Real-time digital twins of multiphysics and turbulent flows

By A. Nóvoa† AND L. Magri†‡

Developing real-time digital twins for multiphysics systems is challenging due to the intricate and nonlinear interactions between subsystems, and the high-dimensionality and heterogeneity of the data. Further, multiphysics systems are often affected by unpredictable turbulent dynamics. In this paper, we propose two digital twins. First, we develop a digital twin of acoustics–flame instabilities in a hydrogen-based annular combustor, which combines raw acoustic pressure data, a physics-based reduced order model (ROM) and a reservoir computer that infers the model bias and measurement shift. The three elements are statistically combined by the regularized bias-aware ensemble Kalman filter (r-EnKF). The r-EnKF enables (i) simultaneous on-the-fly inference of all the physical model parameters and states and (ii) the real-time estimation of errors in both the model and the data. Second, we develop a digital twin of a turbulent wake. We propose a data-driven ROM of a large eddy simulation (LES) of the turbulent wake by leveraging proper orthogonal decomposition and an echo state network (POD-ESN). We show that the r-EnKF can adaptively update the POD-ESN prediction of the turbulent wake in real time using sparse data on a small subset of the domain. This works opens opportunities for real-time digital twining of multiphysics and turbulent flows.

1. Introduction

Multiphysics flows are the backbone of applications from sustainable aviation through health care to energy conversion systems (Magri 2021). Examples of multiphysics flows are structures interacting with fluids (Gomes & Palacios 2022), acoustics interacting with chemistry (Magri 2019), and electromagnetic waves interacting with particles (Lalli & Giusti 2023). Predicting multiphysics systems accurately in time is challenging because (i) different models are used for each subsystem, which have different levels of accuracy; (ii) the interactions between the subsystems are intricate and nonlinear; and (iii) multiphysics systems are typically turbulent, thus they are unpredictable in space and time, as any small perturbation to the system grows exponentially in time (the butterfly effect). To model multiphysics systems on the fly, we need a synergistic and principled approach that combines physics-based models, data from sensors, bias-aware data assimilation and a machine learning tool to estimate the errors present in both data and models (Nóvoa *et al.* 2024*a*). This is a real-time digital twin.

First, a "good" physical model should provide a time-accurate representation of the system at low computational cost, i.e., a reduced order model (ROM). ROMs can be qualitatively accurate because they capture some key physical dynamics, but they are

† Department of Aeronautics, Imperial College London, United Kingdom.
‡ The Alan Turing Institute, London, United Kingdom.

quantitatively inaccurate due to model assumptions (Magri & Doan 2020), which may result in systematic model errors (also known as model bias). This situation is exacerbated by turbulence because of its spatiotemporal chaotic amplification of errors (Duraisamy *et al.* 2019). Second, a digital twin requires data from sensors. The data are accurate, but, typically, they are sparse in time and space and are affected by aleatoric noise (e.g., environmental noise), and systematic errors (e.g., shifts). Data can be used to improve the quantitative accuracy of ROMs by calibrating model parameters or providing a closure model. These techniques are typically offline and not adaptive in real time (e.g., Lassila *et al.* 2014; Pawar *et al.* 2020; Lavadera *et al.* 2024). Nonetheless, a reliable digital twin must adapt on the fly to changes to its physical counterpart. Thus, we need a strategy for the ROM to self-correct with data. Third, in a real-time context, the ROM and the sensors may be affected by systematic errors. Machine learning has been deployed to infer and remove the biases either from the data (e.g., Liang *et al.* 2023) or from the model (e.g., da Silva & Colonius 2020; Nóvoa & Magri 2022), but not both biases at once. It is paramount to design strategies to separate between ROM biases and errors in the sensors. Fourth, to seamlessly combine ROMs and data, we need a rigorous uncertainty-aware technique. Bias-aware data assimilation methods are designed to improve model accuracy by statistically introducing any available data (e.g., pressure and wire measurements) into the model while accounting for biases in the framework.

State-of-the-art computational methods for the time-accurate prediction of multiphysics flows are based on direct numerical simulations, LES and unsteady Reynolds-averaged Navier-Stokes, all of which cannot be deployed in real-time applications because they are computationally expensive (with distinct levels of complexity). In the context of nonlinear ROMs, most available methods are offline, i.e., they are not adaptive in real time [e.g., spectral POD (Arun *et al.* 2023)], or they may be difficult to interpret physically [e.g., adaptive basis modeling (Huang *et al.* 2022)]. The combination of data assimilation with machine learning and data-driven methods has shown potential to predict complex dynamics (Geer 2021). We recently proposed the first real-time bias-aware data assimilation method for combined state, parameters and model error inference (Nóvoa *et al.* 2024*b*). Here, we generalize the framework to infer simultaneously the ROM bias and the data shift, and we apply the framework to two fluid dynamics problems: multiphysics fluids (aeroengines) and turbulence (wakes) to deliver practical digital twins that interface with experimental data.

The paper is structured as follows. Section 2 introduces mathematically multi-physical systems. Section 3 provides the theoretical background to digital twins. Sections 4 and 5 include the design and testing of their digital twin of the acoustics–flame system and the turbulent wake, respectively. Section 6 closes the paper with conclusions.

2. Multiphysics dynamical systems

We view a multiphysics problem with D interacting subsystems as a dynamical system

$$\frac{d}{dt}\begin{bmatrix}\boldsymbol{\phi}_1\\ \boldsymbol{\phi}_2\\ \vdots\\ \boldsymbol{\phi}_D\end{bmatrix}=\begin{bmatrix}\mathbf{G}_1(\boldsymbol{\phi}_1;\boldsymbol{\alpha}_{G,1})\\ \mathbf{G}_2(\boldsymbol{\phi}_2;\boldsymbol{\alpha}_{G,2})\\ \vdots\\ \mathbf{G}_D(\boldsymbol{\phi}_D;\boldsymbol{\alpha}_{G,D})\end{bmatrix}+\begin{bmatrix}\mathbf{H}_1(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\ldots,\boldsymbol{\phi}_D;\boldsymbol{\alpha}_{H,1})\\ \mathbf{H}_2(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\ldots,\boldsymbol{\phi}_D;\boldsymbol{\alpha}_{H,2})\\ \vdots\\ \mathbf{H}_D(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\ldots,\boldsymbol{\phi}_D;\boldsymbol{\alpha}_{H,D})\end{bmatrix}, \tag{2.1}$$

where $\mathbf{G}_i$ is the set of equations that governs the physical subsystem i whose physical state is $\boldsymbol{\phi}_i$, $\mathbf{H}_i$ is the set of interaction terms that couple the different physical quantities (e.g.,

temperature or pressure) and $\boldsymbol{\alpha}_i = [\boldsymbol{\alpha}_{G,i}; \boldsymbol{\alpha}_{H,i}]$ is the set of parameters of subsystem i†. If the governing equations are partial differential equations, the operators $\mathbf{G}_i$ and $\mathbf{H}_i$ are the spatial discretization of the equations, which encapsulate the boundary conditions. The equations in Eq. (2.1) are closed by the initial conditions $\boldsymbol{\phi}_{i,0} = \boldsymbol{\phi}_i(t = 0)$.

A system is multiphysics when $\mathbf{H}_i \neq \mathbf{0}$ for at least one i. We refer to a multiphysics system with D subsystems as a D-way-coupled multiphysics system. For example, a flying flexible wing is a fluid–structure interaction problem with two subsystems, i.e., the wing's structural mechanics and the fluid mechanics around it; and the combustor of an aeroengine, there are $D = 3$ subsystems interacting with each other: the heat released by chemical reactions, turbulent fluid mechanics and acoustics, i.e., it is a three-way-coupled multiphysics system. The degree of coupling and the number of subsystems at play may be a modeling choice. In Section 4, we design a digital twin of the acoustics–flame subsystems in an combustor, i.e., we simplify the problem by focusing on two of the three subsystems. Then, we design a digital twin of a turbulent wake in Section 5.

3. Real-time digital twin: The theoretical components

In this section, we introduce the theoretical formulation of real-time digital twins. In a digital twin framework, we have three sources of information:

(1) *Reduced order model*: Numerical models are the first source of information to estimate a physical system $\boldsymbol{t}_i = \mathcal{M}_i(\boldsymbol{\phi}_i) + \boldsymbol{b} + \boldsymbol{\varepsilon}_\phi$, where $\boldsymbol{t}_i$ is the true state, $\mathcal{M}_i$ is the measurement operator, which maps the state variables into the observable space, $\boldsymbol{b}$ is the model bias, and $\boldsymbol{\varepsilon}_\phi$ is the aleatoric noise. High-fidelity models provide predictions with small bias, but their high computational cost makes them not suitable for real-time applications. Theefore, we employ ROMs, which provide qualitatively accurate estimates with only a few degrees of freedom $\dim\{\boldsymbol{\phi}\} \sim \mathcal{O}(1 - 1000)$.

(2) *Data stream*: The second source of information is the data from sensors, $\boldsymbol{d}_i$. In real time, data are processed on the fly, i.e., one data point at a time. Because the sensors collect raw data, they record the truth subject to aleatoric and epistemic errors, i.e., $\boldsymbol{t}_i = \boldsymbol{d}_i + \boldsymbol{b}_d + \boldsymbol{\varepsilon}_d$, where $\boldsymbol{\varepsilon}_d$ is the aleatoric noise and $\boldsymbol{b}_d$ is the measurement shift.

(3) *Bias and uncertainties:* So far, we have two known quantities: the model $\mathcal{M}_i(\boldsymbol{\phi}_i)$ and the observations $\boldsymbol{d}_i$, and four unknowns: $\boldsymbol{b}$, $\boldsymbol{b}_d$, $\boldsymbol{\varepsilon}_\phi$ and $\boldsymbol{\varepsilon}_d$. First, we assume that the aleatoric noise terms present in the framework are normally distributed, i.e., $\boldsymbol{\varepsilon}_\phi \sim \mathcal{N}(\boldsymbol{0}, \mathbf{C}_{\phi\phi})$ and $\boldsymbol{\varepsilon}_d \sim \mathcal{N}(\boldsymbol{0}, \mathbf{C}_{dd})$, where $\mathcal{N}(\boldsymbol{0}, \mathbf{C})$ is a normal distribution with zero mean and covariance $\mathbf{C}$. Second, we assume that the sensors independent and well-calibrated, such that $\mathbf{C}_{dd}$ is a diagonal matrix with identical entries. Second, we assume that our ROM is qualitatively accurate such that the model bias $\boldsymbol{b}$ is small. In Section 4, we employ reservoir computers to model both $\boldsymbol{b}_d$ and $\boldsymbol{b}$.

To combine the three sources of information, we use data assimilation. Specifically, we use real-time bias-aware data assimilation to simultaneously estimate the model state and parameter and the errors present in the assimilation (Nóvoa *et al.* 2024*b*). First, we define the augmented state vector $\boldsymbol{\psi} = [\boldsymbol{\phi}_1; \ldots; \boldsymbol{\phi}_D; \boldsymbol{\alpha}_1; \ldots; \boldsymbol{\alpha}_D; \mathcal{M}_i(\boldsymbol{\phi}_i); \ldots; \mathcal{M}_D(\boldsymbol{\phi}_D)]$, which we write as $\boldsymbol{\psi} = [\boldsymbol{\phi}; \boldsymbol{\alpha}; \mathcal{M}(\boldsymbol{\phi})]$ for brevity. Second, we model the uncertainty as Gaussian noise, $\boldsymbol{\alpha}^t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_\alpha$. Third, we consider an ensemble approach to estimating the first two moments of the probability density function (PDF) of the solution. Each of the m members in the ensemble are propagated independently and, at the time of assimilation (i.e., the analysis step), the mean and variance of the PDF are estimated

† We employ column vectors and the operator [;] indicates vertical concatenation

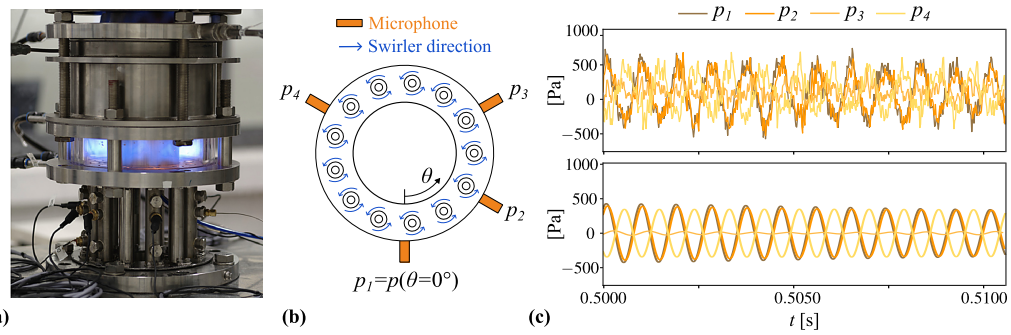

FIGURE 1. (a) Experimental setup of the annular combustor (Indlekofer *et al.* 2021). (b) Diagram of the cross-section area of the combustor with 12 swirling flames and 4 microphones. (c) Example of (top) the raw data recorded by the four microphones, which is assimilated in real time, and (bottom) the band-passed pressure data, which is unavailable in real time and we refer to as the "presumed ground truth".

with the ensemble statistics $\bar{\boldsymbol{\psi}} = \frac{1}{m}\sum_j \boldsymbol{\psi}_j$ and $\mathbf{C}_{\psi\psi} = \frac{1}{m-1}\sum_j(\boldsymbol{\psi}_j - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi}_j - \bar{\boldsymbol{\psi}})^{\mathrm{T}}$, respectively. Thus, we write the compact notation of Eq. (2.1) as

$$\begin{cases} \dfrac{d\boldsymbol{\psi}_j}{dt} = \left[\mathbf{F}(\boldsymbol{\phi}_j;\boldsymbol{\alpha}_j;\boldsymbol{\varepsilon}_\phi;\boldsymbol{\varepsilon}_\alpha);\mathbf{0}\right], \\ \mathbf{M}\boldsymbol{\psi}_j = \mathcal{M}(\boldsymbol{\phi}_j) \end{cases} \quad \text{for } j = 1,\dots,m, \tag{3.1}$$

where $\mathbf{F}$ encapsulates both the subsystems' governing equations and their interactions, and the transformed measurement operator $\mathbf{M} = [\mathbf{0};\mathbf{I}]$ is linear, with $\mathbf{I}$ the identity matrix. Lastly, the ensemble is updated at the analysis step via the r-EnKF. Following Nóvoa *et al.* (2024*b*), we weight the model bias with the data error matrix $\mathbf{C}_{dd}$ and use the regularization parameter γ to tune the norm, such that the r-EnKF reads

$$\boldsymbol{\psi}_j^{\mathrm{a}} = \boldsymbol{\psi}_j^{\mathrm{f}} + \mathbf{K}_r\left[(\mathbf{I}+\mathbf{J})^{\mathrm{T}}\left(\boldsymbol{d}_j + \boldsymbol{b}_d - \mathbf{M}\boldsymbol{\psi}_j^{\mathrm{f}} - \boldsymbol{b}\right) - \gamma\mathbf{J}^{\mathrm{T}}\boldsymbol{b}\right], \quad \text{for } j = 1,\dots,m, \tag{3.2a}$$

where the superscripts 'a' and 'f' stand for analysis and forecast, respectively; $\mathbf{J} = \mathrm{d}\boldsymbol{b}/\mathrm{d}\mathbf{M}\boldsymbol{\psi}$ is the Jacobian of the bias estimator (i.e., the Jacobian of the echo state network in Section 4); and

$$\mathbf{K}_r = \mathbf{C}_{\psi\psi}^{\mathrm{f}}\mathbf{M}^{\mathrm{T}}\left[\mathbf{C}_{dd} + (\mathbf{I}+\mathbf{J})^{\mathrm{T}}(\mathbf{I}+\mathbf{J})\mathbf{M}\mathbf{C}_{\psi\psi}^{\mathrm{f}}\mathbf{M}^{\mathrm{T}} + \gamma\mathbf{J}^{\mathrm{T}}\mathbf{J}\mathbf{M}\mathbf{C}_{\psi\psi}^{\mathrm{f}}\mathbf{M}^{\mathrm{T}}\right]^{-1} \tag{3.2b}$$

is the bias-regularized Kalman gain matrix. The r-EnKF is an approximate Monte Carlo sequential data assimilation method, which generalizes the ensemble Kalman filter (EnKF) to account for errors in the model and/or the data. An analysis state given by (3.2) is "good" if the unbiased state $\mathbf{M}\boldsymbol{\psi} + \boldsymbol{b}$ matches the corrected observations $\boldsymbol{d} + \boldsymbol{b}_d$ (i.e., the truth) and the norm of the bias $\boldsymbol{b}$ is small. If there are no systematic errors, i.e., $\boldsymbol{b} = \boldsymbol{b}_d = \boldsymbol{0}$, Eq. (3.2) simplify to the EnKF.

4. Real-time digital twin of an acoustics–flame system

In this section, we develop a real-time digital twin of the acoustics–flame system shown in Figure 1. The available data in real time (Indlekofer *et al.* 2021) are pressure measurements from microphones located at $\theta = \{0°, 60°, 120°, 240°\}$ [Figure 1(b)]. The data are noisy and shifted, i.e, the raw pressure is not centered at zero [Figure 1(c)]. We employ

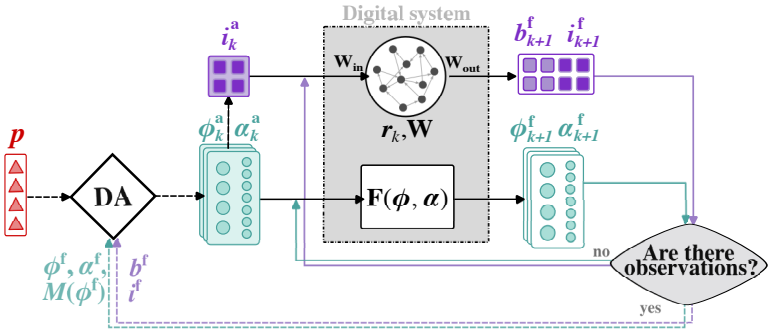

FIGURE 2. Schematic of the digital twin of the acoustics–flame system. The input to the ESN is innovation $\boldsymbol{i} = \boldsymbol{d} - \mathbf{M}\bar{\boldsymbol{\psi}}$, which is an estimate of the sum of the bias and the shift, i.e., $\boldsymbol{i} \approx \boldsymbol{b} + \boldsymbol{b}_d$. The Jacobian term in Eqs. (3.2) is hence $\mathbf{J} = -\mathrm{d}\boldsymbol{b}_{k+1}^{\mathrm{f}}/\mathrm{d}\boldsymbol{i}_k^{\mathrm{a}}$ (see Nóvoa *et al.* 2024*b*, for details).

a well-established ROM of the acoustics–flame system (e.g., Indlekofer *et al.* 2021). The dynamics of the system can be qualitatively modeled by the nonlinear wavelike equation

$$\frac{\partial^2 p}{\partial t^2} + \zeta\frac{\partial p}{\partial t} - \left[1 + \epsilon\cos\left(2(\theta - \Theta_\epsilon)\right)\right]\omega^2\frac{\partial^2 p}{\partial \theta^2} = \frac{\partial \dot{q}}{\partial t}, \tag{4.1}$$

$$\text{with} \quad \frac{\partial \dot{q}}{\partial t} = \beta\left[1 + c_2\cos\left(2(\theta - \Theta_\beta)\right)\right]p - \kappa p^3, \tag{4.2}$$

where p is the acoustic pressure and $\dot{q}$ is the heat release rate. The model parameters are the acoustic damping ζ; the acoustic period $2\pi/\omega$; the reactive symmetry amplitude ϵ and phase Θ_ϵ; the magnitude of nonlinear saturation of the flame response κ; the heat release strength β; the direction of maximum acoustic pressure, Θ_β; and the asymmetry intensity c_2. We project the acoustic pressure on the first degenerate pair of eigenmodes of the homogeneous wave equation $p(t,\theta) = \eta_a(t)\cos\theta + \eta_b(t)\sin\theta$, where η_a and η_b are the acoustic velocity modes. The discretized equations yield the model state vector $\boldsymbol{\phi} = [\eta_a; \dot{\eta}_a; \eta_b; \dot{\eta}_b]$, where $\dot{\eta}_a$ and $\dot{\eta}_b$ are the acoustic pressure modes, and the set of parameters to infer $\boldsymbol{\alpha} = [\nu; \kappa; c_2\beta; \omega; \Theta_\beta; \epsilon; \Theta_\epsilon]$, where $\nu = \frac{1}{2}(\beta - \alpha)$ is the growth rate (Nóvoa *et al.* 2024*a*). Finally, we implement an echo state network (ESN) to estimate the model bias and the measurement shift as depicted in Figure 2. ESNs are a type of reservoir computer, which are generalized nonlinear autocorrelation functions (Aggarwal 2018) and whose training consists of solving the computationally cheap linear regression problem (Lukoševičius 2012)

$$(\mathbf{R}\mathbf{R}^{\mathrm{T}} + \lambda_T\mathbf{I})\mathbf{W}_{\mathrm{out}}^{\mathrm{T}} = \mathbf{R}\mathbf{D}^{\mathrm{T}}, \tag{4.3}$$

where the weights of $\mathbf{W}_{\mathrm{out}}$ are the trainable parameters, $\mathbf{D}$ is the training dataset, $\mathbf{R}$ are the ESN reservoir states and λ_T is the Tikhonov regularization factor† The predictions of the ESN and the model $\mathbf{F}$ come together when microphone data become available (Figure 2). The r-EnKF updates the state and parameters of the model, which in turn provide an updated estimate of the input to the ESN.

4.1. *Results*

We employ an ensemble with $m = 20$ and we assimilate data approximately every 0.7 acoustic periods. The ESN is trained with 50 neurons and 10 washout steps for initialization. Figure 3 shows the inference of the acoustic pressure at one of the four microphones,

† For details on the ESN design and training, the reader is referred to Nóvoa *et al.* (2024*a*).

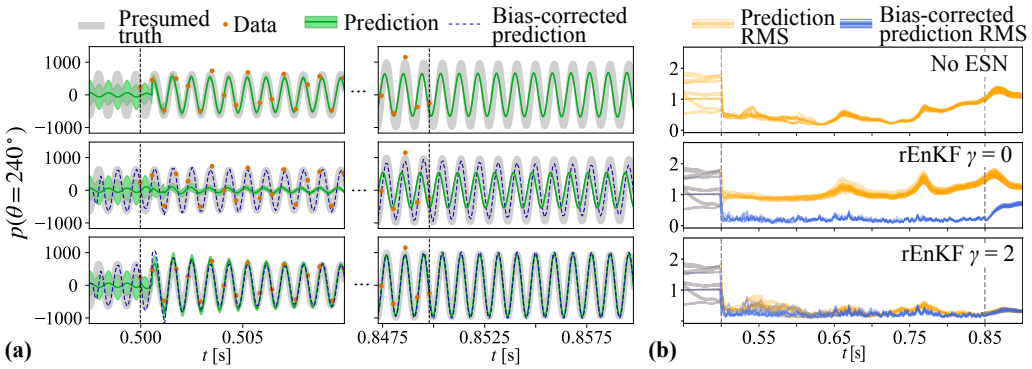

FIGURE 3. Comparison between the digital twin performance using (top) EnKF, i.e., no bias estimation, (middle) r-EnKF with $\gamma = 0$ and (bottom) r-EnKF with $\gamma = 2$. (a) Acoustic pressure at $\theta = 240^\circ$ and (b) RMS error. The vertical lines indicate the assimilation window.

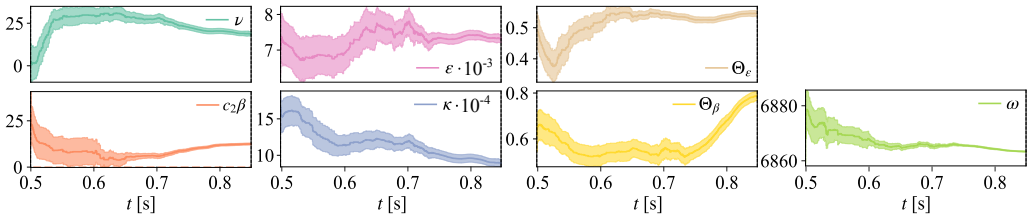

FIGURE 4. Time evolution of the acoustics–flame model parameters throughout assimilation.

and the evolution of the root mean square (RMS) error for three setups: (i) a digital twin without modeling the model bias, i.e., using the bias-unaware EnKF, (ii) a nonregularized bias-aware assimilation and (iii) the proposed r-EnKF with $\gamma = 2$. The proposed bias-regularized method successfully recovers the acoustic pressure with a small-norm bias, and the RMS error remains small after the assimilation has stopped. Therefore, we conclude that the r-EnKF infers in real time a set of model parameters, state and modeling errors that accurately characterize the dynamics of the multiphysics system. The evolution of the model parameters for the regularized case is shown in Figure 4. In contrast to offline ROMs, the proposed digital twin infers simultaneously and in real time all the parameters in the model, the model state and, importantly, the model error and measurement shift.

5. Real-time digital twin of a turbulent wake

In this section, we develop a real-time digital twin of the turbulent wake of flow past a mesh. We employ the data from a LES from Bekoglu *et al.* (2024). The simulations were carried out on the Fortran solver Xcompact3D (Bartholomew *et al.* 2020) at $Re = 5000$ with a homogeneous mesh and zero ambient turbulent intensity. The objective is to estimate the full velocity flow fields by assimilating data from sparse probes in real time. We select the optimal placement of the probes through QR decomposition of a downsampled region (Figure 5). In this scenario, we do not have a well-established physical ROM, in contrast to the acoustic-flame model in Section 4. We propose a POD-ESN model as a non-intrusive ROM. First, we apply snapshot POD to our dataset $\mathbf{U} = [\boldsymbol{u}_{x,0}, \ldots, \boldsymbol{u}_{x,N_t}; \boldsymbol{u}_{y,0}, \ldots, \boldsymbol{u}_{y,N_t}]$, where N_t is the number of training snapshots. The

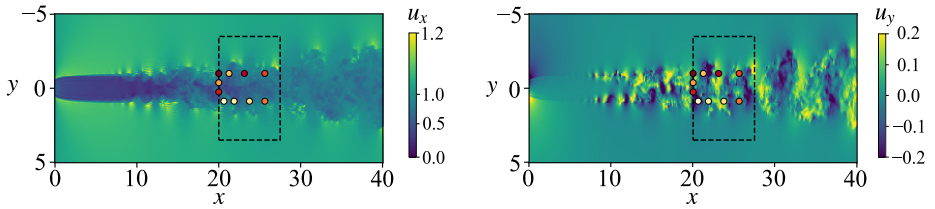

FIGURE 5. Turbulent wake streamwise u_x and spanwise u_y velocity fields from LES simulations (Bekoglu *et al.* 2024). The dashed rectangles shows the measurement region, and the 10 probes are colored from white to dark red according to the QR selection.

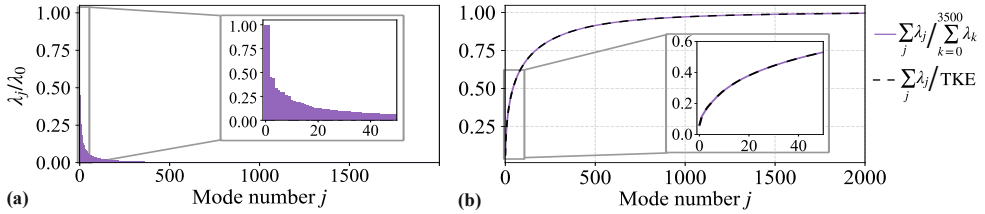

FIGURE 6. Spectrum of the eigenvalues from POD decomposition of the turbulent wake. (a) Eigenvalues normalized by the dominant λ_0. (b) Cumulative energy of the system. The POD decomposition has converged as the cumulative sum relative to the total kinetic energy (TKE) coincides with the cumulative sum relative to the sum of all POD eigenvalues.

snapshot POD decomposition reads

$$\mathbf{U}(\boldsymbol{x}, t) = \mathbf{A}(t)\boldsymbol{\Sigma}\boldsymbol{\Psi}^{\mathrm{T}}(\boldsymbol{x}), \tag{5.1}$$

where $\boldsymbol{x} = [x; y]$ are the spatial coordinates, $\mathbf{A}$ are the POD coefficients, $\boldsymbol{\Sigma}$ is the matrix of singular values σ_j and $\boldsymbol{\Psi}$ are the POD modes (the orthonormal basis). The spectrum of POD eigenvalues $\lambda_j = \sigma_j^2$, which measure the energy level contained in each mode, is shown in Figure 6. Second, we train an ESN to learn the time evolution of the first N_{mode} POD temporal coefficients $\boldsymbol{a}$. We select the $N_{\mathrm{mode}} = 50$ to retain at least 50% of the system's energy in the POD-ESN model. We train a reservoir with $N_r = 1000$ neurons. The POD coefficients are forecast with the ESN equations

$$\begin{aligned} \boldsymbol{a}_{k+1} &= \mathbf{W}_{\mathrm{out}}\left[\boldsymbol{r}_{k+1}; 1\right] \\ \boldsymbol{r}_{k+1} &= \tanh\left(\sigma_{\mathrm{in}}\mathbf{W}_{\mathrm{in}}\left[\boldsymbol{a}_k \odot \boldsymbol{g}; \delta_{\mathrm{r}}\right] + \rho\mathbf{W}\boldsymbol{r}_k\right), \end{aligned} \tag{5.2}$$

where the inputs and outputs of the ESN reservoir are the POD coefficients at time t_k ($\boldsymbol{a}_k$) and t_{k+1} ($\boldsymbol{a}_{k+1}$), respectively. The input matrix $\mathbf{W}_{\mathrm{in}}$ (fixed sparse and randomly generated), is weighted by the input factor σ_{in}, and maps the state normalized by the factor $\boldsymbol{g}$ ($\odot$ is the Hadamard product) into the ESN. The reservoir is characterized by the matrix $\mathbf{W}$ (fixed sparse and randomly generated), which is scaled by the spectral radius ρ and acts as the numerical memory, and the time-varying N_r-dimensional state $\boldsymbol{r}$. The $\tanh(\cdot)$ operation is performed element-wise, and δ_{r} is a symmetry-breaking constant. The output matrix $\mathbf{W}_{\mathrm{out}}$ maps the forecast $\boldsymbol{r}_{k+1}$ back into the physical dimension. The weights of $\mathbf{W}_{\mathrm{out}}$ are trained by solving the linear regression system in Eq. (4.3) and the hyperparameters $\rho = 0.95$ and $\sigma_{\mathrm{in}} = 4.7 \cdot 10^{-7}$ are optimized through a recycle validation and Bayesian optimization (Racca & Magri 2021). We neglect biases in the model and data to simplify the analysis (this assumption will be relaxed in future work). Finally, we apply data assimilation to test the digital twin of the turbulent wake. The POD-ESN prediction is updated with the EnKF (i.e., the limiting case of the r-EnKF

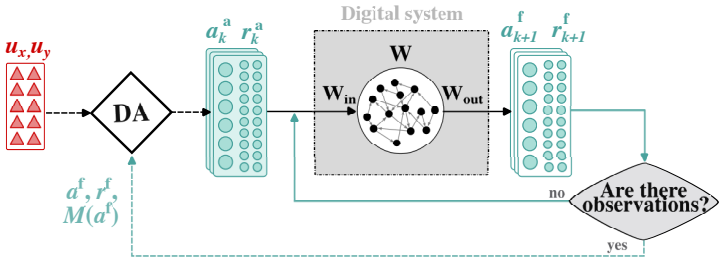

FIGURE 7. Schematic of the digital twin of the turbulent wake system.

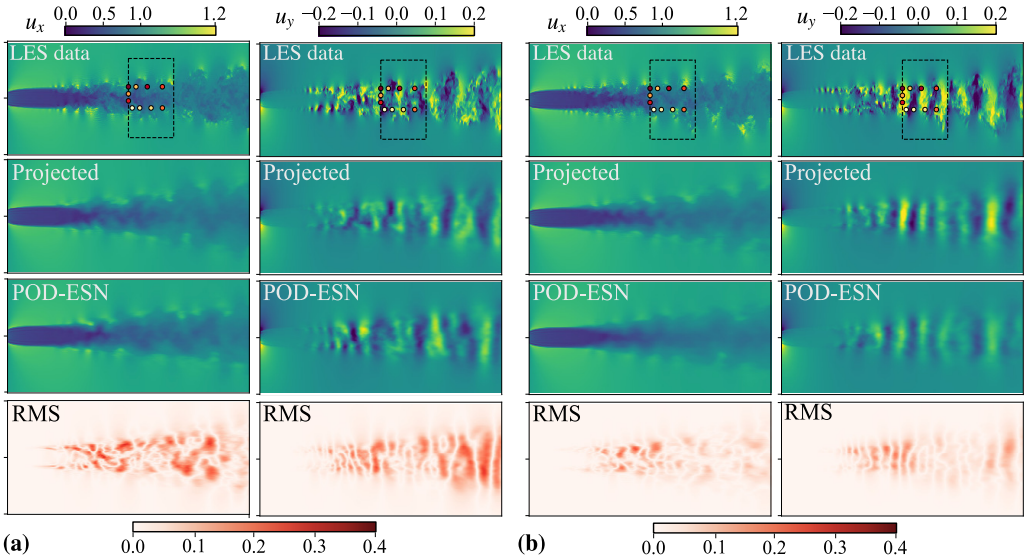

FIGURE 8. Comparison of the velocity fields of the LES data (Bekoglu *et al.* 2024), the LES projection of the POD basis and the POD-ESN reconstruction (a) before and (b) after assimilation. The RMS is computed between the projected LES and the POD-ESN reconstruction.

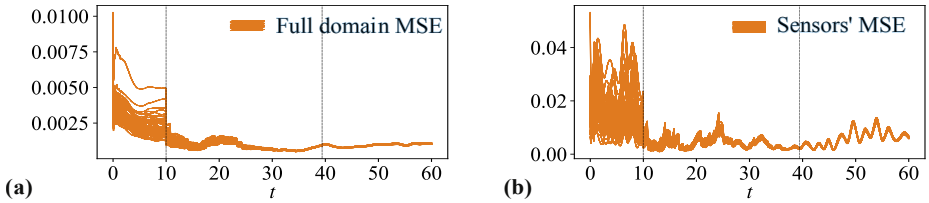

FIGURE 9. Time evolution of the mean squared error (MSE) between the POD-ESN ensemble prediction and the projection of the LES on the POD basis. The MSE is computed (a) for the full domain and (b) for the sensor locations. The vertical lines indicate the assimilation window.

when $\boldsymbol{b} = \boldsymbol{b}_d = \boldsymbol{0}$) when data from the observation probes become available (Figure 7). The state to infer in the POD-ESN is $\boldsymbol{\phi} = [\boldsymbol{a}; \boldsymbol{r}]$, i.e., the temporal coefficients and the reservoir state. Updating $\boldsymbol{r}$ and $\boldsymbol{a}$ simultaneously allows us to keep the reservoir and the physical states synchronized in time.

5.1. *Results*

We employ an ensemble with $m = 50$ and perform 35 analysis steps. Figure 8 shows

the velocity fields at two snapshots: before and after data assimilation. The EnKF correctly nudges the POD-ESN dynamics toward the correct physical state in time [see the matching large/small u_y regions in Figure 8(b)]. Further, the RMS error significantly decreases with the assimilation. The time evolution of the mean squared error (MSE) is shown in Figure 9, where each line corresponds to one ensemble member. The MSE systematically decreases throughout the assimilation but it increases after assimilation. This is because of turbulent effects, as well as due to the non-optimal characteristics of the EnKF solution in the presence of model bias, as discussed in Section 4. Current work focuses on the introduction of a bias estimator into the turbulent wake digital twin.

6. Conclusions

Developing real-time digital twins for multiphysics flows is challenging due to nonlinear interactions between subsystems and the unpredictability of turbulence. We propose a digital twin framework that exploits real-time data assimilation to integrate information from (i) biased but qualitatively accurate reduced-order models (ROMs) and (ii) sparse yet quantitatively accurate data. First, we develop a real-time digital twin for a hydrogen-fueled laboratory annular combustor, which integrates raw pressure data, a physics-based ROM and a reservoir computer that infers model biases and data shifts. These elements are statistically combined via the regularized bias-aware ensemble Kalman filter (r-EnKF) (Nóvoa *et al.* 2024*b*). Unlike offline ROMs, the r-EnKF enables the dynamic inference of the model states and parameters, as well as biases in both the model and data stream. Second, we address the turbulent nature of multiphysics flows with a digital twin of a turbulent wake. A data-driven ROM of the wake, based on large eddy simulation (LES), is constructed using proper orthogonal decomposition and an echo state network (POD-ESN). We use the r-EnKF to update the POD-ESN prediction in real time using sparse data from a subset of the LES domain. Integrating data into the POD-ESN model enhances the time prediction accuracy. The proposed real-time digital twin framework addresses challenges in multiphysics modeling, such as model bias, data uncertainty and real-time adaptation. Future work will focus on designing a bias estimator for the POD-ESN. As in the acoustics–flame system, the r-EnKF will infer modeling errors in real time, providing an optimal and unique estimate of the turbulent system from the ROM.

Acknowledgments

We would like to thank Nicolas Noiray and James Dawson for the experimental data of the annular combustor and Nikos Bempedelis for the LES of the turbulent wake.

REFERENCES

Aggarwal, C. C. 2018 *Neural Networks and Deep Learning*. 2nd ed. Springer Nature.

Arun, R., Bae, H. J. & McKeon, B. J. 2023 Towards real-time reconstruction of velocity fluctuations in turbulent channel flow. *Phys. Rev. Fluids* **8**, 064612.

Bartholomew, P., Deskos, G., Frantz, R. A., Schuch, F. N., Lamballais, E. & Laizet, S. 2020 Xcompact3d: An open-source framework for solving turbulence problems on a cartesian mesh. *SoftwareX* **12**.

Bekoglu, E., Bempedelis, N. & Steiros, K. 2024 On the formation of the primary and secondary vortex street instabilities. In *Thirteen. Turbul. Sheer Flow Phenom.*

Duraisamy, K., Iaccarino, G. & Xiao, H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377.

Geer, A. J. 2021 Learning earth system models from observations: machine learning or data assimilation? *Philos. T. R. Soc. A* **379**, 20200089.

Gomes, P. & Palacios, R. 2022 Aerostructural topology optimization using high fidelity modeling. *Struct. Multidiscip. O.* **65**, 137.

Huang, C., Duraisamy, K. & Merkle, C. 2022 Component-based reduced order modeling of large-scale complex systems. *Front. Phys.* **10**, 900064.

Indlekofer, T., Faure-Beaulieu, A., Noiray, N. & Dawson, J. R. 2021 The effect of dynamic operating conditions on the thermoacoustic response of hydrogen rich flames in an annular combustor. *Combust. Flame* **223**, 284–294.

Lalli, N. S. & Giusti, A. 2023 Coherence effects on the interference colors of soap films. *J. Appl. Phys.* **134** , 093103.

Lassila, T., Manzoni, A., Quarteroni, A. & Rozza, G. 2014 Model order reduction in fluid dynamics: challenges and perspectives. In *Reduc. Order Methods Model. Comput. Reduct.*eds. A. Quarteroni & G. Rozza pp. 235–273.Springer Nature.

Lavadera, M. L., Kamal, M. M., Sharma, S., Donato, L., Galletti, C., Coussement, A. & Parente, A. 2024 A combined experimental, numerical, and data consistency approach for the characterization of temperature distribution in a mild combustion furnace. *Appl. Therm. Eng.* **243**, 122625.

Liang, J., Terasaki, K. & Miyoshi, T. 2023 A machine learning approach to the observation operator for satellite radiance data assimilation. *J. Meteorol. Soc. Jpn.* **101**, 79–95.

Lukoševičius, M. 2012 A practical guide to applying echo state networks. In *Neural Netw. Tricks Trade*2nd ed. (eds. G. Montavon, G. B. Orr & K.-R. Muller), pp. 659–686. Springer Nature.

Magri, L. 2019 Adjoint methods as design tools in thermoacoustics. *Appl. Mech. Rev.* **71** (2).

Magri, L. 2021 Phyco: Physics-constrained adaptive learning for multiphysics optimization. *EXCELLENT SCIENCE - ERC Grant agreement 949388* .

Magri, L. & Doan, N. A. K. 2020 Physics-informed data-driven prediction of turbulent reacting flows with Lyapunov analysis and sequential data assimilation. In *Data Anal. Direct Numer. Simul. Turbul. Combust.* (eds. H. Pitsch & A. Attili), pp. 177–196.Springer Nature.

Nóvoa, A. & Magri, L. 2022 Real-time thermoacoustic data assimilation. *J. Fluid Mech.* **948**, A35.

Nóvoa, A., Noiray, N., Dawson, J. R. & Magri, L. 2024*a* A real-time digital twin of azimuthal thermoacoustic instabilities. arXiv:2404.18793 [physics.flu-dyn].

Nóvoa, A., Racca, A. & Magri, L. 2024*b* Inferring unknown unknowns: regularized bias-aware ensemble Kalman filter. *Comput. Method. Appl. M.* **418**, 116502.

Pawar, S., Ahmed, S. E., San, O. & Rasheed, A. 2020 Data-driven recovery of hidden physics in reduced order modeling of fluid flows. *Phys. Fluids* **32**, 036602.

Racca, A. & Magri, L. 2021 Robust optimization and validation of echo state networks for learning chaotic dynamics. *Neural Netw.* **142**, 252–268.

da Silva, A. F. C. & Colonius, T. 2020 Flow state estimation in the presence of discretization errors. *J. Fluid Mech.* **890**, A10.