

Closure Term Estimation in Spatiotemporal Models of Dynamical Systems

Eric Crislip^{1,2}, Mohammad Khalil³, Teresa Portone², Oksana Chkrebtii¹, and Kyle Neal²

¹Department of Statistics, The Ohio State University, Columbus, Ohio, 43210, USA

²Sandia National Laboratories, Albuquerque, New Mexico, 87185, USA

³Sandia National Laboratories, Livermore, California, 94551, USA

Abstract

Closure modeling — the statistical modeling of missing dynamics in the natural sciences and engineering — is a growing and active area of research. Existing methods for closure modeling are often computationally prohibitive, lack uncertainty quantification, or require noise-free observations of the temporal derivatives over the system state. We propose a novel, computationally efficient approach for the modeling and estimation of closure terms over the spatiotemporal domain that provides uncertainty quantification and is effective even when the observations of the system state are sparse or contain moderate levels of noise. The efficacy of our approach is demonstrated in both one and two spatial dimensions through numerical experiments using the Fisher-KPP reaction-diffusion equation and the advection-diffusion equation as exemplars.

1 Introduction

Physics-based models for engineering and the sciences, often taking the form of ordinary or partial differential equations (ODE, PDE), may inadequately represent real-world phenomena due to missing, unknown, or simplified system dynamics. The field of scientific machine learning (SciML) has recently emerged to ameliorate predictive deficiencies or inferential biases in theoretically-derived physics-based models by leveraging tools from machine learning and real-world observations. Popular SciML approaches for improving predictions from physics-based models have included directly modeling the discrepancy between physics-based model output and observed measurements [1], incorporating governing equations as soft constraints in data-driven machine learning models [2, 3], and more recently, directly learning missing terms from the governing equations themselves [4, 5]. The last approach, henceforth referred to as *closure modeling* [6], has several distinct advantages, including greater interpretability, the potential to improve predictions for unmeasured quantities of interest, and the ability to impose physical constraints in the machine learning-corrected physics-based model.

The use of data assimilation tools has attracted some recent attention for estimating the closure term from observed system responses [4, 7, 8, 9]. Data assimilation is the process of synthesizing scientific or engineering physics-based models with real-world data to create estimates of the unknown system states that are evolving over time. The central idea of sequential data assimilation for state estimation is to cyclically *forecast* future values of the state based on an understanding of system dynamics and previously obtained observations, then update

these forecasts by an *analysis* step when new observations become available. If desired, past predictions may be updated with current data to obtain smoother estimates. The Bayesian interpretation of data assimilation represents the forecast and analysis cycle probabilistically through the forecast and analysis probability distributions, which inherently allows for uncertainty quantification [10]. When new data arrive, the forecast distribution may be updated to the analysis distribution by applying Bayes' law, while the distributions for past states can be updated to smoothing distributions via backwards recursion [11].

Among the most widely studied methods for data assimilation is the Kalman filter, which provides exact closed-form expressions for the analysis and forecast distributions when both the dynamics and observation process are linear and Gaussian [12, 13]. The assumption of linear dynamics makes the naive Kalman filter unsuitable for nonlinear dynamical systems that are often encountered in practice. A popular extension of the Kalman filter that allows for nonlinear evolution models is the ensemble Kalman filter (EnKF), which represents forecast and analysis probability distributions using a set of Monte Carlo particles known as ensemble members [14]. In the EnKF, the ensemble members evolve forward in time to provide an approximation of the forecast distribution, which are then updated to an approximation of the analysis distribution using the closed-form expression of the Kalman filter. These closed-form expressions are derived from the Kalman filter's assumption of joint Gaussianity for the state and observations, which generally does not hold for nonlinear evolution models, yet the approximations induced in the forecast and analysis distributions by the closed-form analysis step of the EnKF often provide reasonable results owing the correspondence with Bayesian linear estimation theory [15, 16, 17]. Furthermore, the EnKF is fairly robust to deviations from the Gaussianity assumption for the data and extensions to the EnKF exist to accommodate strongly non-Gaussian observations. The EnKF is often combined with *state augmentation*, where the state vector is augmented with model parameters from the governing equations [18, 14]. State augmentation allows the model parameters to be inferred from the data through the filtering process as part of the analysis step. Parameter learning via state augmentation has been widely successful for many applications, though it has been shown to fail for certain variance parameters [19]. Our proposed method utilizes state augmentation to provide computationally efficient estimates of the closure term, with any auxiliary parameters estimated by maximum marginal likelihood (empirical Bayes under some conditions).

While data assimilation tools have previously been applied in closure modeling, the aim of this work is to extend previous work on estimating closure terms with data assimilation for systems described by ODEs to PDEs and other spatiotemporal models [4, 7, 8, 9]. To facilitate the use of data assimilation in capturing the unknown spatial closure term, we use basis expansion methods for function approximation [20]. As a means of dimensionality reduction, basis expansion of the spatial closure component increases computational efficiency and reduces the variability in the closure term estimates relative to closure estimation at each spatial point over a discrete spatial grid, as well as mitigate potential adverse effects from spurious correlations in the EnKF procedure. Furthermore, basis expansion defines a desired level of spatial smoothness in the closure term estimates. Basis expansion methods have a long and rich history in the fields of approximation theory and nonlinear regression. We will give particular attention to Gaussian kernel bases, which are known to be universal approximators for continuous functions [21, 22, 23], and B-splines, which are a particularly popular choice of basis for univariate regression [24, 25].

Current work for closure modeling in physics-based models governed by PDEs primarily focuses on the setting where observations of the temporal derivatives are assumed to be available [26] or attempts to estimate a parametric model for the functional relationship between the discretized state and closure term through an expensive calibration procedure [5]. Our method for closure term estimation can be combined with machine learning tools to model the functional relationship between the state and closure term as part of a two-staged approach [27] or on its own as a standalone means of model validation and diagnostics, sensitivity analysis, and diagnostics. The two-stage approach is as follows: in the first stage, the value of the system state and closure term are estimated by data assimilation using our methodology. In the second stage, the estimates of the system state and closure term are collected as input-output pairs and used as "data" to train a machine learning model for the functional form of the

closure. We conjecture that the two-staged approach to closure modeling will have several advantages. Namely, the decoupling of the closure term estimation and training of the machine learning closure model ensures that only a fixed number of runs of the expensive physics-based model will be utilized. Furthermore, data assimilation, when viewed from a Bayesian perspective, comes with built-in uncertainty quantification, which may be propagated through the machine learning mode. However, the focus of this paper is on the estimation of the closure term, and subsequent second-stage modeling will be the focus of future work.

This paper is organized as follows. Section 2 reviews the ensemble Kalman filter and describes our proposed methodology. Section 3 details numerical experiments on (1) the 1D Fisher-KPP equation, (2) the 1D advection-diffusion equation, (3) the 2D Fisher-KPP equation, and (4) the 2D advection-diffusion equation. Finally, the conclusions are presented in Section 4.

2 Methodology

Next, we describe the proposed methodology for closure term estimation in spatiotemporal models. Section 2.1 introduces data assimilation and some required notation. Section 2.2 provides a review of the stochastic ensemble Kalman filter for state estimation. Section 2.3 outlines existing work on closure term estimation using the EnKF for systems based on ordinary differential equations. In Section 2.4, we propose an extension of the EnKF-based closure term estimation methodology to PDE models. Section 2.5 provides a discussion on the choice of spatial basis functions used in our proposed methodology, while Section 2.6 details a maximum likelihood selection criterion for estimating the hyperparameters required to implement our method in practice.

2.1 Review of data assimilation for state estimation

We begin with a review of the data assimilation for state estimation and some notation. Suppose we collect observations of the state at times $0 \leq t_1 < t_2 < \dots < t_{N_o} \leq T$. Let us denote the spatially discretized state vector at time $t \in [0, T]$ by $\mathbf{u}_t \in \mathbb{R}^{N_x}$. The dimension N_x may refer to a finite number of state variables or a vectorized, discrete representation of one or more fields. If we are interested in the scalar field $u : \Omega \times [0, T] \rightarrow \mathbb{R}$ and let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_x}$ be a set of spatial grid points over the spatial domain $\Omega \subseteq \mathbb{R}^p$, then

$$\mathbf{u}_t = [u(\mathbf{x}_1, t), u(\mathbf{x}_2, t), \dots, u(\mathbf{x}_{N_x}, t)]^\top. \quad (1)$$

Now suppose the data at time t_i , $\mathbf{y}_{t_i} \in \mathbb{R}^{N_{t_i}}$, consists of N_{t_i} measurements and is obtained from the observation model $\mathbf{y}_{t_i} = \mathcal{H}_{t_i}(\mathbf{u}_{t_i}, \boldsymbol{\epsilon}_{t_i})$ with measurement noise $\boldsymbol{\epsilon}_{t_i}$. In the context of the ensemble Kalman filter (EnKF), the observation model is typically assumed to be linear with additive Gaussian noise,

$$\mathbf{y}_{t_i} = \mathbf{H}_{t_i} \mathbf{u}_{t_i} + \boldsymbol{\epsilon}_{t_i}, \quad \boldsymbol{\epsilon}_{t_i} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{t_i}). \quad (2)$$

Here $\mathbf{H}_{t_i} \in \mathbb{R}^{N_{t_i} \times N_x}$ is a linear observation matrix and $\{\boldsymbol{\epsilon}_{t_i}\}_{i=1}^{N_o}$ are temporally independent random variables. Modifications to the EnKF exist that can accommodate nonlinear observation operators $\mathcal{H}_{t_i}(\cdot, \cdot)$ [14], but we will assume the observation process is linear in this work. For simplicity, we will also assume a diagonal observation error covariance matrix $\mathbf{R}_{t_i} = \gamma \mathbf{I}_{N_{t_i}}$ with known noise parameter γ in all of our numerical experiments. This assumption corresponds to spatially independent observation errors, but the method is still applicable when the observations are spatially correlated.

The discrete-time physics-based model for the state u may be represented by the equations

$$\mathbf{u}_{t_{i+1}} = \mathcal{M}_{t_i}(\mathbf{u}_{t_i}), \quad i = 1, \dots, N_o, \quad (3)$$

in which the function $\mathcal{M}_{t_i} : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$ is often referred to as the *evolution model* that describes the time-evolution of the state and could be deterministic or stochastic. Implicitly, we assume future values of the state only depend on the state's current value, i.e. the process is first-order Markovian. For example, \mathcal{M}_{t_i} could represent the flow map for a discretized system of deterministic PDEs.

The Bayesian interpretation of data assimilation combines the evolution model and observation model to obtain probabilistic representations of unknown quantities of interest. Two primary objectives under this paradigm are to obtain the *forecast* distribution with density

$$p(\mathbf{u}_t | \mathbf{y}_{t_1}, \mathbf{y}_{t_2}, \dots, \mathbf{y}_{t_i}), \quad (4)$$

which represents degrees of belief about the state at a future time $t > t_i$ given past data, and the *analysis* distribution with density

$$p(\mathbf{u}_{t_i} | \mathbf{y}_{t_1}, \mathbf{y}_{t_2}, \dots, \mathbf{y}_{t_i}), \quad (5)$$

which represents belief about the system state at the time an observation is collected, given all the data available at the time of analysis. These distributions can be constructed from the evolution and observation models after eliciting a prior distribution $p(\mathbf{u}_0)$ on the state at the initial time, which should reflect *a priori* knowledge about the initial system state and may be chosen to obey known physical constraints. If the initial condition is known, then this prior would be a point mass distribution on the initial value. The forecast and analysis distributions follow from the prior distribution through the repeated application of Bayes' law.

When the evolution model is nonlinear, the forecast and analysis distributions are typically unavailable in closed-form [28], but may be estimated in some cases from a Monte Carlo sample. These methods generate individual particles, or *ensemble members*, that form an approximate sample from the required distributions. For many complex physical systems, characterization of the full forecast and analysis probability distributions is computationally infeasible, as typical sequential Monte Carlo methods such as particle filters may require sample sizes many orders of magnitude greater than the dimension of the associated state space [29]. An alternative to Monte Carlo methods that approximate the full forecast and analysis distributions is the ensemble Kalman filter, which merely represents the aforementioned probability distributions by their mean and covariance. The ensemble mean of the ensemble Kalman filter has enjoyed great success as an estimator of the system state in many applications with relatively small ensemble sizes [30, 31], thus it is our preferred technique for data assimilation.

2.2 The stochastic ensemble Kalman filter for state estimation

With the evolution and observation models and the prior distribution on the state at the initial time specified, we may present an algorithmic treatment of the ensemble Kalman filter. The EnKF begins by initializing ensemble members as Monte Carlo realizations from the prior on the state at the initial time

$$\mathbf{u}_{0,\ell} \stackrel{iid}{\sim} p(\mathbf{u}_0), \quad \ell = 1, 2, \dots, N_{ens}. \quad (6)$$

Each ensemble member is then pushed through the evolution model $\mathbf{u}_{t_1,\ell}^f = \mathcal{M}_0(\mathbf{u}_{0,\ell}^f)$ to obtain samples from the forecast distribution at time t_1 . To update the forecast ensemble at time t_i to the analysis ensemble at time t_i using the available data \mathbf{y}_{t_i} , the forecast ensemble mean and covariance must be first estimated by

$$\hat{\mathbf{m}}_{t_i}^f = \frac{1}{N_{ens}} \sum_{\ell=1}^{N_{ens}} \mathbf{u}_{t_i,\ell}^f, \quad (7)$$

$$\hat{\mathbf{P}}_{t_i}^f = \frac{1}{N_{ens} - 1} \sum_{\ell=1}^{N_{ens}} (\mathbf{u}_{t_i,\ell}^f - \hat{\mathbf{m}}_{t_i}^f)^T (\mathbf{u}_{t_i,\ell}^f - \hat{\mathbf{m}}_{t_i}^f). \quad (8)$$

Each ensemble member is then updated from a sample of the forecast distribution at time t_i to a sample from the (approximate) analysis distribution at time t_i according to

$$\mathbf{u}_{t_i,\ell}^a = \hat{\mathbf{K}}_{t_i}(\mathbf{y}_{t_i} - \mathbf{H}_{t_i}\mathbf{u}_{t_i,\ell}^f + \mathbf{e}_{t_i,\ell}), \quad \ell = 1, \dots, N_{ens}, \quad (9)$$

where $\mathbf{e}_{t_i,\ell} \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{R}_{t_i})$ are *measurement perturbations* and $\hat{\mathbf{K}}_{t_i}$ is the *Kalman gain* given by

$$\hat{\mathbf{K}}_{t_i} = \hat{\mathbf{P}}_{t_i}^f \mathbf{H}_{t_i}^T (\mathbf{H}_{t_i} \hat{\mathbf{P}}_{t_i}^f \mathbf{H}_{t_i}^T + \mathbf{R}_{t_i})^{-1}. \quad (10)$$

When the forecast distribution is Gaussian (e.g. the assumptions of the Kalman filter hold) and the forecast mean $\mathbf{m}_{t_i}^f$ and covariance $\mathbf{P}_{t_i}^f$ are known, the analysis step updates the forecast distribution ensemble exactly to samples from the analysis distribution

$$\mathcal{N}\left(\mathbf{m}_{t_i} + \mathbf{K}_{t_i}(\mathbf{y}_{t_i} - \mathbf{H}_{t_i}\mathbf{m}_{t_i}), (\mathbf{I}_{N_x} - \mathbf{K}_{t_i}\mathbf{H}_{t_i})\mathbf{P}_{t_i}\right), \quad (11)$$

where the optimal Kalman gain \mathbf{K}_{t_i} is found by using the true parameters in equation (10). Otherwise, the analysis step in the EnKF can be interpreted as falsely assuming the forecast distribution is Gaussian and approximating the true analysis distribution with (11), which may also induce approximations in subsequent forecast distributions. Alternatively, one can view the analysis step of the EnKF through the lens of Bayesian linear estimation theory [15], which does not require distributional assumptions for the forecast distribution beyond the mean and covariance. However, both interpretations of the EnKF are only ever approximate due to the Monte Carlo error in equations (7-8).

The addition of measurement perturbations as part of the analysis step is the distinguishing feature of the stochastic ensemble Kalman filter, as it ensures the ensemble variance is not underestimated [32, 33]. Deterministic variants of the ensemble Kalman filter exist and circumvent the need for measurement perturbations, but these methods may be more sensitive to deviations from Gaussianity in the forecast distributions than their stochastic counterparts [34].

After the ensemble is updated to the analysis distribution, we can then move forward in time through the evolution model and obtain the forecast ensemble

$$\mathbf{u}_{t_{i+1},\ell}^f = \mathcal{M}_{t_i}(\mathbf{u}_{t_i,\ell}^a), \quad \ell = 1, 2, \dots, N_{ens}. \quad (12)$$

If desired, the evolution model may be applied to the ensemble members repeatedly to obtain forecasts in-between observation times. The algorithm proceeds iteratively, alternating between forecast and analysis steps as new data arrives.

2.3 EnKF closure estimation for systems based on ODEs

We will now show how the ensemble Kalman filter can be modified to estimate the missing dynamics of the governing equation for simple models with a single state variable, as in [4, 9]. Suppose our governing equation is

$$\frac{d}{dt}u(t) = F(u(t)) + \phi(u(t)), \quad (13)$$

where $F : \mathbb{R} \rightarrow \mathbb{R}$ represents our incomplete understanding of the system and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is the unknown system dynamics. We assume this system admits a unique solution. The additive nature of the unknown dynamics is not a restrictive assumption, as for any F , we have that $\phi(u(t)) = \frac{d}{dt}u(t) - F(u(t))$. Additive model corrections are commonly found in the literature, see e.g. [4, 5, 6].

Let $\xi_t := (\phi \circ u)(t)$ be the true value of the closure term at a given time t . Since we do not have access to the missing dynamics $\phi(\cdot)$, we represent our uncertainty about ξ_t with the prior probability distribution

$$\xi_t \stackrel{\mathcal{D}}{=} \sigma W(t), \quad (14)$$

where $W(t)$ is a Brownian motion, σ is a scale parameter, and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. This probability model describes our belief about the time-evolution of the closure term and may be used in the context of data assimilation to form a forecast distribution for ξ_t . Brownian motion can be thought of as a continuous-time analogue to a symmetric random walk, hence this model may be interpreted as expressing ignorance about the true evolution of the closure term. Moreover, the increasing variance of Brownian motion ensures that even if the value of the closure term is known at a given point in time, uncertainty about the closure term will increase with the distance from that time point. However, standard Brownian motion starts at a deterministic initial value of $W(0) = 0$, which corresponds to complete certainty in the absence of missing dynamics at time $t = 0$, and therefore may be an unreasonable modeling choice. Instead, we may choose to work with a translated Brownian motion $\tilde{W}(t) = W(t) + \mu$, which will allow us to encode our belief about the initial missing dynamics into a prior distribution with density $p(\mu)$.

With a model for the time-evolution of the closure term established, we may proceed to estimate ξ_t using past and current measurements of the state through the ensemble Kalman filter and the method of state augmentation. That is, we form a new state vector $\mathbf{z}_{t_i} = \begin{bmatrix} u_{t_i} \\ \xi_{t_i} \end{bmatrix}$ to be used as the state in the EnKF. If our original observation model is

$$y_{t_i} = h_{t_i} u_{t_i} + \epsilon_{t_i}, \quad \epsilon_{t_i} \sim \mathcal{N}(0, \gamma^2), \quad (15)$$

then our augmented state vector can be associated with an augmented linear observation matrix $\mathbf{G}_{t_i} = [h_{t_i} \ 0]$ to obtain the equivalent observation model

$$y_{t_i} = \mathbf{G}_{t_i} \mathbf{z}_{t_i} + \epsilon_{t_i}, \quad \epsilon_{t_i} \sim \mathcal{N}(0, \gamma^2). \quad (16)$$

The discrete-time evolution model $\mathcal{M}_{t_i}^z(\mathbf{z}_{t_i})$ can be found by solving the system of stochastic differential equations,

$$\begin{aligned} du &= (F(u) + \xi_t) dt, \\ d\xi_t &= \sigma dW. \end{aligned} \quad (17)$$

Samples from the approximate solution to (17) can be obtained by using a stochastic solver, e.g. the Euler-Maruyama method.

Given the (augmented) observation model and evolution model, the EnKF proceeds exactly as outlined in Section 2.1 and returns approximate samples for the forecast and analysis distributions over ξ_t .

2.4 EnKF closure estimation for systems based on PDEs

The contribution of this work is to extend previous closure estimation methodology from scalar problems to the setting where the closure term is an entire spatial field. Suppose our governing equation is

$$\partial_t u(\mathbf{x}, t) = \mathcal{F}(u)(\mathbf{x}, t) + \phi(u)(\mathbf{x}, t) \quad (18)$$

where $\mathcal{F}(u)$ represents our incomplete understanding of the system and $\phi(u)$ is the missing system dynamics. At a given time $t \in [0, T]$, the closure term $\xi_t(\cdot) := (\phi \circ u)(\cdot, t)$ is no longer a single real number, but is now an entire scalar field over the spatial domain $\Omega \subseteq \mathbb{R}^p$. In many cases, a naive application of the previous approach

to closure term estimation is inadequate, as an independent treatment of the closure at each spatial point would produce everywhere discontinuous closure terms, that may suffer from high uncertainty in the analysis distribution. Furthermore, a discrete representation of the closure term that is allowed to grow with the dimension of the state space may quickly become computationally prohibitive in high-dimensional models.

To circumvent the potential pitfalls of the naive approach, we propose reducing the infinite-dimensional estimation problem of the spatial closure term to the estimation of a finite set of basis expansion coefficients. Analogously to the scalar estimation setting, we can represent our uncertainty about ξ_t with the prior probability distribution

$$\xi_t(\mathbf{x}) \stackrel{\mathcal{D}}{=} \sum_{k=0}^M \sigma_k \eta_k(t) b_k(\mathbf{x}), \quad (19)$$

where $\{\eta_k\}_{k=0}^M$ are independent Brownian motion processes, $\{\sigma_k\}_{k=0}^M$ are scale parameters, $\{b_k\}_{k=0}^M$ are a set of basis functions over the spatial domain, and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. Our probability model for ξ is separated into two distinct components: a temporal component determined by the coefficients η_k and a spatial component determined by the choice of bases b_k . This space-time decoupling allows for a parsimonious representation of the closure term, ensuring that only a select few coefficients be estimated at each analysis step. Similarly to the scalar setting, when viewed as a continuous analogue to the random walk, the Brownian motion distributions of the coefficients express uncertainty about the temporal evolution of the closure term. The choice of basis, however, encodes explicit knowledge regarding the spatial structure of the underlying closure field variable and should have sufficient flexibility to adequately model the true spatial closure term. A more in-depth discussion of the choice of basis functions can be found in Section 2.4.

The basis coefficients of interest $\{\eta_k\}_{k=0}^M$ can again be estimated as part of the EnKF through the method of state augmentation. To do so, we will augment the state vector at time t_i with the vector of coefficients $\boldsymbol{\eta}_{t_i}$ at time t_i to form the new augmented state $\mathbf{z}_{t_i} = \begin{bmatrix} \mathbf{u}_{t_i} \\ \boldsymbol{\eta}_{t_i} \end{bmatrix}$. If the true observation process is described by equation (2), then we can let $\mathbf{G}_{t_i} = [\mathbf{H}_{t_i} \quad \mathbf{0}_{N_{t_i} \times (M+1)}]$ be our augmented linear observation matrix and obtain the equivalent observation model

$$\mathbf{y}_{t_i} = \mathbf{G}_{t_i} \mathbf{z}_{t_i} + \boldsymbol{\epsilon}_{t_i}, \quad \boldsymbol{\epsilon}_{t_i} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{t_i}). \quad (20)$$

The discrete-time evolution model $\mathcal{M}_{t_i}^z(\mathbf{z}_{t_i})$ for the augmented state may be found by solving the system of stochastic differential equations

$$\begin{aligned} \partial_t u &= \mathcal{F}(u) + \sum_{k=0}^M \sigma_k \eta_k b_k, \\ d\eta_0 &= dW_0, \\ d\eta_1 &= dW_1, \\ &\vdots \\ d\eta_M &= dW_M. \end{aligned} \quad (21)$$

Samples from the approximate solution to (21) can be obtained by e.g. applying central differences and using a stochastic solver such as the Euler-Maruyama method.

Given the augmented observation model and time-discrete evolution model, the ensemble Kalman filter provides approximations for the forecast and analysis distributions of the coefficients $\{\eta_k(t)\}_{k=0}^M$, which induce forecast and analysis distributions for the closure term through equation (19).

2.5 Choice of spatial basis functions for spatial closure term estimation

The choice of basis functions should be determined by the practitioner's knowledge of the system and belief about the missing physics. For example, the user may wish to enforce periodicity with a Fourier basis or discontinuities with splines or step functions. For our numerical experiments, we have found that a suitable default choice of basis in one spatial dimension is the cubic B-Spline basis, which generates functions that are continuous in their first two spatial derivatives. This modeling choice produces estimates that are visually quite smooth. An example of a cubic B-spline basis is given in Figure 1a. The splines near the boundary have small support and therefore induce a large posterior variance when the observations are equally spaced and have identically distributed errors; however, they may be necessary to ensure the basis has sufficient flexibility to estimate the function in this region. This issue of increased bias and variance of nonparametric function estimators near the boundary of the domain is common in the nonparametric function estimation literature [35]. The bases presented in Figure 1a have equally spaced knots, but if there is prior knowledge on the behavior of the closure term, the knots can be concentrated in areas where the closure term is known to vary rapidly in space.

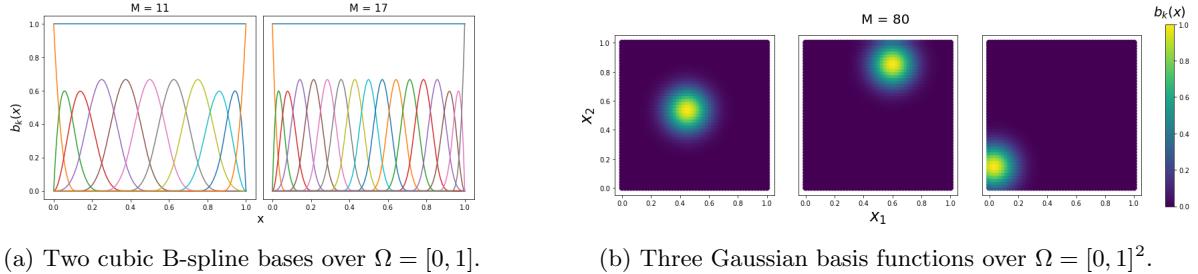


Figure 1: Examples of spatial basis functions.

In higher dimensions, we have found success with a radial basis function (RBF) basis using Gaussian kernels,

$$b_k(\mathbf{x}) = e^{-\frac{\|\mathbf{x} - \mathbf{c}_k\|_2^2}{h^2}}, \quad k = 1, \dots, M. \quad (22)$$

This basis typically performs well on standard multivariate function approximation problems and has the universal approximation property, though it may produce estimates that are visually slightly "bumpy" owing to the peaks of the individual Gaussian kernels. We follow [22] and select the centroids $\{\mathbf{c}_k\}_{k=0}^M$ by k-means clustering over the discretized spatial mesh. We have similarly found success by following [36] and setting the bandwidth parameter to

$$h = \frac{d_{max}}{N_x}, \quad (23)$$

where $d_{max} := \max_{i,j} \|\mathbf{c}_i - \mathbf{c}_j\|$ is the maximum pairwise distance between the centroids. Figure 1b displays three Gaussian kernels from a set of $M = 80$ basis functions. The spatial mesh for clustering was chosen to be a two-dimensional regular grid over a rectangular domain. The Gaussian kernel basis lacks a natural way of modeling the closure term on the boundary, but the size of the kernels offers a nice balance between imposing smoothness on the closure term estimate and modeling local behavior. A major limitation of the current methodology is the lack of regularization for the basis expansion coefficients. Thus, bases with high frequency components that are typically used in conjunction with hard-thresholding, such as wavelet or polynomial bases, may be unsuitable in the current framework. However, smooth bases with local or approximately local support, such as B-splines or Gaussian RBFs, appear to work quite well in our contexts.

2.6 Estimation of EnKF hyperparameters with maximum likelihood

Our method requires the specification of the number of basis functions M , as well as the scale parameters $\{\sigma_k\}_{k=0}^M$ that control the degree of variability in our forecast model for the closure term. Larger values of σ_k and M lead to more uncertainty in our analysis distribution for ξ and potentially unstable estimates, while smaller values of σ_k and M may cause our estimates of ξ to be rigid and unresponsive to the data. These hyperparameters can be difficult to specify a priori, as the time-evolution of the closure term depends on both the state and the unknown relationship between the state and closure. Furthermore, estimation of scale and correlation parameters through state augmentation can often fail owing to nonlinear relationships between the parameters and the data [37]. An alternative approach to estimation via state augmentation is the method of maximum marginal likelihood, also known as empirical Bayes or evidence maximization. However, for many nonlinear systems the marginal likelihood is intractable and must be approximated. A common approach in the EnKF literature is to approximate the marginal distribution of the data with a Gaussian distribution [19, 38], which can be derived mathematically when interpreting the EnKF as approximating the forecast distribution itself with a Gaussian distribution. Suppose we have auxiliary parameters $\boldsymbol{\theta}$, e.g. $\boldsymbol{\theta} = (M, \sigma_0, \sigma_1, \dots, \sigma_M)$, that we wish to estimate using maximum marginal likelihood. If we let $\hat{\mathbf{m}}_{t_i, \boldsymbol{\theta}}^f$ be the estimated forecast mean and $\hat{\mathbf{P}}_{t_i, \boldsymbol{\theta}}^f$ be the estimated forecast covariance for a given value of $\boldsymbol{\theta}$, we can write our optimization problem as

$$\boldsymbol{\theta}_{MML} = \operatorname{argmax}_{\boldsymbol{\theta}} \log p(\mathbf{y}_{t_1}, \mathbf{y}_{t_2}, \dots, \mathbf{y}_{t_{N_o}} | \boldsymbol{\theta}) \quad (24)$$

$$= \operatorname{argmax}_{\boldsymbol{\theta}} \sum_{i=1}^{N_o} \log p(\mathbf{y}_{t_i} | \mathbf{y}_{t_1}, \mathbf{y}_{t_2}, \dots, \mathbf{y}_{t_{i-1}}, \boldsymbol{\theta}) \quad (25)$$

$$\approx \operatorname{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^{N_o} [\det(\mathbf{H}_{t_i} \hat{\mathbf{P}}_{t_i, \boldsymbol{\theta}}^f \mathbf{H}_{t_i}^T) + (\mathbf{y}_{t_i} - \hat{\mathbf{m}}_{t_i, \boldsymbol{\theta}}^f)^T (\mathbf{H}_{t_i} \hat{\mathbf{P}}_{t_i, \boldsymbol{\theta}}^f \mathbf{H}_{t_i}^T)^{-1} (\mathbf{y}_{t_i} - \hat{\mathbf{m}}_{t_i, \boldsymbol{\theta}}^f)]. \quad (26)$$

The approximation in the last step is a result of both the Gaussian approximation for the marginal likelihood and Monte Carlo error in the estimated forecast means and covariances.

Maximum marginal likelihood turns the estimation of $\boldsymbol{\theta}$ into a stochastic optimization problem. Each evaluation of the objective function requires a complete run of the ensemble Kalman filter to estimate the forecast mean $\hat{\mathbf{m}}_{t_i, \boldsymbol{\theta}}^f$ and covariance $\hat{\mathbf{P}}_{t_i, \boldsymbol{\theta}}^f$. These mean and covariance estimates are random due to finite ensemble sizes and the probabilistic nature of the measurement perturbations, Brownian motion sample paths, and random initializations of the ensemble members. Hence, the objective function is also random and must be treated with care.

Though running the entire filter on a large-scale system may be computationally expensive, the expense of the objective function could be mitigated by the use of specialized optimization routines such as Bayesian Optimization, which can be modified to accommodate noise in the objective function [39]. In this work, we take a crude approach by using large ensemble sizes, fixing the stochastic elements of the optimization function, and performing grid search optimization as if the function were deterministic. Another potential pitfall of maximum marginal likelihood is that it does not necessarily guarantee the best estimate of the closure term, as the loss function is focused directly on the state-space predictions. However, our numerical experiments suggest promising results with maximum marginal likelihood as an estimation method for the auxiliary parameters. An investigation into the relationship between the marginal likelihood and the error of the EnKF estimates will be presented in Section 3. Future work could be dedicated to developing more computationally or statistically efficient methods for hyperparameter estimation.

3 Numerical Experiments

In this section, we illustrate the performance of our method using numerical experiments. In the examples presented, we learn the closure term from data simulated via the Fisher-KPP reaction-diffusion equation and advection-diffusion equation. For the reaction-diffusion equation, we examine the model under partial misspecification of the reaction term. The incomplete model for the advection-diffusion equation is examined under completely missing physics due to the omission of the advection term. We study both equations in both one and two spatial dimensions, as the simple, one-dimensional case is advantageous for visualization and the relative computational efficiency makes it an ideal setting for parametric studies, while the two-dimensional case showcases the method's performance under additional complexity. For the one-dimensional case, we perform a parametric study to determine the sensitivity of our method to the choice of a common scale parameter and the number of basis functions. We also investigate the influence of the prior distribution on our method, as well as the method's efficacy under two experimental settings: 1) low signal-to-noise ratio and spatially dense data, and 2) high signal-to-noise ratio and spatially sparse data. In the two-dimensional case, we simply report estimates after mild parameter tuning for brevity.

3.1 Fisher-KPP equation

Our first example is the problem of closure estimation for a system governed by the Fisher-KPP reaction-diffusion equation [40, 41], which has both a linear diffusion term and logistic reaction term

$$\partial_t u = D \nabla^2 u + u \left(1 - \frac{u}{K}\right). \quad (27)$$

This system has two model parameters: K , which controls the carrying capacity of the population density u , and the diffusivity D , which controls the speed of spatial spread for that population. The Fisher-KPP equation has enjoyed widespread use in the modeling of invasive populations and traveling waves and has diverse applications in cell biology, ecology, and combustion theory [42]. We chose to use this equation in our example as it is a simple, well-studied nonlinear PDE that is commonly applied to real-world problems.

Suppose we lack full knowledge of the reaction term. Using the closure modeling framework, we can write our governing equation as

$$\partial_t u = D \nabla^2 u + u(1-u) + \phi. \quad (28)$$

We will take the unknown closure term to be $\phi = (1 - \frac{1}{K})u^2$, which happens to correspond to a misspecified carrying capacity in equation 27. The nonlinear nature of this closure term makes for an interesting case-study, as the ensemble Kalman filter analysis updates are linear in the observations. We will estimate this closure term for the Fisher-KPP equation under two experimental settings for the i.i.d. Gaussian measurements, investigate the influence of the prior for the initial closure term on subsequent closure estimates, and perform a parametric study to determine the sensitivity of the closure term estimates to the choice of method hyperparameters.

In all of the one-dimensional numerical experiments, we solve the system over an equally-spaced discretization of the spatial domain $\Omega = [0, 1]$ with $N_x = 1000$ grid points. We use Neumann boundary conditions $\partial_x u(0, t) = \partial_x u(1, t) = 0$ and initial condition $u(x, 0) = \sin(4\pi x) + 1$. Figure 2 shows the solution of the system when $(D, K) = (0.01, 2/3)$ as it evolves over time. These model parameters were chosen so that the missing closure term would have a moderate but notable effect on the system dynamics. We can see that the solution is attracted towards the steady state value at $u = 2/3$, rather than the theorized $u = 1$ from the incomplete model (28). Figure 2 also shows observations of the state under two conditions: spatially dense with high noise $(N_{t_i}, \gamma) = (200, 0.01)$ and spatially sparse with low noise $(N_{t_i}, \gamma) = (15, 0.001)$. For the sparse observation case, the plot of linearly interpolated observations is overlaid with a scatterplot of the observations to aid in visualization.

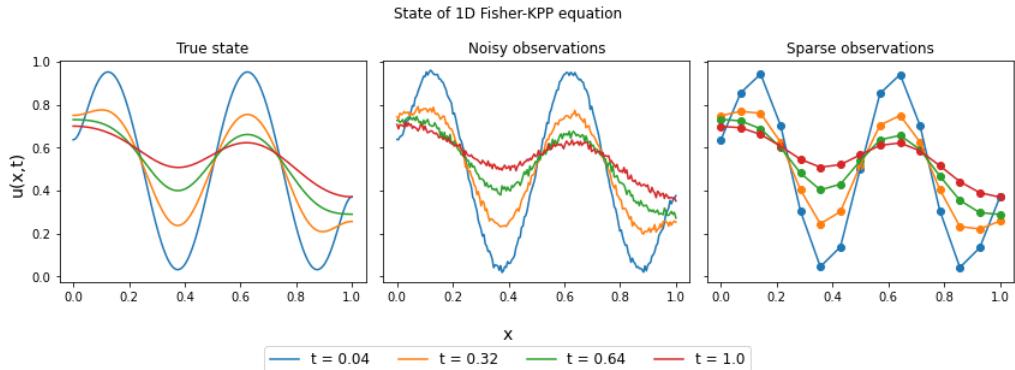


Figure 2: True value of the state (left) compared to spatially dense, high noise observations (middle) and spatially sparse, high signal-to-noise ratio observations (right). The state and observations are plotted at four distinct time points, which are denoted by different colors.

For estimating the unknown closure term, we assume equal Brownian motion scale parameters $\sigma_0 = \sigma_1 = \dots = \sigma_M =: \sigma$ for the basis expansion coefficients to limit the dimension of the hyperparameter optimization problem. For the one-dimensional examples, we also use the spline basis shown in Figure 1a as our choice of spatial basis functions. Estimates of the missing closure term with parameters chosen by grid-search optimization can be observed in Figure 3. The observations used for these estimates were taken at $N_o = 25$ time points $t_i = i/25$. For each time point shown in the figure, the ensemble Kalman filter analysis mean and approximate 95% credible intervals directly after an analysis update are plotted against the true closure term. The credible intervals are centered at the EnKF mean with half-widths of $z_{0.975}$ times the EnKF standard deviation, where $z_{0.975}$ is the 97.5-th percentile of the standard normal distribution. Notice that our method is able to track the evolution of the missing term over time with reasonable posterior coverage: for all but two of the scenarios shown, the true closure term is entirely contained by the pointwise credible intervals. The two remaining cases are the noisy setting at $t = 0.32$ and dense setting at $t = 0.04$, for which the pointwise credible intervals capture the true closure term at 99% and 91% of the spatial grid points respectively. While our EnKF analysis means do not exactly match the true closure terms, they are reasonable estimates given the data noise and sparsity, as well as the approximate nature of the data assimilation algorithm.

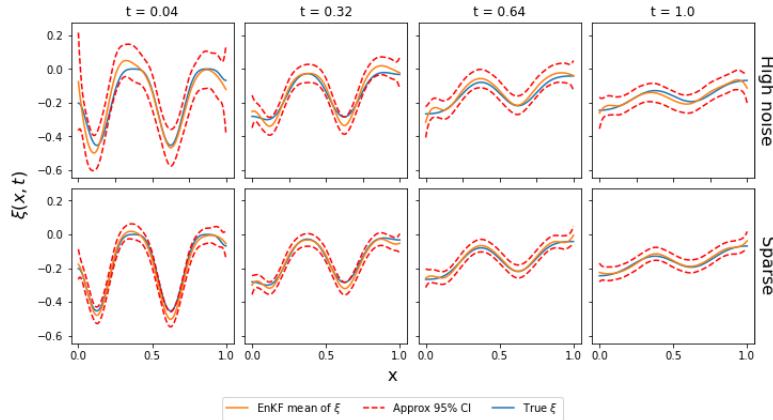


Figure 3: Estimated closure term using the high noise measurements (top) and sparse measurements (bottom). The EnKF hyperparameters used are $(\sigma, M, N_{ens}) = (0.1, 11, 1000)$.

To determine the sensitivity of our method to the choice of spatial basis and common scale parameter σ , as well as assess the relationship between the marginal likelihood and error of the closure term estimates, a parametric study is provided in Figure 4. The measurements were taken in the spatially dense, noisy setting with $N_o = 25$. In this figure, we plot the negative of the (approximate) log likelihood given in (26), as well as a discrete approximation to the normalized L^2 error of the EnKF mean $\hat{\xi}$, which is defined to be

$$\text{StandardizedL2Error} = \sqrt{\frac{\iint (\hat{\xi} - \xi)^2 dx dt}{\iint (\xi - \iint \xi dx dt)^2 dx dt}}, \quad (29)$$

which is a function of σ and the number of splines M . L^2 loss is a common choice for error assessment in machine learning models, which we elected to normalize due to the small scale of the closure term. We see that the minimizers of both the L^2 error and negative log-likelihood loss surfaces coincide over the coarse grid, suggesting that the negative log-likelihood is a reasonable loss function for closure term estimation.

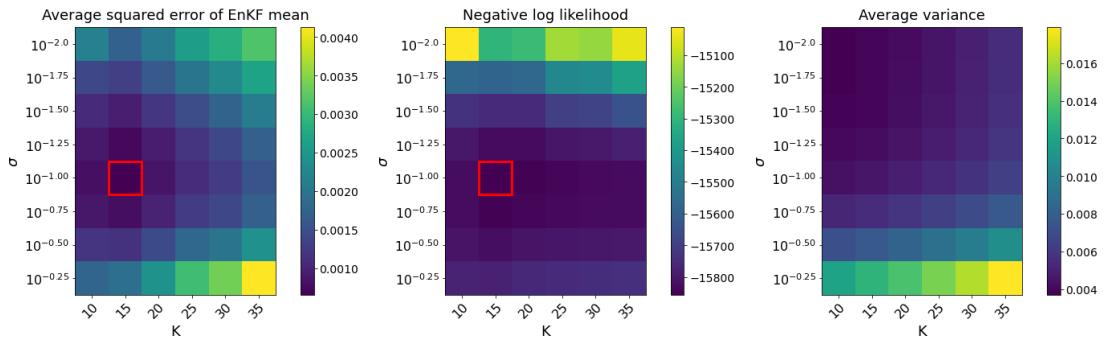


Figure 4: Comparison of the normalized L2-error of the EnKF mean as an estimator of the true closure (left) and the negative marginal log-likelihood of the data (right) as functions of the scale parameter σ and number of basis functions M .

Another matter of interest is the influence of the prior distribution for the spatial basis coefficients of the closure term at the initial time on the subsequent forecast and analysis estimates of the closure term. This is of particular import, as the functional form of the closure term may be completely unknown. In Figure 5, the estimates obtained from utilizing a prior that places low density on the closure term at the initial time are contrasted with estimates obtained from a prior that is centered around coefficients which minimize the squared distance from the true closure term. After assimilating three observations under the spatially dense, high noise experimental setting, the differences are nearly visually imperceptible. As this result suggests the estimates are not sensitive to the choice of prior, for all other numerical experiments we use prior distributions that are centered around the least-squares optimal closure term at the initial time.

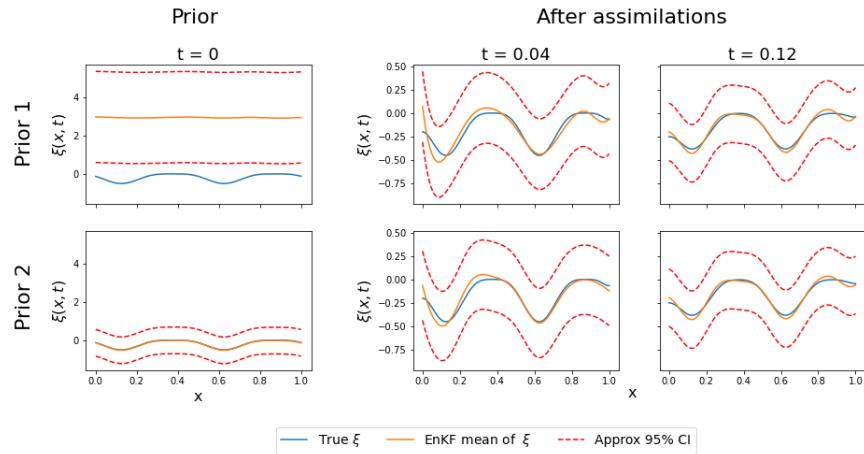


Figure 5: Comparison of ξ estimation without (top) and with (bottom) high density near the true value in the prior of $\xi(\mathbf{x}, 0)$. We show the initial prior distribution (left), as well as the EnKF estimates after one (middle) and three (right) analysis updates are completed. Note the y-axis is changed after observations are taken.

To demonstrate the performance of our method in higher spatial dimensions, we also present estimates of the closure term in the two-dimensional case. The true state at four selected time points with model parameters $(D, K) = (0.01, 2/3)$ is given in Figure 6. The system was solved via forward-time, central-space finite differences over a 50×50 discretization of the spatial domain $\Omega = [0, 1]^2$. The boundary condition used is $\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}, t) = 0$ for all $\mathbf{x} \in \partial\Omega$, where $\frac{\partial u}{\partial \mathbf{n}}$ is the normal derivative of u . The initial condition was generated from a scaled, exponentiated Gaussian process with squared-exponential covariance and is assumed to be known and fixed.

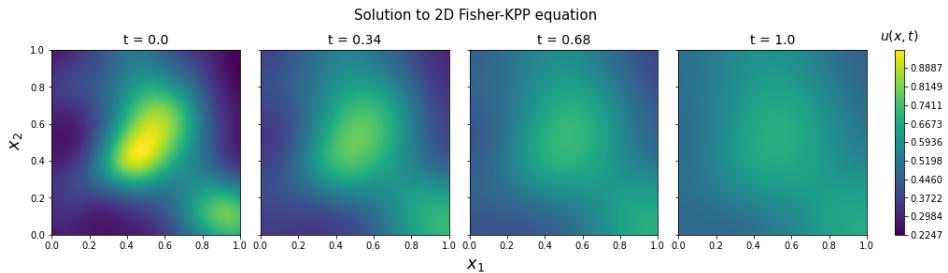


Figure 6: True solution of the 2D Fisher-KPP equation at four selected time points.

Figure 7 shows contour plots for the analysis mean estimates of the closure term at four specific time points. These estimates were obtained using $N_o = 50$ temporal observations at $t_i = i/50$ with experimental settings $(N_{t_i}, \gamma) = (2500, 0.0025)$. We see that the EnKF mean is capable of tracking the evolution of the closure term, as there are only small errors visible from the contours and the shape of the closure term is largely preserved. Error plots are given in Figure 8 that show the difference between the true closure term and the EnKF analysis mean estimate at four time points. Besides a slight underestimation of the closure at the lower right-hand corner, there does not seem to be any obvious systematic deviations in our estimates, and the method seems to perform quite well.

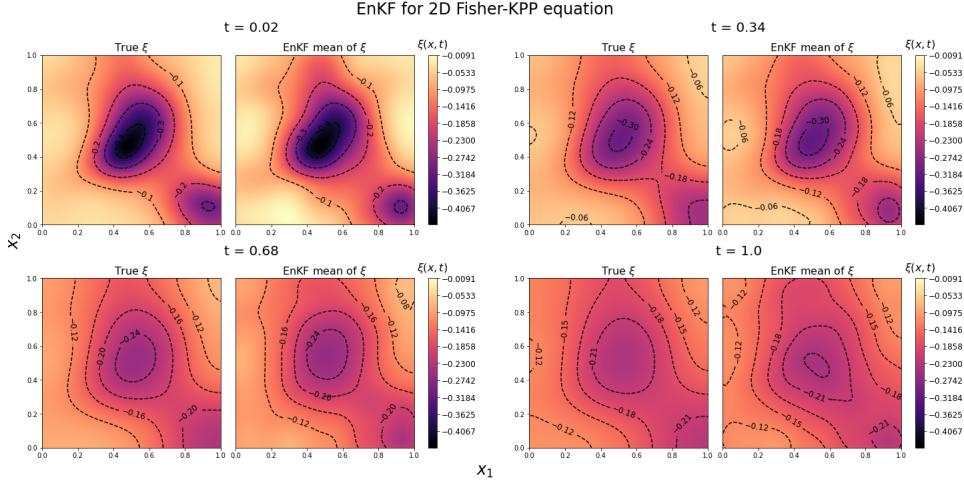


Figure 7: Contour plots for analysis mean estimates of the closure term for the 2D Fisher-KPP equation when $M = 80, N_{ens} = 1000$ and $\sigma = 0.025$.

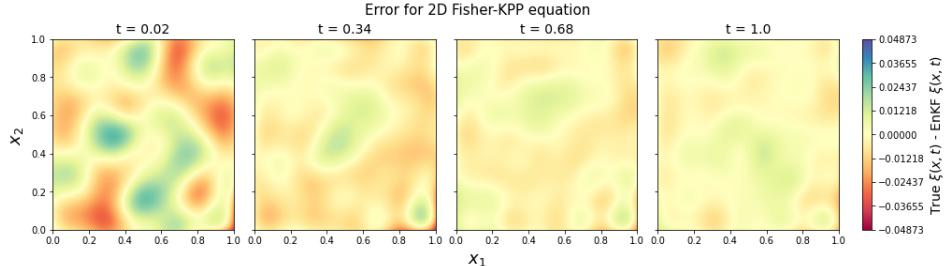


Figure 8: Residual error of the analysis mean from the ground truth closure term at four different time points for the 2D Fisher-KPP equation when $M = 80, N_{ens} = 1000$, and $\sigma = 0.025$.

3.2 Advection-diffusion equation

Our next physics-based model exemplar is the advection-diffusion equation

$$\partial_t u = D \nabla^2 u - \nabla \cdot (cu). \quad (30)$$

This system has two model parameters: the diffusivity D , which controls the rate of diffusion, and velocity c , which controls the rate of advection. The advection-diffusion equation has been used to model the transport of various quantities, including mass, energy, or heat, in diverse fields such as hydrology, chemical engineering, atmospheric science, and biology [43]. For example, the advection-diffusion equation has been used to model the dispersion of pollutants in rivers and the forced cooling of windings in turbo generators by passing fluids [43].

Suppose our incomplete knowledge of the system is

$$\partial_t u = D \nabla^2 u + \phi. \quad (31)$$

Then $\phi = -\nabla \cdot (cu)$ is our closure term, which we will again estimate using our proposed methodology. In this example, we have completely missing physics in our incomplete understanding of the system. We specify a Gaussian peak initial condition for this system in all experiments, which will enforce interesting spatial heterogeneity in the closure term. Specifically, some regions of space will have numerically zero closure term values, while other regions experience quite rapid temporal changes at early points in the trajectory. As such, one might expect the spatial heterogeneity of this closure term to pose a particular challenge when a shared scale parameter is used for all basis coefficients. As in the previous example, in the one-dimensional case we estimate the closure term using two experimental conditions: 1) high noise and spatially dense observations, and 2) low noise and spatially sparse observations. We furthermore perform a similar parametric study on the sensitivity of our estimates to the number of spatial basis functions M and shared scale parameter σ . Specific to the advection-diffusion exemplar, we also apply our method in the case of severe model specification by considering an advection-dominated problem, for which we report results in both one and two spatial dimensions.

In all one-dimensional numerical experiments, we solve this system over an equally-spaced discretization of the spatial grid $\Omega = [0, L]$ with $N_x = 1000$ spatial grid points. We use Dirichlet boundary condition $u(0, t) = u(L, t) = 0$ and initial condition $u(x, 0) = e^{-200(x-0.25)^2}$. The solution of the complete system with parameters $(c, D) = (0.1, 0.01)$ is shown in Figure 9, together with measurements under spatially dense with high noise $(N_{t_i}, \gamma) = (200, 0.01)$ and spatially sparse with low noise $(N_{t_i}, \gamma) = (20, 0.001)$ experimental settings. Slightly more spatial observations were used in the sparse setting for the advection-diffusion equation than the Fisher-KPP equation, as there are large differences in the spatial observations near the Gaussian peak.

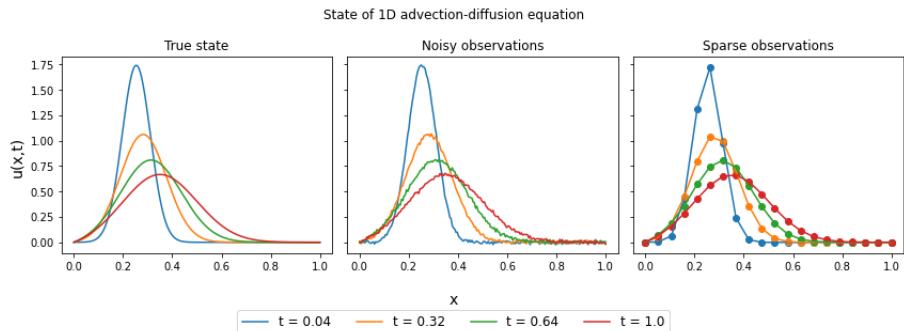


Figure 9: True value of the state (left) compared to dense, noisy observations (middle) and sparse, clean observations (right).

Analysis mean estimates of the closure term in the $c = 0.01$ case are given in figure 10. Like the Fisher-KPP equation, a spline basis and shared scale parameter is used for the prior model. The closure estimates were obtained with $N_o = 25$ temporal observations at $t_i = i/25$ and the hyperparameter estimates were obtained by grid search.

The pointwise 95% credible intervals contain the true closure at 91.5% of the spatial grid points at time $t = 0.04$ in the dense setting. For the sparse setting, the 95% credible intervals contain the closure at 87.5% of the spatial grid points at $t = 0.04$ and 99.1% of the spatial grid points at $t = 0.64$. For both settings, the estimated closure term appears to have high error near the peaks of the function at $t = 0.04$, which are in spatial regions that experience rapid changes at early time points in the trajectory. This suggests an underestimation of σ_k for the bases in this region (due to the assumption of a shared σ parameter) or possibly a need to have time-varying scale coefficients.

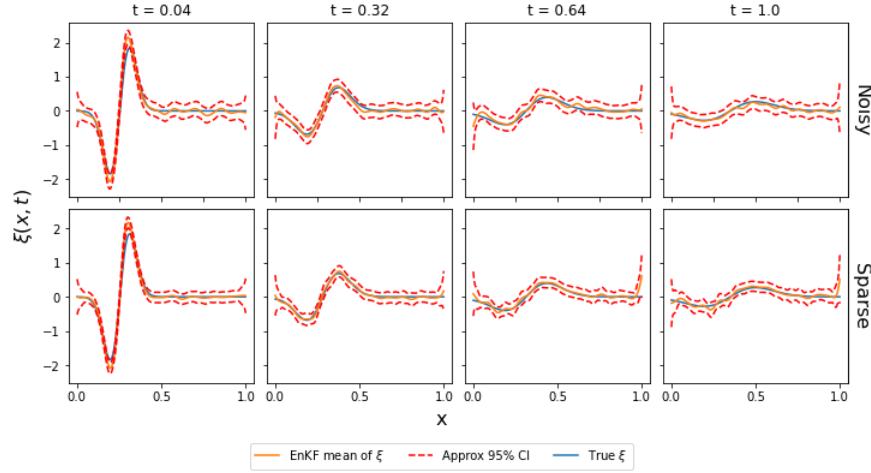


Figure 10: Estimated closure term in the $c = 0.1$ case using the high noise measurements (top) and sparse measurements (bottom). The EnKF hyperparameters used are $(\sigma, M, N_{ens}) = (0.1, 11, 1000)$.

Another parametric study was performed in Figure 11 that supports the findings from Figure 4. Once again, the observations were collected in the spatially dense, high noise setting with $N_o = 25$. From the figure, we see the optimal σ value in terms of log-likelihood is close to, though does not quite match, the optimal value in terms of normalized L^2 error, while the optimal number of bases M is the same in both cases. Once again, the log-likelihood appears to be a reasonable criterion for hyperparameter selection.

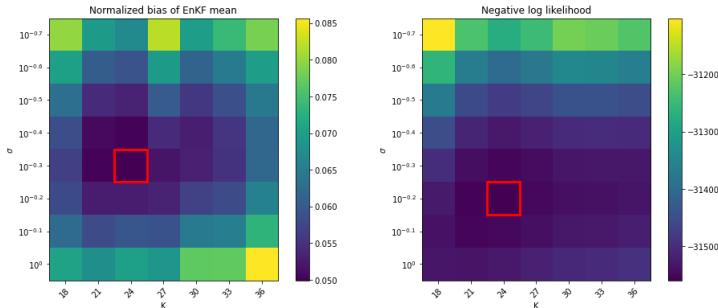


Figure 11: Comparison of the bias of the EnKF mean as an estimator of the true closure (left) and the marginal likelihood of the data (right) as functions of the scale parameter σ and number of basis functions M .

We also consider the advection-diffusion equation in two spatial dimensions. For model parameters $(c, D) =$

$(0.1, 0.01)$, we solve the system with solved via forward-time, central-space finite differences over a 50×50 discretization of the spatial domain $\Omega = [0, 1]^2$. The boundary condition used is $u(\mathbf{x}, t) = 0$ for all $\mathbf{x} \in \partial\Omega$, and the initial condition is a Gaussian peak $u(\mathbf{x}, 0) = e^{-100||\mathbf{x}-0.25\mathbf{1}_2||_2^2}$. The solution to the system at four time point is given in Figure 12.

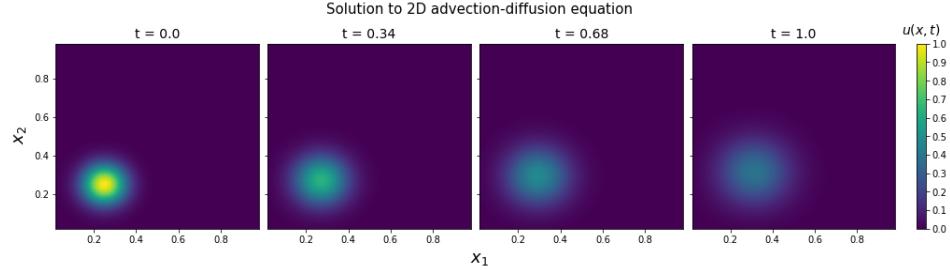


Figure 12: True solution of the 2D advection-diffusion equation at four selected time points.

The EnKF analysis mean estimates of the closure term are compared with the true closure term in Figure 13. The analysis estimates were computed using N_o observations were taken at time points $t_i = i/50$ with $(N_{t_i}, \gamma) = (2500, 0.0025)$. The analysis mean estimates appear to be noisier than the estimates for the Fisher-KPP equation, with visible small-scale deviations present throughout space at each analysis step presented. However, the closure term estimates appear to track the true closure term quite well at the areas of high magnitude. Error plots from Figure 14 tell a slightly different story. These plots demonstrate that the error of the analysis mean is highest near the Gaussian peak after the first analysis step. This suggests an underestimation of the scale parameters σ_k for the bases in that spatial region, which may be caused by the assumption of a shared σ value for all basis coefficients.

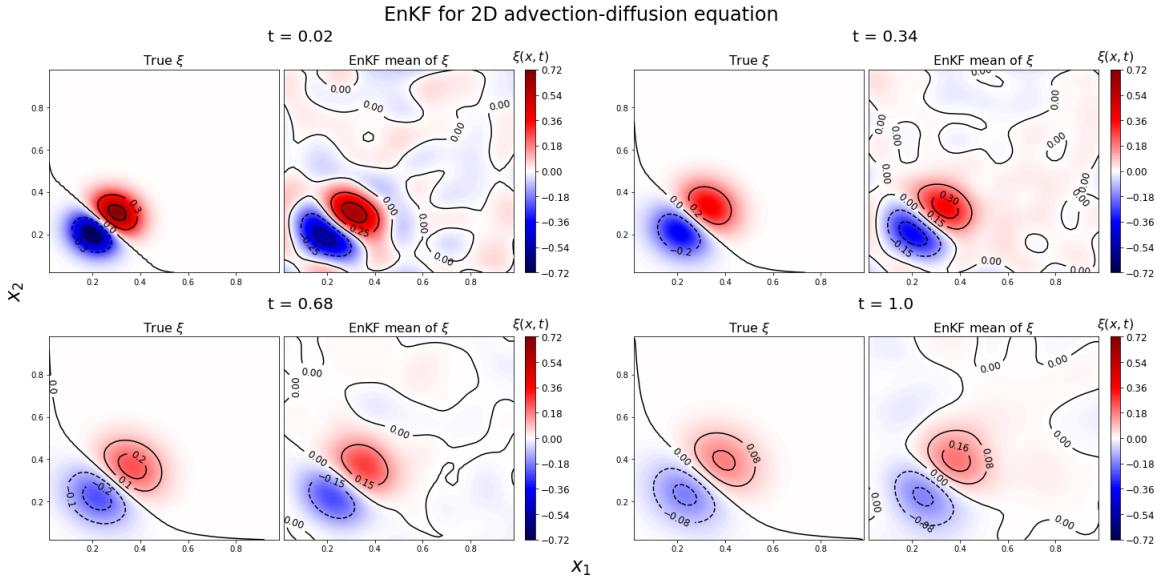


Figure 13: Contour plots for estimates of the closure term for the 2D advection-diffusion equation when $M = 81$, $N_{ens} = 1000$ and $\sigma = 0.1$.

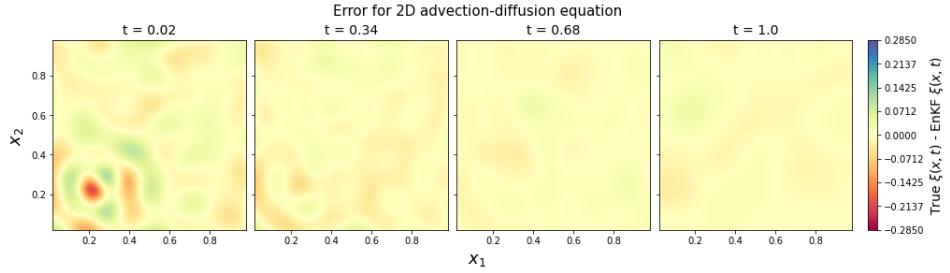


Figure 14: Residual error for the closure term of the 2D advection-diffusion equation at four different time points when $\gamma = 0.0025$, $M = 80$, $N_{obs} = 50$ and $\sigma = 0.1$.

A potential concern in the previous examples is that the problem is not advection-dominant and therefore may not be viewed as significantly different from the reaction-diffusion example. To demonstrate that our method remains functional even in the advection-dominant case, we increase the advection coefficient from $c = 0.1$ to $c = 0.5$, and since the increase in advection would cause the solution to approach the boundary of the previous domain, consider the new domain $\Omega = [0, 1.5]^d$ for $d = 1, 2$. We also double the temporal sampling rate and increase the spatial grid size and spatial sampling rate by a multiplicative factor of 1.5. The EnKF estimates are shown in Figure 15. The estimates appear reasonable, indicating that the success of the method is not limited to a narrow range of system dynamics.

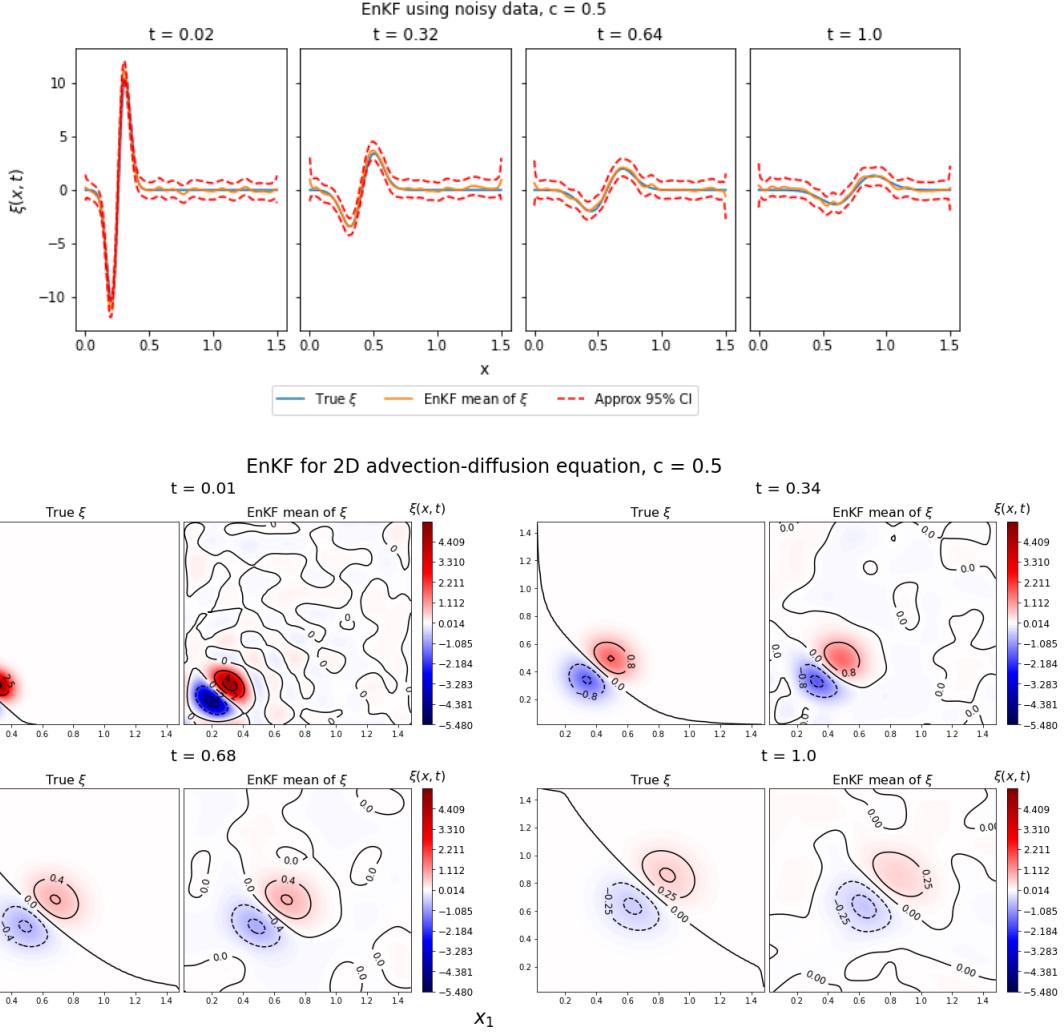


Figure 15: Estimates for the closure term of the advection-diffusion example when $c = 0.5$.

4 Conclusion

In this study, we have introduced a novel methodology for estimating closure terms in spatiotemporal models of dynamical systems, leveraging the ensemble Kalman filter (EnKF) framework. Our approach addresses the significant challenges associated with closure modeling, particularly in the context of partial differential equations (PDEs) where traditional methods often fall short due to computational demands or assumptions of noise-free observations.

Through extensive numerical experiments involving the Fisher-KPP reaction-diffusion equation and the advection-diffusion equation, we demonstrated the efficacy of our method in accurately estimating closure terms even under conditions of sparse and noisy data. The incorporation of basis expansion techniques not only enhanced computa-

tional efficiency but also provided a structured means to capture the spatial dynamics of the closure terms, leading to improved uncertainty quantification.

Our findings indicate that the proposed two-stage approach, which decouples closure term estimation from subsequent machine learning modeling, offers significant advantages in terms of interpretability and computational resource management. The results suggest that our method is robust across different experimental settings, including varying signal-to-noise ratios and spatial configurations.

Future work will focus on refining the hyperparameter estimation process and exploring the integration of advanced machine learning techniques to further enhance the predictive capabilities of our closure modeling framework. Additionally, we aim to extend our methodology to more complex dynamical systems and investigate its applicability in real-world scenarios, thereby contributing to the broader field of scientific machine learning and its applications in the natural sciences and engineering.

Acknowledgments

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia , LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <https://www.energy.gov/downloads/doe-public-access-plan>.

References

- [1] Marc C. Kennedy and Anthony O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- [2] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- [3] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, January 2021.
- [4] Abhinav Subramanian and Sankaran Mahadevan. Error estimation in coupled multi-physics models. *Journal of Computational Physics*, 395:19–37, 2019.
- [5] Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning, 2021.

- [6] Benjamin Sanderse, Panos Stinis, Romit Maulik, and Shady E. Ahmed. Scientific machine learning for closure models in multiscale problems: a review, 2024.
- [7] Shailesh Garg, Souvik Chakraborty, and Budhaditya Hazra. Physics-integrated hybrid framework for model form error identification in nonlinear dynamical systems. *Mechanical Systems and Signal Processing*, 173:109039, 2022.
- [8] Sahil Kashyap, Timothy J. Rogers, and Rajdip Nayek. A gaussian-process assisted model-form error estimation in multiple-degrees-of-freedom systems. *Mechanical Systems and Signal Processing*, 216:111474, 2024.
- [9] Kyle D Neal, Mohammad Khalil, and Teresa Portone. Investigating model form error estimation for sparse data. In *Annual Conference of the PHM Society*, volume 16, 2024.
- [10] Timothy John Sullivan. *Introduction to uncertainty quantification*, volume 63. Springer, 2015.
- [11] Simon J Godsill, Arnaud Doucet, and Mike West. Monte carlo smoothing for nonlinear time series. *Journal of the american statistical association*, 99(465):156–168, 2004.
- [12] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.
- [13] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer Berlin Heidelberg, 2006.
- [14] Geir Evensen. The ensemble kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29(3):83–104, 2009.
- [15] JA Hartigan. Linear bayesian methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(3):446–454, 1969.
- [16] Michael Goldstein and David Wooff. *Bayes linear statistics: Theory and methods*. John Wiley & Sons, 2007.
- [17] Mike West and Jeff Harrison. *Bayesian forecasting and dynamic models*. Springer Science & Business Media, 2006.
- [18] J. L. Anderson. An ensemble adjustment kalman filter for data assimilation. *Mon. Wea. Rev.*, 129:2884–2903, 2001.
- [19] J. R. Stroud, M. L. Stein, B. M. Lesht, D. J. Schwab, and D. Beletsky. An ensemble kalman filter and smoother for satellite data assimilation. *J. Amer. Stat. Assoc.*, 105:978–990, 2010.
- [20] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.
- [21] Martin Dietrich Buhmann. Radial basis functions. *Acta numerica*, 9:1–38, 2000.
- [22] K L Du and MNS Swamy. Radial basis function networks. *Neural networks in a softcomputing framework*, pages 251–294, 2006.
- [23] Jooyoung Park and Irwin W Sandberg. Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257, 1991.
- [24] Carl De Boor and Carl De Boor. *A practical guide to splines*, volume 27. springer New York, 1978.

- [25] James O Ramsay and Bernard W Silverman. *Applied functional data analysis: methods and case studies*. Springer, 2002.
- [26] Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.
- [27] Abhinav Subramanian and Sankaran Mahadevan. Probabilistic physics-informed machine learning for dynamic systems. *Reliability Engineering & System Safety*, 230:108899, 2023.
- [28] Christopher K. Wikle and L. Mark Berliner. A bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1):1–16, 2007. Data Assimilation.
- [29] Thomas Bengtsson, Peter Bickel, and Bo Li. *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, page 316–334. Institute of Mathematical Statistics, 2008.
- [30] Yuhua Zhou, Dennis McLaughlin, and Dara Entekhabi. Assessing the performance of the ensemble kalman filter for land surface data assimilation. *Monthly Weather Review*, 134(8):2128 – 2142, 2006.
- [31] Steven Gillijns, O Barrero Mendoza, Jaganath Chandrasekar, BLR De Moor, Dennis S Bernstein, and A Ridley. What is the ensemble kalman filter and how well does it work? In *2006 American control conference*, pages 6–pp. IEEE, 2006.
- [32] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.
- [33] Peter L Houtekamer and Herschel L Mitchell. Data assimilation using an ensemble kalman filter technique. *Monthly weather review*, 126(3):796–811, 1998.
- [34] Jing Lei, Peter Bickel, and Chris Snyder. Comparison of ensemble kalman filters under non-gaussianity. *Monthly Weather Review*, 138(4):1293–1306, 2010.
- [35] Cha Kyung-Joon and William R. Schucany. Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference*, 66(2):289–304, 1998.
- [36] David Lowe and D Broomhead. Multivariable functional interpolation and adaptive networks. *Complex systems*, 2(3):321–355, 1988.
- [37] Jonathan R Stroud and Thomas Bengtsson. Sequential state and variance estimation within the ensemble kalman filter. *Monthly weather review*, 135(9):3194–3208, 2007.
- [38] Herschel L. Mitchell and P. L. Houtekamer. An adaptive ensemble kalman filter. *Monthly Weather Review*, 128(2):416 – 433, 2000.
- [39] Peter I. Frazier. A tutorial on bayesian optimization, 2018.
- [40] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of eugenics*, 7(4):355–369, 1937.
- [41] A Kolmogoroff, I Petrovsky, and N Piscounoff. Study of the diffusion equation with growth of the quantity of matter and its application to a biology problem. In *Dynamics of curved fronts*, pages 105–130. Elsevier, 1988.
- [42] Maud El-Hachem, Scott W McCue, Wang Jin, Yihong Du, and Matthew J Simpson. Revisiting the fisher–kolmogorov–petrovsky–piskunov equation to interpret the spreading–extinction dichotomy. *Proceedings of the Royal Society A*, 475(2229):20190378, 2019.

- [43] Mehdi Dehghan. Numerical solution of the three-dimensional advection–diffusion equation. *Applied Mathematics and Computation*, 150(1):5–19, 2004.
- [44] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005.
- [45] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, March 2016.
- [46] Erin Acuesta. Adapting verification and validation principles to a credibility process for scientific machine learning. Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), 06 2021.
- [47] Megan R. Ebers, Katherine M. Steele, and J. Nathan Kutz. Discrepancy modeling framework: Learning missing physics, modeling systematic residuals, and disambiguating between deterministic and random effects, 2023.
- [48] Ziyou Wu, Steven L. Brunton, and Shai Revzen. Challenges in dynamic mode decomposition. *Journal of The Royal Society Interface*, 18(185):20210686, 2021.
- [49] Matthias Katzfuss, Jonathan R Stroud, and Christopher K Wikle. Understanding the ensemble kalman filter. *The American Statistician*, 70(4):350–357, 2016.
- [50] Jacob Stevens-Haas, Yash Bhangale, Aleksandr Aravkin, and Nathan Kutz. Learning nonlinear dynamics using kalman smoothing, 2024.
- [51] Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.
- [52] Joshua S. North, Christopher K. Wikle, and Erin M. Schliep. A bayesian approach for spatio-temporal data-driven dynamic equation discovery, 2022.
- [53] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.
- [54] Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.
- [55] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.
- [56] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [57] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [58] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.