, for new parameter configurations of interest, we compute the LF solution \mathbf{x}_{LF} and project it onto the POD basis to obtain the corresponding reduced states given by the projection coefficients $\mathbf{x}_{\text{LF}}^{\text{coef}}$. These coefficients are passed as inputs to the MF LSTM model and hence mapped to the sequences of POD coefficients towards the HF level, obtaining $\hat{\mathbf{x}}_{\text{HF}}^{\text{coef}}$, from which the high-resolution fields $\hat{\mathbf{x}}_{\text{HF}}$ can be reconstructed. Thus, the proposed reduced-order surrogate model allows to provide an approximation of HF simulations at a cheap online price of LF run-times and, moreover, to infer the future states of HF solutions from the LF solutions evaluated forward in time. This enables reliable long-term forecast without querying the expensive full-order model at all.

2.3. Metrics for performance evaluation

In the following sections we will assess the performance of MF-POD in several numerical examples. To highlight the advantages of MF-POD over HF solvers in saving computational costs, as well as the gain in accuracy in comparison to LF solvers, we report here the metrics for performance evaluation in terms of computational time and errors.

- Computational time is recorded for the complete time evolution of interest, averaged over the testing parameter instances. For LF and HF solutions, this means the run time of respective solvers, while for the MF model, it refers to the evaluations in the *online testing* procedure. Percentages are computed with respect to the HF time.
- Relative error with respect to the HF reference solution \mathbf{x}_{HF} is evaluated for both the MF solution $\hat{\mathbf{x}}_{\text{HF}}$ and the *lifted LF input* \mathbf{x}_{LF} , as an average over the test set:

$$\text{err}_{\text{MF-POD}}\% = \frac{100\%}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \frac{\|\mathbf{x}_{\text{HF}}(t_i, \mu_i) - \hat{\mathbf{x}}_{\text{HF}}(t_i, \mu_i)\|_2}{\|\mathbf{x}_{\text{HF}}(t_i, \mu_i)\|_2}, \quad \text{err}_{\text{LF}}\% = \frac{100\%}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \frac{\|\mathbf{x}_{\text{HF}}(t_i, \mu_i) - \mathbf{x}_{\text{LF}}(t_i, \mu_i)\|_2}{\|\mathbf{x}_{\text{HF}}(t_i, \mu_i)\|_2}, \quad (6)$$

where N_{test} is the number of time-parameter combinations in the test set.

3. Numerical example (I): Reaction-diffusion problem

We consider a lambda-omega reaction-diffusion system governed by the following equations

$$\begin{aligned}\dot{u} &= \left(1 - (u^2 + v^2)\right) u + \mu (u^2 + v^2) v + d (u_{xx} + u_{yy}), \\ \dot{v} &= -\mu (u^2 + v^2) u + \left(1 - (u^2 + v^2)\right) v + d (v_{xx} + v_{yy}),\end{aligned}\quad (7)$$

defined over a spatial domain $(x, y) \in [-L, L]^2$ for $L = 20$ and a time span $t \in [0, T]$ for $T = 80$, where μ and d are parameters that respectively regulate the reaction and diffusion behaviors of the system. We prescribe periodic boundary conditions, and the initial condition is defined as

$$u(x, y, 0) = v(x, y, 0) = \tanh\left(\sqrt{x^2 + y^2} \cos\left((x + iy) - \sqrt{x^2 + y^2}\right)\right).$$

The solution $[u, v]^T(x, y, t)$ to problem (7) represents two oscillating modes which generate spiral waves, representing an attracting limit cycle in the state space.

Our goal is to approximate the solution components u and v as functions of the varying reaction parameter $\mu \in \mathcal{P} = [0.5, 1.5]$ with a fixed diffusion coefficient $d = 0.05$. We employ the MF-POD method to efficiently evaluate high-resolution solutions over the whole time span with certain parametric variation, with the aid of cheaply obtained, low-resolution LF approximations. To reduce offline computational costs, we train the MF-POD model with a small amount of expensive HF solution data computed at a limited set of parameter locations over a shorter time horizon $T_{\text{train}} = 40 < T$, while leveraging LF solutions generated on a coarse spatial mesh with a corrupted value of diffusion coefficient.

3.1. Multi-fidelity setting

Both HF and LF solution datasets are constructed by solving the PDEs (7) using the Fourier spectral method [74] with time step $\Delta t = 0.05$. The two fidelity levels are defined as follows, and the difference between LF and HF solutions is shown in Fig. 4.

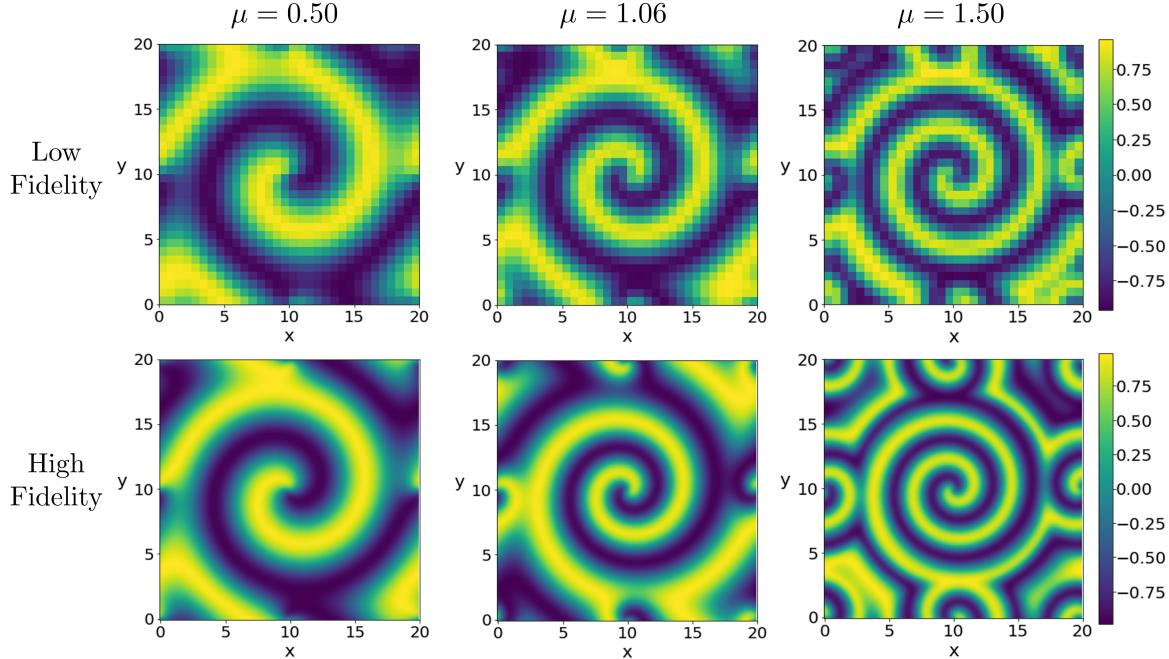


Figure 4: LF (above) and HF (below) solution snapshots of u at three parameter locations and $t = T_{\text{train}} = 40$ in example (I). Besides the difference in spatial resolution, the corrupted diffusion coefficient in the LF model leads to different shapes of spiral waves than those in the HF solutions.

- LF solution data are generated on a coarse equispaced spatial grid with $n_{\text{LF}} = 32$ points in each direction, while a fine grid with $n_{\text{HF}} = 100$ is adopted for the HF data.
- LF solutions are evaluated at a corrupted diffusion coefficient $d_{\text{LF}} = 0.1$, instead of $d_{\text{HF}} = d = 0.05$. This represents a bias in the LF modeling in terms of the physical property of viscosity.
- HF data are only available over a limited time window $[0, T_{\text{train}}]$ with $T_{\text{train}} = 40 < T = 80$. We hence aim to extrapolate for a same-length time window beyond that covered by the HF training data.
- Training data on both fidelity levels are computed for a small number of parameter instances $\mu \in \mathcal{P} = [0.5, 1.5]$. In particular, $N_\mu = 10$ μ -values are selected over an equispaced grid of \mathcal{P} .

To apply the MF-POD method, we perform the POD reduction on the HF snapshots and retain the first $N_{\text{POD}} = 9$ modes. LF data are lifted to the HF spatial dimensionality via nearest-neighbor interpolation. For both fidelity levels, POD coefficients are computed by projecting the data onto the reduced basis and then fed to the LSTM neural network for training. An LSTM network with 1 layer, 69 nodes and subsequences of length $K = 133$ has been considered, and these hyperparameters are optimized by Hyperopt [75]. Once the training phase is concluded, we run the LF solvers to efficiently evolve LF solutions over the whole time window $[0, T]$, and test the MF-POD accuracy in estimating the HF solutions over an unseen set of $N_\mu^{\text{test}} = 25$ equispaced parameter locations over \mathcal{P} .

3.2. Results

For a few extrapolated time instances at unseen testing values of the reaction coefficient μ , we show in Fig. 5 the reconstruction of entire solution field predicted by MF-POD, compared with both HF reference and LF input. The proposed MF strategy is able to recover high resolution and correct the prediction of system behavior from inaccurate LF solutions. Table 1 provides a quantitative comparison of computational time and relative errors among LF, HF, and MF solutions, highlighting the good performance of MF-POD.

The combination of POD reduction and MF regression with LSTM proves to be effective in this example. The POD at the HF level extracts a global basis that represents predominant spatial patterns of the spiral wave propagation, and the LSTM network model approximates the corresponding expansion coefficients that describe how these spatial modes evolve over time. The expressive power of LSTM neural networks is crucial not only for the approximation of POD coefficients' strongly nonlinear dependency on time and parameters, but also for the correlation modeling between the LF and HF levels, especially considering that such an unknown correlation is parameter-dependent and potentially highly complex. In Fig. 6, we depict the evolution of several POD coefficients predicted by the LSTM network. We notice that MF-POD provides very accurate estimation of the time-dependent POD coefficients at unseen parameter locations, not only in the training time interval $[0, 40]$, but also for the future states over $[40, 80]$, as the MF-predicted coefficients match very well with the HF reference and present a significant improvement compared to the LF level. These results over twice the length of the training time span indicate the capability of MF-POD in long-term forecast.

As a final remark, we analyze how the proposed algorithm's performance relies on the *quantity* of HF training data and the *quality* of the LF data. The amount of HF data used in the training stage represents the main contribution to the offline computational costs, while the quality of LF data determines the degree of correlation between LF and HF levels. Fig. 7 shows how the relative test error of the MF-POD model decreases as the number of HF data increases, when the quality of LF model also varies. In particular, the LF data quality is regulated by corrupting the existing LF samples with Gaussian noises whose intensities are 20% and 40% of the LF signal's amplitude, respectively. We note that the proposed MF-POD method is robust with respect to the quality of the LF data, because the performance with noisy data is comparable to the noise-free case. We also observe that, even with a single HF training datum, the MF method outperforms all LF models; with only 10 HF samples, our method effectively reduces the relative error to less than 20%, considering that the LF relative errors exceed 100%.

3.3. The need of LSTM networks

Although the presented results have shown the proposed method's capability in mapping LF solutions to HF ones, the need to create this map via an LSTM neural network, instead of alternative regression

techniques, has not been made clear yet. To show more evidence in this regard, we report further results obtained by replacing the LSTM neural network with a “static” feed-forward (FF) neural network (i.e., without t in the inputs) in the MF regression step (Step 4 offline). In this scenario, the LF POD coefficients are mapped to their HF counterparts instant by instant, instead of being processed as time series. In Fig. 8, we report such computed spatial reconstructions and their absolute errors at the same time-parameter test locations considered before. We note a clear worsening in reconstruction accuracy and a significant increase in approximation errors, compared with the predictions by MF-POD with LSTM networks reported in Fig. 5. This highlights that the MF regression with LSTM networks allows a better detection of temporal patterns in time series and nonlinear correlations between datasets at different fidelity levels, guaranteeing an improved predictive performance. Note that incorporating time as an input of the FF network would merely establish an explicit dependency on time, yet this adjustment cannot enhance the regression performance meaningfully. Indeed FF networks lack the capability to retain memory across time steps, and thus they are unable to preserve the temporal dependency within the data. We refer to [62] for further discussions on these aspects.

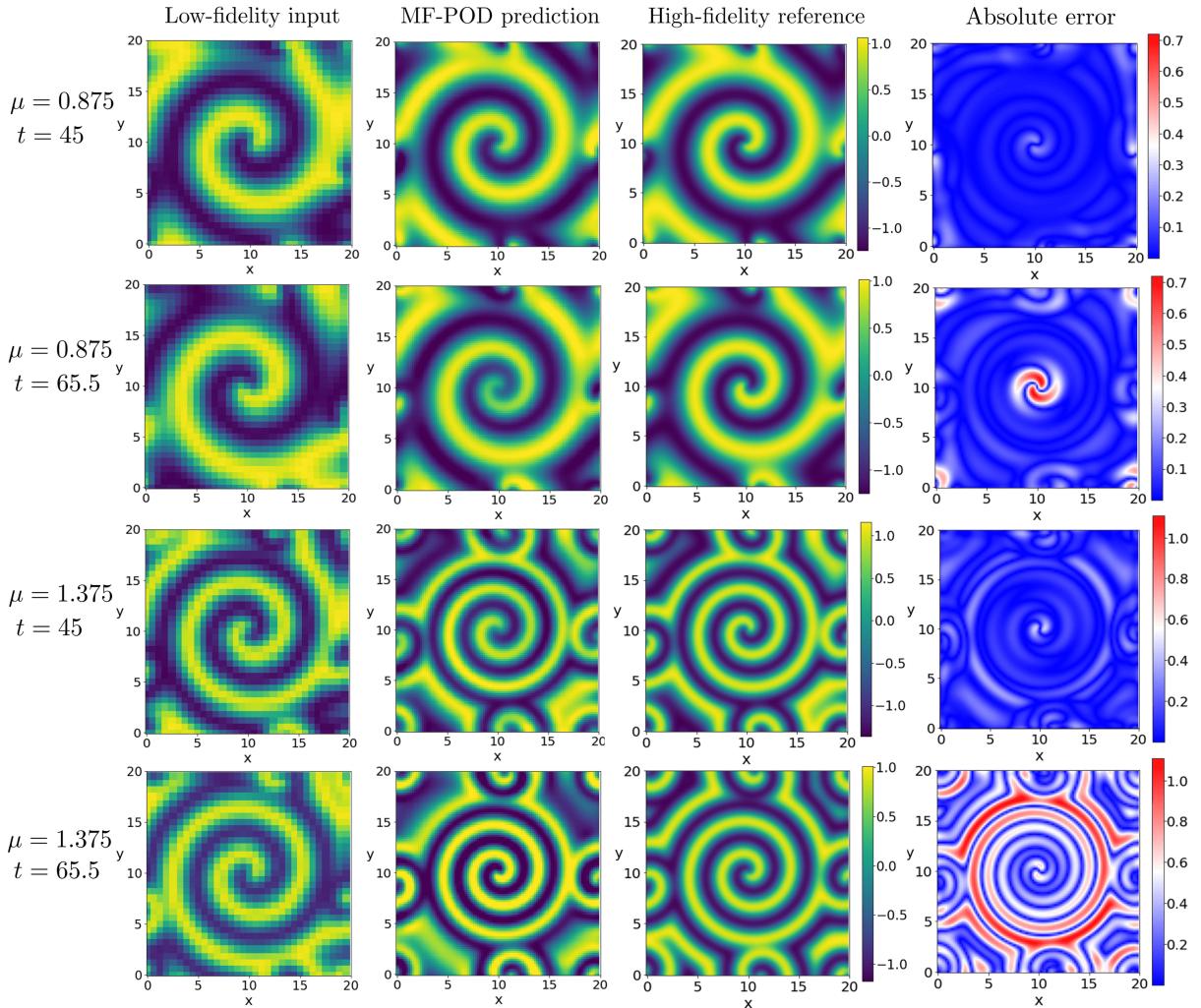


Figure 5: Comparison of solution fields in example (I) among the approximation by MF-POD, the corresponding LF input, and the HF ground truth (used as reference). The snapshots refer to two extrapolated time instances $t \in \{45, 65.5\}$ (being $T_{\text{train}} = 40$ the end of HF training coverage) for two testing values of reaction parameters $\mu \in \{0.875, 1.375\}$ that are unseen during the training. Absolute error shows the discrepancy between the MF solution and the HF reference.

Table 1: Comparison of computational time and accuracy in example (I). See section 2.3 for the definition of metrics.

	Low-fidelity input	MF-POD predicted	High-fidelity reference
Computational time	0.58s (2.70%)	1.89s (8.79%)	21.49s (100%)
Relative error	111%	16%	-

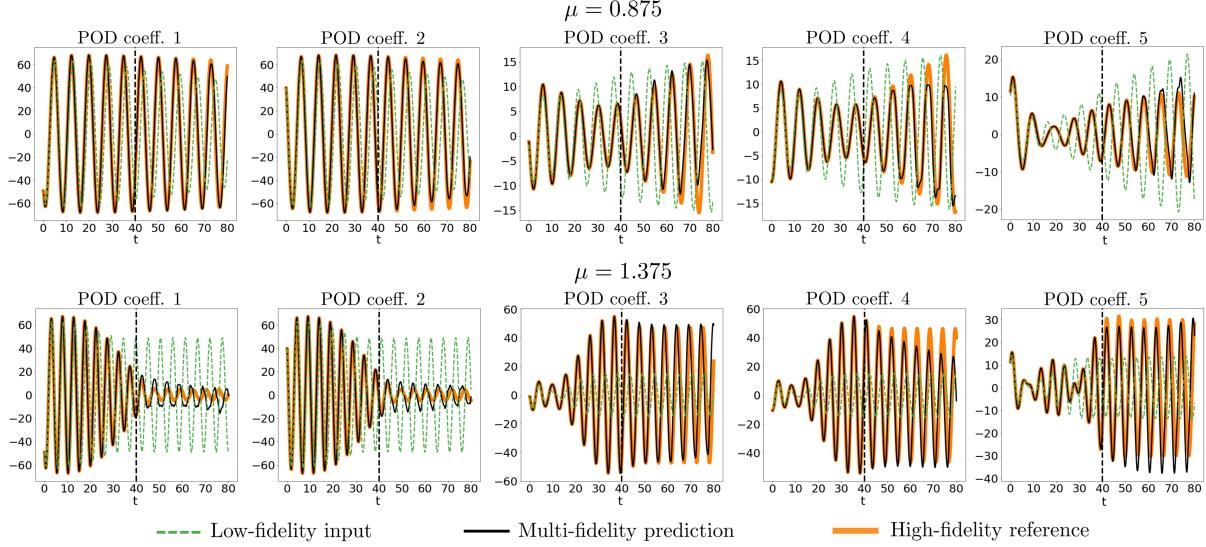


Figure 6: Comparison of the time evolution of POD coefficients in example (I) among MF solution, LF input, and HF reference at testing parameter instances $\mu = 0.875$ (above) and 1.375 (below). The dashed line at $T_{\text{train}} = 40$ indicates the end time of HF data coverage in training, i.e., no HF information is available over $t \in [T_{\text{train}}, T] = [40, 80]$.

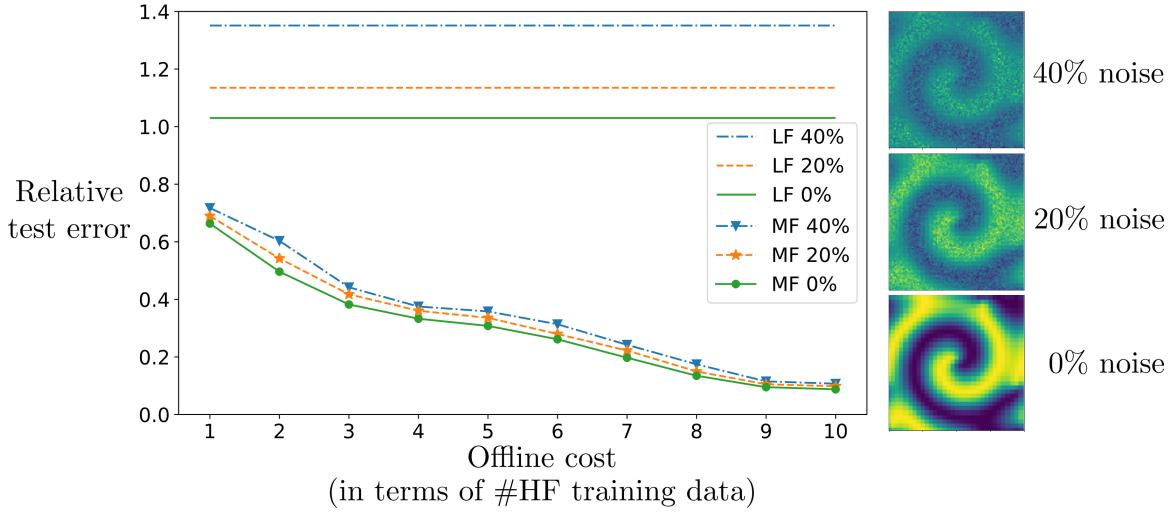


Figure 7: Relative test errors versus the quantity of HF data used during training (up to $N_\mu = 10$). The MF-POD model is retrained when an additional HF datum is integrated into the training set. Three LF datasets are considered: noise-free (in green), and those corrupted by 20% (in orange) and 40% noise (in black). The horizontal lines represent the corresponding LF test errors, which are independent of the number of HF data. Additionally, LF snapshots with different noise levels are also presented.

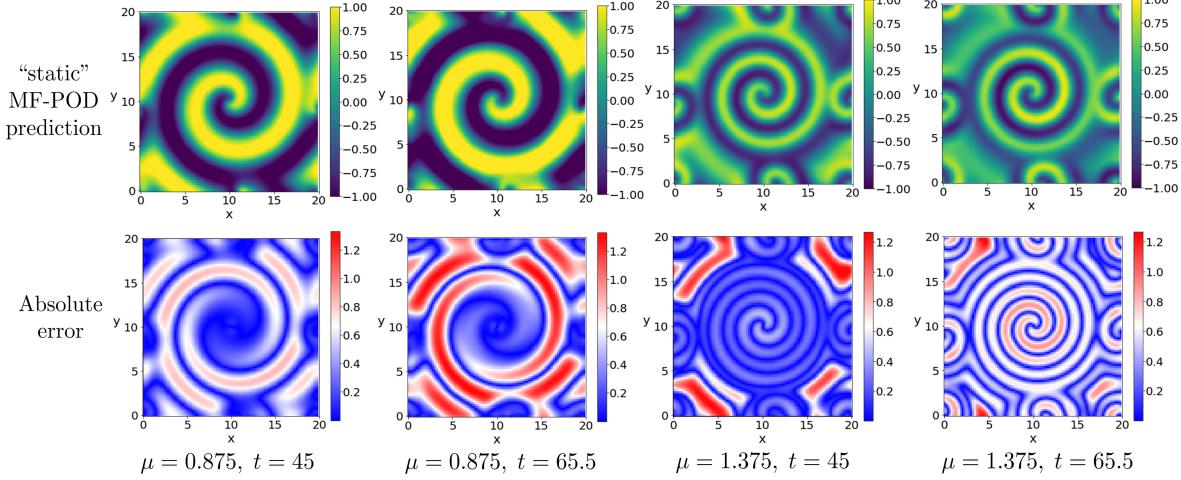


Figure 8: Predicted solution fields in example (I) by MF-POD with “static” feed-forward neural networks, instead of LSTM networks. The snapshots refer to two extrapolated time instances $t \in \{45, 65.5\}$ for two testing parameter values $\mu \in \{0.875, 1.375\}$. Absolute error shows the discrepancy between the “static” MF solution and the HF reference.

4. Numerical example (II): Advection-diffusion in shallow water

In this example, we consider an advection-diffusion problem describing a fluid motion in the shallow water limit [76] given by

$$\frac{\partial \omega}{\partial t} + \mu \left(\frac{\partial \psi}{\partial x} \frac{\partial \omega}{\partial y} - \frac{\partial \psi}{\partial y} \frac{\partial \omega}{\partial x} \right) = d \nabla^2 \omega, \quad (8a)$$

$$\nabla^2 \psi = \omega, \quad (8b)$$

defined over a spatial domain $(x, y) \in [-L, L]^2$ and a time span $t \in [0, T]$. Here $\omega(x, y, t)$ and $\psi(x, y, t)$ represent the vorticity and streamfunction, respectively, $\nabla^2 = \partial_x^2 + \partial_y^2$ is the two-dimensional Laplacian, $d = 0.001$ is the diffusion coefficient, and we take $L = 10$ and $T = 20$. We assume periodic boundary conditions and a stretched Gaussian function as the initial condition of vorticity:

$$\omega(x, y, 0) = \exp \left(-2x^2 - \frac{y^2}{20} \right), \quad (x, y) \in [-L, L]^2. \quad (9)$$

We are interested in approximating the time-dependent vorticity field ω as the parameter μ varies over $\mathcal{P} = [1, 5]$. The general procedure for solving the system (8) numerically is to (i) compute the streamfunction $\psi(x, y, 0)$ at $t = 0$ by solving the elliptic equation (8b) given the initial condition of vorticity (9), (ii) use a time-stepper on (8a) to advance the vorticity ω by one step Δt , and (iii) repeat (i) and (ii) starting with the updated $\omega(x, y, t + \Delta t)$ until the final time T is reached. A small time-step Δt implies a large number of iterations, and a fine spatial discretization requires solving a large linear system in (i). Moreover, this procedure must be repeated for each instance of parameter μ , which makes HF simulations impractical. We use the proposed MF-POD method to relieve these heavy computational burdens. In particular, we consider LF solutions on a coarser spatial grid with larger time-steps, while only evaluating a limited number of HF solution data at several parameter locations for MF-POD training.

4.1. Multi-fidelity setting

As in the previous example, we adopt two fidelity levels that differ in the spatial resolution of discretization, but also consider larger steps of time integration in the LF model than its HF counterpart. Specifically, the HF (resp. LF) solution data are generated via the Fourier spectral method on an equispaced spatial grid of $n_{\text{HF}} = 200$ (resp. $n_{\text{LF}} = 50$) nodal points along each direction with time-step size $\Delta t_{\text{HF}} = 0.25$ (resp. $\Delta t_{\text{LF}} = 1.00$). The bi-fidelity training data cover a limited time window $[0, T_{\text{train}}] = [0, 12]$ with $T_{\text{train}} < T = 20$, only at a small number ($N_\mu = 5$) of parameters locations equispaced over \mathcal{P} . Thereafter,

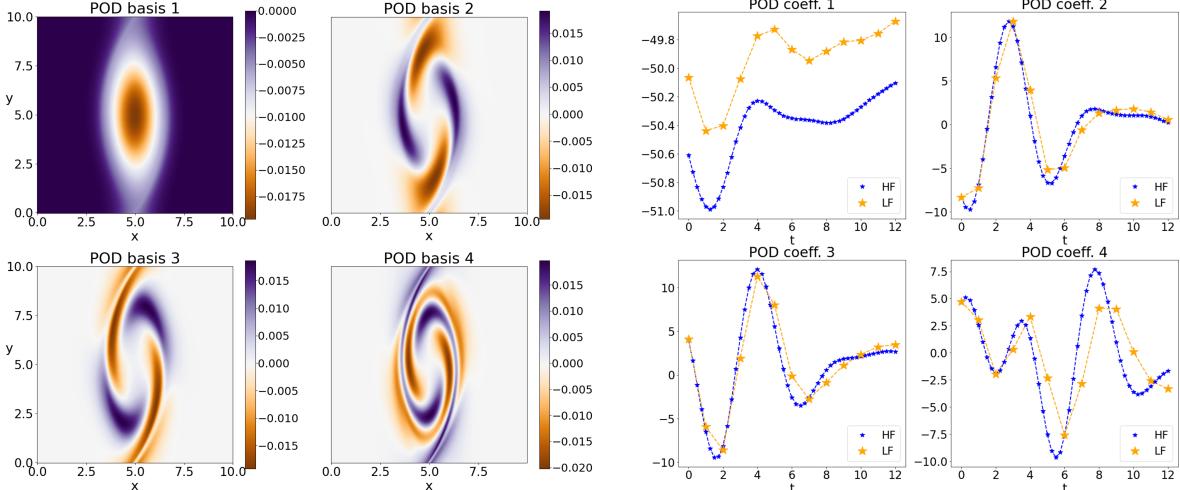


Figure 9: The first four spatial modes identified by the POD (left) and the bi-fidelity data of corresponding expansion coefficients (right) in example (II) ($\mu = 3$). Note that the training data only cover a limited time window $[0, T_{\text{train}}] = [0, 12]$ with $T_{\text{train}} < T = 20$.

the training data are fed to the offline algorithm of MF-POD. In this example, we retain the first $N_{\text{POD}} = 17$ POD modes and consider linear interpolation both in time and space to lift LF data to the HF dimensionality for the POD projection. The first four POD modes and the corresponding time-dependent expansion coefficients at $\mu = 3$ are depicted in Fig. 9. Regarding the network architecture, an LSTM with 1 layer, 52 nodes and $K = 39$ is employed. The predictive performance of MF-POD is tested for $N_{\mu}^{\text{test}} = 4$ unseen parameters values $\mu \in \{1.5, 2.5, 3.5, 4.5\} \subset \mathcal{P}$.

4.2. Results

The MF-POD method allows us to create an efficient, reliable, low-dimensional surrogate model for the advection-diffusion in shallow water. As presented in Table 2, parametric solutions can be evaluated with MF-POD at a computational cost comparable to that of the LF, yet with a significant improvement in accuracy as highlighted by the substantial reduction in predictive error.

This example is challenging for data-driven surrogate modeling, especially in terms of the approximation of POD coefficients describing a parametric coherent structure that propagates over time (i.e., the wave-type phenomena). If only the time-parameter inputs are accounted for, the regression may very likely suffer from difficulties with limited data, leading to poor generalization performance, and extrapolation beyond the training time window would thus be barely possible. The proposed MF-POD method mitigates such technical risks by incorporating physically meaningful LF solutions, which can capture unseen dynamical characteristics and inform the solution approximation via the LF-to-HF mapping on POD coefficients.

To illustrate the effectiveness of MF-POD in predicting forward in time and for new parameter instances simultaneously, we depict in Fig. 10 the evolution of vorticity ω at an unseen testing parameter value $\mu = 3.5$. The physical behavior is accurately predicted by the MF-POD approach, even with very low resolution on the LF level. As time evolves, the wave propagation presents finer spatial patterns, and the temporal extrapolation becomes more complex, especially considering that no HF information is available after $t = T_{\text{train}} = 12$, and the LF input does not have sufficient resolution to perfectly detect fine patterns. Nevertheless, the proposed method allows to accurately extrapolate (see, e.g., the prediction at $t = 15$ in

Table 2: Comparison of computational time and accuracy in example (II).

	Low-fidelity input	MF-POD predicted	High-fidelity reference
Computational time	3.23s (1.03%)	4.12s (1.32%)	311s (100%)
Relative error	12.8%	3.35%	-

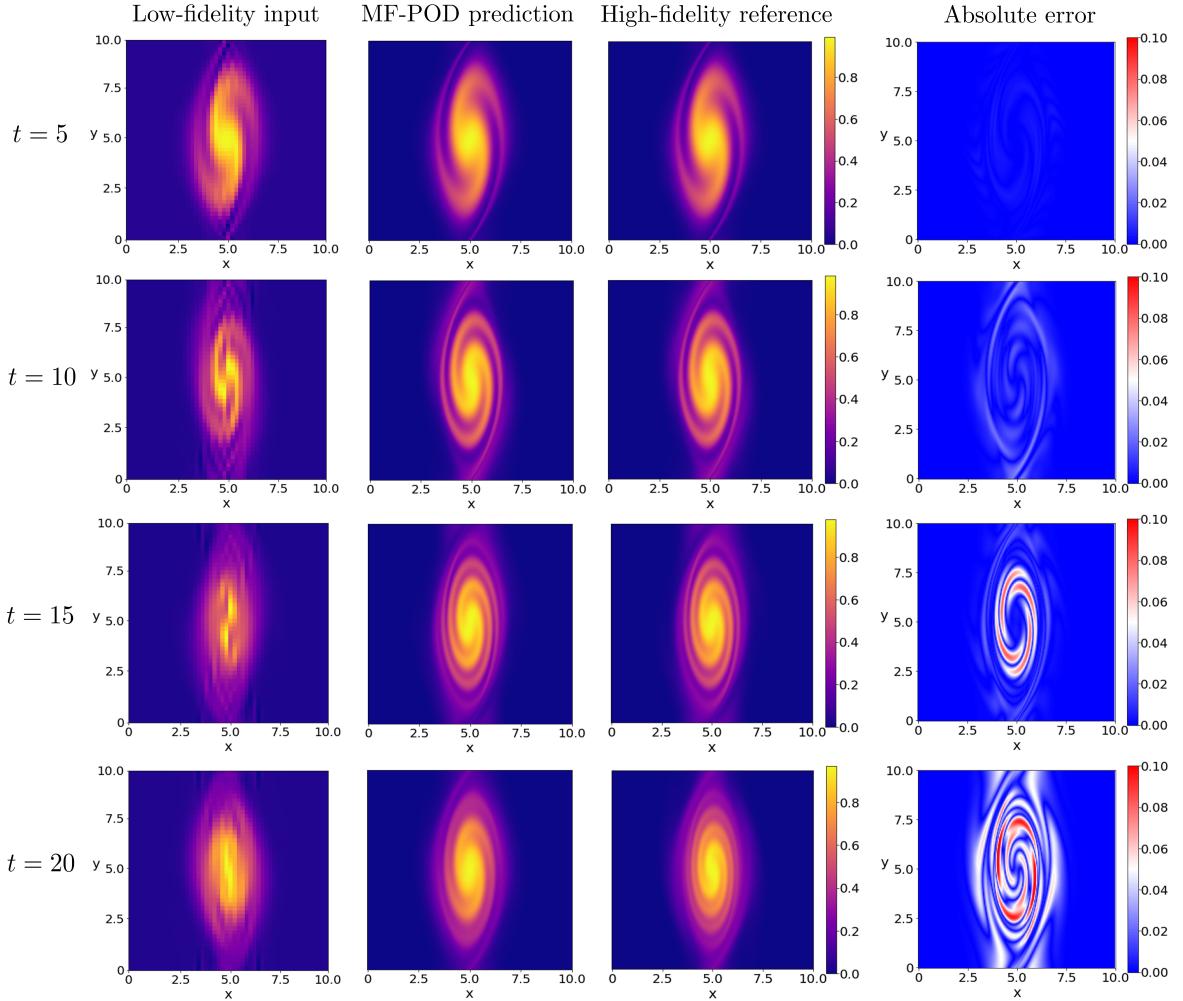


Figure 10: Comparison of solution fields in example (II) among the approximation by MF-POD, the corresponding LF input, and the HF ground truth (used as reference). The snapshots refer to two training time instances $t \in \{5, 10\}$ and two extrapolated time instances $t \in \{15, 20\}$ (being $T_{\text{train}} = 12$ the end of HF training coverage) for the testing parameter $\mu = 3.5$ that is unseen during the training. Absolute error shows the discrepancy between the MF solution and the HF reference.

Fig. 10) up to time $T = 20$, showing the power of MF data fusion, though such power is not unlimited as the predictive error in wave propagation is no longer negligible after T .

5. Numerical example (III): Navier-Stokes equations

For the last example, we consider a two-dimensional fluid flow around a cylinder — a benchmark problem in computational fluid dynamics. Our goal is to efficiently approximate the velocity and pressure fields of a viscous, incompressible Newtonian fluid flow as its Reynolds number varies. The problem is governed by the following Navier-Stokes equations

$$\begin{aligned} \rho \frac{\partial \mathbf{v}}{\partial t} - \rho \mathbf{v} \cdot \nabla \mathbf{v} - \nabla \cdot \boldsymbol{\sigma}(\mathbf{v}, p) &= \mathbf{0}, & (\mathbf{x}, t) \in \Omega \times (0, T), \\ \nabla \cdot \mathbf{v} &= 0, & (\mathbf{x}, t) \in \Omega \times (0, T), \end{aligned} \quad (10)$$

where $\mathbf{v}(\mathbf{x}, t)$ and $p(\mathbf{x}, t)$ represent the velocity and pressure field, respectively, and $\rho = 1.0 \text{ kg/m}^3$ is the fluid density, $\boldsymbol{\sigma}(\mathbf{v}, p) = -p\mathbf{I} + 2\nu\boldsymbol{\epsilon}(\mathbf{v})$ is the stress tensor with $\boldsymbol{\epsilon}(\mathbf{v})$ denoting the strain tensor. The kinematic viscosity is defined as $\nu = 1/Re$ [77], in which the Reynolds number, Re , is the system parameter of interest.

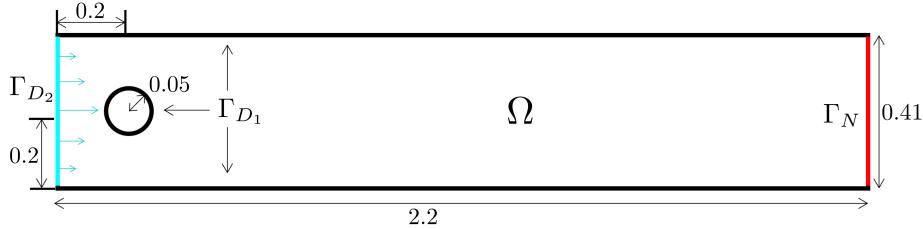


Figure 11: Geometry for a 2D channel flow around a cylinder. All lengths are measured in meters.

As initial conditions we consider the fluid at rest

$$\mathbf{v}(\mathbf{x}, 0) = \mathbf{0}, \quad \mathbf{x} \in \Omega,$$

and we provide the following boundary conditions for the domain $\Omega = (0, 2.2) \times (0, 0.41) \setminus B_r(0.2, 0.2)$ ($r = 0.05$), representing a 2D channel with a cylindrical obstacle (see Fig. 11):

$$\begin{aligned} \mathbf{v} &= \mathbf{0}, & (\mathbf{x}, t) &\in \Gamma_{D_1} \times (0, T), \\ \mathbf{v} &= \mathbf{h}, & (\mathbf{x}, t) &\in \Gamma_{D_2} \times (0, T), \\ \boldsymbol{\sigma}(\mathbf{v}, p)\mathbf{n} &= \mathbf{0}, & (\mathbf{x}, t) &\in \Gamma_N \times (0, T), \end{aligned}$$

which include a no-slip condition on Γ_{D_1} , a parabolic inflow

$$\mathbf{h}(\mathbf{x}, t) = \left(\frac{4U(t)x_2(0.41 - x_2)}{0.41^2}, 0 \right), \quad U(t) = \begin{cases} 0.75(1 - \cos(\pi t)), & t < 1 \\ 1.5, & t \geq 1 \end{cases}$$

on the inlet Γ_{D_2} , and an open boundary condition at the outlet Γ_N .

In the present study, we consider $\mu = Re \in \mathcal{P} = [30, 100]$. When $Re < 49$, the flow presents a *steady* behavior; for larger values of Reynolds number, the flow transitions to an *unsteady* state and a pair of vortices form in the wake of the cylinder, oscillating periodically between top and bottom sides [78, 79]. As Re varies across steady and unsteady ranges, we use the MF-POD method to construct efficient MF surrogate models for the velocity and pressure fields up to $T = 18$ s, at which time the fluid is fully developed and presents a periodic behavior.

5.1. Multi-fidelity setting

Solution data are generated through a finite element approximation of (10) with the backward differentiation formula provided by the MATLAB library `redbKIT` [80]. We consider a training set computed over a short time window $T_{\text{train}} = 12$ s $< T = 18$ s at $N_\mu = 15$ Reynolds numbers equispaced over \mathcal{P} . In this case, the difference between the fidelity levels incorporates all the characteristics considered so far in the previous examples, namely the choice of spatial mesh size and time step, as well as whether the physical characteristic Re is corrupted.

For the HF data, the total number of spatial degrees of freedom is $N_{\text{HF}}^{\text{dof}} = 73131$, obtained with quadratic finite elements for the velocity field and linear finite elements for the pressure field over a mesh with 16478 triangular elements and 8239 vertices, while the temporal discretization is with step size $\Delta T_{\text{train}} = 5$ ms. Instead, LF data are computed with a larger time step $\Delta T_{\text{LF}} = 50$ ms over a coarser mesh consisting of 7789 triangular elements and 3899 nodes, thus resulting in snapshots with $N_{\text{LF}}^{\text{dof}} = 34439$. The LF solutions can be computed with significantly reduced time in comparison with the HF ones. Moreover, we consider a corruption factor of $\alpha = 0.95$ that multiplies Re in the generation of LF data. For example, when $Re = 60$ for the HF level, the LF solution is evaluated at $\tilde{Re} = \alpha Re = 57$.

Once solution data are prepared, POD reduction is applied to the HF snapshots and the first $N_{\text{POD}} = 32$ modes are retained. LF data are lifted by linear interpolation over the spatial and temporal domains, and the POD coefficients, obtained by projecting the HF and lifted LF data onto the reduced basis, are then

passed to the MF LSTM network for training. In this example, an LSTM network with 1 layer, 64 nodes and subsequences of length $K = 158$ is considered.

5.2. Results

The MF-POD method is tested for approximating the fluid velocity and pressure fields at unseen parameter values $Re \in \{37, 48, 63, 78, 92\}$, while simultaneously extrapolating into the time window $[T_{\text{train}}, T] = [12 \text{ s}, 18 \text{ s}]$ over which no training data are provided. For a given new value of Re , the LF solution is evolved up to the final time of interest $T = 18 \text{ s}$, then the LF POD coefficients are computed and mapped through the LSTM network to their HF counterparts, from which the whole velocity and pressure fields are reconstructed.

In Table 3, we report the computational costs and predictive errors evaluated over the test set. Once again, we observe that the MF-POD method achieves advantages over the HF and LF models, by drastically reducing computational time while preserving a good accuracy, respectively. Moreover, in Fig. 12, we illustrate the MF predictive solutions for two testing parameter values $Re \in \{48, 63\}$ and one training value $Re = 50$ in comparison with their LF inputs and HF references. The testing values $Re \in \{48, 63\}$ correspond to two different regimes of the fluid flow — steady and unsteady, respectively. In both cases, the proposed model is able to capture correct fluid behaviors with a good accuracy. For Reynolds numbers close to the bifurcation ($Re = 49$) between the two regimes, a coarse approximation may lead to significant errors in simulating the dynamical behaviors of the fluid. For example, when $Re = 48$, we observe that the LF solution exhibits the onset of unsteady oscillatory phenomena, which are, instead, absent in the reference HF solution, which exhibits the expected steady behavior. This is even more evident by observing the temporal evolution of POD coefficients as shown in Fig. 13, where we notice that the LF solution features a non-negligible oscillatory contribution from the third and fourth POD coefficients; these latter are instead almost vanishing for the reference HF solution. We note that MF-POD is able to accurately recover the correct HF physical characteristics. In Fig. 12, the solutions for the training instance $Re = 50$ are also presented. This parameter value falls within the unsteady regime just past the bifurcation at $Re = 49$, in comparison with the aforementioned steady solution with $Re = 48$. The predicted MF-POD solution accurately portrays the onset of oscillatory phenomenon in the unsteady regime. This further demonstrates the proposed method's ability to preserve physical characteristics in the observational data.

Table 3: Comparison of computational time and accuracy in example (III).

	Low-fidelity input	MF-POD predicted	High-fidelity reference
Computational time	108s (4.05%)	111s (4.17%)	2664s (100%)
Relative error	18.5%	3.48%	-

6. Conclusions

We have developed a new reduced-order surrogate modeling method that relies on solution data from different fidelity levels to reduce computational costs while preserving predictive accuracy. Like traditional model reduction, the MF-POD architecture employs POD to extract a low-dimensional basis that approximates the solution manifold from a limited amount of HF data. An MF LSTM model is subsequently trained to infer the temporal evolution of the HF solution on the POD manifold from its LF counterpart. Once trained *offline*, the proposed model can be deployed *online* to generate new solutions that approach HF accuracy at the computational cost of LF evaluations. The advantages of the MF POD technique has been compared to both HF and LF methods directly through a diverse number of example PDEs with the results summarized in Fig. 14.

By incorporating physically meaningful LF data, MF-POD addresses the potential lack of physical consistency in purely data-driven methods, thereby ensuring interpretability and reliability in parametric generalization and temporal forecasting while still benefiting from its non-intrusive nature. A limitation of the proposed framework is the assumption of a strict hierarchy among fidelity levels, which implies that LF solutions are required at parameter configurations in which we are interested on the HF level. A further limitation is that this method has not fully benefited from the large availability of LF data. In fact, the

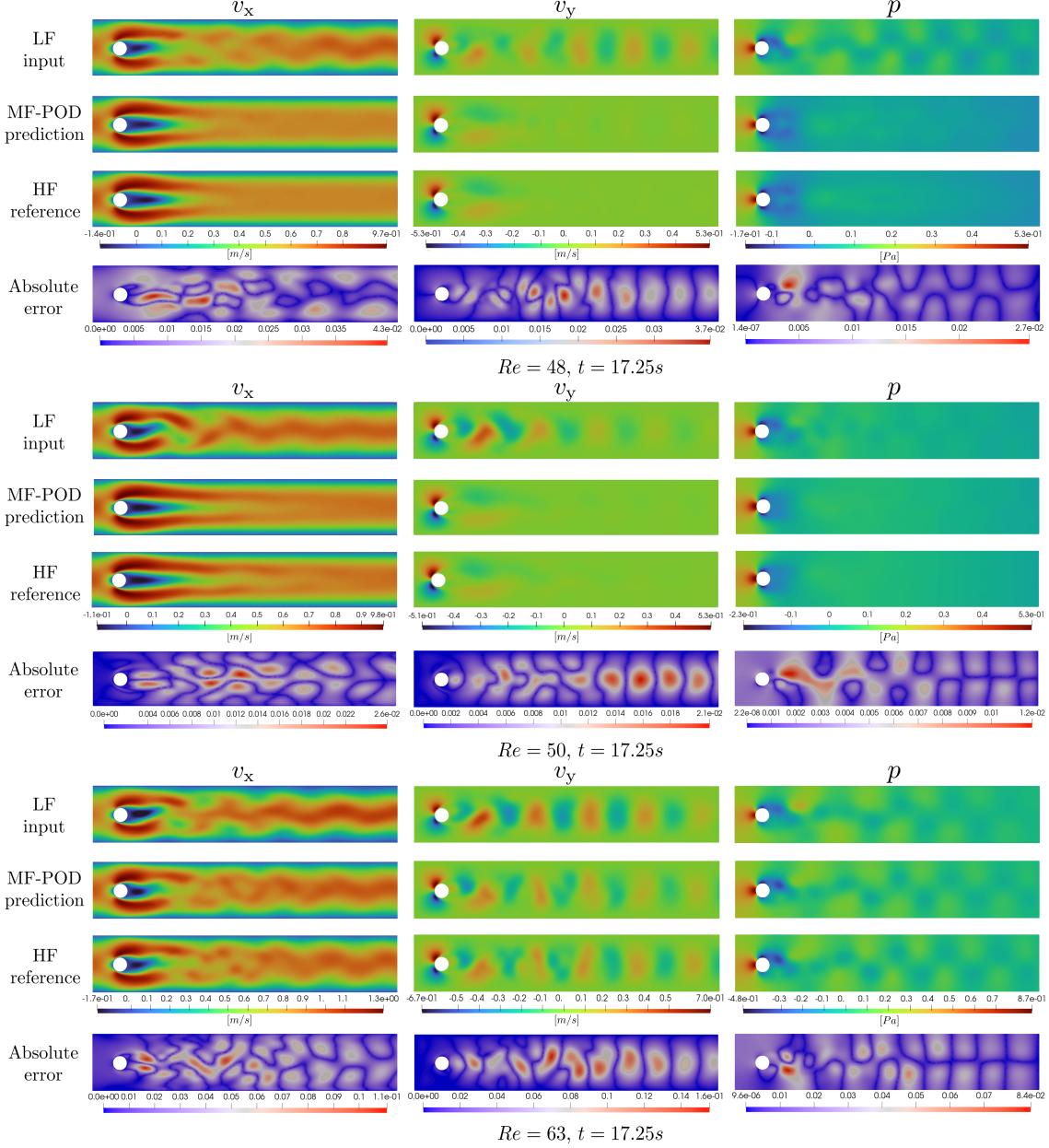


Figure 12: Comparison of solution fields in example (III) among the approximation by MF-POD, the corresponding LF input, and the HF ground truth (used as reference). The snapshots refer to the extrapolated time instance $t = 17.25$ s (being $T_{\text{train}} = 12$ s the end of HF training coverage) for two testing values of Reynolds number $Re = 48$ (top) and 64 (bottom), as well as for the training value $Re = 50$ (middle). Note that $Re = 48$ and 50 respectively represent the values right before and after the bifurcation (at $Re = 49$) between steady and unsteady regimes. The closest training Re values to the bifurcation are 45 and 50 . Absolute error shows the discrepancy between the MF solution and the HF reference.

MF LSTM networks are trained on the same amount of LF and HF data to learn the LF-to-HF mapping, sampled at the same parameter locations (see Fig. 3). Moreover, the low-dimensional basis is constructed solely with the HF data. Therefore, potential LF information could be additionally exploited to improve both the construction of reduced basis and the MF regression for the expansion coefficients.

However, the non-intrusiveness of the proposed method guarantees remarkable flexibility in numerical implementations, and enables seamless adaptation, extension, and enhancement of its individual components. For instance, the use of other techniques for reduced basis construction, which may be better tailored to suit specific applications of interest, can replace the POD reduction. Moreover, alternative recurrent neural

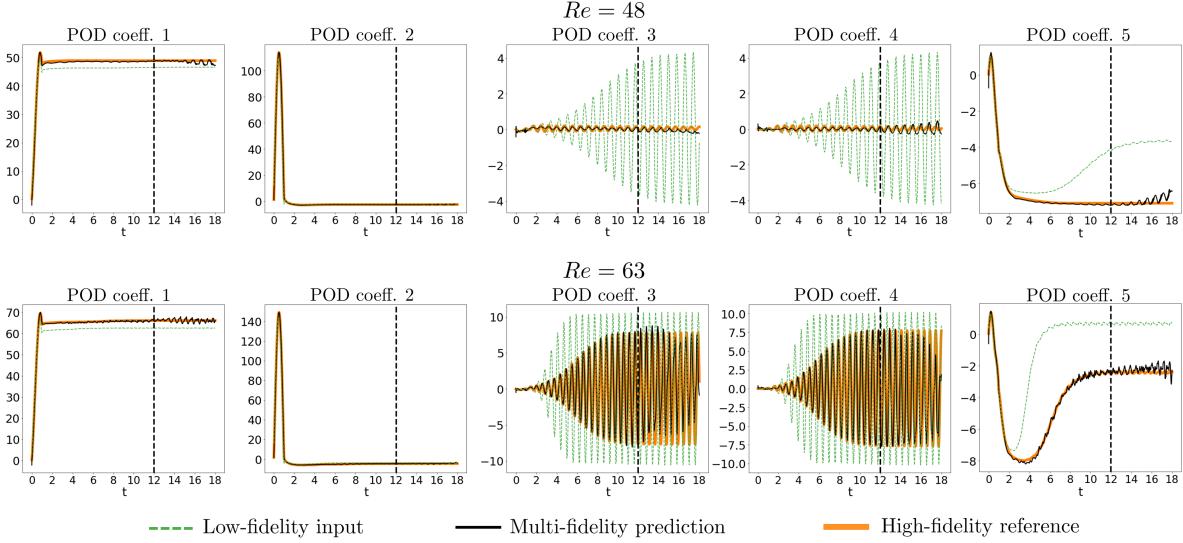


Figure 13: Comparison of the time evolution of POD coefficients in example (III) among MF solution, LF input, and HF reference at testing parameter instances $Re = 48$ (above) and 63 (below). The dashed line at $T_{\text{train}} = 12$ s indicates the end time of HF data coverage in training, i.e., no HF information is available over $t \in [T_{\text{train}}, T] = [12 \text{ s}, 18 \text{ s}]$.

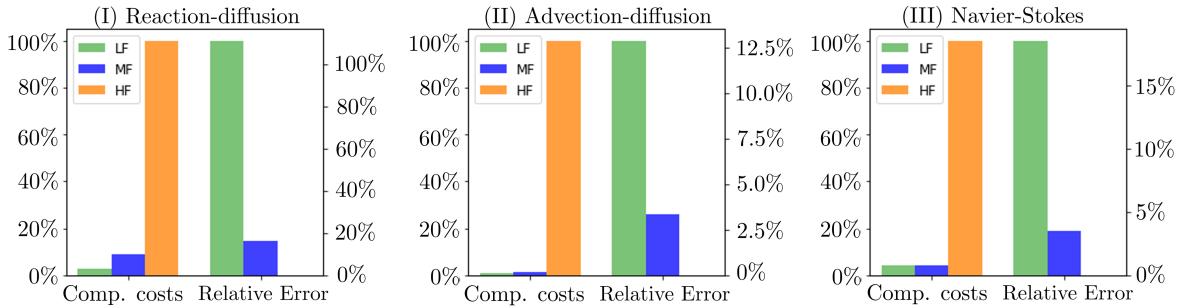


Figure 14: Comparison of the LF, HF, and the proposed MF approaches in terms of computational times and accuracy. For each considered example, we provide a plot with the visualizations of the values collected in Tables 1-2-3. Online computational costs are evaluated on a workstation with an MD Ryzen 9 5950X 16-core processor. Values of the computational costs are indicated on the left vertical axis as percentages with respect to the computational costs of running HF solvers, while relative prediction errors with respect to the HF reference solutions are indicated on the vertical axis on the right. We observe how the proposed MF method allows for a significant improvement in predictive accuracy with respect to the LF approach, while maintaining extremely limited online computational costs, in particular as compared to the HF ones.

network or other emerging architectures (e.g., transformers [81]) can be employed instead of the LSTM layers in the MF regression task. Such refinements will further empower the proposed method to evolve and excel in diverse physical simulations in computational science and engineering.

In particular, as a direction for future development, the presented framework can be naturally extended to real-life applications, in which experimental measurement data can be fused with synthetic (LF) simulation data to create surrogate models that produce accurate real-time predictions beyond the experimental data coverage.

Data Availability

The source code of the proposed method is made available in the following public repository [67]:
https://github.com/ContiPaolo/MultiFidelity_POD.

The full training and testing datasets are available at <https://doi.org/10.5281/zenodo.8316324> [82].

Acknowledgements

PC is supported under the JRC STEAM STM-Politecnico di Milano agreement. MG receives support from Sectorplan Bèta (the Netherlands) under the focus area *Mathematics of Computational Science*. The present research has been partially supported by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence) and by MUR, grant Dipartimento di Eccellenza 2023-2027. PC, SLB, and JNK acknowledge generous funding support from the National Science Foundation AI Institute in Dynamic Systems (grant number 2112085).

References

- [1] Bui-Thanh T, Willcox K, Ghattas O. 2008 Parametric reduced-order models for probabilistic analysis of unsteady aerodynamic applications. *AIAA journal* **46**, 2520–2529.
- [2] Sudret B, Der Kiureghian A. 2000 *Stochastic finite element methods and reliability: a state-of-the-art report*. Department of Civil and Environmental Engineering, University of California.
- [3] Negri F, Rozza G, Manzoni A, Quarteroni A. 2013 Reduced basis method for parametrized elliptic optimal control problems. *SIAM Journal on Scientific Computing* **35**, A2316–A2340.
- [4] Sinigaglia C, Quadrelli DE, Manzoni A, Braghin F. 2022 Fast active thermal cloaking through PDE-constrained optimization and reduced-order modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **478**, 20210813.
- [5] Manzoni A, Quarteroni A, Rozza G. 2012 Shape optimization for viscous flows by reduced basis methods and free-form deformation. *International Journal for Numerical Methods in Fluids* **70**, 646–670.
- [6] Frangos M, Marzouk Y, Willcox K, van Bloemen Waanders B. 2010 Surrogate and reduced-order modeling: A comparison of approaches for large-scale statistical inverse problems. In Biegler L, Biros G, Ghattas O, Heinkenschloss M, Keyes D, Mallick B, Marzouk Y, van Bloemen Waanders B, Willcox K, editors, *Large-Scale Inverse Problems and Quantification of Uncertainty*, pp. 123–150. John Wiley & Sons, Ltd.
- [7] Cui T, Marzouk Y, Willcox K. 2015 Data-driven model reduction for the Bayesian solution of inverse problems. *International Journal for Numerical Methods in Engineering* **102**, 966–990.
- [8] Maday Y, Patera A, Penn J, Yano M. 2015 A parametrized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering* **102**, 933–965.
- [9] Haik W, Maday Y, Chamoin L. 2023 A real-time variational data assimilation method with data-driven model enrichment for time-dependent problems. *Computer Methods in Applied Mechanics and Engineering* **405**, 115868.
- [10] Rubio PB, Chamoin L, Louf F. 2021 Real-time data assimilation and control on mechanical systems under uncertainties. *Advanced Modeling and Simulation in Engineering Sciences* **8**, 1–25.
- [11] Quarteroni A, Manzoni A, Negri F. 2016 *Reduced Basis Methods for Partial Differential Equations. An Introduction*. Springer International Publishing.
- [12] Benner P, Gugercin S, Willcox K. 2015 A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review* **57**, 483–531.
- [13] Antoulas AC. 2005 A survey of model reduction methods for large-scale systems. *Contemporary Mathematics* **280**, 193–219.
- [14] Noack BR, Schlegel M, Morzynski M, Tadmor G. 2011a *Galerkin method for nonlinear dynamics*. Springer.

- [15] Noack BR, Morzynski M, Tadmor G. 2011b *Reduced-order modelling for flow control* vol. 528. Springer Science & Business Media.
- [16] Carlberg KT, Tuminaro R, Boggs P. 2015 Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics. *SIAM Journal of Scientific Computing* **37**, B153–B184.
- [17] Hesthaven JS, Rozza G, Stamm B. 2016 *Certified reduced basis methods for parametrized partial differential equations*. Springer International Publishing.
- [18] Schmid PJ. 2010 Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics* **656**, 5–28.
- [19] Brunton SL, Kutz JN. 2022 *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- [20] Peherstorfer B, Willcox K. 2016 Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering* **306**, 196–215.
- [21] Qian E, Kramer B, Peherstorfer B, Willcox K. 2020 Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena* **406**, 132401.
- [22] Ghattas O, Willcox K. 2021 Learning physics-based models from data: perspectives from inverse problems and model reduction. *Acta Numerica* **30**, 445–554.
- [23] Guo M, McQuarrie SA, Willcox KE. 2022 Bayesian operator inference for data-driven reduced-order modeling. *Computer Methods in Applied Mechanics and Engineering* **402**, 115336.
- [24] Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **113**, 3932–3937.
- [25] Champion K, Lusch B, Kutz JN, Brunton SL. 2019 Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences* **116**, 22445–22451.
- [26] Bakarji J, Champion K, Kutz JN, Brunton SL. 2022 Discovering governing equations from partial measurements with deep delay autoencoders. *arXiv preprint arXiv:2201.05136*.
- [27] Schaeffer H. 2017 Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**, 20160446.
- [28] Conti P, Gobat G, Fresca S, Manzoni A, Frangi A. 2023 Reduced order modeling of parametrized systems through autoencoders and SINDy approach: continuation of periodic solutions. *Computer Methods in Applied Mechanics and Engineering* **411**, 116072.
- [29] Mars Gao L, Kutz JN. 2022 Bayesian autoencoders for data-driven discovery of coordinates, governing equations and fundamental constants. *arXiv e-prints* pp. arXiv-2211.
- [30] Lee K, Carlberg KT. 2020 Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics* **404**, 108973.
- [31] Fresca S, Dede L, Manzoni A. 2021 A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs. *Journal of Scientific Computing* **87**, 1–36.
- [32] Fresca S, Manzoni A. 2022 POD-DL-ROM: enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition. *Computer Methods in Applied Mechanics and Engineering* **388**, 114181.
- [33] Otto SE, Rowley CW. 2019 Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems* **18**, 558–593.
- [34] Zhuang Q, Lorenzi JM, Bungartz HJ, Hartmann D. 2021 Model order reduction based on Runge–Kutta neural networks. *Data-Centric Engineering* **2**, e13.

- [35] Guo M, Hesthaven JS. 2019 Data-driven reduced order modeling for time-dependent problems. *Computer methods in applied mechanics and engineering* **345**, 75–99.
- [36] Botteghi N, Guo M, Brune C. 2022 Deep kernel learning of dynamical models from high-dimensional noisy data. *Scientific Reports* **12**, 21530.
- [37] Bonneville C, Choi Y, Ghosh D, Belof JL. 2024 GPLaSDI: Gaussian Process-based interpretable Latent Space Dynamics Identification through deep autoencoder. *Computer Methods in Applied Mechanics and Engineering* **418**, 116535.
- [38] Grosjean E, Maday Y. 2021 Error estimate of the non-intrusive reduced basis method with finite volume schemes. *ESAIM: Mathematical Modelling and Numerical Analysis* **55**, 1941–1961.
- [39] Wang Z, Xiao D, Fang F, Govindan R, Pain CC, Guo Y. 2018 Model identification of reduced order fluid dynamics systems using deep learning. *International Journal for Numerical Methods in Fluids* **86**, 255–268.
- [40] Fu R, Xiao D, Navon I, Fang F, Yang L, Wang C, Cheng S. 2023 A non-linear non-intrusive reduced order model of fluid flow by auto-encoder and self-attention deep learning methods. *International Journal for Numerical Methods in Engineering*.
- [41] Kennedy MC, O'Hagan A. 2000 Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13.
- [42] Álvarez MA, Rosasco L, Lawrence ND. 2012 Kernels for Vector-Valued Functions: A Review. *Foundations and Trends® in Machine Learning* **4**, 195–266.
- [43] Lam R, Allaire DL, Willcox KE. 2015 Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources. In *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference* p. 0143.
- [44] Meng X, Karniadakis GE. 2020 A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. *Journal of Computational Physics* **401**, 109020.
- [45] Meng X, Babaee H, Karniadakis GE. 2021 Multi-fidelity Bayesian neural networks: Algorithms and applications. *Journal of Computational Physics* **438**, 110361.
- [46] Liu D, Wang Y. 2019 Multi-Fidelity Physics-Constrained Neural Network and Its Application in Materials Modeling. *Journal of Mechanical Design* **141**.
- [47] Motamed M. 2020 A multi-fidelity neural network surrogate sampling method for uncertainty quantification. *International Journal for Uncertainty Quantification* **10**.
- [48] Guo M, Manzoni A, Amendt M, Conti P, Hesthaven JS. 2022 Multi-fidelity regression using artificial neural networks: efficient approximation of parameter-dependent output quantities. *Computer methods in Applied Mechanics and Engineering* **389**, 114378.
- [49] Peherstorfer B, Willcox K, Gunzburger M. 2018 Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review* **60**, 550–591.
- [50] Raissi M, Perdikaris P, Karniadakis GE. 2017 Inferring solutions of differential equations using noisy multi-fidelity data. *Journal of Computational Physics* **335**, 736–746.
- [51] Pepper N, Gaymann A, Sharma S, Montomoli F. 2021 Local bi-fidelity field approximation with knowledge based neural networks for computational fluid dynamics. *Scientific Reports* **11**, 1–11.
- [52] Kast M, Guo M, Hesthaven JS. 2020 A non-intrusive multifidelity method for the reduced order modeling of nonlinear problems. *Computer Methods in Applied Mechanics and Engineering* **364**, 112947.

- [53] Torzoni M, Manzoni A, Mariani S. 2023 A multi-fidelity surrogate model for structural health monitoring exploiting model order reduction and artificial neural networks. *Mechanical Systems and Signal Processing* **197**, 110376.
- [54] Geneva N, Zabaras N. 2020 Multi-fidelity generative deep learning turbulent flows. *Foundations of Data Science* **2**, 391–428.
- [55] Ahmed SE, San O, Kara K, Younis R, Rasheed A. 2021 Multifidelity computing for coupling full and reduced order models. *Plos one* **16**, e0246092.
- [56] Hochreiter S, Schmidhuber J. 1997 Long short-term memory. *Neural Computation* **9**, 1735–1780.
- [57] Gers FA, Schmidhuber J, Cummins F. 2000 Learning to forget: Continual prediction with LSTM. *Neural Computation* **12**, 2451–2471.
- [58] Graves A, Fernández S, Schmidhuber J. 2005 Bidirectional LSTM networks for improved phoneme classification and recognition. In *International Conference on Artificial Neural Networks* pp. 799–804. Springer.
- [59] Graves A, Jaitly N, Mohamed Ar. 2013 Hybrid speech recognition with deep bidirectional LSTM. In *2013 IEEE workshop on automatic speech recognition and understanding* pp. 273–278. IEEE.
- [60] Wang S, Jiang J. 2015 Learning natural language inference with LSTM. *arXiv preprint arXiv:1512.08849*.
- [61] Zhang R, Chen Z, Chen S, Zheng J, Büyüköztürk O, Sun H. 2019 Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures* **220**, 55–68.
- [62] Conti P, Guo M, Manzoni A, Hesthaven JS. 2023 Multi-fidelity surrogate modeling using long short-term memory networks. *Computer methods in applied mechanics and engineering* **404**, 115811.
- [63] Vlachas PR, Byeon W, Wan ZY, Sapsis TP, Koumoutsakos P. 2018 Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **474**, 20170844.
- [64] Vlachas PR, Arampatzis G, Uhler C, Koumoutsakos P. 2022 Multiscale simulations of complex systems by learning their effective dynamics. *Nature Machine Intelligence* **4**, 359–366.
- [65] Nakamura T, Fukami K, Hasegawa K, Nabae Y, Fukagata K. 2021 Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow. *Physics of Fluids* **33**, 025116.
- [66] Maulik R, Lusch B, Balaprakash P. 2021 Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. *Physics of Fluids* **33**, 037106.
- [67] Conti P. 2023 MultiFidelity_POD. https://github.com/ContiPaolo/MultiFidelity_POD.
- [68] Wang C, Mahadevan S. 2009 A general framework for manifold alignment. In *2009 AAAI Fall Symposium Series*.
- [69] Perron C, Rajaram D, Mavris D. 2020 Development of a multi-fidelity reduced-order model based on manifold alignment. In *AIAA Aviation 2020 Forum* p. 3124.
- [70] Kingma DP, Ba J. 2014 Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [71] Olah C. 2015 Understanding LSTM Networks. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Accessed: April 21, 2023.
- [72] Bergstra J, Bardenet R, Bengio Y, Kégl B. 2011 Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems* **24**.

- [73] Bergstra J, Yamins D, Cox D. 2022 Hyperopt: Distributed asynchronous hyper-parameter optimization. *Astrophysics Source Code Library* pp. ascl–2205.
- [74] Trefethen LN. 2000 *Spectral methods in MATLAB*. SIAM.
- [75] Bergstra J, Yamins D, Cox DD. 2013 Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28* ICML’13 pp. 115–123. JMLR.org.
- [76] Kutz JN. 2013 *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press.
- [77] Fresca S, Manzoni A. 2021 Real-time simulation of parameter-dependent fluid flows through deep learning-based reduced order models. *Fluids* **6**, 259.
- [78] Rajani B, Kandasamy A, Majumdar S. 2009 Numerical simulation of laminar flow past a circular cylinder. *Applied Mathematical Modeling* **33**, 1228 – 1247.
- [79] Zdravkovich MM. 1997 *Flow around circular cylinders: Volume 2: Applications* vol. 2. Oxford university press.
- [80] Negri F. 2016 redbKIT Version 2.2. <http://redbkit.github.io/redbKIT/>.
- [81] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017 Attention is all you need. *Advances in Neural Information Processing Systems* **30**.
- [82] Conti P. 2023 MultiFidelity_POD dataset. (10.5281/zenodo.8316324)

MOX Technical Reports, last issues

Dipartimento di Matematica

Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 82/2024** Rosafalco, L.; Conti, P.; Manzoni, A.; Mariani, S.; Frangi, A.
EKF-SINDy: Empowering the extended Kalman filter with sparse identification of nonlinear dynamics
- 80/2024** Crippa, B.; Scotti, A.; Villa, A
Numerical Solution of linear drift-diffusion and pure drift equations on one-dimensional graphs
- 79/2024** Baioni, P.J.; Benacchi, T.; Capone, L.; de Falco, C.
Portable, Massively Parallel Implementation of a Material Point Method for Compressible Flows
- 78/2024** Ziarelli, G.; Pagani, S.; Parolini, N.; Regazzoni, F.; Verani, M.
A model learning framework for inferring the dynamics of transmission rate depending on exogenous variables for epidemic forecasts
- 77/2024** Piersanti, R.; Bradley, R.; Ali, S.Y.; Quarteroni A.; Dede', L; Trayanova, N.A.
Defining myocardial fiber bundle architecture in atrial digital twins
- 75/2024** Cattarossi, L.; Sacco, F.; Giuliani, N.; Parolini, N.; Mola, A.
A geometry aware arbitrary order collocation Boundary Element Method solver for the potential flow past three dimensional lifting surfaces
- 74/2024** Crippa, B., Scotti, A.; Villa, A
A mixed-dimensional model for the electrostatic problem on coupled domains
- 73/2024** Liverotti, L.; Ferro, N.; Soli, L.; Matteucci, M.; Perotto, S.
Using SAR Data as an Effective Surrogate for Optical Data in Nitrogen Variable Rate Applications: a Winter Wheat Case Study
- 72/2024** Patanè, G.; Bortolotti, T.; Yordanov, V.; Biagi, L. G. A.; Brovelli, M. A.; Truong, A. Q; Vantini, S.
An interpretable and transferable model for shallow landslides detachment combining spatial Poisson point processes and generalized additive models
- 71/2024** Zhang, L.; Pagani, S.; Zhang, J.; Regazzoni, F.
Shape-informed surrogate models based on signed distance function domain encoding