

Bayesian operator inference for data-driven reduced-order modeling

Mengwu Guo^{a,*}, Shane A. McQuarrie^b, Karen E. Willcox^b

^a*Department of Applied Mathematics, University of Twente*

^b*Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin*

Abstract

This work proposes a Bayesian inference method for the reduced-order modeling of time-dependent systems. Informed by the structure of the governing equations, the task of learning a reduced-order model from data is posed as a Bayesian inverse problem with Gaussian prior and likelihood. The resulting posterior distribution characterizes the operators defining the reduced-order model, hence the predictions subsequently issued by the reduced-order model are endowed with uncertainty. The statistical moments of these predictions are estimated via a Monte Carlo sampling of the posterior distribution. Since the reduced models are fast to solve, this sampling is computationally efficient. Furthermore, the proposed Bayesian framework provides a statistical interpretation of the regularization term that is present in the deterministic operator inference problem, and the empirical Bayes approach of maximum marginal likelihood suggests a selection algorithm for the regularization hyperparameters. The proposed method is demonstrated on two examples: the compressible Euler equations with noise-corrupted observations, and a single-injector combustion process.

Keywords: data-driven reduced-order modeling, uncertainty quantification, operator inference, Bayesian inversion, Tikhonov regularization, single-injector combustion

1. Introduction

The objective of this work is to enable uncertainty quantification for data-driven reduced-order modeling. A reduced-order model (ROM) is a low-dimensional surrogate for a high-dimensional physical system, such as the many natural and engineering systems that are governed by partial differential equations. Ideally, the computational cost of a ROM should be reduced compared to that of a high-fidelity model of the original system, yet without compromising the solution accuracy. Model reduction strategies typically construct a ROM by identifying a predominant, low-dimensional subspace that well approximates the solution manifold, then representing the state dynamics in the subspace [4, 43]. Classical reduced-order models are constructed by explicitly projecting high-fidelity operators onto the reduced subspace. Such models are called *intrusive* because the high-fidelity operators must be extracted from a numerical solution code. In contrast, *non-intrusive* data-driven ROMs learn low-dimensional representations of systems primarily from solution data, and do not require intrusive access to a high-fidelity simulator. Non-intrusive ROMs are advantageous for applications with well-established legacy codes that can be readily executed but not easily modified. On the other hand, the extensive use of data introduces uncertainties beyond those originating from the modeling decisions of low-dimensional approximation and makes it difficult to obtain classical *a posteriori* error estimates as in [21]. In particular, measurement noise in the data and misspecification of the model structure may impact the accuracy and robustness of non-intrusive ROMs. It is therefore crucial that validation efforts for data-driven ROMs quantify the errors and uncertainties associated with non-intrusive learning.

*Corresponding author.

Email addresses: m.guo@utwente.nl (Mengwu Guo), shanemcq@utexas.edu (Shane A. McQuarrie), kwillcox@oden.utexas.edu (Karen E. Willcox)

There are several data-driven methods for constructing ROMs, including dynamic mode decomposition [45], operator inference to learn a ROM with polynomial structure inspired by known governing equations [15, 40], sparse regression to identify reduced-order latent dynamics [8, 44], manifold learning using deep auto-encoders [30], data-driven approximation of time-integration schemes [56], and Gaussian process modeling for parametric reduced basis methods [17]. This paper focuses on operator inference (OpInf), a non-intrusive, projection-based model reduction technique introduced in [40]. OpInf learns a ROM by inferring reduced-order operators through a deterministic linear regression. The regression requires solution data of the states and their corresponding time derivatives; if needed, the time derivatives are typically estimated from the states with finite differences. Importantly, numerical regularization is often required in the inference problem to address ill-conditioning, over-fitting, and/or model inadequacy [31, 48], but the regularization selection remains a subject of further investigation. Errors introduced by time derivative approximation, noise in the solution data, and/or numerical regularization have not been directly addressed in the literature, with the exception of [50], which considers an active sampling strategy for overcoming noise pollution in the training data. Even without noise in the data, establishing error bounds or uncertainty guarantees in OpInf is challenging because the learning is driven by data, although [49] establishes probabilistic error estimates for linear OpInf ROMs in the case where the reduced-order operators can be ensured to be the same as those determined by explicit projection as in [39].

The main contribution of this work is to establish a probabilistic formulation of OpInf, thereby facilitating uncertainty quantification for non-intrusive ROMs of time-dependent systems. Our strategy accounts for approximation errors and data noise, endows the resulting ROM with uncertainty bounds, and provides statistical insight into the numerical regularization. Specifically, the task of inferring reduced-order operators is posed as a Bayesian inversion problem [6], which incorporates prior knowledge and observational information to update the probabilistic distribution of unknown system parameters. Bayesian approaches are especially useful in data-driven modeling and analysis [9] and, in particular, have recently shown promise in estimating the operators in differential equations, see, e.g., [22, 38, 55]. In our case, multivariate normal distributions are adopted to define the prior and likelihood when Bayes' rule is applied, yielding a Bayesian linear regression to estimate the reduced-order operators. Different from the deterministic OpInf workflow that gives point estimates only, the proposed Bayesian version derives posterior distributions of the learned reduced-order operators, whose uncertainties are subsequently propagated to solutions through time integration. In other words, endowing the reduced-order operators with probabilistic descriptions results in ROM predictions with quantifiable uncertainties. Additionally, our probabilistic treatment of the regression leads to a selection algorithm for the regularization hyperparameters.

We present this work in honor of the lifetime academic achievements of Professor J. Tinsley Oden, especially his seminal work in highlighting the importance of *verification & validation* [1, 2, 37] and *uncertainty quantification* [35, 36] to achieve predictive models in computational engineering and sciences [34]. He laid the mathematical foundation of predictive computational science [35] using Bayesian methods and has utilized the framework in a substantial range of scientific, medical and engineering applications [13, 20, 27, 41]. Professor Oden's pioneering, groundbreaking research on predictive computational science sets the stage for this work and inspires us to enable the uncertainty quantification and model validation of data-driven ROMs through a Bayesian approach.

The paper is structured as follows. Section 2 introduces OpInf in its established deterministic framework. Section 3 endows OpInf with a Bayesian perspective by establishing a posterior distribution for the operators of a reduced-order system, and Section 4 discusses the implications for the numerical regularization that is needed in the ROM learning formulation. Numerical results for two examples are presented in Section 5, and Section 6 makes concluding remarks.

2. Deterministic operator inference

We consider problems governed by partial differential equations, written after spatial discretization in the form

$$\frac{d}{dt}\mathbf{q}(t) = \mathbf{F}(\mathbf{q}(t), \mathbf{u}(t)), \quad \mathbf{q}(t_0) = \mathbf{q}_0, \quad t \in [t_0, t_f], \quad (2.1)$$

in which $\mathbf{q}(t) \in \mathbb{R}^n$ is the state vector collecting values of n degrees of freedom at time t , $\mathbf{u}(t) \in \mathbb{R}^m$ collects m input values at time t (e.g., encoding boundary conditions or time-dependent source terms), $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ maps state/input pairs to the state evolution, t_0 and t_f are the initial and final time, respectively, and \mathbf{q}_0 is the prescribed initial condition. A projection-based ROM approximates the full-order system (2.1) in a subspace with reduced dimensionality $r \ll n$, spanned by a collection of r basis vectors $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_r] \in \mathbb{R}^{n \times r}$. One widely used approach to construct the basis \mathbf{V} is the proper orthogonal decomposition (POD) [5, 16, 46]. For a collection of known state snapshots $\mathbf{q}_j := \mathbf{q}(t_j)$ and inputs $\mathbf{u}_j := \mathbf{u}(t_j)$, $j = 0, 1, \dots, k-1$, at k time instances $t_0 < t_1 < \cdots < t_{k-1}$, organized as $\mathbf{Q} = [\mathbf{q}_0 \ \mathbf{q}_1 \ \cdots \ \mathbf{q}_{k-1}] \in \mathbb{R}^{n \times k}$ and $\mathbf{U} = [\mathbf{u}_0 \ \mathbf{u}_1 \ \cdots \ \mathbf{u}_{k-1}] \in \mathbb{R}^{m \times k}$, POD takes advantage of the singular value decomposition $\mathbf{Q} = \mathbf{\Phi}\mathbf{\Xi}\mathbf{\Psi}^\top$ by using the first r columns of $\mathbf{\Phi}$ as the r -dimensional reduced basis, i.e., $\mathbf{V} = \mathbf{\Phi}_{:,1:r}$. Projection-based methods, such as those based on POD, have the key advantage of producing ROMs that inherit structure from the full-order system (2.1) [16]. This observation motivates us to consider, for a wide class of problems, ROMs with the following polynomial structure:

$$\frac{d}{dt}\hat{\mathbf{q}}(t) = \hat{\mathbf{A}}\hat{\mathbf{q}}(t) + \hat{\mathbf{H}}[\hat{\mathbf{q}}(t) \otimes \hat{\mathbf{q}}(t)] + \hat{\mathbf{B}}\mathbf{u}(t) + \hat{\mathbf{c}}, \quad \hat{\mathbf{q}}(t_0) = \mathbf{V}^\top \mathbf{q}_0, \quad t \in [t_0, t_f], \quad (2.2)$$

where $\hat{\mathbf{q}}(t) \in \mathbb{R}^r$ is the reduced state, \otimes denotes the Kronecker product [51], and $\hat{\mathbf{A}} \in \mathbb{R}^{r \times r}$, $\hat{\mathbf{H}} \in \mathbb{R}^{r \times r^2}$, $\hat{\mathbf{B}} \in \mathbb{R}^{r \times m}$, and $\hat{\mathbf{c}} \in \mathbb{R}^r$ are reduced-order operators that together comprise a low-dimensional representation of \mathbf{F} . The full- and reduced-order states are related by a linear approximation $\mathbf{q}(t) \approx \mathbf{V}\hat{\mathbf{q}}(t)$.

The data-driven *operator inference* (OpInf) method [40] employs the projection-based setting of (2.2), but utilizes a regression problem to find the reduced-order operators that best match the snapshot data in a minimum-residual sense. For non-polynomial problems, training data may be pre-processed with variable transformations to induce an exact or approximate polynomial structure as needed [42]. Such a non-intrusive approach avoids entailing intrusive queries and access to source code. The regression problem to infer the reduced-order operators is written as

$$\min_{\hat{\mathbf{A}}, \hat{\mathbf{H}}, \hat{\mathbf{B}}, \hat{\mathbf{c}}} \left\{ \sum_{j=0}^{k-1} \left\| \hat{\mathbf{A}}\hat{\mathbf{q}}_j + \hat{\mathbf{H}}[\hat{\mathbf{q}}_j \otimes \hat{\mathbf{q}}_j] + \hat{\mathbf{B}}\mathbf{u}_j + \hat{\mathbf{c}} - \dot{\hat{\mathbf{q}}}_j \right\|_2^2 + \mathcal{P}([\hat{\mathbf{A}} \ \hat{\mathbf{H}} \ \hat{\mathbf{B}} \ \hat{\mathbf{c}}]) \right\}, \quad (2.3)$$

in which the first term is a loss function defined by the residual of (2.2) at the k time instances $t_0 < t_1 < \cdots < t_{k-1}$, and the second term \mathcal{P} represents a regularization. Here the reduced-state dataset $\{\hat{\mathbf{q}}_j\}_{j=0}^{k-1}$ is constructed by evaluating the projection of known snapshots onto the reduced basis \mathbf{V} , i.e., $\hat{\mathbf{q}}_j = \mathbf{V}^\top \mathbf{q}_j$, and the time-derivative data $\{\dot{\hat{\mathbf{q}}}_j\}_{j=0}^{k-1}$ are then typically approximated using finite differences. The regularization term \mathcal{P} penalizes the entries of the learned quantities, which is often necessary to induce stability in the resulting ROM [31]. The training time instances are contained in $[t_0, t_{k-1}]$, a subset of the full time interval of interest $[t_0, t_f]$, and the reduced-order solution over $[t_k, t_f]$ will be predictive through the inferred ROM.

Remark 1: The quadratic interaction $\hat{\mathbf{q}} \otimes \hat{\mathbf{q}} = \text{vec}(\hat{\mathbf{q}}\hat{\mathbf{q}}^\top) \in \mathbb{R}^{r^2}$ has only $r(r+1)/2 < r^2$ degrees of freedom due to symmetry, i.e., the upper-triangular elements of $\hat{\mathbf{q}}\hat{\mathbf{q}}^\top$ completely prescribe $\hat{\mathbf{q}} \otimes \hat{\mathbf{q}}$. Hence, in practice we may learn an $r \times r(r+1)/2$ quadratic operator $\hat{\mathbf{H}}$ as opposed to an $r \times r^2$ quadratic operator.

The least-squares problem (2.3) can be rewritten in a matrix form and decoupled into r least-squares

problems, one for each ordinary differential equation in the reduced system (2.2), i.e.,

$$\min_{\hat{\mathbf{O}}} \left\{ \left\| \mathbf{D}\hat{\mathbf{O}}^\top - \mathbf{R}^\top \right\|_F^2 + \mathcal{P}(\hat{\mathbf{O}}) \right\} = \sum_{i=1}^r \min_{\hat{\mathbf{o}}_i} \left\{ \left\| \mathbf{D}\hat{\mathbf{o}}_i - \mathbf{r}_i \right\|_2^2 + \mathcal{P}_i(\hat{\mathbf{o}}_i) \right\}, \quad (2.4)$$

where

$$\begin{aligned} \hat{\mathbf{O}} &:= [\hat{\mathbf{A}} \ \hat{\mathbf{H}} \ \hat{\mathbf{B}} \ \hat{\mathbf{c}}] := [\hat{\mathbf{o}}_1 \ \cdots \ \hat{\mathbf{o}}_r]^\top \in \mathbb{R}^{r \times d(r,m)}, \\ \mathbf{D} &:= [\hat{\mathbf{Q}}^\top \ (\hat{\mathbf{Q}} \odot \hat{\mathbf{Q}})^\top \ \mathbf{U}^\top \ \mathbf{1}_k] \in \mathbb{R}^{k \times d(r,m)}, \\ \hat{\mathbf{Q}} &:= [\hat{\mathbf{q}}_0 \ \hat{\mathbf{q}}_1 \ \cdots \ \hat{\mathbf{q}}_{k-1}] \in \mathbb{R}^{r \times k}, \\ \mathbf{R} &:= [\hat{\mathbf{q}}_0 \ \hat{\mathbf{q}}_1 \ \cdots \ \hat{\mathbf{q}}_{k-1}] := [\mathbf{r}_1 \ \cdots \ \mathbf{r}_r]^\top \in \mathbb{R}^{r \times k}, \quad \text{and} \\ \mathbf{U} &:= [\mathbf{u}_0 \ \mathbf{u}_1 \ \cdots \ \mathbf{u}_{k-1}] \in \mathbb{R}^{m \times k}, \end{aligned} \quad (2.5)$$

in which $d(r, m) = r + r(r + 1)/2 + m + 1$, $\mathbf{1}_k \in \mathbb{R}^k$ is a length- k vector with all entries set to 1, \odot is the Khatri-Rao product (the column-wise Kronecker product [28, 29]), and $\hat{\mathbf{o}}_i \in \mathbb{R}^{d(r,m)}$ and $\mathbf{r}_i \in \mathbb{R}^k$ are the i -th rows of $\hat{\mathbf{O}}$ and \mathbf{R} , respectively. Here an individual regularization \mathcal{P}_i is defined for each $\hat{\mathbf{o}}_i$, i.e., $\mathcal{P}(\hat{\mathbf{O}}) = \sum_{i=1}^r \mathcal{P}_i(\hat{\mathbf{o}}_i)$. Thus the OpInf comprises of r independent regressions with respect to the unknown reduced-order operators $\hat{\mathbf{o}}_i$. Note that while (2.4) is a linear regression, the corresponding ROM (2.2) defined by $\hat{\mathbf{o}}_i$ is a nonlinear dynamical system.

We usually have $k > d(r, m)$; in that case, and with $\mathcal{P}_i(\hat{\mathbf{o}}_i) = 0$, the point estimate $\tilde{\mathbf{o}}_i$ of $\hat{\mathbf{o}}_i$ through the least squares (2.4) can be yielded in an explicit form as

$$\tilde{\mathbf{o}}_i := \mathbf{D}^+ \mathbf{r}_i = (\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{r}_i, \quad (2.6)$$

\mathbf{D}^+ denoting the Moore–Penrose inverse of the matrix \mathbf{D} . Here we assume that \mathbf{D} has full column rank; the cases with ill-conditioned $\mathbf{D}^\top \mathbf{D}$ will be discussed in Section 4.

3. Bayesian operator inference

In this section, the regression for learning the reduced-order operators $\hat{\mathbf{O}} = [\hat{\mathbf{A}} \ \hat{\mathbf{H}} \ \hat{\mathbf{B}} \ \hat{\mathbf{c}}]$ is posed as a Bayesian inference problem [6, 9]. We define a probabilistic ROM through the resulting posterior distribution of $\hat{\mathbf{O}}$ substituted into the form (2.2), which enables uncertainty estimates through Monte-Carlo sampling. We abbreviate the proposed *Bayesian operator inference* method as BayesOpInf.

3.1. Bayesian inference of reduced-order operators

The quadratic structure of the ROM (2.2) introduces a model misspecification error whenever the full-order system (2.1) includes non-quadratic dynamics. Additionally, the quality of the least-squares estimation (2.4) may suffer from approximation error in the estimated time derivatives of the training snapshots and/or noise in the snapshots themselves. To account for these uncertainties, we use Bayesian inference to formulate the estimators of the reduced-order operators $\hat{\mathbf{o}}_i$ as probabilistic distributions, $i = 1, \dots, r$.

For the definition of a likelihood distribution, we model the residual error in the data-driven scheme (2.4) as independent Gaussian noise, i.e., we consider

$$\mathbf{r}_i = \mathbf{D}\hat{\mathbf{o}}_i + \boldsymbol{\epsilon}_i, \quad p(\boldsymbol{\epsilon}_i) = \mathcal{N}(\boldsymbol{\epsilon}_i | \mathbf{0}_k, \sigma_i^2 \mathbf{I}_k), \quad (3.1)$$

in which $\mathbf{0}_k$ is the k -dimensional zero vector and \mathbf{I}_k denotes the $k \times k$ identity matrix. Here the noise $\boldsymbol{\epsilon}_i \in \mathbb{R}^k$ is a random vector consisting of the error terms in the observations at k time instances. All these noise measurements are assumed to follow independent normal distributions with zero mean and variance σ_i^2 . Hence the likelihood distribution for the training data pair $(\mathbf{D}, \mathbf{r}_i)$ is written as

$$p(\mathbf{r}_i | \mathbf{D}, \hat{\mathbf{o}}_i, \sigma_i^2) = \mathcal{N}(\mathbf{r}_i | \mathbf{D}\hat{\mathbf{o}}_i, \sigma_i^2 \mathbf{I}_k). \quad (3.2)$$

The choice to model the entries of $\boldsymbol{\epsilon}_i$ as independent random variables (i.e., assuming the noise is not correlated across time) is convenient but not a requirement. An alternative, kernel-based model of the residual error is provided in Appendix 1, which accounts for correlations over time and constructs a surrogate model of the ROM closure.

We also define the prior distribution of $\hat{\boldsymbol{o}}_i$ as a multivariate normal distribution,

$$p(\hat{\boldsymbol{o}}_i | \sigma_i^2, \boldsymbol{\lambda}_i) = \mathcal{N}(\hat{\boldsymbol{o}}_i | \boldsymbol{\beta}_i, \sigma_i^2 \text{diag}(\boldsymbol{\lambda}_i)^{-1}), \quad (3.3)$$

where $\boldsymbol{\beta}_i \in \mathbb{R}^{d(r,m)}$ is a pre-defined mean vector and $\boldsymbol{\lambda}_i \in \mathbb{R}^{d(r,m)}$ is a vector of positive numbers hyper-parametrizing the prior variances of $\hat{\boldsymbol{o}}_i$. Note that the prior covariance is a diagonal matrix, implying that the prior does not impose correlations between the components of $\hat{\boldsymbol{o}}_i$.

Conditioning on the training data, Bayes' rule gives the posterior distribution $p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i, \sigma_i^2, \boldsymbol{\lambda}_i)$ as follows:

$$\begin{aligned} p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i, \sigma_i^2, \boldsymbol{\lambda}_i) &\propto p(\mathbf{r}_i | \mathbf{D}, \hat{\boldsymbol{o}}_i, \sigma_i^2) p(\hat{\boldsymbol{o}}_i | \sigma_i^2, \boldsymbol{\lambda}_i) \\ &= \mathcal{N}(\mathbf{r}_i | \mathbf{D}\hat{\boldsymbol{o}}_i, \sigma_i^2 \mathbf{I}_k) \mathcal{N}(\hat{\boldsymbol{o}}_i | \boldsymbol{\beta}_i, \sigma_i^2 \text{diag}(\boldsymbol{\lambda}_i)^{-1}) \\ &= \mathcal{N}(\mathbf{D}\hat{\boldsymbol{o}}_i | \mathbf{r}_i, \sigma_i^2 \mathbf{I}_k) \mathcal{N}(\hat{\boldsymbol{o}}_i | \boldsymbol{\beta}_i, \sigma_i^2 \text{diag}(\boldsymbol{\lambda}_i)^{-1}) \\ &= \mathcal{N}(\mathbf{r}_i | \mathbf{D}\boldsymbol{\beta}_i, \sigma_i^2 \mathbf{D} \text{diag}(\boldsymbol{\lambda}_i)^{-1} \mathbf{D}^\top + \sigma_i^2 \mathbf{I}_k) \mathcal{N}(\hat{\boldsymbol{o}}_i | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad \text{i.e.,} \\ p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i, \sigma_i^2, \boldsymbol{\lambda}_i) &= \mathcal{N}(\hat{\boldsymbol{o}}_i | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \end{aligned} \quad (3.4)$$

in which

$$\begin{aligned} \boldsymbol{\Sigma}_i &= \boldsymbol{\Sigma}_i(\mathbf{D}, \sigma_i^2, \boldsymbol{\lambda}_i) = \sigma_i^2 [\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{D}]^{-1}, \\ \boldsymbol{\mu}_i &= \boldsymbol{\mu}_i(\mathbf{r}_i, \mathbf{D}, \boldsymbol{\lambda}_i) = [\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{D}]^{-1} (\text{diag}(\boldsymbol{\lambda}_i) \boldsymbol{\beta}_i + \mathbf{D}^\top \mathbf{r}_i) \\ &= \boldsymbol{\beta}_i + \underbrace{[\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{D}]^{-1} \mathbf{D}^\top (\mathbf{r}_i - \mathbf{D}\boldsymbol{\beta}_i)}_{\delta \boldsymbol{\mu}_i(\mathbf{r}_i, \mathbf{D}, \boldsymbol{\lambda}_i)}, \end{aligned} \quad (3.5)$$

all consistent with the results of Bayesian linear regression [6].

To marginalize $(\sigma_i^2, \boldsymbol{\lambda}_i)$ in the distribution $p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i, \sigma_i^2, \boldsymbol{\lambda}_i)$ and obtain $p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i)$ as the inference of $\hat{\boldsymbol{o}}_i$, we adopt an empirical Bayes method—maximum log marginal likelihood [54]—to estimate a set of optimal values of $(\sigma_i^2, \boldsymbol{\lambda}_i)$. Mathematically, the maximization problem is written as

$$\begin{aligned} (\sigma_i^{*2}, \boldsymbol{\lambda}_i^*) &= \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i} \log p(\mathbf{r}_i | \mathbf{D}, \sigma_i^2, \boldsymbol{\lambda}_i) \\ &= \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i} \log \int p(\mathbf{r}_i | \mathbf{D}, \hat{\boldsymbol{o}}_i, \sigma_i^2) p(\hat{\boldsymbol{o}}_i | \sigma_i^2, \boldsymbol{\lambda}_i) d\hat{\boldsymbol{o}}_i \\ &\stackrel{\text{from (3.4)}}{=} \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i} \log \mathcal{N}(\mathbf{r}_i | \mathbf{D}\boldsymbol{\beta}_i, \sigma_i^2 (\mathbf{D} \text{diag}(\boldsymbol{\lambda}_i)^{-1} \mathbf{D}^\top + \mathbf{I}_k)) \\ &= \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i} -\frac{1}{2\sigma_i^2} \|\mathbf{r}_i - \mathbf{D}\boldsymbol{\mu}_i(\mathbf{r}_i, \mathbf{D}, \boldsymbol{\lambda}_i)\|_2^2 - \frac{1}{2\sigma_i^2} \left\| \text{diag}(\boldsymbol{\lambda}_i)^{1/2} \delta \boldsymbol{\mu}_i(\mathbf{r}_i, \mathbf{D}, \boldsymbol{\lambda}_i) \right\|_2^2 \\ &\quad - \frac{1}{2} \log |\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{D}| + \frac{1}{2} \log(\boldsymbol{\lambda}_i)^\top \mathbf{1}_{d(r,m)} - \frac{k}{2} \log(\sigma_i^2) - \frac{k}{2} \log(2\pi), \end{aligned} \quad (3.6)$$

where $|\cdot|$ indicates the matrix determinant. Thus the final posterior distribution $p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i)$ is approximated as

$$p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i) = p(\hat{\boldsymbol{o}}_i | \mathbf{D}, \mathbf{r}_i, \sigma_i^{*2}, \boldsymbol{\lambda}_i^*) = \mathcal{N}(\hat{\boldsymbol{o}}_i | \boldsymbol{\mu}_i(\mathbf{r}_i, \mathbf{D}, \boldsymbol{\lambda}_i^*), \boldsymbol{\Sigma}_i(\mathbf{D}, \sigma_i^{*2}, \boldsymbol{\lambda}_i^*)). \quad (3.7)$$

Furthermore, by taking partial derivative of the objective function in (3.6) with respect to σ_i^2 and setting it

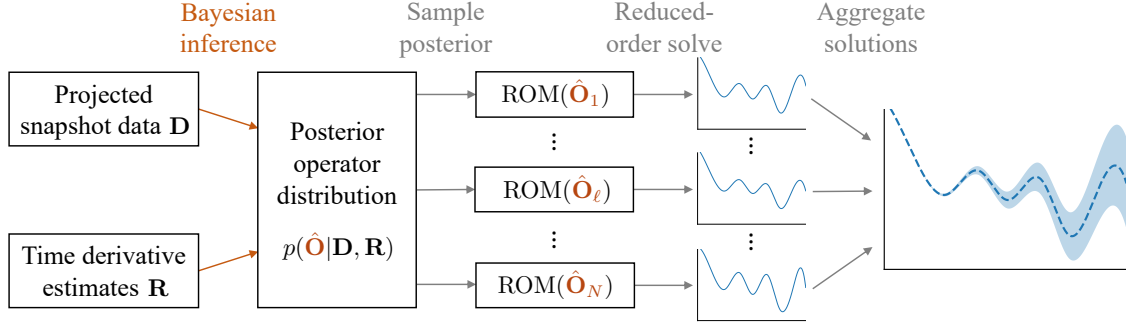


Figure 1: The BayesOpInf workflow. To derive the probabilistic ROM, infer the multivariate normal posterior distribution for the operator matrix $\hat{\mathbf{O}}$ from snapshot and time derivative data. This is a one-time, offline cost. To evaluate the probabilistic ROM in the online stage, draw samples from the posterior distribution of $\hat{\mathbf{O}}$ and solve the corresponding ROM realizations, then calculate the statistics of these ROM solutions.

to zero, we see that the maximum is achieved at

$$\sigma_i^{*2} = \frac{\|\mathbf{r}_i - \mathbf{D}\boldsymbol{\mu}_i\|_2^2 + \left\| \text{diag}(\boldsymbol{\lambda}_i^*)^{1/2} \delta\boldsymbol{\mu}_i \right\|_2^2}{k}. \quad (3.8)$$

Remark 2: A pragmatic choice is to set $\boldsymbol{\beta}_i = \mathbf{0}$ and thus $\delta\boldsymbol{\mu}_i = \boldsymbol{\mu}_i$. When we assume that the prior variance values are very large ($\boldsymbol{\lambda}_i \rightarrow \mathbf{0}$), i.e., the prior of $\hat{\mathbf{o}}_i$ is uninformative, we have $\boldsymbol{\mu}_i = \tilde{\mathbf{o}}_i = \mathbf{D}^+ \mathbf{r}_i$, $\boldsymbol{\Sigma}_i = \sigma_i^{*2} (\mathbf{D}^T \mathbf{D})^{-1}$, and $\sigma_i^{*2} = \|\mathbf{r}_i - \mathbf{D}\tilde{\mathbf{o}}_i\|_2^2 / k$. In this case, the posterior mean vector of $\hat{\mathbf{o}}_i$ coincides with the original least squares estimate $\tilde{\mathbf{o}}_i$ as in deterministic OpInf, and σ_i^{*2} is the mean squared error of the linear regression.

Remark 3: If the pre-defined prior mean vector of $\hat{\mathbf{o}}_i$ is taken as its point estimate $\tilde{\mathbf{o}}_i$, i.e., $\boldsymbol{\beta}_i = \tilde{\mathbf{o}}_i = \mathbf{D}^+ \mathbf{r}_i$, then we have $\delta\boldsymbol{\mu}_i = \mathbf{0}$ and $\boldsymbol{\mu}_i = \tilde{\mathbf{o}}_i$, as well as $\sigma_i^{*2} = \|\mathbf{r}_i - \mathbf{D}\tilde{\mathbf{o}}_i\|_2^2 / k$. Moreover, by simplifying (3.6) the estimate of $\boldsymbol{\lambda}_i$ is given by

$$\begin{aligned} \boldsymbol{\lambda}_i^* &= \arg \max_{\boldsymbol{\lambda}_i} -\frac{1}{2} \log |\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^T \mathbf{D}| + \frac{1}{2} \log |\text{diag}(\boldsymbol{\lambda}_i)| \\ &= \arg \max_{\boldsymbol{\lambda}_i} -\frac{1}{2} \log \left| \mathbf{I} + \text{diag}(\boldsymbol{\lambda}_i)^{-\frac{1}{2}} \mathbf{D}^T \mathbf{D} \text{diag}(\boldsymbol{\lambda}_i)^{-\frac{1}{2}} \right|. \end{aligned} \quad (3.9)$$

Since $\text{diag}(\boldsymbol{\lambda}_i)^{-\frac{1}{2}} \mathbf{D}^T \mathbf{D} \text{diag}(\boldsymbol{\lambda}_i)^{-\frac{1}{2}}$ is semi-positive definite, $\inf_{\boldsymbol{\lambda}_i} \left| \mathbf{I} + \text{diag}(\boldsymbol{\lambda}_i)^{-\frac{1}{2}} \mathbf{D}^T \mathbf{D} \text{diag}(\boldsymbol{\lambda}_i)^{-\frac{1}{2}} \right| = 1$, which is achieved as $\text{diag}(\boldsymbol{\lambda}_i)^{-1} \rightarrow \mathbf{0}$ ($\boldsymbol{\lambda}_i \rightarrow \infty^{d(r,m)}$), i.e., when the prior of $\hat{\mathbf{o}}_i$ is deterministic. In this case, the Bayesian formulation reduces to deterministic OpInf, indicating that the deterministic version essentially forms an optimal solution, provided that the Gram matrix $\mathbf{D}^T \mathbf{D}$ is well-conditioned.

3.2. A probabilistic reduced-order model

Because the OpInf regression naturally decouples along the rows of the reduced operator matrix $\hat{\mathbf{O}} = [\hat{\mathbf{A}} \ \hat{\mathbf{H}} \ \hat{\mathbf{B}} \ \hat{\mathbf{c}}]$, we likewise treat the Bayesian inference of each row $\hat{\mathbf{o}}_i$ independently, $1 \leq i \leq r$. Hence $\hat{\mathbf{O}}$ is described by the posterior distribution

$$p(\hat{\mathbf{O}} | \mathbf{D}, \mathbf{R}) = \prod_{i=1}^r p(\hat{\mathbf{o}}_i | \mathbf{D}, \mathbf{r}_i). \quad (3.10)$$

Given $\hat{\mathbf{O}}$, we solve for the reduced state $\hat{\mathbf{q}}(t)$ through the corresponding reduced-order system (2.2) over $[t_0, t_f]$. The reduced state can therefore be viewed as a stochastic process because it depends on the random

in which $\delta_{\text{Dirac}}(\cdot)$ denotes the Dirac-delta function. This prior Gaussian process is consistent with the likelihood and prior definitions adopted in (3.2) and (3.3), respectively, for the Bayesian inference (3.4).

3.3. Equivalence to ridge regression

Recall the posterior mean vector of $\hat{\mathbf{o}}_i$ given in (3.5). The correction term $\delta\boldsymbol{\mu}_i = \boldsymbol{\mu}_i - \boldsymbol{\beta}_i$ can be expressed alternatively as

$$\delta\boldsymbol{\mu}_i = \arg \min_{\boldsymbol{\eta} \in \mathbb{R}^{d(r,m)}} \left\{ \|\mathbf{r}_i - \mathbf{D}(\boldsymbol{\beta}_i + \boldsymbol{\eta})\|_2^2 + \|\text{diag}(\boldsymbol{\lambda}_i)^{\frac{1}{2}} \boldsymbol{\eta}\|_2^2 \right\}, \quad (3.14)$$

implying that the posterior mean is the solution of a *ridge regression* problem, i.e., a linear least-squares problem with a Tikhonov regularization (in this case $\|\text{diag}(\boldsymbol{\lambda}_i)^{\frac{1}{2}} \boldsymbol{\eta}\|_2^2$). The use of Tikhonov regularization in OpInf has been reported in [31, 48]. Thus another source of modeling uncertainty is introduced to the ROM by the Tikhonov regularization defined with OpInf. Based on the equivalence between the posterior mean of the inferred operators and ridge regression, it can be claimed that the uncertainty introduced by Tikhonov regularization is quantified through the proposed Bayesian framework as well.

On the other hand, the selection of the regularization parameters $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_r$ has been shown to be important in addressing ill-conditioning, over-fitting, or model inadequacy in OpInf [31, 48]. The proposed Bayesian method in this work not only formulates a framework of uncertainty quantification, but also provides a new perspective for understanding the regularization selection process for deterministic OpInf. We address this further in Section 4.

4. Regularization selection

Without regularization, the OpInf point estimate (2.6) is represented with the Moore–Penrose inverse of matrix \mathbf{D} and requires that \mathbf{D} has full column rank, i.e., the Gram matrix $\mathbf{D}^T \mathbf{D}$ is invertible. In practice, however, the condition number of $\mathbf{D}^T \mathbf{D}$ may be large, which can lead to a poor estimate of the reduced operators and compromise the stability of the corresponding ROM. In this case, making use of a proper regularization can be critical for stabilizing the OpInf scheme and improving the robustness of the ROM [31, 48].

As specified in Section 3.3, the posterior mean of inferred operators is equivalent to an estimator through ridge regression. The penalty coefficients $\boldsymbol{\lambda}_i$ in the Tikhonov regularization correspond to the prior variances of the operators $\hat{\mathbf{o}}_i$. As given in (3.6), the values of the hyperparameters $\boldsymbol{\lambda}_i$ can be determined by an empirical Bayes method of maximizing marginal likelihood,

$$\begin{aligned} \mathcal{L}_i(\sigma_i^2, \boldsymbol{\lambda}_i) &= \log \mathcal{N}(\mathbf{r}_i | \mathbf{D}\boldsymbol{\beta}_i, \sigma_i^2 (\mathbf{D}\text{diag}(\boldsymbol{\lambda}_i)^{-1} \mathbf{D}^T + \mathbf{I}_k)) \\ &= -\frac{1}{2\sigma_i^2} \|\mathbf{r}_i - \mathbf{D}\boldsymbol{\mu}_i\|_2^2 - \frac{1}{2\sigma_i^2} \left\| \text{diag}(\boldsymbol{\lambda}_i)^{1/2} \delta\boldsymbol{\mu}_i \right\|_2^2 \\ &\quad - \frac{1}{2} \log |\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^T \mathbf{D}| + \frac{1}{2} \log(\boldsymbol{\lambda}_i)^T \mathbf{1}_{d(r,m)} - \frac{k}{2} \log(\sigma_i^2) - \frac{k}{2} \log(2\pi). \end{aligned} \quad (4.1)$$

In this sense, the empirical method provides guidance for determining the penalty parameters $\boldsymbol{\lambda}_i$ in regularized OpInf.

Taking partial derivatives of \mathcal{L}_i with respect to σ_i^2 and all the entries of $\boldsymbol{\lambda}_i = [\lambda_{i,1}, \dots, \lambda_{i,d(r,m)}]^T$, the extremum conditions are

$$\frac{\partial \mathcal{L}_i}{\partial \sigma_i^2} = 0 \quad \implies \quad \sigma_i^{*2} = \frac{\|\mathbf{r}_i - \mathbf{D}\boldsymbol{\mu}_i\|_2^2 + \left\| \text{diag}(\boldsymbol{\lambda}_i^*)^{1/2} \delta\boldsymbol{\mu}_i \right\|_2^2}{k},$$

which has already been given in (3.8), and

$$\frac{\partial \mathcal{L}_i}{\partial \lambda_{i,j}} = 0 \quad \implies \quad \lambda_{i,j}^* (\delta\mu_{i,j}^2 + \Sigma_{i,jj}) = \sigma_i^{*2}, \quad 1 \leq j \leq d(r,m), \quad (4.2)$$

in which $\delta\mu_{i,j}$ and $\Sigma_{i,jj}$ are the j -th entry of $\delta\boldsymbol{\mu}_i$ and the j -th diagonal entry of $\boldsymbol{\Sigma}_i$, respectively.¹ Note that (4.2) is an implicit representation of $\boldsymbol{\lambda}_i^*$, the penalty coefficients of Tikhonov regularization.

For each reduced system mode $i = 1, \dots, r$, the problem of maximum marginal likelihood in (4.1) is non-convex with $1 + d(r, m)$ unknowns. Solving such a problem using gradient-based methods is possible, but inconvenient for moderately sized r and not guaranteed to find a global optimum. To alleviate complexity, we impose the constraint $\lambda_{i,j} = \lambda_i$ for all $1 \leq j \leq d(r, m)$, hence $\boldsymbol{\lambda}_i = [\lambda_i \ \lambda_i \ \dots \ \lambda_i]^\top \in \mathbb{R}^{d(r, m)}$. In this case, the regularization term in (2.4), as well as in steps 4 and 5 of Algorithm 1, is an L_2 -regularization with penalty coefficient $\lambda_i > 0$, i.e.,

$$\mathcal{P}_i(\boldsymbol{\mu}_i) = \mathcal{P}_i(\boldsymbol{\beta}_i + \delta\boldsymbol{\mu}_i) = \left\| \text{diag}(\boldsymbol{\lambda}_i)^{1/2} \delta\boldsymbol{\mu}_i \right\|_2^2 = \lambda_i \|\delta\boldsymbol{\mu}_i\|_2^2. \quad (4.3)$$

This means that the operator entries corresponding to the reduced system mode \hat{q}_i , i.e., the i th rows of $\hat{\mathbf{c}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$, and $\hat{\mathbf{B}}$, are regularized equally, while operator entries corresponding to different system modes are regularized separately. Summing (4.2) over all $1 \leq j \leq d(r, m)$ yields

$$\lambda_i^* (\|\delta\boldsymbol{\mu}_i\|_2^2 + \text{tr}(\boldsymbol{\Sigma}_i)) = d(r, m) \sigma_i^{*2}, \quad (4.4)$$

in which $\text{tr}(\boldsymbol{\Sigma}_i)$ can alternatively be written as follows [54]:

$$\text{tr}(\boldsymbol{\Sigma}_i) = \sigma_i^{*2} \sum_{l=1}^{d(r, m)} \frac{1}{\lambda_i^* + g_l}, \quad (4.5)$$

where $g_1, \dots, g_{d(r, m)} \geq 0$ are the non-negative eigenvalues of the Gram matrix $\mathbf{D}^\top \mathbf{D}$. This representation of $\text{tr}(\boldsymbol{\Sigma}_i)$ has the computational advantage of avoiding explicit matrix inversion in (3.5) to evaluate $\boldsymbol{\Sigma}_i$. We then have

$$\lambda_i^* = \frac{\sigma_i^{*2}}{\|\delta\boldsymbol{\mu}_i\|_2^2} \underbrace{\sum_{l=1}^{d(r, m)} \frac{g_l}{\lambda_i^* + g_l}}_{:=\gamma_i} = \frac{\gamma_i \sigma_i^{*2}}{\|\delta\boldsymbol{\mu}_i\|_2^2}, \quad (4.6)$$

in which $0 < \gamma_i < d(r, m)$. The following equality shows the connections among the noise σ_i^{*2} and the two terms in the loss function of the ridge regression (3.14):

$$\frac{\lambda_i^* \|\delta\boldsymbol{\mu}_i\|_2^2}{\gamma_i} = \frac{\|\mathbf{r}_i - \mathbf{D}\boldsymbol{\mu}_i\|_2^2}{k - \gamma_i} = \sigma_i^{*2}. \quad (4.7)$$

Substituting (3.8) into (4.6), λ_i^* is expressed by the posterior mean $\delta\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, both being dependent on λ_i^* as well. Thus an implicit representation of λ_i^* is given by $\lambda_i^* = \mathcal{F}(\lambda_i^*)$, where \mathcal{F} denotes the right-hand side of (4.6). In this work, we suggest a fixed-point iterative strategy, $(\lambda_i^*)_{s+1} = \mathcal{F}((\lambda_i^*)_s)$, $s = 0, 1, \dots$, as a practical approach to simultaneously satisfy (3.8) and (4.6) and approximately maximize (4.1). The initial values $\{(\lambda_i^*)_0\}_{i=1}^r$ can be set to a selection of regularization coefficients that result in a stable ROM, but which are not necessarily optimal. This iterative regularization update procedure is given by Algorithm 2, which considers a zero prior mean of reduced operators, i.e., $\boldsymbol{\beta}_i = \mathbf{0}$ and $\boldsymbol{\mu}_i = \delta\boldsymbol{\mu}_i$. We note that while Algorithm 2 is computationally efficient, it does not include stability information for the reduced-order ODE system (2.2) because of its completely non-intrusive nature.

¹The evaluation of $\frac{\partial \mathcal{L}_i}{\partial \lambda_{i,j}}$ uses the identity $\partial_\theta (\log \det \mathbf{A}(\theta)) = \text{tr}(\mathbf{A}(\theta)^{-1} \partial_\theta \mathbf{A}(\theta))$, in which \mathbf{A} is an invertible matrix parametrized by θ .

Algorithm 2 Iterative regularization update for BayesOpInf with $\lambda_{i,j} = \lambda_i$.

```

1: procedure BAYESOPINFREGULARIZATION( Data matrix  $\mathbf{D} \in \mathbb{R}^{k \times d(r,m)}$ , projected time derivatives
    $\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_r]^\top \in \mathbb{R}^{r \times k}$ , initial regularization parameters  $\lambda_1, \dots, \lambda_r \geq 0$ , convergence tolerance  $\varepsilon > 0$ 
)
2:    $g_1, \dots, g_{d(r,m)} \leftarrow \mathbf{eig}(\mathbf{D}^\top \mathbf{D})$  # Get eigenvalues of the data Gram matrix.
3:    $\boldsymbol{\lambda}^{(0)} \leftarrow [\lambda_1 \cdots \lambda_r]$ 
4:   for  $\ell = 0, 1, \dots$  do
5:     for  $i = 1, \dots, r$  do
6:        $\boldsymbol{\mu}_i \leftarrow \arg \min_{\boldsymbol{\mu}_i} \|\mathbf{D}\boldsymbol{\mu}_i - \mathbf{r}_i\|_2^2 + \lambda_i \|\boldsymbol{\mu}_i\|_2^2$  # Update operator mean.
7:        $\sigma_i^{*2} \leftarrow \frac{1}{k} \left( \|\mathbf{D}\boldsymbol{\mu}_i - \mathbf{r}_i\|_2^2 + \lambda_i \|\boldsymbol{\mu}_i\|_2^2 \right)$  # Update operator variance.
8:        $\lambda_i^* \leftarrow \frac{\sigma_i^{*2}}{\|\boldsymbol{\mu}_i\|_2^2} \sum_{l=1}^{d(r,m)} \frac{g_l}{\lambda_i + g_l}$  # Update regularization.
9:      $\boldsymbol{\lambda}^{(\ell+1)} \leftarrow [\lambda_1^* \cdots \lambda_r^*]$ 
10:    if  $\|\boldsymbol{\lambda}^{(\ell+1)} - \boldsymbol{\lambda}^{(\ell)}\|_2 / \|\boldsymbol{\lambda}^{(\ell)}\|_2 < \varepsilon$  then # Check for convergence.
11:      return  $\lambda_1^*, \dots, \lambda_r^*$ 
12:     $\lambda_1, \dots, \lambda_r \leftarrow \lambda_1^*, \dots, \lambda_r^*$  # Continue iterating if not converged.

```

5. Numerical examples

Section 5.1 demonstrates the proposed method on one-dimensional Euler equations where the state solution data are polluted by Gaussian noise. In Section 5.2, we examine a combustion application in which the regularizing terms play a crucial role for learning stable ROMs from data. The code for these experiments can be found at <https://github.com/Willcox-Research-Group/ROM-OpInf-Combustion-2D/tree/cmame2022>.

5.1. Noised Euler Equations

Consider the conservative one-dimensional compressible Euler equations for an ideal gas,

$$\frac{\partial}{\partial t} [\rho] = -\frac{\partial}{\partial x} [\rho u], \quad \frac{\partial}{\partial t} [\rho u] = -\frac{\partial}{\partial x} [\rho u^2 + p], \quad \frac{\partial}{\partial t} [\rho e] = -\frac{\partial}{\partial x} [(\rho e + p)u], \quad (5.1)$$

where u is velocity [m/s], ρ is density [kg/m³], p is pressure [Pa], ρu is specific momentum [kg/m²s], and ρe is total energy [kJ/m³]. The state variables are related via the ideal gas law $\rho e = \frac{p}{\gamma-1} + \frac{1}{2}\rho u^2$, where $\gamma = 1.4$ is the heat capacity ratio. Our goal is to learn a reduced model of (5.1) from noisy observations of the conservative variables $(\rho, \rho u, \rho e)$ and to quantify the uncertainties induced by the noise. The ideal gas law enables a change of variables $\tau : (\rho, \rho u, \rho e) \mapsto (u, p, 1/\rho)$ that transforms (5.1) into the purely quadratic system

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - \zeta \frac{\partial p}{\partial x}, \quad \frac{\partial p}{\partial t} = -\gamma p \frac{\partial u}{\partial x} - u \frac{\partial p}{\partial x}, \quad \frac{\partial \zeta}{\partial t} = -u \frac{\partial \zeta}{\partial x} + \zeta \frac{\partial u}{\partial x}, \quad (5.2)$$

where $\zeta = 1/\rho$ is the specific volume [m³/kg]. In the model (5.2), every term is the product of exactly two state variables (u, p, ζ) or their spatial derivatives, hence for this problem we seek a reduced model with quadratic terms only:

$$\frac{d}{dt} \hat{\mathbf{q}}(t) = \hat{\mathbf{H}}[\hat{\mathbf{q}}(t) \otimes \hat{\mathbf{q}}(t)]. \quad (5.3)$$

The OpInf operator matrix of (2.5) in this case is $\hat{\mathbf{O}} = \hat{\mathbf{H}} \in \mathbb{R}^{r \times d(r)}$, with column dimension $d(r) = r(r+1)/2$.

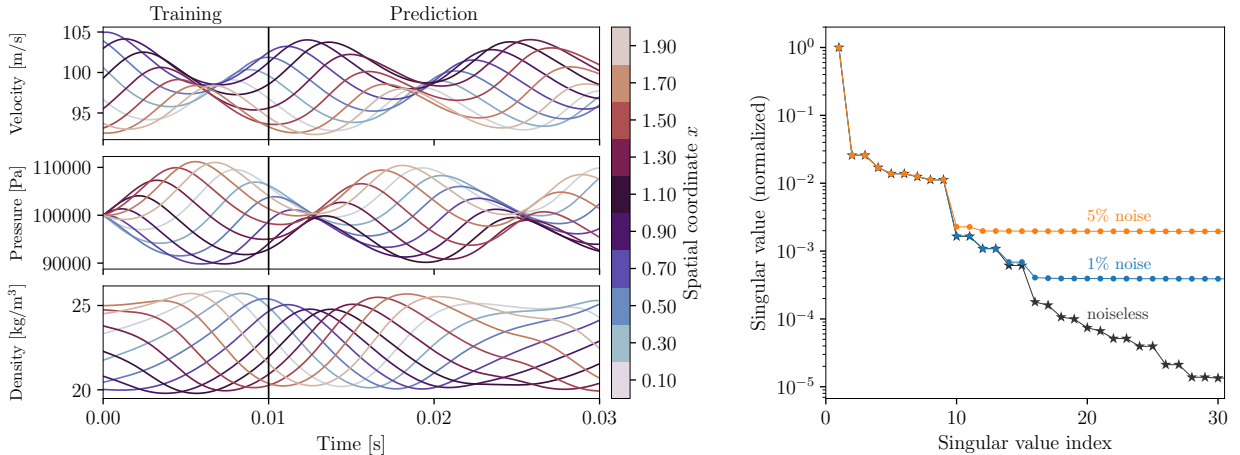


Figure 2: One of the sixty-four training trajectories at various spatial coordinates before noise corruption (left) and singular value decay of the entire training set with various noise levels (right). Training data consists of noised snapshots for $0 \leq t < 0.01$.

We consider the spatial domain $[0, 2)$ and the time domain $[t_0, t_f] = [0, 0.03]$, with periodic boundary conditions imposed at the spatial boundaries. Training data is generated over the shorter time domain $[0, 0.01]$ for multiple initial conditions: the initial pressure is set to 10^5 Pa everywhere, and initial velocity and density profiles are constructed via cubic spline interpolation by fixing values at the spatial points $x \in \{0, \frac{2}{3}, \frac{4}{3}\}$ to either 95 or 105 m/s for velocity and either 20 or 24 kg/m³ for density. We therefore have $2^{3 \times 2} = 64$ initial conditions (three interpolation nodes with two choices of values for two variables). For each initial condition, we solve the conservative system (5.1) by discretizing the spatial domain via first-order finite differences with spatial resolution $\delta x = 0.01$, then stepping forward in time via the first-order explicit Euler iteration with time step $\delta t = 10^{-5}$. The result is $k = 64 \times 1000 = 64,000$ total training snapshots, each of size $n = 200 \times 3 = 600$. See Figure 2 for an example trajectory in the training set. Before processing the snapshot data for learning, each snapshot entry is polluted with Gaussian noise proportional to the range of the associated variable to represent observation error. For example, to each snapshot entry corresponding to density we add error drawn from $\mathcal{N}(0, \varsigma^2)$ where $\varsigma = \xi(\max_{i,j}\{\rho_{i,j}\} - \min_{i,j}\{\rho_{i,j}\})$, $\xi = 5\%$ is the noise level, and $\rho_{i,j}$ is the i th entry of the j th density training snapshot. The noisy snapshots are then transformed from the conservative states of (5.1) to the states of the quadratic system (5.2) via the mapping τ and non-dimensionalized, yielding training data $\mathbf{q}_0, \dots, \mathbf{q}_{k-1} \in \mathbb{R}^n$ to be used in OpInf. Note that the noise in the lifted variables is not necessarily Gaussian since τ is a nonlinear mapping.

From the lifted, noisy snapshots we compute a POD basis $\mathbf{V} \in \mathbb{R}^{n \times r}$ and project the training data to the corresponding low-dimensional linear subspace as $\hat{\mathbf{q}}_j = \mathbf{V}^T \mathbf{q}_j$. The dimension $r = 9$ is selected based on where the decay of the singular values flattens out due to the noise pollution as seen in Figure 2. It has been widely observed that POD has a smoothing effect (see, e.g., [11, 12, 52]). The projected training data, an example of which is shown in Figure 3, exhibits relatively little noise when compared to the observed training data (see also Figure 4) up to $r = 9$ modes. However, the OpInf regression requires an estimate of the time derivatives $\dot{\hat{\mathbf{q}}}_j$, which will be highly inaccurate if computed as finite differences of the noisy projected data. We therefore apply the noise-resistant strategy of [10], which uses implicit local polynomial regression to estimate derivatives with minimal variance, to estimate each $\dot{\hat{\mathbf{q}}}_j$. Therefore, uncertainty in the OpInf regression stems from *i*) noise in the observed data, *ii*) the truncated POD representation (model form error), and *iii*) error in the estimates of the time derivatives of the observed data.

The noise in the training data necessitates some regularization in order to achieve stable reduced models, though in this example among stable models the learned OpInf ROM is not highly sensitive to the choice of regularization. We initially set $\lambda_{i,j} = 50$ for all i, j and use Algorithm 2 to iterate until the relative change (as measured in step 10) drops below $\varepsilon = 0.1\%$, which occurs after only nine iterations. We then draw 100 samples of $\hat{\mathbf{O}}$ from the posterior operator distribution defined in (3.7) and integrate the corresponding

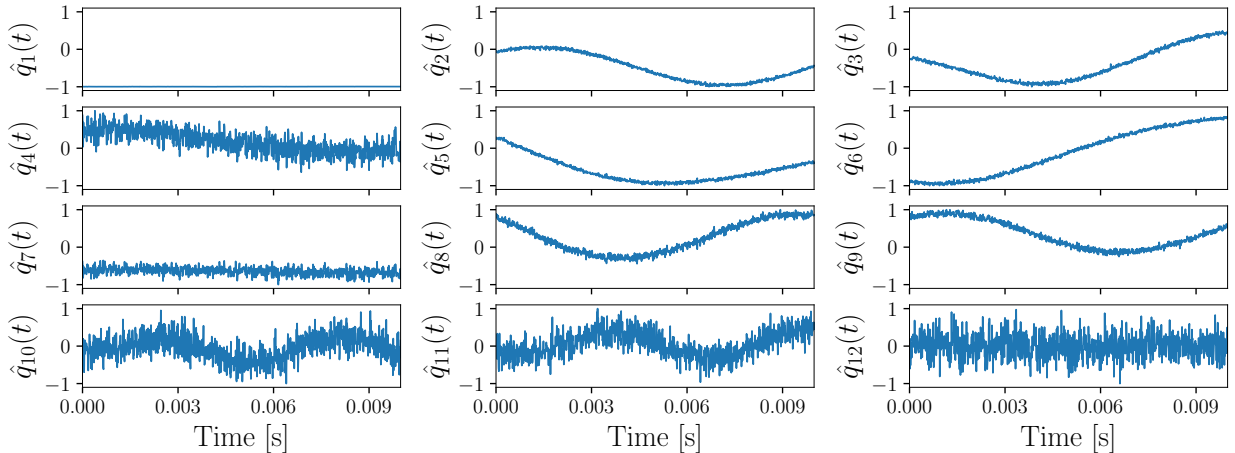


Figure 3: The first several modes of the (noisy) projected training data, normalized to the range $[-1, 1]$, for the training trajectory shown in Figure 2. The POD projection successfully filters much, but not all, of the noise on the first $r = 9$ modes.

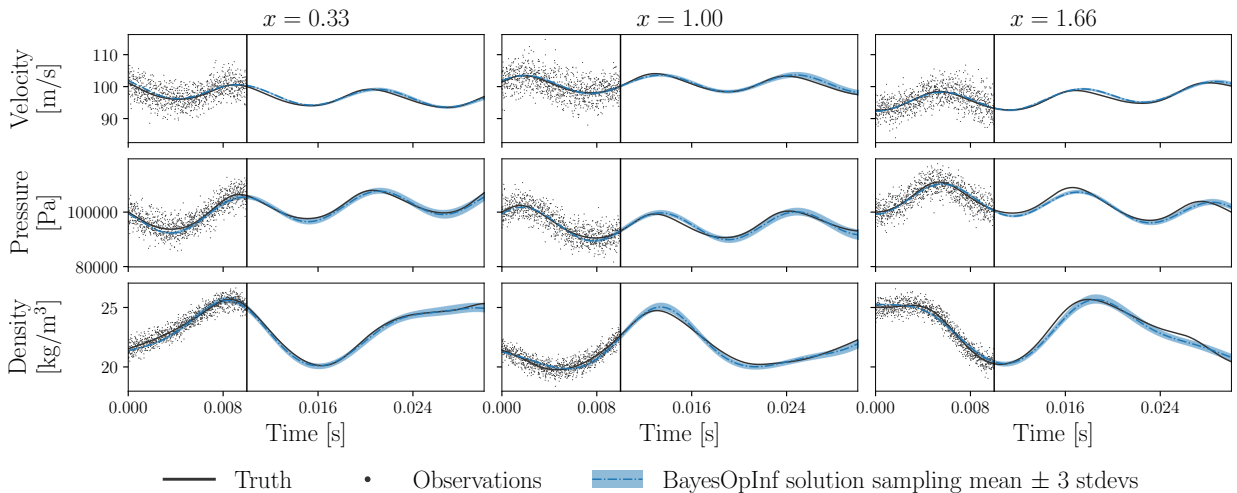


Figure 4: Traces of full-order and ROM solutions to the Euler problem at three spatial locations, with the same initial condition as in Figure 2. The noisy training data (noise level $\xi = 5\%$) is also shown as individual points over the training interval. The BayesOpInf ROM uses $r = 9$ modes, with the regularization chosen via Algorithm 2, and 100 draws are made from the posterior distribution to sample the mean and standard deviations.

ROMs over the full time domain $[t_0, t_f]$ for each of the (noisy) initial conditions in the training set. Figure 4 shows the sample mean of these draws and three standard deviations from the sample mean at several spatial locations for the trajectory shown in Figure 2; the results are representative for the set of all initial conditions considered. Note that the width of three standard deviations, which slightly increases with time, is relatively small compared to the observed data. This is because the uncertainty band shows a credible interval, which quantifies the accuracy of model fits, i.e., the learned posterior distribution describes the uncertainty relative to the true solution. Note that, over the training interval $[0, 0.01]$, the credible interval is by definition not taking into account the measurement noise in the snapshot data².

The regularization selection strategy of Algorithm 2 only approximates the maximizers of the marginal likelihood (4.1), however the cost of each iteration in Algorithm 2 is dominated only by the OpInf regres-

²The band that takes both the propagated uncertainty from the posterior model and the measurement errors (noise) into account is often referred to as a prediction interval, see chapter 9 of [47].

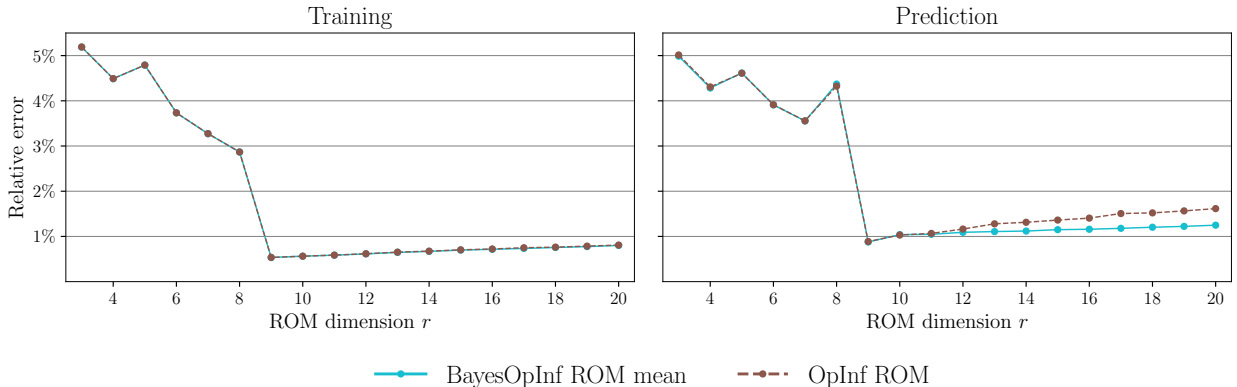


Figure 5: Average relative error over the set of 64 initial conditions for increasing basis sizes, over the training regime (left) and the prediction regime (right), for ROMs where the regularization is chosen via Algorithm 2 (BayesOpInf) and by Algorithm 3 (OpInf).

sion (determining each μ_i and Σ_i). To compare this strategy to existing methods, Figure 5 repeats the experiments described above with basis sizes $r = 3, 4, \dots, 20$ using Algorithm 2 to select the regularization, and computes the average relative error of the ROM defined by the posterior operator mean μ_i over the set of all initial conditions for both the training regime $t \in [0, 0.01]$ (with the error computed with respect to the true trajectories, not the noisy observations) and the prediction regime $t \in [0.01, 0.03]$. The same results are shown for the more involved regularization selection strategy of [31] detailed in Algorithm 3 (see Appendix 2). This strategy chooses a scalar regularization value (i.e., $\lambda_{i,j} = \lambda$ for all i, j) by solving an optimization problem based on two criteria: *i*) that the resulting ROM remains stable throughout the full time horizon $t \in [0, 0.03]$ with each of the considered initial conditions, and *ii*) that the reconstruction error of the training data is minimized. Note that Algorithm 3 chooses the scalar regularization that minimizes the difference between the integrated ROM state and the *noisy* training data. Figure 5 shows that the errors for each selection method are highly similar, especially over the training regime, but that the BayesOpInf approach is slightly more accurate in the prediction regime for $r > 11$. This is likely because Algorithm 2 treats the regularization of each dynamic mode \hat{q}_i separately, while Algorithm 3 uses a single global regularizer for all terms of the learned operators. The key takeaway is that when the stability of the learned model is not highly sensitive to the regularization hyperparameters, Algorithm 2 provides a lightweight option for selecting the regularization. The next section details an example where the choice of regularization is critical for model stability, in which case Algorithm 2 is a less appropriate option for choosing the λ_i .

5.2. Single-injector Combustion Process

We now consider a combustion application on a two-dimensional domain in which model form error, and hence the choice of regularization, plays a significant role in the OpInf procedure. This problem presents a challenging application for model reduction and has been used as a test problem for both intrusive methods [23–25] and non-intrusive OpInf [26, 31, 48]. The governing dynamics are the nonlinear conservation laws

$$\frac{\partial \vec{q}_c}{\partial t} + \nabla \cdot (\vec{K} - \vec{K}_v) = \vec{S}, \quad (5.4)$$

where $\vec{q}_c = (\rho, \rho v_x, \rho v_y, \rho e, \rho Y_1, \rho Y_2, \rho Y_3, \rho Y_4)$ are the conservative variables, \vec{K} is the inviscid flux, \vec{K}_v is the viscous flux, and \vec{S} contains source terms. Here ρ is the density [$\frac{\text{kg}}{\text{m}^3}$], v_x and v_y are the x and y velocities [$\frac{\text{m}}{\text{s}}$], e is the total energy [$\frac{\text{J}}{\text{m}^3}$], and Y_1, Y_2, Y_3, Y_4 are the mass fractions for the chemical species $\text{CH}_4, \text{O}_2, \text{H}_2\text{O}$, and CO_2 , respectively, which follow the global one-step reaction $\text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O}$ [53]. The top and bottom walls of the combustor are assigned no-slip boundary conditions, the upstream boundary is constrained to have a constant mass flow at the inlets, and at the downstream end a non-reflecting boundary

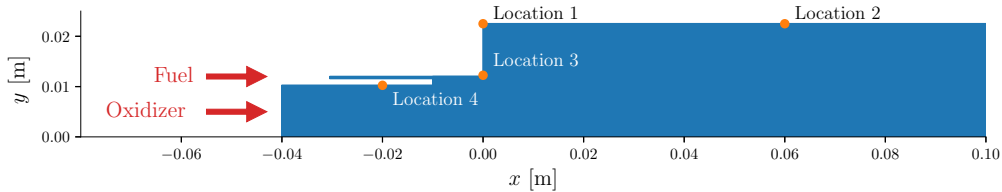


Figure 6: Two-dimensional domain of the single-injector combustor. The solution is monitored pointwise at sensor locations 1–4 and reported in Figures 8–11.

condition is imposed to maintain the chamber pressure via

$$p_{\text{back}}(t) = p_{\text{back,ref}} (1 + 0.1 \sin(2\pi ft)), \quad (5.5)$$

where $p_{\text{back,ref}} = 10^6$ Pa and $f = 5,000$ Hz. See [19] for more details on the governing equations.

As observed in [48], the governing conservative equations (5.4) take on a nearly quadratic form when written in a specific volume form with variables $\vec{q} = (p, v_x, v_y, T, \zeta, c_1, c_2, c_3, c_4)$, where p is the pressure [Pa], T is the temperature [K], $\zeta = 1/\rho$ is the specific volume [$\frac{\text{m}^3}{\text{kg}}$], and c_1, \dots, c_4 are the species molar concentrations [$\frac{\text{kmol}}{\text{m}^3}$] given by $c_\ell = \rho Y_\ell / M_\ell$ with M_ℓ the molar mass of the ℓ th species [$\frac{\text{g}}{\text{mol}}$]. In [31], it was shown that including pressure, specific volume, and temperature led to improved ROM performance, although only two of the three are needed to fully specify the high-fidelity model, with the equation of state defining the third. Hence, we construct a fully quadratic ROM of the form (2.2), with the understanding that this structure only approximates the structure of the governing equations.

In this experiment, we consider a single set of initial conditions and the corresponding trajectory. Our objective is to predict in time beyond the available training data. We solve (5.4) via the finite-volume based General Equation and Mesh Solver (GEMS), see [19] for details. The spatial domain is discretized with $n_x = 38,523$ cells (hence each snapshot has $8n_x = 308,184$ entries) and the solution is computed for 50,000 time steps with step size $\delta t = 10^{-7}$ s, from $t_0 = 0.015$ s to $t_f = 0.020$ s. We retain the first $k = 20,000$ snapshots for training and reserve the remaining 30,000 for testing. The training snapshots are lifted to the specific volume variables \vec{q} , then scaled to $[-1, 1]$ by first subtracting out the mean profile in the temperature, pressure, and specific volume, then normalizing each variable by its maximum absolute entry. These processed training snapshots $\mathbf{q}_0, \dots, \mathbf{q}_{k-1}$ have dimension $n = 9n_x = 346,000$. We compute the POD basis \mathbf{V} of the training snapshots using a randomized singular-value decomposition [18] and retain $r = 38$ modes based on the associated singular value decay. As the training data is smooth (i.e., not corrupted by observational noise), we estimate time derivatives of the projected training data $\hat{\mathbf{q}}_j = \mathbf{V}^T \mathbf{q}_j$ with fourth-order finite differences. This estimation introduces a further source of error, though the predominant error is due to the model misspecification by assuming a quadratic ROM form.

Previous applications of OpInf to this problem have shown that regularization is essential to obtain stable ROMs [31, 48]. We therefore utilize the regularization selection strategy of [31] (Algorithm 3) to choose the regularization parameters $\lambda_{i,j}$. In this context, we select two scalars: one to penalize the terms of the non-quadratic operators $\hat{\mathbf{c}}$, $\hat{\mathbf{A}}$, and $\hat{\mathbf{B}}$, and one for $\hat{\mathbf{H}}$ alone (see Appendix 2 for details). Thus the posterior mean μ_i of the operator $\hat{\mathbf{o}}_i$ given in (3.7) is equivalent to the ROM inference studied in [31]. We additionally draw 100 operator samples from (3.7) and integrate the resulting ROMs to quantify the uncertainties in the learned model. The computational cost of integrating 100 ROMs is trivial compared to the high-dimensional code: computing the 50,000 unprocessed snapshots with GEMS costs approximately 1,000 CPU hours, while constructing and integrating a single ROM from the posterior takes ~ 0.3 CPU seconds. Even when the mean operators μ_i of the posterior distribution (3.7) define a stable ROM, there is no guarantee that operators drawn from the distribution will result in stable ROMs. In this experiment, 96 of the 100 posterior draws yield stable ROMs. This is an indicator of the complexity of the problem at hand and highlights the constraint-free nature of the posterior formulation.

Figure 7 plots the dominant modes of the reduced-order state as a function of time together with the deviation defined by the posterior draws. Both the trajectory corresponding to the mean operators μ_i and

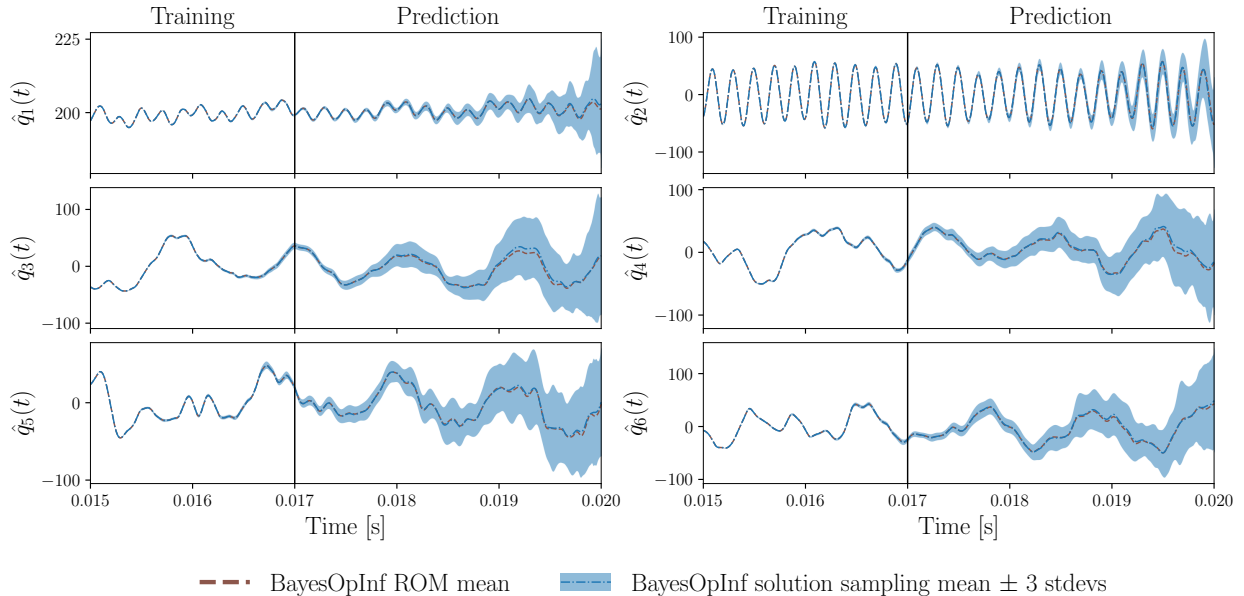


Figure 7: Evaluations of the first six modes (i.e., the first six entries of $\hat{\mathbf{q}}(t)$) of 100 draws from the posterior distribution of the Bayesian OpInf ROM. The dashed lines show the trajectory of the ROM with posterior mean operators ($\hat{\mathbf{o}}_i = \boldsymbol{\mu}_i$); the shaded areas are the regions within three standard deviations of the sample mean. Training data is observed for $t \in [0.015, 0.017)$.

the sample mean of the 96 stable draws are shown, but they differ very little. We emphasize that, while the entries of the model operators are assumed to follow a Gaussian distribution, there is no reason for the integrated trajectories of the corresponding quadratic model to be Gaussian in nature. Notably, the uncertainty bands remain tight until the end of the training regime, and widen increasingly with time. The band is thinnest in the second mode, which corresponds to the pressure signal imparted by (5.5).

Figures 8 and 9 show the pressure and x -velocity, respectively, of the reconstructed ROM trajectory at the sensor locations, together with the data from the high-dimensional solver. We examine these variables in particular because they exhibit low projection error in the prediction regime with respect to the underlying POD basis \mathbf{V} . The full-order trajectory computed by GEMS is also shown for comparison. As in Figure 7, the uncertainty bands remain quite tight in the training regime but start to noticeably increase in width shortly after the transition to the prediction regime. The bands are widest at the peaks and troughs of the pressure signal, reflecting that the model more accurately captures the frequency than the amplitude. While the sample mean is sometimes point-wise inaccurate, the true solution is generally well captured within three standard deviations of the sample mean. Figures 10 and 11 show this more precisely for pressure and x -velocity, respectively, by plotting the absolute error and three sample standard deviation (half the width of the uncertainty bands in Figures 8 and 9) as a function of time at the sensor locations. Sharp drops in the sample mean error indicate that the sample mean crosses the true values. The width of the uncertainty band generally exceeds the absolute sample mean error in the prediction regime, especially for $t > 0.018$ s. The sample mean error and the standard deviation are positively correlated in time, i.e., the uncertainty in the prediction generally increases with the error in the model.

6. Concluding remarks

The proposed BayesOpInf method non-intrusively propagates uncertainties from training data noise and model form error to learned low-dimensional operators of a ROM. Specifically, a linear Bayesian inference driven by projected state data and Gaussian priors leads to Gaussian posterior distributions for the reduced operators. Because constructing the posterior distributions for the reduced operators and integrating particular realizations of the ROM are relatively inexpensive operations, we can efficiently equip the resulting

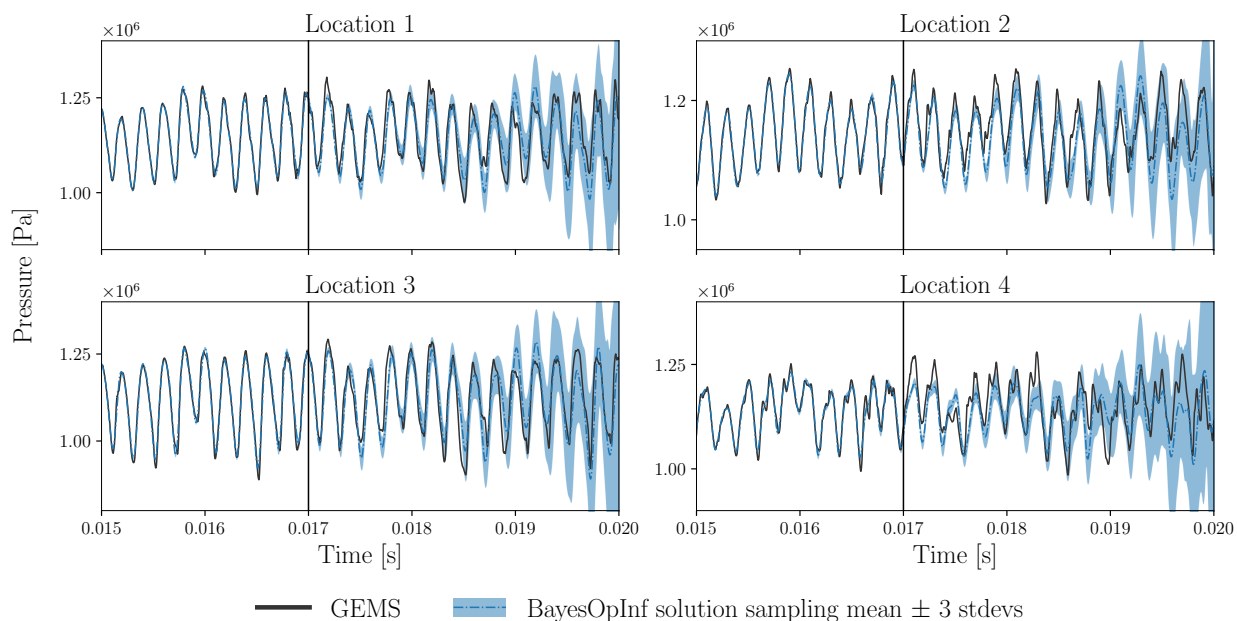


Figure 8: Full-order GEMS data and Bayesian OpInf ROM reconstructions of the pressure at the four sensor locations marked in Figure 6. The shaded areas are the regions within three standard deviations of the sample mean, using the same 100 draws as in Figure 7. Training data is observed for $t \in [0.015, 0.017)$.

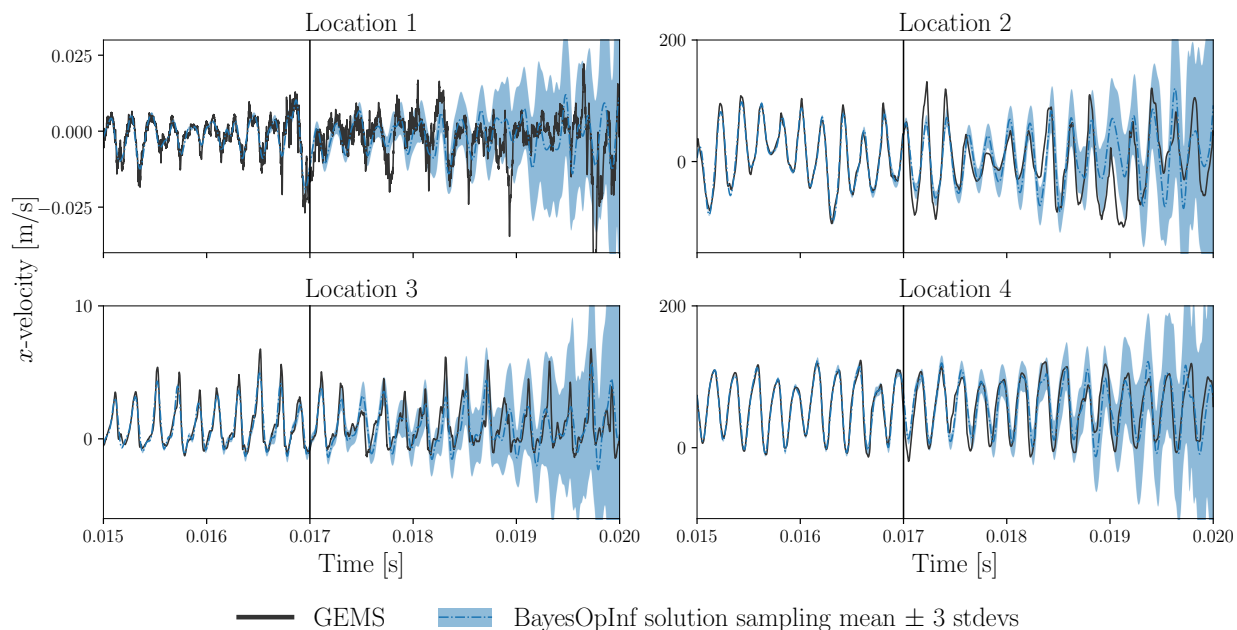


Figure 9: Full-order GEMS data and Bayesian OpInf ROM reconstructions of the horizontal velocity at the four sensor locations marked in Figure 6. The shaded areas are the regions within three standard deviations of the sample mean, using the same 100 draws as in Figure 7. Training data is observed for $t \in [0.015, 0.017)$.

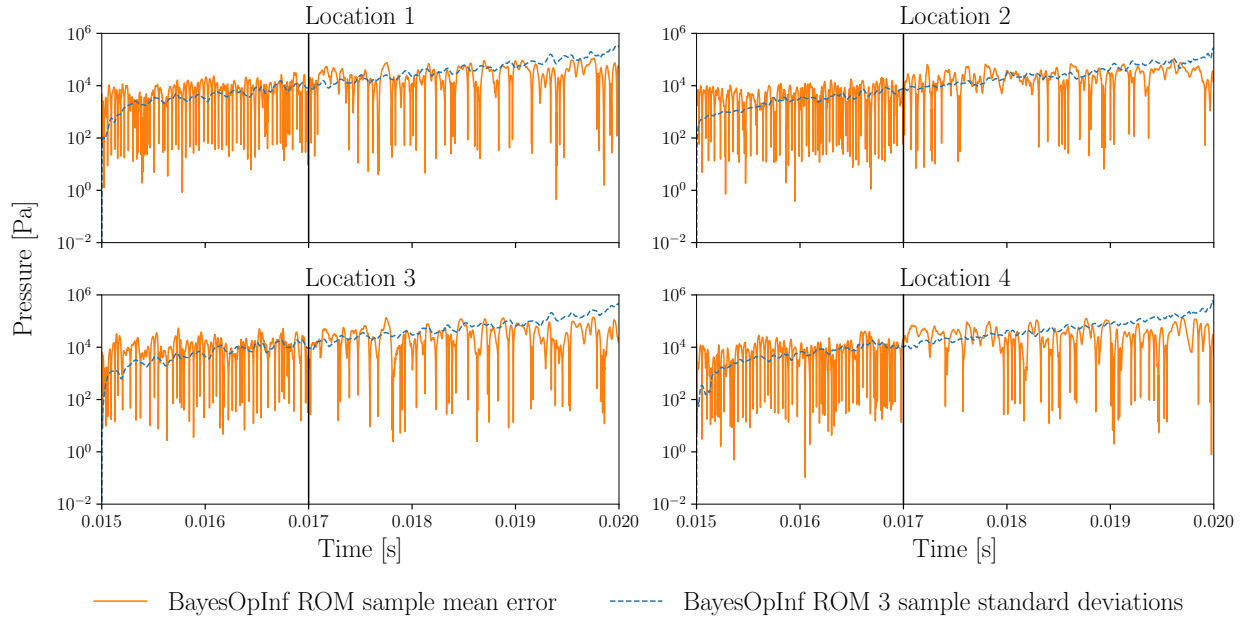


Figure 10: Absolute error of the sample mean and three sample standard deviations for the pressure at the locations in Figure 6, computed from 100 draws from the posterior distribution of the Bayesian OpInf ROM. Note the logarithmic scale of the y -axis.

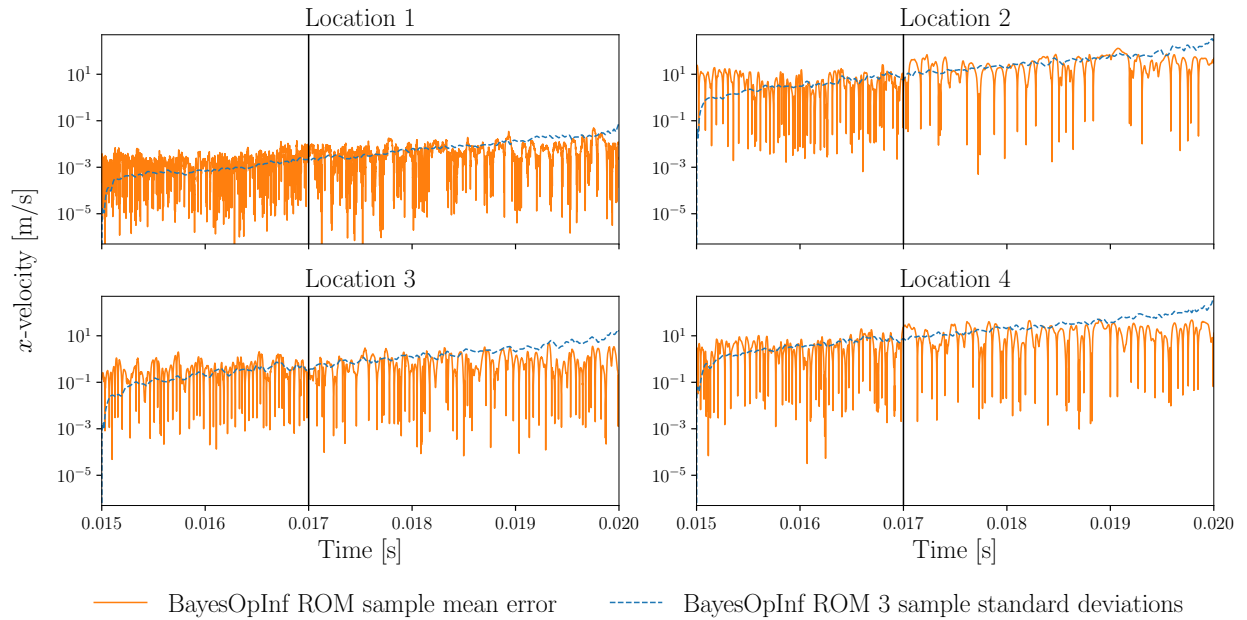


Figure 11: Absolute error of the sample mean and three sample standard deviations for the horizontal velocity at the locations in Figure 6, computed from 100 draws from the posterior distribution of the Bayesian OpInf ROM. Note the logarithmic scale of the y -axis.

ROM solutions with uncertainty bands through Monte Carlo sampling.

BayesOpInf is a natural generalization of the deterministic OpInf framework. Since the inference problem is generic and carried out only using the solution snapshots and their time-derivative data, BayesOpInf is a promising tool for more specialized applications of OpInf, including localized methods [14], problems with detailed structure [3], or parametric settings [32]. Furthermore, because the posterior mean of BayesOpInf is equivalent to Tikhonov-regularized OpInf, the approach is applicable for any Tikhonov-type regularization strategy. Hierarchical Bayes models may be a good candidate for formulating an implicit, non-Tikhonov regularization in the future. For now, we observe that the choice of the Tikhonov regularization remains an important issue for the stability of ROMs, but coupling the state-of-the-art error-based regularization selection of [31] with BayesOpInf leads to strong performance. Even with carefully chosen regularization, reduced models drawn from the posterior distribution may be unstable. Ensuring ROM stability—and, more generally, embedding desired model structure such as symmetry or spectral properties—through the posterior distribution remains an important topic for future work.

Our Bayesian approach has an inherent connection to Gaussian processes (GPs). A further discussion on this fact is provided in Appendix 1, wherein a GP surrogate model is used for the likelihood definition in the Bayesian inference. One open question is whether additional known information about the problem structure can be used to more strongly inform the Bayesian inference, for example by imposing a non-Gaussian distribution to model the reduced-order residual. Another potential extension of this work is utilizing the uncertainty indicators in the ROM solutions to define acquisition functions for active learning, for example adaptively selecting new training snapshots over time and parameter domains.

In summary, BayesOpInf quantifies the modeling uncertainties stemming from the data noise, the discretization errors of time derivatives, and the model misspecification error when the full-order model generating the snapshots does not match the assumed form of the ROM, as well as the uncertainties introduced by the use of Tikhonov regularization. As demonstrated in the numerical examples, BayesOpInf is effective in constructing accurate ROMs from noisy snapshot data and is capable of capturing complex physics while simultaneously providing meaningful credible intervals.

Acknowledgments

This work was supported in part by the ARPA-E DIFFERENTIATE Program under award DE-AR0001208; the U.S. Department of Energy, National Nuclear Security Administration under award DE-NA0003969; the U.S. Department of Energy AEOLUS MMICC center under award DE-SC0019303; and the Air Force Center of Excellence on Multi-Fidelity Modeling of Rocket Combustor Dynamics (Air Force Office of Scientific Research), award FA9550-17-1-0195. The first author also acknowledges the financial support from Sectorplan Bèta (the Netherlands) under the focus area *Mathematics of Computational Science*. The authors express their appreciation to Dr. Nicole Aretz for her valuable feedback.

Appendix 1: Alternative definition of the likelihood

In (3.1), an independent Gaussian noise was considered for the modeling of residual errors. If one would like to take into account the correlation over time and/or construct a surrogate model of the ROM closure, a Gaussian process can be used to replace the noise term, i.e.,

$$\frac{d}{dt}\hat{q}_i(t) = \mathbf{d}(\hat{\mathbf{q}}(t), \mathbf{u}(t))^\top \hat{\mathbf{o}}_i + \epsilon_i(t), \quad \epsilon_i(t) \sim \mathcal{GP}(0, \sigma_i^2 \kappa(t, t' | \ell_i)), \quad (\text{A.1})$$

in which $\kappa(t, t' | \ell_i)$ is a given kernel function with hyperparameters ℓ_i . In this case the likelihood is alternatively defined as

$$p(\mathbf{r}_i | \mathbf{D}, \hat{\mathbf{o}}_i, \sigma_i^2, \ell) = \mathcal{N}(\mathbf{r}_i | \mathbf{D} \hat{\mathbf{o}}_i, \sigma_i^2 \mathbf{K}_i(\ell_i)), \quad (\text{A.2})$$

where the covariance matrix $\mathbf{K}_i(\ell_i) \in \mathbb{R}^{k \times k}$ is defined to be $\mathbf{K}_i(\ell_i) = \kappa(\mathcal{T}, \mathcal{T} | \ell_i)$ and assumed to be positive definite. Here \mathcal{T} denotes the set of k time instances $\{t_0, t_1, \dots, t_{k-1}\}$ corresponding to the training data.

Conditioning on the training data $(\mathbf{D}, \mathbf{r}_i)$, the posterior distribution $p(\hat{\boldsymbol{\mu}}_i | \mathbf{D}, \mathbf{r}_i, \sigma_i^2, \boldsymbol{\lambda}_i, \boldsymbol{\ell}_i) = \mathcal{N}(\hat{\boldsymbol{\mu}}_i | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ thus has new formulations of mean vector and covariance matrix as follows:

$$\begin{aligned}\boldsymbol{\Sigma}_i &= \sigma_i^2 [\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{K}_i^{-1} \mathbf{D}]^{-1} \\ \boldsymbol{\mu}_i &= [\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{K}_i^{-1} \mathbf{D}]^{-1} (\text{diag}(\boldsymbol{\lambda}_i) \boldsymbol{\beta}_i + \mathbf{D}^\top \mathbf{K}_i^{-1} \mathbf{r}_i) \\ &= \boldsymbol{\beta}_i + \underbrace{[\text{diag}(\boldsymbol{\lambda}_i) + \mathbf{D}^\top \mathbf{K}_i^{-1} \mathbf{D}]^{-1} \mathbf{D}^\top \mathbf{K}_i^{-1} (\mathbf{r}_i - \mathbf{D} \boldsymbol{\beta}_i)}_{:= \delta \boldsymbol{\mu}_i}.\end{aligned}\tag{A.3}$$

Note that the correction term $\delta \boldsymbol{\mu}_i = \boldsymbol{\mu}_i - \boldsymbol{\beta}_i$ is then the solution to the following generalized least squares problem with a Tikhonov regularization

$$\delta \boldsymbol{\mu}_i = \arg \min_{\boldsymbol{\eta} \in \mathbb{R}^{d(r,m)}} \left\{ \|\mathbf{r}_i - \mathbf{D}(\boldsymbol{\beta}_i + \boldsymbol{\eta})\|_{\mathbf{K}_i^{-1}}^2 + \|\text{diag}(\boldsymbol{\lambda}_i)^{\frac{1}{2}} \boldsymbol{\eta}\|_2^2 \right\}.\tag{A.4}$$

Here the norm $\|\cdot\|_{\mathbf{K}_i^{-1}}$ is defined as $\|\mathbf{r}\|_{\mathbf{K}_i^{-1}} = \sqrt{\mathbf{r}^\top \mathbf{K}_i^{-1} \mathbf{r}}$, $\mathbf{r} \in \mathbb{R}^k$. In the meantime, the estimate of hyperparameters $(\sigma_i^2, \boldsymbol{\lambda}_i, \boldsymbol{\ell}_i)$ by maximizing marginal likelihood is reformulated as

$$\begin{aligned}(\sigma_i^{*2}, \boldsymbol{\lambda}_i^*, \boldsymbol{\ell}_i^*) &= \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i, \boldsymbol{\ell}_i} \log p(\mathbf{r}_i | \mathbf{D}, \sigma_i^2, \boldsymbol{\lambda}_i, \boldsymbol{\ell}_i) \\ &= \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i, \boldsymbol{\ell}_i} \log \int p(\mathbf{r}_i | \mathbf{D}, \hat{\boldsymbol{\mu}}_i, \sigma_i^2, \boldsymbol{\ell}_i) p(\hat{\boldsymbol{\mu}}_i | \sigma_i^2, \boldsymbol{\lambda}_i) d\hat{\boldsymbol{\mu}}_i \\ &= \arg \max_{\sigma_i^2, \boldsymbol{\lambda}_i, \boldsymbol{\ell}_i} \log \mathcal{N}(\mathbf{r}_i | \mathbf{D} \boldsymbol{\beta}_i, \sigma_i^2 (\mathbf{D} \text{diag}(\boldsymbol{\lambda}_i)^{-1} \mathbf{D}^\top + \mathbf{K}_i(\boldsymbol{\ell}_i))).\end{aligned}\tag{A.5}$$

The model (A.1) is equivalent to a universal kriging with the following Gaussian process prior:

$$\frac{d}{dt} \hat{q}_i(t) \sim \mathcal{GP} \left(\mathbf{d}(\hat{\mathbf{q}}(t), \mathbf{u}(t))^\top \boldsymbol{\beta}_i, \sigma_i^2 \mathbf{d}(\hat{\mathbf{q}}(t), \mathbf{u}(t))^\top \text{diag}(\boldsymbol{\lambda}_i)^{-1} \mathbf{d}(\hat{\mathbf{q}}(t'), \mathbf{u}(t')) + \sigma_i^2 \kappa(t, t' | \boldsymbol{\ell}_i) \right).\tag{A.6}$$

Accordingly, the posterior prediction of $\frac{d}{dt} \hat{q}_i(t)$ is again a Gaussian process whose mean and covariance functions take the form of

$$\begin{aligned}\mathbb{E} \left[\frac{d}{dt} \hat{q}_i(t) \middle| \mathbf{D}, \mathbf{r}_i \right] &= \mathbf{d}(\hat{\mathbf{q}}(t), \mathbf{u}(t))^\top \boldsymbol{\mu}_i + \kappa(t, \mathcal{T} | \boldsymbol{\ell}_i) \mathbf{K}_i^{-1} (\mathbf{r}_i - \mathbf{D} \boldsymbol{\mu}_i), \quad \text{and} \\ \text{Cov} \left[\frac{d}{dt} \hat{q}_i(t), \frac{d}{dt} \hat{q}_i(t') \middle| \mathbf{D}, \mathbf{r}_i \right] &= \sigma_i^2 \kappa(t, t' | \boldsymbol{\ell}_i) - \sigma_i^2 \kappa(t, \mathcal{T} | \boldsymbol{\ell}_i) \mathbf{K}_i^{-1} \kappa(\mathcal{T}, t' | \boldsymbol{\ell}_i) \\ &\quad + [\mathbf{d}(\hat{\mathbf{q}}(t), \mathbf{u}(t))^\top - \kappa(t, \mathcal{T} | \boldsymbol{\ell}_i) \mathbf{K}_i^{-1} \mathbf{D}] \boldsymbol{\Sigma}_i [\mathbf{d}(\hat{\mathbf{q}}(t'), \mathbf{u}(t')) - \mathbf{D}^\top \mathbf{K}_i^{-1} \kappa(\mathcal{T}, t' | \boldsymbol{\ell}_i)],\end{aligned}\tag{A.7}$$

respectively. Note that the second term in the mean function, i.e., $\kappa(t, \mathcal{T} | \boldsymbol{\ell}_i) \mathbf{K}_i^{-1} (\mathbf{r}_i - \mathbf{D} \boldsymbol{\mu}_i)$, formulates a surrogate of the closure in the linear model (2.2).

Appendix 2: Error-based regularization selection

Algorithm 3 details the OpInf procedure where the regularization parameters are selected with an optimization routine [31]. Regularization parameters candidates are assessed by solving the corresponding deterministic OpInf problem (2.4), then using the resultant ROM to reconstruct the training data for $L \geq 1$ initial conditions and make predictions in time beyond the training regime. The projected training snapshots are organized by initial condition, i.e., $\hat{\mathbf{Q}}^{(\ell)}$ contains all projected snapshots for the ℓ th initial condition.

In the results for Section 5.1 (Figure 5), Algorithm 3 is tailored to the purely quadratic ROM (5.3). In this setting, there are no external inputs \mathbf{U} , the data matrix of step 2 is given by $\mathbf{D} = (\hat{\mathbf{Q}} \odot \hat{\mathbf{Q}})^\top$, and the

Algorithm 3 Operator Inference with error-based regularization selection (Algorithm 1 of [31]).

```

1: procedure OPINF( Projected training snapshots  $\hat{\mathbf{Q}} = [ \hat{\mathbf{Q}}^{(1)} \dots \hat{\mathbf{Q}}^{(L)} ] \in \mathbb{R}^{r \times k}$ , training inputs
    $\mathbf{U} \in \mathbb{R}^{m \times k}$ , projected snapshot time derivatives  $\mathbf{R} = [ \mathbf{r}_1 \dots \mathbf{r}_r ]^\top \in \mathbb{R}^{r \times k}$ , final time  $t_f > t_0$ , bound
   margin  $\tau \geq 1$  )
2:    $\mathbf{D} \leftarrow \text{data\_matrix}(\hat{\mathbf{Q}}, \mathbf{U})$  # Construct data matrix as in (2.5).
3:    $B \leftarrow \tau \max_{i,j,\ell} |\hat{\mathbf{Q}}_{i,j}^{(\ell)}|$  # Set a bound for the ROM stability criterion.
4:   # Subprocedure for assessing regularization parameters.
5:   procedure OPINFERROR( regularizer  $\boldsymbol{\lambda} \in \mathbb{R}^{d(r,m)}$  )
6:     for  $i = 1, \dots, r$  do # OpInf with current regularization.
7:        $\hat{\mathbf{o}}_i \leftarrow \arg \min_{\boldsymbol{\mu}_i} \| \mathbf{D} \boldsymbol{\mu}_i - \mathbf{r}_i \|_2^2 + \| \text{diag}(\boldsymbol{\lambda})^{1/2} \boldsymbol{\mu}_i \|_2^2$ 
8:     for  $\ell = 1, \dots, L$  do
9:        $\tilde{\mathbf{Q}} \leftarrow \text{integrate (2.2) with } \hat{\mathbf{O}} = [ \hat{\mathbf{o}}_1 \dots \hat{\mathbf{o}}_r ]^\top \text{ from } \hat{\mathbf{q}}_0 = \hat{\mathbf{Q}}_{:,0} \text{ over } [t_0, t_f]$ 
10:      if  $\max_{i,j,\ell} |\tilde{\mathbf{Q}}_{i,j}^{(\ell)}| > B$  then # Ensure the ROM is stable on  $[t_0, t_f]$  for all ICs.
11:        return  $\infty$ 
12:      return  $\frac{1}{L} \sum_{\ell=1}^L \| \hat{\mathbf{Q}}^{(\ell)} - \tilde{\mathbf{Q}}_{:,k}^{(\ell)} \|$  # Compare ROM outputs to training data.
13:   # Minimize the subprocedure to choose the regularization parameters.
14:   return  $\arg \min_{\boldsymbol{\lambda}} \text{OPINFERROR}(\boldsymbol{\lambda})$ 

```

operator matrix of step 9 is $\hat{\mathbf{O}} = \hat{\mathbf{H}}$. The regularization parameters $\boldsymbol{\lambda}$ are globally parameterized by a single scalar hyperparameter $\lambda \geq 0$, i.e., $\boldsymbol{\lambda} = [\lambda \dots \lambda]^\top$. Hence, the minimization in step 14 can be carried out with a one-dimensional optimization on λ via, e.g., Brent's method [7].

For the combustion problem of Section 5.2, we have a single initial condition ($L = 1$) and the ROM is (2.2) with data matrix $\mathbf{D} = [\hat{\mathbf{Q}}^\top (\hat{\mathbf{Q}} \odot \hat{\mathbf{Q}})^\top \mathbf{U}^\top \mathbf{1}_k]$ and reduced-order operators $\hat{\mathbf{O}} = [\hat{\mathbf{A}} \hat{\mathbf{H}} \hat{\mathbf{B}} \hat{\mathbf{c}}]$. The regularization parameters are parameterized by two scalars $\lambda_1, \lambda_2 \geq 0$:

$$\boldsymbol{\lambda} = [\underbrace{\lambda_1, \dots, \lambda_1}_{r \text{ times}}, \underbrace{\lambda_2, \dots, \lambda_2}_{\frac{r(r+1)}{2} \text{ times}}, \lambda_1, \lambda_1]^\top. \quad (\text{B.1})$$

This choice of $\boldsymbol{\lambda}$ penalizes the entries of the quadratic operator $\hat{\mathbf{H}}$ by λ_2 and the entries of the remaining operators $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{c}}$ by λ_1 . More precisely, the regularization term as written in (2.3) is given by

$$\mathcal{P}([\hat{\mathbf{A}} \hat{\mathbf{H}} \hat{\mathbf{B}} \hat{\mathbf{c}}]) = \lambda_1 \left(\| \hat{\mathbf{A}} \|_F^2 + \| \hat{\mathbf{B}} \|_F^2 + \| \hat{\mathbf{c}} \|_2^2 \right) + \lambda_2 \| \hat{\mathbf{H}} \|_F^2. \quad (\text{B.2})$$

The optimization in step 14 of Algorithm 3 is carried out with a coarse grid search in the principal quadrant of \mathbb{R}^2 , followed by the Nelder-Mead method [33].

References

- [1] M. Ainsworth and J. T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*, volume 37. John Wiley & Sons, 2011.
- [2] I. Babuska and J. T. Oden. Verification and validation in computational engineering and science: basic concepts. *Computer Methods in Applied Mechanics and Engineering*, 36(193):4057–4066, 2004.
- [3] P. Benner, P. Goyal, J. Heiland, and I. P. Duff. Operator inference and physics-informed learning of low-dimensional models for incompressible flows. *Electronic Transactions on Numerical Analysis*, 56:28–51, 2022.
- [4] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.
- [5] G. Berkooz, P. Holmes, and J. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.

- [6] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, 2011.
- [7] R. P. Brent. *Algorithms for Minimization Without Derivatives*. Dover Publications, New York, NY, 2002.
- [8] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.
- [9] O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami. Bayesian solution uncertainty quantification for differential equations. *Bayesian Analysis*, 11(4):1239–1267, 2016.
- [10] K. De Brabanter, J. De Brabanter, B. De Moor, and I. Gijbels. Derivative estimation with local polynomial fitting. *Journal of Machine Learning Research*, 14(1), 2013.
- [11] D. del Castillo-Negrete, D. A. Spong, and S. P. Hirshman. Proper orthogonal decomposition methods for noise reduction in particle-based transport calculations. *Physics of Plasmas*, 15(9):092308, 2008.
- [12] B. P. Epps and A. H. Techet. An error threshold criterion for singular value decomposition modes extracted from PIV data. *Experiments in fluids*, 48(2):355–367, 2010.
- [13] K. Farrell, J. T. Oden, and D. Faghihi. A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems. *Journal of Computational Physics*, 295:189–208, 2015.
- [14] R. Geelen and K. Willcox. Localized non-intrusive reduced-order modeling in the operator inference framework. *Philosophical Transactions of the Royal Society A*, to appear, 2022.
- [15] O. Ghattas and K. Willcox. Learning physics-based models from data: perspectives from inverse problems and model reduction. *Acta Numerica*, 30:445–554, 2021.
- [16] W. R. Graham, J. Peraire, and K. Y. Tang. Optimal control of vortex shedding using low-order models. Part I—Open-loop model development. *International Journal for Numerical Methods in Engineering*, 44(7):945–972, 1999.
- [17] M. Guo and J. S. Hesthaven. Data-driven reduced order modeling for time-dependent problems. *Computer Methods in Applied Mechanics and Engineering*, 345:75–99, 2019.
- [18] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.
- [19] M. E. Harvazinski, C. Huang, V. Sankaran, T. W. Feldman, W. E. Anderson, C. L. Merkle, and D. G. Talley. Coupling between hydrodynamics, acoustics, and heat release in a self-excited unstable combustor. *Physics of Fluids*, 27(4):045102, 2015.
- [20] A. Hawkins-Daarud, S. Prudhomme, K. G. van der Zee, and J. T. Oden. Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth. *Journal of Mathematical Biology*, 67(6):1457–1485, 2013.
- [21] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified reduced basis methods for parametrized partial differential equations*, volume 590. Springer, 2016.
- [22] S. M. Hirsh, D. A. Barajas-Solano, and J. N. Kutz. Sparsifying priors for Bayesian uncertainty quantification in model discovery. *arXiv preprint arXiv:2107.02107*, 2021.
- [23] C. Huang, K. Duraisamy, and C. Merkle. Challenges in reduced order modeling of reacting flows. In *2018 Joint Propulsion Conference*, Cincinnati, OH, 2018. Paper AIAA-2018-4675.
- [24] C. Huang, K. Duraisamy, and C. L. Merkle. Investigations and improvement of robustness of reduced-order models of reacting flow. *AIAA Journal*, 57(12):5377–5389, 2019.
- [25] C. Huang, J. Xu, K. Duraisamy, and C. Merkle. Exploration of reduced-order models for rocket combustion applications. In *2018 AIAA Aerospace Sciences Meeting*, Orlando, FL, 2018. Paper AIAA-2018-1183.
- [26] P. Jain, S. McQuarrie, and B. Kramer. Performance comparison of data-driven reduced models for a single-injector combustion process. In *AIAA Propulsion and Energy 2021 Forum*, Virtual Event, 2021. Paper AIAA-2021-3633.
- [27] P. K. Jha, L. Cao, and J. T. Oden. Bayesian-based predictions of COVID-19 evolution in Texas using multispecies mixture-theoretic continuum models. *Computational Mechanics*, 66(5):1055–1068, 2020.
- [28] C. Khatri and C. R. Rao. Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 167–180, 1968.
- [29] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [30] K. Lee and K. T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, 2020.
- [31] S. A. McQuarrie, C. Huang, and K. E. Willcox. Data-driven reduced-order models via regularised operator inference for a single-injector combustion process. *Journal of the Royal Society of New Zealand*, 51(2):194–211, 2021.
- [32] S. A. McQuarrie, P. Khodabakhshi, and K. E. Willcox. Non-intrusive reduced-order models for parametric partial differential equations via data-driven operator inference. Oden Institute Report 21-17, University of Texas at Austin, 2021.
- [33] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
- [34] J. T. Oden. Adaptive multiscale predictive modelling. *Acta Numerica*, 27:353–450, 2018.
- [35] J. T. Oden, I. Babuška, and D. Faghihi. Predictive computational science: Computer predictions in the presence of uncertainty. *Encyclopedia of Computational Mechanics Second Edition*, pages 1–26, 2017.
- [36] J. T. Oden, I. Babuška, F. Nobile, Y. Feng, and R. Tempone. Theory and methodology for estimation and control of errors due to modeling, approximation, and uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 194(2-5):195–204, 2005.
- [37] J. T. Oden and S. Prudhomme. Estimation of modeling error in computational mechanics. *Journal of Computational Physics*, 182(2):496–515, 2002.
- [38] W. Pan, Y. Yuan, J. Gonçalves, and G.-B. Stan. A sparse Bayesian approach to the identification of nonlinear state-space systems. *IEEE Transactions on Automatic Control*, 61(1):182–187, 2015.
- [39] B. Peherstorfer. Sampling low-dimensional Markovian dynamics for pre-asymptotically recovering reduced models from data with operator inference. *SIAM Journal on Scientific Computing*, 42(5):A3489–A3515, 2020.

- [40] B. Peherstorfer and K. Willcox. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, 306:196–215, 2016.
- [41] E. Prudencio, P. Bauman, D. Faghihi, K. Ravi-Chandar, and J. T. Oden. A computational framework for dynamic data-driven material damage control, based on Bayesian inference and model selection. *International Journal for Numerical Methods in Engineering*, 102(3-4):379–403, 2015.
- [42] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox. Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401, 2020.
- [43] A. Quarteroni, G. Rozza, et al. *Reduced order methods for modeling and computational reduction*, volume 9. Springer, 2014.
- [44] H. Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2197):20160446, 2017.
- [45] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.
- [46] L. Sirovich. Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987.
- [47] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM, 2013.
- [48] R. Swischuk, B. Kramer, C. Huang, and K. Willcox. Learning physics-based reduced-order models for a single-injector combustion process. *AIAA Journal*, 58(6):2658–2672, 2020.
- [49] W. I. T. Uy and B. Peherstorfer. Probabilistic error estimation for non-intrusive reduced models learned from data of systems governed by linear parabolic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 55(3):735–761, 2021.
- [50] W. I. T. Uy, Y. Wang, Y. Wen, and B. Peherstorfer. Active operator inference for learning low-dimensional dynamical-system models from noisy data. *arXiv preprint arXiv:2107.09256*, 2021.
- [51] C. F. Van Loan. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*, 123(1-2):85–100, 2000.
- [52] D. Venturi. On proper orthogonal decomposition of randomly perturbed fields with applications to flow past a cylinder and natural convection over a horizontal plate. *Journal of Fluid Mechanics*, 559:215–254, 2006.
- [53] C. K. Westbrook and F. L. Dryer. Simplified reaction mechanisms for the oxidation of hydrocarbon fuels in flames. *Combustion Science and Technology*, 27(1-2):31–43, 1981.
- [54] C. K. Williams and C. E. Rasmussen. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [55] S. Zhang and G. Lin. Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2217):20180305, 2018.
- [56] Q. Zhuang, J. M. Lorenzi, H.-J. Bungartz, and D. Hartmann. Model order reduction based on Runge-Kutta neural networks. *Data-Centric Engineering*, 2, 2021.