A dynamic mode decomposition-based Kalman filter for Bayesian inverse problem of nonlinear dynamic systems

Yuming Ba[1], Qiuqi Li*, Zhiwen Zhang**

Abstract

Ensemble Kalman filter (EnKF) has been widely used in parameter estimation of the dynamic models. When the forward model is computationally intensive, such as nonlinear parameterized partial differential equations (PDEs), a direct application of EnKF forecast with full order model would be computationally prohibitive. For nonlinear parameterized PDEs, constructing an accurate reduced order model (ROM) has always been a challenging problem. As a ROM technique, dynamic mode decomposition (DMD) has gained significant popularity. However, it is an efficient data-driven method for ROM of time-dependent problems and has the limitation for the parameterized problems. In this paper, a new ROM is proposed based on DMD coupled with the weighted & interpolated nearest-neighbors algorithm (wiNN). The wiNN can approximate the solution for any given parameter value by choosing and computing the weighted average of the n nearest DMD solutions. This extends the applicability of DMD to parameterized problems. The weights are obtained by the distances between the given parameter and training parameters. Moreover, a low rank approximation of Kalman gain is used to EnKF and avoid explicitly computing the second-order derivative of the forward model. This can accelerate the EnKF update. We apply the proposed method to nonlinear parameterized PDEs for the two-dimensional fluid flow and investigate their Bayesian inverse problems. The results are presented to show the applicability and efficiency of the proposed EnKF with DMD-wiNN method by taking account of parameters in nonlinear diffusion functions, nonlinear reaction functions and source functions. For the predictive ability, the EnKF based on DMD-wiNN outperforms that based on FEM and gPC.

Keywords: stochastic nonlinear PDEs; dynamic mode decomposition; wiNN; ensemble filter methods; Bayesian inverse problems

*Corresponding author. School of Mathematics, Hunan University, Changsha 410082, China.

**Corresponding author. Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region of China.

Email addresses: yumingba@gpnu.edu.cn (Yuming Ba), qli28@hnu.edu.cn (Qiuqi Li), zhangzw@hku.hk (Zhiwen Zhang)

[1]School of Mathematics and Systems Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China.

1. Introduction

Many complex real-world models in science and engineering, such as optimization, control, or uncertainty quantification, are commonly simulated by parametric partial differential equations (PDEs). These parameters can originate from physical properties, geometric configurations, initial or boundary conditions, among others, and may be full of uncertainty due to the incomplete knowledge. The uncertainties propagate through the model and this can significantly affect the prediction of the model. Thus efficient and effective parameter estimation in complex dynamic systems is an important task. Measurement data is often necessary for the estimation problem. In general, the data is noisy and limited. They may be obtained by measuring a function of the system response. This leads to inverse problems. To circumvent the ill-posedness of inverse problems, the Bayesian regularisation method [38, 37] is often used and it combines the prior knowledge with given measurement data. In the paper, the parameter estimation problem is considered in time-dependent nonlinear diffusion-reaction models.

In general, the Bayesian inference is often applied to solve the inverse problems. The Bayesian methods incorporate uncertainties in incomplete measurement data and prior information, and then give the posterior distribution of unknowns. The most often used approach is a Markov chain Monte Carlo (MCMC) method [11]. It requires costly computation to achieve a good convergence and the samples may be not independent. To accelerate the MCMC method, the stochastic spectral methods have been proposed in [25, 46]. The unknowns are expanded based on the prior information by the polynomial choas [41] or Karhunen-Loève [26] expansion. Ensemble method is another Bayesian approach. In the ensemble method, the posterior mean and covariance matrix are approximated by ensemble mean and ensemble covariance, respectively. Ensemble methods, such as ensemble Kalman filter (EnKF) [9, 2] and ensemble smoother (ES) [40], are proposed not only for data assimilation, but also for the estimation of unknowns. The EnKF method based on polynomial chaos expansion (PCE) has been proposed in [33]. However, the number of stochastic basis functions grows with an exponential rate as the dimension of unknowns increases. To reduce the number of polynomial basis functions, the generalized polynomial chaos based stochastic collocation methods have been proposed in [21, 20, 47]. This method uses a few number of samples and reduces the expensive computation of solving the forward model. For deriving a linear Bayesian update for the estimation of parameter and state, a sampling-free method is proposed in [32, 29]. It adopts the Bayesian update of the PCE based on the prior information to a posterior distribution without any sampling. This is a direct algebraic way and only the coefficients obtained by PCE are estimated. The original KF [17] is shown to be the low-order part of the sampling-free methods. The above sampling-free methods are obtained by PCE. Due to the exponential increasing of PCE terms as the dimension of unknowns increases, a large number of coefficients need to be estimated in the sampling-free method. To elimilate the issue of high dimensionality, a variable-separation method for solving Bayesian inverse problems is proposed in [28, 3]. The stochastic basis functions are derived from the model and can represent the model in a compact form. However, the compact form for the nonlinear PDEs may be difficult to build.

Repeatedly calculating the forward problems for each sample is inevitable and time-consuming in ensemble methods. Traditional full-order techniques, such as finite element

or finite volume methods, are often too computationally expensive, especially when dealing with nonlinear, multiphysics, and time-dependent phenomena. To overcome these challenges, many model order reduction (MOR, [13, 31, 27]) have been proposed to reduce the computation complexity and improve computation efficiency. These methods construct an approximate model with lower dimensionality but still describes important aspects of the original model. The projection-based method [5, 15, 7, 8] is one of the model order reduction methods, arising from the ffeld of structure mechanics, fluid dynamics. The success of these methods relies on the assumption that the solution manifold can be embedded in a low-dimensional space. However, the important class of problems given by parametric dynamical systems usually induce rough solution manifold with slowly decaying Kolmogorov n-widths. This implies that traditional MOR methods are generally not effective. In this contribution, we try to combine dynamic mode decomposition with the weighted & interpolated nearest-neighbors algorithm (wiNN) to construct an efficient and reliable approximation of input-output relationship (i.e. surrogate model) for parametric dynamical systems defined on complex geometries.

Over the past decade, the dynamic mode decomposition (DMD) method has gained significant popularity as a reduced-order modeling technique. Originally introduced by Schmid in the field of fluid dynamics [36], the DMD method has quickly become a standard algorithm for approximating the Koopman operator [19] from data. Rowley et al. established the initial connection between the DMD method [39, 18, 42, 43, 30] and the Koopman operator. Koopman operator is an infinite dimensional linear operator acting on the observation function space. The spectral decomposition of the Koopman can capture linear systems in the observation function space. For numerical computation, it is necessary to approximate the Koopman operator in a finite dimensional subspace. DMD is used to approximate Koopman eigenvalues and eigenvectors in the subspace are spanned by a set of observation functions. DMD [36] describes the dynamical system in an equation-free manner and can be used for prediction and control. DMD is a spatio-temporal matrix decomposition method that connects spatial dimensionality-reduction technology (POD) and Fourier transforms in time. In the standard DMD method [36], identity functions are used as a finite dimensional set of observation functions for approximate the Koopman operator. The DMD mode calculated using standard DMD is a projection approximation. Tu et al. [39] proposed that the exact DMD method can obtain accurate DMD modes. Williams et al. [42] developed the extended dynamic mode decomposition (EDMD) method by applying the Koopman operator to a given dictionary of observation functions. When these dictionary functions are sufficiently rich, the matrix obtained by EDMD converges to the Koopman operator [19]. In addition, there are a number of DMD variant methods, such as sparsity promoting DMD [16], compressed DMD [4], Hankel-DMD [1] and so on. The DMD method has proven to be highly efficient in analyzing and extracting coherent structures from dynamic systems. For dynamical systems with random parameters, combining K-nearest-neighbors regression [12] or parametric interpolation [14, 22] with DMD, and including the stacked parametric dynamic mode decomposition method [35].

In this work, we aim to apply DMD to reduce computational costs of solving forward models while accurately capturing the most important features of the original parametric system. Huhn et al.[14] analyzed the stacked-snapshots approach and proposed two new methods based on the interpolation of the Koopman eigen-pair and the components of the

reduced Koopman operator matrix. It shows that the reduced Koopman operator interpolation method is more applicable to actual models. To this end, we propose a more general method to predict the reduced DMD matrix and SVD-modes for a given parameter value. To improve computation efficiency, we combine DMD with wiNN ([6, 44]), and develop a new model reduction method. The whole computation consists of an offline stage and an online stage. The online stage has high computational efficiency and its computational cost is completely independent of spatial discretization. In the offline stage, we firstly solve parameterized equations for each parameter in training set to generate the snapshots and get the data matrixes. Then singular value decomposition (SVD) is performed to get the reduced operator matrixes, and the wiNN is adopted to construct the efficient surrogate models of the reduced operator matrixes (including reduced Koopman operator matix, SVD-modes). In the online stage, for each parameter, we just perform operations on low-dimensional matrixes to obtain the efficient surrogate model. Besides, to handle problems with complex geometries or evolving boundaries without relying on a predefined mesh, some meshless methods are considered. Meshless methods, including radial basis function-based approaches, share the common feature of approximating or interpolating solutions using dispersed nodes or particles, without explicitly defining connectivity between them. Radial basis functions, in particular, have attracted attention due to their inherent meshless nature and their ability to achieve spectral accuracy when interpolating or approximating solutions at dispersed nodes. By combining the radial basis functions with the scattered node set, radial basis function-based methods provide a powerful tool for solving problems with complex geometries. The dispersed nodes can be placed directly on the boundary or within the domain, and the radial basis functions allow the solution to be accurately interpolated or approximated throughout the entire computational domain.

To the end, a low rank EnKF using DMD-wiNN model is proposed to take care of the challenges and concerns mentioned above. Inference of model parameters is one step in most uncertainty quantifications often ending in predictions that support decision in the form of design or control. Thus, we also focus on the predictive ability of DMD for nonlinear parametrized PDEs within the training time region and outside of the training time region. The wiNN method is used to extend the applications of DMD to parametrized PDEs. In DMD-wiNN, DMD is applied to predict the outputs of PDEs for the training parameter set and then wiNN is used to generate the approximate outputs of PDEs for each new sample. The main idea of wiNN is to find the n nearest neighbors of the new sample with some metric measure, and then using the weighted average of these n nearest neighbor to predict the reduced DMD matrix. To further reduce cost, we apply the low-rank approximation to generate the second-order derivative of the forward model in the inversion process. This is not only improve the efficiency, but avoid the rank-deficient when ensemble covariance is singular. One of the advantage is EnKF is performed by sequential measurements. The surrogate model can be rebuilded according to the practice problems, where the dynamic model is sensitivity for the time. This increases flexibility in the Bayesian inference. To extend the proposed EnKF to non-Gaussian models, we can integrate this method with normal-score transformation [2] to broaden the applicability.

The rest of this paper is arranged as follows. Section 2 describes the problem tackled in this work. In section 3, the dynamic mode decomposition method is introduced. In section 4, we introduce the low rank EnKF using DMD-wiNN model. In section 5, a few

numerical experiments are presented to illustrate the performance of the proposed DMD-wiNN based EnKF with applications to the nonlinear dynamic systems. Finally, we make some conclusions and comments in Section 6.

2. Problem setting

Let $\mathbb{N}$ be a generic forward operator on a Hilbert space $\mathcal{U}$ for a dynamic system. The forward operator describes the relation of model coefficient k, state u and source term f, i.e.,

$$\mathbb{N}(u;k) = f, \tag{2.1}$$

where $u \in \mathcal{U}$, $f \in \mathcal{U}^*$, the dual space of $\mathcal{U}$. In practical modeling, the coefficient function k and source term function f may be unknown. Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space, where Ω is the sample space, $\mathcal{B}$ is the σ- algebra on Ω, $\mathbb{P}$ is the probability measure on $\mathcal{B}$. We assume that $\theta : \Omega \to \mathbb{R}^{n_\theta}$, i.e., $\theta(\omega) \in \mathbb{R}^{n_\theta}$. We define

$$L^2(\mathbb{R}^{n_\theta}) := L^2(\Omega, \mathcal{B}, \mathbb{P}; \mathbb{R}^{n_\theta}) = \Big\{\theta(\omega) : \int_\Omega |\theta(\omega)|^2 \mathbb{P}(d\omega) < \infty\Big\}.$$

2.1. Bayesian inference using ensemble-based filter

For Bayesian inference, the unknowns k and f are assumed to be parameterized as $k \approx k(\theta_k)$ and $f \approx f(\theta_f)$, respectively. The estimated parameter θ will be $(\theta_k, \theta_f)^T$ when f and k are both unknown, where $[\cdot]^T$ is the transpose operator. Let $\mathbb{H}$ be the observation operator mapping the state $u(\theta)$ to the observation space $\mathcal{Y}$, i.e.,

$$d = \mathbb{H}\big(u(\theta)\big) + \varepsilon := \mathbb{H}(\theta) + \varepsilon, \tag{2.2}$$

where ε is assumed to be additive Gaussian noise for measurement and independent of θ. In practical setting, the measurement data is in a finite-dimensional space, i.e., $d \in \mathbb{R}^{n_d}$, where n_d is the dimension of measurements. Let ε be an independent and identically distributed (i.i.d.) Gaussian random vector with zero mean and variance σ^2, i.e.,

$$\varepsilon \sim N(0, \sigma^2 \boldsymbol{I}),$$

where $\boldsymbol{I}$ is the $n_d \times n_d$ identity matrix.

The goal of parameter estimation is to find an appropriate solution to minimize the misfit between the measurements and outputs of the forward model. In general, the minimization problem is ill-posed. For ensuring the well-posedness, some regularization term on the misfit are necessary. In this paper, we use Bayesian inference to identify the unknown parameters by some given noisy measurements of (2.2). For Bayesian context, both θ and d are random variables. Thus Bayes rule gives the posterior probability density for θ by

$$p(\theta|d) \propto p(d|\theta)p(\theta), \tag{2.3}$$

where $p(\theta)$ is the prior distribution before the data is observed. The data enters the Bayesian inference through the likelihood function $p(d|\theta)$. Then we estimate the posterior mean or the maximum a posteriori (MAP) of unknowns. Let Υ be the parameter space. If a

Gaussian prior is used, then the MAP estimate is equivalent to the solution of the following minimization problem

$$\min_{\theta\in\Upsilon}\Big(\frac{\big\|d-\mathbb{H}(\theta)\big\|^2}{2\sigma^2}+\frac{\big(\theta-\theta^b\big)^T(\boldsymbol{P}^b)^{-1}\big(\theta-\theta^b\big)}{2}\Big), \tag{2.4}$$

where $\|\cdot\|$ is the Euclidean norm, θ^b and $\boldsymbol{P}^b$ are the mean and covariance matrix of the prior (background information) for θ.

In practice, the analytical expression of the posterior distribution in Eq. (2.3) is generally unavailable and the high dimension integration involved in posterior expectation is a great challenge. The Monte Carlo method is often used to approximate the integration. The samples in Monte Carlo can be obtained by ensemble methods, which can avoid explicit calculating the gradient and adjoint equation of forward model. In the paper, we use filter methods to get the samples. Assume θ^f and θ^a be the forecast and analysis of θ, respectively. Let $\mathbb{F}$ be the forecast operator mapping the current prior into the forecast space, i.e.,

$$\theta^f=\mathbb{F}(\theta^a).$$

Algorithm 1 describes the main steps of ensemble-based filter methods.

Algorithm 1 Ensemble-based filter methods

Input: {Given the prior μ_0, the sample size N_e, the forecast operator $\mathbb{F}$ and the measurement data d}.
Output: $\{\theta_i^a\}_{i=1}^{N_e}$
1. **Initial ensemble**: Generate samples $\{\theta_i^a\}_{i=1}^{N_e}$ of θ from μ_0
2. **Forecast**: Update the initial ensemble $\{\theta_i^a\}_{i=1}^{N_e}$ by $\theta_i^f=\mathbb{F}(\theta_i^a),\quad i=1,\cdots,N_e$.
   Generate samples of the model response: $\big\{\mathbb{H}(\theta_i^f)\big\}_{i=1}^{N_e}$.
3. **Analysis**: $\{\theta_i^a\}_{i=1}^{N_e}$ is obtained by updating the forecast ensemble.
   The analysis step utilizes θ_i^f, $\mathbb{H}(\theta_i^f)$ and d together.

Let D be a given bounded physical domain with Lipschitz continuous boundary ∂D. In this paper, Eq. (2.1) is the most general parametric partial differential equations as follows

$$\begin{cases}\dfrac{\partial}{\partial t}\mathbf{u}(t,\mathbf{x};\theta)=\mathbf{f}(\mathbf{u},\nabla\mathbf{u},\nabla^2\mathbf{u},\cdots;\theta),\\ \mathbf{u}(t,\mathbf{x};\theta)=\mathbf{u}_0(\mathbf{x};\theta),\end{cases} \tag{2.5}$$

where $\mathbf{u}(t,\mathbf{x};\theta)\in\mathbb{R}^{N_h}$ is the solution vector with N_h components at time $t\in[0,T]$, at spatial coordinates $\mathbf{x}\in D$, and a parameter vector $\theta\in\mathbb{R}^{n_\theta}$. The function $\mathbf{f}\in\mathbb{R}^{N_h}$ represents a linear or nonlinear function involving $\mathbf{u}$ and the spatial derivatives of $\mathbf{u}$. $\mathbf{u}_0(\mathbf{x};\theta)$ is the initial condition.

The full model (e.g., finite element method, finite difference method, spectral method or radial basis function methods, etc) requires a large number of degrees of freedom and a huge number of time steps and iterations. Therefore, the computational cost is prohibitive when solving complex problems, especially for many-query problems such as quantifying the effects of parameters and estimating unknown parameter values. We note that the forecast

step and the analysis step of Algorithm 1 are repeated in a sequential manner. For the posterior exploration, the forward model needs to be solved repeatedly online. Our goal is to construct a surrogate model $\hat{\mathbf{u}}$ of $\mathbf{u}$ so that the output of the system (2.5) can be evaluated without solving the full model. Based on the surrogate model $\hat{\mathbf{u}}$, we can estimate the statistics of $\mathbf{u}$ efficiently. More generally, given a function $G(\mathbf{u})$, the expectation of $G(\mathbf{u})$ can be approximated using Monte Carlo as follows

$$\mathbb{E}[G(\mathbf{u})] \approx \mathbb{E}[G(\hat{\mathbf{u}})] = \int_{\Omega} G(\hat{\mathbf{u}}(\theta(\omega)))\mathbb{P}(\mathrm{d}\omega) \approx \frac{1}{N_e}\sum_{i=1}^{N_e} G\left(\hat{\mathbf{u}}\left(\theta_i\right)\right).$$

The dynamic mode decomposition (DMD) method has gained popularity due to its ability to extract coherent structures and dynamic features from data without requiring detailed knowledge of the underlying system equations. Its simplicity and effectiveness have made it a valuable tool for data-driven analysis and modeling in various scientific and engineering fields. In this work, we attempt to extend DMD to the parametric dynamical systems. In the next section, we briefly introduce the radial basis function finite difference method.

3. Dynamic mode decomposition

We consider the following continuous time dynamical system to present the dynamic mode decomposition

$$\begin{cases} \dfrac{\partial}{\partial t}\mathbf{u}(t,\mathbf{x}) = \mathbf{f}(\mathbf{u}, \nabla\mathbf{u}, \nabla^2\mathbf{u}, \cdots), \\ \mathbf{u}(t,\mathbf{x}) = \mathbf{u}_0(\mathbf{x}). \end{cases} \tag{3.1}$$

Using the finite element method to solve equation (3.1), we get the following discrete dynamical system with external inputs

$$\mathbf{u}_{n+1} = F(\mathbf{u}_n, \boldsymbol{z}_n), \quad \mathbf{u} \in \mathcal{M}, \quad \boldsymbol{z} \in \mathcal{N}, \tag{3.2}$$

where $\mathbf{u}_n = \mathbf{u}(t_0 + n\Delta t)$ with Δt being the size of the time step, $\mathcal{M}$ denotes the state space, and F is a map from $\mathcal{M}$ to itself, $\boldsymbol{z}_n = \boldsymbol{z}(t_0 + n\Delta t)$ represents external input, which can be introduced by the source item and boundary conditions. When the dynamical system (3.1) has no external inputs, i.e., the source term equals zero and the boundary conditions are homogeneous, then we have

$$\mathbf{u}_{n+1} = F(\mathbf{u}_n), \quad \mathbf{u} \in \mathcal{M}. \tag{3.3}$$

F can be a linear mapping or nonlinear mapping. When F is a nonlinear mapping, solving nonlinear dynamical system takes a large amount of computing resources. Koopman operator is a linear operator acting on the observation function space. Koopman operator can transform a nonlinear dynamical system into a linear system in the observation function space. For discrete dynamical systems (3.3), the Koopman operator is defined as follows.

Definition 3.1. *(Koopman operator) Let $\boldsymbol{G}(\mathcal{M})$ is an infinite dimensional observation function space for any scalar-valued observable function $g : \mathcal{M} \to \mathbb{R}$. The Koopman operator $\mathcal{K} : \boldsymbol{G}(\mathcal{M}) \to \boldsymbol{G}(\mathcal{M})$ is defined by*

$$\mathcal{K}g(\mathbf{u}) := g\big(F(\mathbf{u})\big), \quad \forall g \in \boldsymbol{G}(\mathcal{M}).$$

The nonlinear dynamical system (3.3) can be lifted to the following linear problem,

$$\mathbf{u}_{n+1} = F(\mathbf{u}_n) \quad \Rightarrow \quad g(\mathbf{u}_{n+1}) = \mathcal{K}g(\mathbf{u}_n).$$

$\mathcal{K}$ is an infinite-dimensional operator. We restrict the infinite-dimensional system to a finite-dimensional invariant subspace $\mathcal{G}(\mathcal{M}) \subseteq \boldsymbol{G}(\mathcal{M})$ for the numerical simulation. Suppose that there exists an invariant subspace $\mathcal{G}$ of $\mathcal{K}$, i.e.,

$$\mathcal{K}g \in \mathcal{G}, \quad \forall g \in \mathcal{G}.$$

Let a set of observation functions $\{g_1, \ldots, g_q\}(q < \infty)$ span $\mathcal{G}$. We restrict $\mathcal{K}$ to $\mathcal{G}$ and denote it by $\mathcal{K}|_{\mathcal{G}}$. Then $\mathcal{K}|_{\mathcal{G}}$ has a matrix-form representation $\boldsymbol{K}$ with respect to $\{g_1, \ldots, g_q\}$, i.e.,

$$\boldsymbol{g}(F(\mathbf{u})) \quad = \quad \begin{bmatrix} g_1(F(\mathbf{u})) \\ g_2(F(\mathbf{u})) \\ \vdots \\ g_q(F(\mathbf{u})) \end{bmatrix} = \begin{bmatrix} \mathcal{K}g_1(\mathbf{u}) \\ \mathcal{K}g_2(\mathbf{u}) \\ \vdots \\ \mathcal{K}g_q(\mathbf{u}) \end{bmatrix} = \boldsymbol{K}\boldsymbol{g}(\mathbf{u}).$$

The spectral decomposition theory of the Koopman operator can give an expression for the observation functions. Therefore, we consider the eigendecomposition of the matrix $\boldsymbol{K}$. Let

$$\boldsymbol{K}\boldsymbol{\psi}_j = \lambda_j \boldsymbol{\psi}_j, \quad \boldsymbol{\omega}_j^T \boldsymbol{K} = \lambda_j \boldsymbol{\omega}_j^T,$$

where λ_j is the eigenvalue, $\boldsymbol{\psi}_j$ is the right eigenvector, $\boldsymbol{\omega}_j$ is the left eigenvector and $\boldsymbol{\omega}_i^T \boldsymbol{\psi}_j = \delta_{ij}$. If the matrix $\boldsymbol{K} \in \mathbb{R}^{q \times q}$ has q linearly independent eigenvectors, then any $\boldsymbol{g} \in \mathcal{G}$ can be expressed by a linear combination of the eigenvectors, i.e.,

$$\boldsymbol{g}(\mathbf{u}) = \sum_{j=1}^{q} v_j(\mathbf{u}) \boldsymbol{\psi}_j, \tag{3.4}$$

where $v_j(\mathbf{u}) = \boldsymbol{\omega}_j^T \boldsymbol{g}(\mathbf{u})$. For a sequential time series, repeatedly applying the Koopman operator to equation (3.4) gives

$$\boldsymbol{g}(\mathbf{u}_n) = \sum_{j=1}^{q} \lambda_j^n v_j(\mathbf{u}_0) \boldsymbol{\psi}_j. \tag{3.5}$$

Therefore, we can use the eigenvalues and eigenvectors of the Koopman operator and $\boldsymbol{g}(\mathbf{u}_0)$ to evaluate the observation function $\boldsymbol{g}(\mathbf{u})$ at any time without knowing the specific expression of F.

Dynamic mode decomposition algorithm only uses the observation data to compute an approximation to the Koopman eigenvalues and eigenvectors. Suppose we have a snapshot sequence of data $\{\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$. Given the observation function $\boldsymbol{g}$, we define the data matrices of observables S_1 and S_2 as follows:

$$S_1 = \begin{bmatrix} | & | & & | \\ \boldsymbol{g}(\mathbf{u}_0) & \boldsymbol{g}(\mathbf{u}_1) & \ldots & \boldsymbol{g}(\mathbf{u}_{m-1}) \\ | & | & & | \end{bmatrix}, \quad S_2 = \begin{bmatrix} | & | & & | \\ \boldsymbol{g}(\mathbf{u}_1) & \boldsymbol{g}(\mathbf{u}_2) & \ldots & \boldsymbol{g}(\mathbf{u}_m) \\ | & | & & | \end{bmatrix}.$$

The DMD matrix can be defined as $L = S_2 S_1^{\dagger}$, where $S_1^{\dagger}$ is the Moore-Penrose pseudoinverse of S_1, then we have

$$\boldsymbol{g}(\mathbf{u}_{n+1}) \approx L\boldsymbol{g}(\mathbf{u}_n).$$

DMD matrix L is an approximation of the Koopman matrix $\boldsymbol{K}$. It may be very expensive to do eigendecomposition directly on the matrix L. Tu et al. [39] proposed DMD, which reconstructs the nonzero eigenvalues and eigenvectors of L by calculating the eigendecomposition of a low-dimensional projection of L. We note that U is termed as the SVD-modes, and the low-dimensional projection of DMD matrix L can be termed as the reduced DMD matrix or reduced Koopman operator matrix, i.e., $L_r = U^* L U$. Here " $*$ " denotes the conjugate transpose of matrixes. Let's briefly summarize the main steps of DMD as follows in Algorithm 2.

Algorithm 2 DMD method

input: Given snapshots $\{\boldsymbol{g}(\mathbf{u}_0), \boldsymbol{g}(\mathbf{u}_1), \cdots, \boldsymbol{g}(\mathbf{u}_m)\}$ and the truncated rank r
output: DMD solution $\mathbf{u}_{DMD}(t)$

1. Obtain data sets S_1 and S_2 from data set.
2. Perform SVD of S_1 as $S_1 = U\Sigma V^*$, $U \in \mathbb{R}^{N\times r}$, $\Sigma \in \mathbb{R}^{r\times r}$, $V \in \mathbb{R}^{m\times r}$.
3. Define a low-dimensional projection of L: $L_r = U^* L U = U^* S_2 V \Sigma^{-1}$.
4. Compute the eigenvalues and eigenvectors of L_r, $L_r W = W\Lambda$.
5. Set the mode of L as $\Phi = UW$.
6. Then the future state in the space of observables is given by $\boldsymbol{g}_{\mathrm{DMD}}(\mathbf{u}_n) = \Phi\Lambda^n \boldsymbol{b}$, with $\boldsymbol{b} = \Phi^{\dagger}\boldsymbol{g}(\mathbf{u}_0)$, the continuous formulation $\boldsymbol{g}_{\mathrm{DMD}}(\mathbf{u}(t)) = \Phi\Lambda^{t/\Delta t}\boldsymbol{b}$.
7. Finally, transform to state space $\mathbf{u}_{DMD}(t) = \boldsymbol{g}^{-1}(\boldsymbol{g}_{\mathrm{DMD}}(\mathbf{u}(t))) = \boldsymbol{g}^{-1}(\Phi\Lambda^{t/\Delta t}\boldsymbol{b})$, where $\boldsymbol{g}^{-1}$ is in the sense of least-squares if $\boldsymbol{g}$ is not invertible.

4. Ensemble-based filter using DMD-wiNN model

For the ensemble-based filter methods presented in Algorithm 1, we need to repeatedly compute the forward model for all ensemble members. This computation is very expensive when the forward model is a complex PDE model and the number of ensemble members is large. In order to significantly accelerate the forward model computation, we construct a surrogate model for the forward model using model reduction methods.

The goal is to approximate a large-scale problem in a low dimensional space. To this end, the key idea is to choose a set of appropriate basis functions, which can span a good approximation space for the solution. If Eq. (2.5) is nonlinear with respect to $\mathbf{u}$, we can derive an algebraic system for Eq. (2.5) as follows by applying suitable discretization method

$$\mathbf{A}(\theta, \mathbf{u})\mathbf{u} = \mathbf{B}(\mathbf{u}), \tag{4.1}$$

where $\mathbf{B} \in \mathbb{R}^{N_h}$ the source vector. The N_h is the number of spatial degree of freedoms and is usually very large if we straightforwardly solve the equation in fine grid. We can use a model reduction method and reduce the number of basis functions to improve the efficiency. Then we can get a reduced algebraic system for Eq. (2.5),

$$\mathbf{A}_r(\theta, \mathbf{u}_r)\mathbf{u}_r = \mathbf{B}_r(\mathbf{u}_r).$$

In order to accelerate evaluations of the posterior density for each updated parameter ensemble, The stochastic response surface methods can be applied to construct surrogate. The solution $\mathbf{u}_r$ of the reduced model can be expressed by stochastic basis functions such as polynomial chaos [45], radial basis functions [34], and wavelet basis functions [23]. In this paper, the surrogate model is constructed through DMD method to reduce the computation cost of solving forward models. In the next subsection, we will introduce a parametric dynamic mode decomposition method, i.e., a model order reduction (MOR) method based on the reduced DMD matrixs to construct a parameter-dependent surrogate model for the nonlinear parametric dynamical systems.

4.1. A MOR method based on the reduced DMD matrixs

Using the finite element method to simulate equation (2.5), we can obtain a system of coupled parametric ODEs such as

$$\mathbf{u}_{n+1}(\theta) = F(\mathbf{u}_n; \theta), \quad \mathbf{u} \in \mathcal{M}. \tag{4.2}$$

By the defination of Koopman operator in 3.1, the nonlinear system (4.2) can be rewritten as

$$\boldsymbol{g}(\mathbf{u}_{n+1}; \theta) = \mathcal{K}(\theta)\boldsymbol{g}(\mathbf{u}_n; \theta). \tag{4.3}$$

When F is a nonlinear mapping, it starves for a large amount of computing resources using the Newton method or Piccard iterative method to solve nonlinear problems. To improve computation efficiency, we combine DMD with the weighted & interpolated nearest-neighbors algorithm (wiNN, [6, 44]) to develop a model reduction method, which is named as DMD-wiNN, and it admits an offline-online procedure. We first define the parameter training set $\Xi_{\text{train}} \subsetneq \Upsilon$, which is a finite subset with the cardinality being $|\Xi_{\text{train}}| = N_s$.

In the offline stage, given training set Ξ_{train}, we perform it as follows in Algorithm 3.

Algorithm 3 The offline stage for DMD-wiNN

input: The parameter training set $\Xi_{\text{train}} \subsetneq \Upsilon$

output: Reduced DMD matrixes $\{L_r(\theta_j)\}_{j=1}^{N_s}$, SVD-modes $\{U(\theta_j)\}_{j=1}^{N_s}$ and initial coefficients $\{\boldsymbol{g}(\mathbf{u}_0; \theta_j)\}_{j=1}^{N_s}$.

1. Generate snapshot data for each parameter $\theta_j \in \Xi_{\text{train}}$ such as $\{\mathbf{u}_0(\theta_j), \mathbf{u}_1(\theta_j), \mathbf{u}_2(\theta_j), \cdots, \mathbf{u}_{m_j}(\theta_j)\}_{j=1}^{N_s}$;
2. Obtain data set $S_1(\theta_j)$ and $S_2(\theta_j)$ from snapshot data;
3. Perform singular value decomposition of $S_1(\theta_j)$ as $S_1(\theta_j) = U(\theta_j)\Sigma(\theta_j)V(\theta_j)^*$, and define the corresponding reduced Koopman operator matrix as $L_r(\theta_j) = U(\theta_j)^* L(\theta_j) U(\theta_j) = U^* S_2 V \Sigma^{-1}(\theta_j), \quad \text{for } 1 \leqslant j \leqslant N_s.$

Once the offline stage is completed, we need storing the reduced DMD matrixes $\{L_r(\theta_j)\}_{j=1}^{N_s}$, SVD-modes $\{U(\theta_j)\}_{j=1}^{N_s}$ for the online stage computation. Subsequently, we focus on constructing the efficient surrogate model of parametric DMD (an approximation of Koopman operator) matrix $L(\theta)$. There are several approaches for the approximation of L_r^{θ}, such as point-wise Lagrangian interpolation [14], Neural network and least square fitting and so on. In this work, we mainly adopt the wiNN algorithm.

The main idea of wiNN is to find the n nearest neighbors of the new sample θ with some metric measure, and then using the weighted average of these n nearest neighbor to predict the reduced DMD matrix L_r^θ. More specifically, let θ^i be the i-th nearest neighbor of the test sample θ among the training data Ξ_{train}, and $w_d(\omega, z)$ be a weight function $\Upsilon \times \Upsilon \to R$ as follows

$$w_d(\theta, \theta^i) = \frac{\frac{1}{\|\theta-\theta^i\|}}{\sum_{j=1}^{n} \frac{1}{\|\theta-\theta^j\|}}, \quad \text{for } i = 1, \cdots, n. \tag{4.4}$$

We note that the weighting scheme satisfies $w_d(\theta, \theta^i) \geq 0$ and $\sum_{i=1}^{n} w_d(\theta, \theta^i) = 1$. According to (4.4), we find the parameter θ_j close to θ has a large weight w^{θ_j}, and parameter θ_j far from θ with a small weight w^{θ_j}.

Then the weighted nearest neighbor scheme for $L_r(\theta)$ is approximated of the form

$$L_r^\theta = \sum_{i=1}^{k} L_r(\theta^i) w_d(\theta, \theta^i). \tag{4.5}$$

Similarly, we approximate SVD-modes U^θ as

$$U^\theta = \sum_{i=1}^{k} U(\theta^i) w_d(\theta, \theta^i). \tag{4.6}$$

For each parameter $\theta \in \Upsilon$, the online stage can be summarized as follows in Algorithm 4.

Algorithm 4 The online stage for DMD-wiNN

input: Reduced DMD matrixes $\{L_r(\theta_j)\}_{j=1}^{N_s}$, SVD-modes $\{U(\theta_j)\}_{j=1}^{N_s}$ and initial coefficients $\{\boldsymbol{g}(\mathbf{u}_0; \theta_j)\}_{j=1}^{N_s}$.

output: Parametric DMD solution $\mathbf{u}_{DMD}(t; \theta)$

1. Find the n nearest neighbor parameter values of θ, denoted as $\{\theta^i\}_{i=1}^{n}$;
2. Construct the reduced DMD matrix L_r^θ by equation (4.5);
3. Perform the eigen-decomposition of L^θ to get the reduced eigen-pair (Λ, W);
4. Construct the SVD-mode U^θ by equation (4.6);
5. Compute the DMD-mode $\Phi(\theta) = U^\theta W$;
6. Compute initial coefficients $\mathbf{b}^\theta = (\Phi^\theta)^\dagger \boldsymbol{g}^\theta(\mathbf{u}_0)$;
7. Then $\boldsymbol{g}_{\text{DMD}}(\mathbf{u}(t; \theta)) = \Phi^\theta \Lambda^{t/\Delta t} \boldsymbol{b}^\theta$;
8. Finally, transform to state space $\mathbf{u}_{DMD}(t; \theta) = \boldsymbol{g}^{-1}(\Phi^\theta \Lambda^{t/\Delta t} \boldsymbol{b}^\theta)$.

4.2. A low-rank EnKF using DMD-wiNN model

The goal of inverse problems is to identify an appropriate solution which can minimize the misfit between the outputs of interest and measurements. Here it is equivalent to solving the minimization problem (2.4). In this paper, we use EnKF methods to solve the minimization problem. Thus, we present a low-rank EnKF using DMD-wiNN model to accelerate posterior

exploration and improve the sequential update performance. The lth update of EnKF [2] is as following

$$\begin{cases} \theta_l^a = \theta_l^f + \boldsymbol{K}_{al}^l(d_l - \mathbb{H}(\theta_l^f)), \\ \boldsymbol{P}_l^a = (\boldsymbol{I} - \boldsymbol{K}_{al}^l \boldsymbol{H})\boldsymbol{P}_l^f, \end{cases}$$

where $\boldsymbol{H}$ is the gradient of operator $\mathbb{H}$, $\boldsymbol{P}_l^f$ and $\boldsymbol{P}_l^a$ are the covariances of forecast and analysis ensemble, respectively, and Kalman gain

$$\boldsymbol{K}_{al}^l = \boldsymbol{P}_l^f \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}_l^f \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1}, \quad \boldsymbol{R} = \sigma^2 \boldsymbol{I}.$$

Algorithm 5 low-rank EnKF with DMD-wiNN model

Input: Initial $\{\theta_{0,s}^a\}_{s=1}^{N_e}$, parameter training set $\Xi_{\text{train}} \subsetneq \Upsilon$, given the truncated rank r and the total artificial time steps I, measurement data $\{d_1, \cdots, d_I\}$

Output: Θ_I^a.

1. Perform algorithm 3 to construct the snapshots: Reduced DMD matrixes $\{L_r(\theta_j)\}_{j=1}^{N_s}$, SVD-modes $\{U(\theta_j)\}_{j=1}^{N_s}$ and initial coefficients $\{\boldsymbol{g}(\mathbf{u}_0;\theta_j)\}_{j=1}^{N_s}$, where $\theta_j \in \Xi_{\text{train}}$.
2. $\Theta_0^a = (\theta_{0,1}^a, \cdots, \theta_{0,N_e}^a)$
3. **for**: $l = 1 : I$
   - (1). Forecast step:
     - (i) $\Theta_l^f = \Theta_{l-1}^a$.
     - (ii) **parfor:** $i = 1 : N_e$
       Perform Algorithm 4 to get $\mathbb{H}(\mathbf{u}(t, \theta_{l,i}^f))$
       endfor
     - (iii) $Z_l = (\mathbb{H}(\mathbf{u}(t, \theta_{l,1}^f)), \cdots, \mathbb{H}(\mathbf{u}(t, \theta_{l,N_e}^f)))$.
   - (2). Compute the deviation $\tilde{Z}_l = \frac{1}{\sigma\sqrt{N_e - 1}}(Z_l - \frac{\sum_{i=1}^{N_e} \mathbb{H}(\mathbf{u}(t,\theta_{l,i}^f))}{N_e})$
   - (3). $[U^e, \Lambda^e, V^e] = \mathrm{SVD}(\tilde{Z}_l)$.
   - (4). Obtain the truncated $\{U_{tr}^e, \Lambda_{tr}^e, V_{tr}^e\}$ by retain the 98% energy.
   - (5). The low-rank $\boldsymbol{K}_{al} = \Theta_l^f V_{tr}^e \Lambda_{tr}^e (\boldsymbol{I} + \Lambda_{tr}^e (\Lambda_{tr}^e)^T)^{-1} (U_{tr}^e)^T$.
   - (6). $Z_l^o = (d_l, \cdots, d_l)$.
   - (7). Analysis step:$\Theta_l^a = \Theta_l^f + \boldsymbol{K}_{al}(Z_l^o - Z_l)$.

endfor

In general, Kalman gain matrix is appproximate using the covariance of ensemble sample and simulated data, i.e.,

$$\boldsymbol{K}_{al}^l \approx \mathrm{Cov}(\Theta_l^f, Z_l)\mathrm{Cov}(Z_l, Z_l)^{-1}, \tag{4.7}$$

where Θ_l^f is the forecast ensemble and Z_l is the ensemble of simulated measurements. EnKF is a sampling method and avoids expensive gradient computation for solving minimization problem. The second-order derivative is included in the form of Kalman gain $\boldsymbol{K}_{al}^l$. Thus, the forecast error covariance matrix and analysis error covariance matrix are not necessary to compute. The true mean and covariance are approximated by ensemble mean and ensemble covariance, respectively. In the paper, we make use of the stochastic analysis ensemble generation method, where the simulated measurements are perturbed by simulated measurement error ε^f. The ε^f is independent of ε.

Although EnKF method can avoid linearization and repeated sampling to explore posterior density, it requires to compute the forward model many times in each forecast step described in algorithm 1. When the dimension of unknown parameters θ is large, we need a large number of ensemble members to estimate θ. This implies that the computation of the forecast is very expensive. For improving the computation efficiency, we construct a surrogate model based on DMD-wiNN to approximately represent the full order model. The reduced model only retains a fewer basis functions. To avoid the singular of (4.7) for a small ensemble size, we apply the SVD method for the deviation of Z_l.

However, EnKF has an inherent constraint assumption for the prior distribution, which must be a Gaussian distribution. As we know, the support of Gaussian distribution is $\mathbb{R}$. If the support of unknown parameters is bounded, its values are uncontrollable during the sequentially update of Kalman filter, ensemble samples may run out of the interval. For this case, we apply normal-score EnKF [2, 49] to perform non-Gaussian models. The objective of the surrogate model is to construct a reduced model that quantifies the primary features of the high-fidelity model while providing the computational efficiency required for uncertainty quantification. In stationary Bayesian inference, we may need to build surrogate model only once. However, EnKF method integrates new measurement data in each analysis step. Thus, the surrogate model can be updated sequentially according to the practice. The outline of low-rank EnKF using DMD is presented in Algorithm 5.

5. Numerical results

In this section, we demonstrate applicability and efficiency of the proposed filter method using DMD-wiNN on several nonlinear time-dependent equations. A few numerical results will be presented for the estimations of the model's unknown parameters and different unknown sources. The prediction errors between the predictive solutions and the reference solutions show the predictive ability of DMD-wiNN outside of the training time region. In Subsection 5.1, we indetify a nonlinear reaction parameter when reaction term is a quadratic polynominal of u. In Subsection 5.2, we estimate the parameter of nonlinear diffusion term when equation (2.5) is a advection diffusion model. In Subsection 5.3, we will recover a nonlinear permeability field and source locations.

For comparison, the effectiveness and effficiency of EnKF without surrogate model and with different order gPC models [2] are computed. The same prior ensemble and measurements as DMD-wiNN in each experiment are used. Measurements are generated synthetically using the finite element method (FEM) for spatial discritization and the explicit Euler method for the time in a fine grid. Furthermore, FEM are applied to solve the forward models when no extra surrogate model is builded. For convenience, FEM in the following figures refers to no extra surrogate, gPC (N=2) denotes the second-order gPC and gPC (N=4) is the fourth-order gPC for each experiment. For the numerical examples, we consider a dimensionless square domain $D = [0, 1] \times [0, 1]$ for spatial variable and a homogeneous Neumman boundary condition

$$\frac{\partial u}{\partial \mathbf{n}} = 0, \qquad \mathbf{x} \in \partial D, \quad t \in [0, T], \quad \omega \in \Omega. \tag{5.1}$$

The final time $T = 1$ for each experiment. The deviation of measurement noise is set as $\sigma = 0.01$.

To assess the estimation by the filter methods based on different surrogate models, we use two indicators of ensemble samples, i.e., root mean square error (RMSE) and spread [48, 24]. They are defined as follows

$$\mathrm{RMSE}_\theta := \sqrt{\frac{1}{n_\theta}\sum_{i=1}^{n_\theta}(\mathbb{E}[\theta_i]-\theta_i^t)^2}, \quad \mathrm{spread}_\theta := \sqrt{\frac{1}{n_\theta}\sum_{i=1}^{n_\theta}\mathrm{Var}(\theta_i)},$$

where $\mathbb{E}[\theta_i]$, θ_i^t and $\mathrm{Var}(\theta_i)$ denote the ensemble mean, reference and ensemble variance of the ith element of the unknown θ, respectively. Here RMSE measures the accuracy of the mean and spread measures the uncertainty in the estimation. The match between the RMSE and the spread provides another indicator [48], defined as

$$\mathrm{RS}_\theta := \mathrm{RMSE}_\theta^2 - \mathrm{spread}_\theta^2 = \frac{2}{n_\theta}\sum_{i=1}^{n_\theta}\Big(\mathbb{E}[\theta_i](\mathbb{E}[\theta_i]-\theta_i^t)\Big) + \frac{1}{n_\theta}\sum_{i=1}^{n_\theta}\Big((\theta_i^*)^2 - \mathbb{E}[\theta_i^2]\Big). \tag{5.2}$$

RS_θ characterizes the mean and uncertainty estimation. When $\mathbb{E}[\theta] \to \theta^t$ and $\mathbb{E}[\theta^2] \to (\theta^t)^2$, RS_θ will tend to zero. Thus the small value of RS_θ indicates an accurate estimation. For the practical application, the model output $\mathbb{H}(\theta)$ is also the key. For providing a quantitative comparison, the average data dismatch (ADM) [48] is defined as

$$\mathrm{ADM} = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}\left[\frac{1}{n_d}\sum_{j=1}^{n_d}(d_{i,j}^s - d_{i,j}^t)^2\right]}$$

where n_t it the total number of measurement steps, $d_{i,j}^s$ is the simulated data of the jth measurement location at the ith measurement step and $d_{i,j}^t$ is the corresponding true data given as the reference.

5.1. Identify the nonlinear reaction term

In this subsection, we consider the following non-linear diffusion-reaction model

$$\frac{\partial u(\mathbf{x},t,\omega)}{\partial t} - \mu\Delta u(\mathbf{x},t,\omega) = f(u,\mathbf{x},t;\omega), \quad \mathbf{x}\in D, \quad t\in(0,T], \quad \omega\in\Omega$$

with the boundary condition given by Eq. (5.1). For this example, we prescribe an initial condition that are a sum of exponential plumes, i.e.,

$$u_0(\mathbf{x},\omega) = \sum_{i=1}^{2}\exp(-\frac{1}{2\alpha_i^2}\|\mathbf{x}-\mathbf{x}_i\|^2),$$

where the standard deviations $\alpha_1 = 0.3$, $\alpha_2 = 0.2$ and centers $\mathbf{x}_1 = (0.2, 0.2)$, $\mathbf{x}_2 = (0.75, 0.8)$. The diffusion coeficient $\mu = 0.2$ and the reaction term is a quadratic polynomial in u, i.e.,

$$f(u,x,t;\omega) = \lambda_1(\omega)u\left(1-\frac{u}{\lambda_2(\omega)}\right)$$

with excitation rate λ_1 and threshold potential λ_2. Here both λ_1 and λ_2 are unknown.

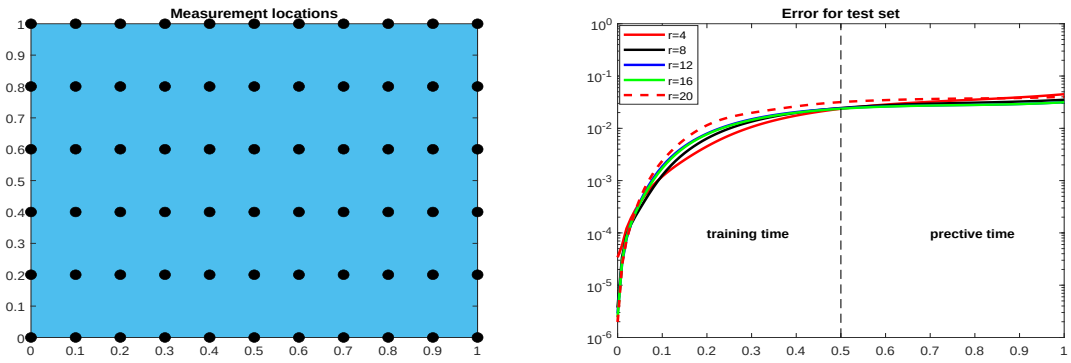

Figure 5.1: The black points are the measurement locations (left). The relative errors for the test set using different truncation DMD basis functions (right).

For this example, the ground true parameters λ_1 and λ_2 are set as 10^{-4} and 2, respectively. The magnitude of λ_1 is far smaller than λ_2. Thus we scale λ_1 using the logarithmic function, i.e., $-\log_{10}\lambda_1 = 4$. To the end, the estimated parameter is $\theta = (-\log_{10}\lambda_1, \lambda_2)$. The number of artificial time steps for data assimilation is set as 12. The measurement locations are distributed on the uniform 11×6 grid of the domain $[0,1] \times [0,1]$ as shown in Fig. 5.1 (left). Measurement data are generated from 100×100 grid with time step $\Delta t = 0.005$. For the inversion based on EnKF, ensemble size is set as 300 and prior ensemble samples are randomly drawn from the standard multivariate normal distribution. For Algorithm 5, the artificial time steps I of EnKF is set as 8.

The forward model is defined on a uniform 100×100 fine grid with time step $\Delta t = 0.01$. 400 training samples with randomly generated from the uniform distribution $[2,10] \times [1,10]$ are used to build the parameterized DMD model. For the applicability and efficiency of DMD-wiNN model, the relative errors for different truncation basis functions are plotted in Fig. 5.1 (right). At the beginning, the error for the test set slightly decreases when the number r of truncation basis functions increases. However, the test error increases for the case of $r = 20$. Figure 5.1 (right) shows it is enough to select $r = 8$ basis functions for the DMD-wiNN method. Furthermore, the training data are obtained by FEM and the training time interval is $[0, 0.5]$ as shown the left region of the vertical dash in Fig. 5.1 (right). For $t \in (0.5, 1]$, it is the predictive error from DMD-wiNN model. Both the test errors of training time and predictive time are steady when 8 basis functions are applied to construct the DMD-wiNN model.

Table 1 demonstrates the CPU time, RS_θ and ADM using DMD-wiNN, FEM and gPC surrogate model. EnKF based on FEM is time-consuming because of no extra surrogate model. Compared with FEM, the surrogate model using DMD-wiNN is faster. The EnKF using gPC (N=2) is easy to perform. However, the RMSE using gPC is bigger than using FEM and DMD-wiNN. Increasing the order number of gPC can not improve the RMSE. Furthermore, the spread values using gPC (N=2) and FEM are smaller than the proposed DMD-wiNN based EnKF method. This is due to the underestimation exists in the ensemble filter methods [10]. In term of the match RS_θ between the RMSE and the spread, DMD-wiNN outperforms the orther surrogate models. The ADM from DMD, FEM and different order gPC are also shown in Table. 1. The proposed EnKF based on DMD-wiNN model

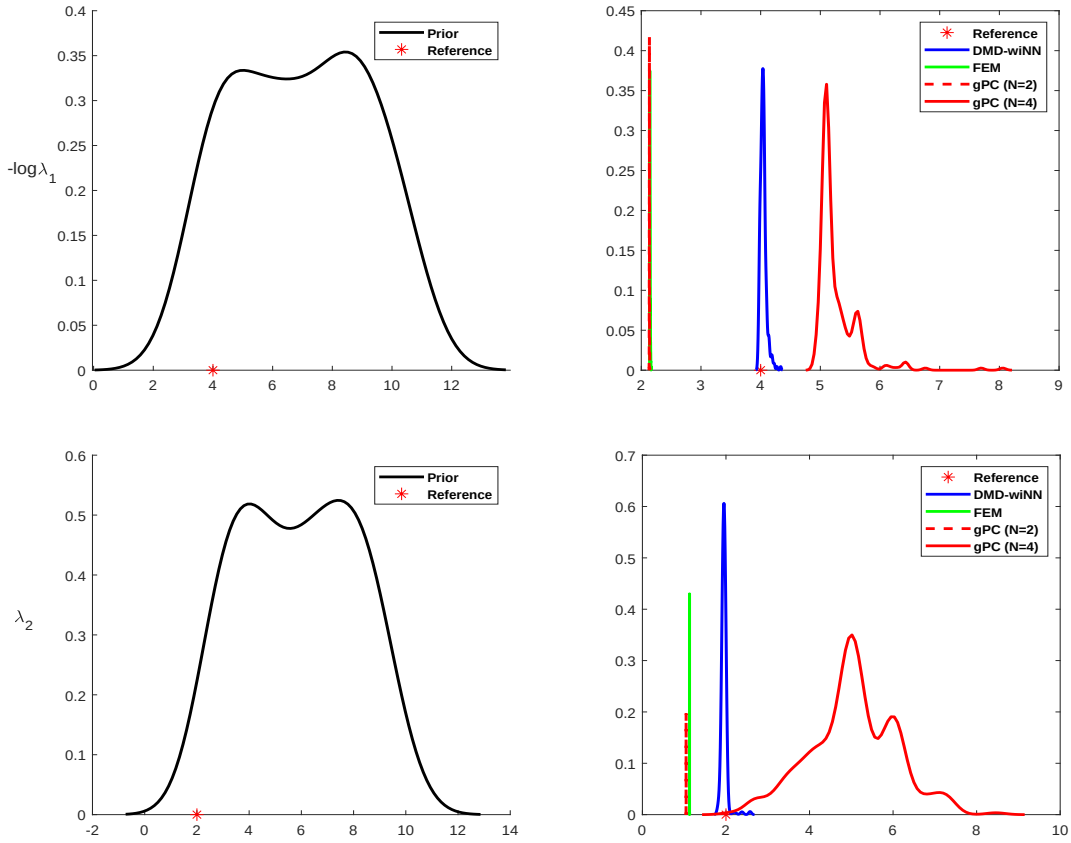

Figure 5.2: The prior and posterior distributions for $-\log\lambda_1$ and λ_2 generated by three sufforgate models (DMD-wiNN, FEM and gPC).

provides the smallest ADM than based on FEM and gPC. The gPC (N=2) gives the largest ADM for the three surrogate models.

In order to measure the estimate accuracy of θ, Fig. 5.2 displays the final marginal posterior density estimations generated by EnKF using FEM, DMD-wiNN and gPC. We note the marginals by the two-stage EnKF are closer to the reference values than the standard EnKF. Although the output of the DMD-wiNN model in the time interval [0.5, 1] is the predictive data, it is sufficiently informative to identify a small range of values for the unknown parameters as plotted in Fig. 5.2 (blue solid line). The important region of marginal densities becomes narrower than the prior as data information gains. The marginal poste-

Table 1: CPU time, RS_θ and ADM using three different surrogates.

	FEM	DMD-wiNN	gPC (N=2)	gPC (N=4)
CPU time (s)	4035.8	215.12	17.30	65.13
RMSE_θ	1.4491	0.0434	1.4803	2.3293
spread_θ	0.0031	0.0649	0.0012	0.7688
$\lvert\text{RS}_\theta\rvert$	2.0999	0.0023	2.1914	4.8344
ADM	0.8345	0.0743	2.5141	1.7549

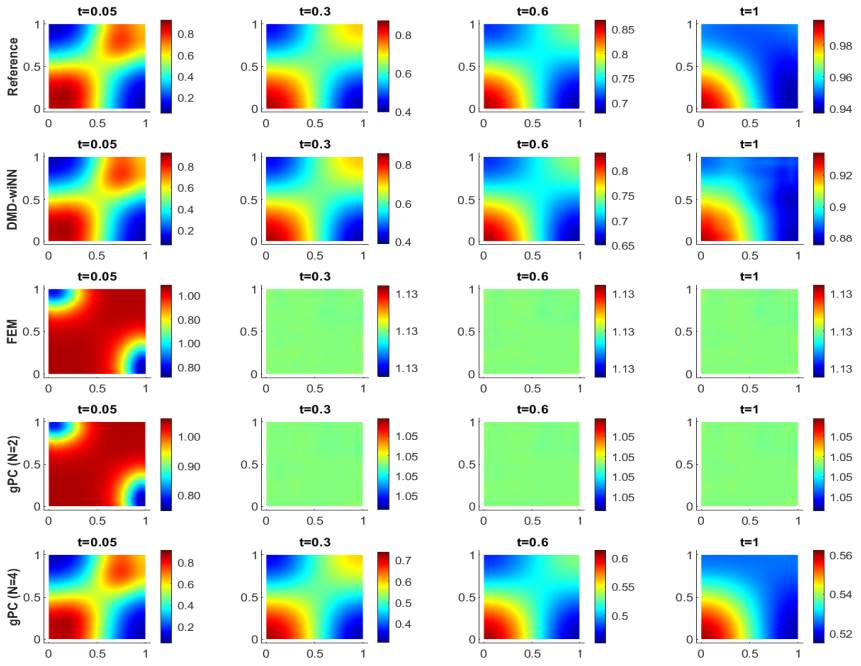

Figure 5.3: The reference solutions (top row) and posterior mean solutions using DMD-wiNN, FEM, gPC (N=2) and gPC (N=4) against the time.

rior densities by FEM and gPC (N=2) are almost concentrated in one point far from the reference. This is incosistent with the small spread in Table. 1. When using gPC (N=4), the uncertainty is overestimation. The marginal posteriors from the proposed EnKF using DMD-wiNN method provides a better estimation of $-\log\lambda_1$ and λ_2.

Because the true parameter of the reaction term is actually unknown in practice, computing the match between the estimated output and the reference is the most straightforward way to measure the performance. Figure. 5.3 (top row) illustrates the structure and value of reference solutions. Both FEM and gPC (N=2) can not capture the main structures. The estimated solutions tend to the same value. The gPC (N=4) improves estimated solutions. The structure is close to that of reference, but the value is smaller than that of reference due to overestimating the uncertainty of θ. For the proposed method, the first two subfigures of second row belong to the training time region. So both the value and structure are the same as the references. In the predictive time region, the value is slightly smaller than that of the reference. It implies that the DMD-wiNN model is effective for state predictions.

5.2. Estimate the nonlinear diffusion coefficient

In this subsection, we consider the following transient nonlinear diffusion model

$$\frac{\partial u(\mathbf{x},t,\omega)}{\partial t}+\mathbf{w}\cdot\nabla u-\nabla\cdot\kappa(u,\omega)\nabla u(\mathbf{x},t,\omega)=f(\mathbf{x},t,\omega),\quad \mathbf{x}\in D,\quad t\in(0,T],\quad \omega\in\Omega \tag{5.3}$$

with the boundary condition in Eq. (5.1). For this example, we prescribe an initial condition

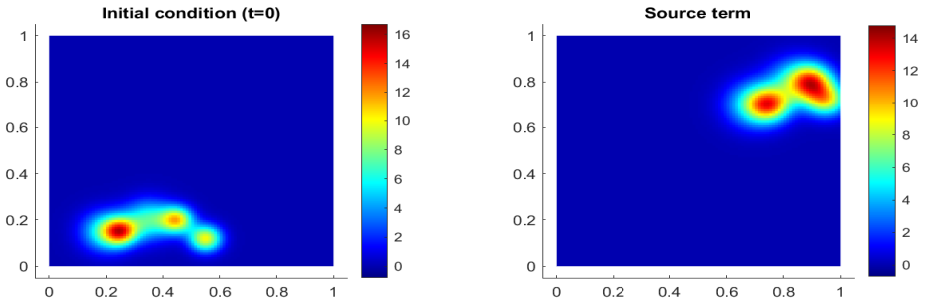

Figure 5.4: The initial condition (left) and source term (right).

Table 2: Standard deviations and center locations for the initial condition in Eq. (5.4). The initial condition is pictured in Fig. 5.4 (left).

i	1	2	3	4	5
α_i	0.06	0.04	0.06	0.04	0.04
x_i	0.2	0.25	0.35	0.45	0.55
y_i	0.15	0.15	0.2	0.2	0.12

Table 3: Standard deviations and center locations forthe source term in Eq. (5.5). The source term is pictured in Fig. 5.4 (right).

i	1	2	3	4	5
β_i	0.07	0.05	0.07	0.05	0.05
x_i	0.7	0.75	0.85	0.9	0.95
y_i	0.7	0.7	0.8	0.8	0.72

as shown in Fig. 5.4 that are a sum of Gaussian plumes, i.e.,

$$u_0(\mathbf{x},\omega) = \sum_{i=1}^{5} \frac{1}{\alpha_i\sqrt{2\pi}} \exp(-\frac{1}{2\alpha_i^2}\|\mathbf{x}-\mathbf{x}_i\|^2), \tag{5.4}$$

where the standard deviations α_i and centers $\mathbf{x}_i$, $i = 1, 2, \cdots, 5$, are given in Table 2. The form of the source term is similar to the initial condition, i.e.,

$$f(\mathbf{x},t,\omega) = \sum_{i=1}^{5} \frac{1}{\beta_i\sqrt{2\pi}} \exp(-\frac{1}{2\beta_i^2}\|\mathbf{x}-\mathbf{x}_i\|^2). \tag{5.5}$$

where the detailed parameters are given in Table 3. The advection velocity for this problem is $\mathbf{w} = 0.8\hat{i} + 0.4\hat{j}$. The thermal conductivity κ is a nonlinear function of temperature u

$$\kappa(u,\omega) = a(\omega) + u^{b(\omega)}. \tag{5.6}$$

where a and b are parameters of the conductivity correlation. Here both a and b are unknown, i.e., $\theta = (a, b)$.

For generated the measurements, the ground true parameters a and b are set as 0.3 and 4, respectively. The artificial time steps I for EnKF in Algorithm 5 is set as 10. The measurement locations are distributed on the uniform 11×6 grid of the domain $[0, 1] \times [0, 1]$ as shown in Fig. 5.1(left). Measurement data are generated from 100×100 grid with time step $\Delta t = 0.01$. For the inversion based on EnKF, ensemble size is set as 300.

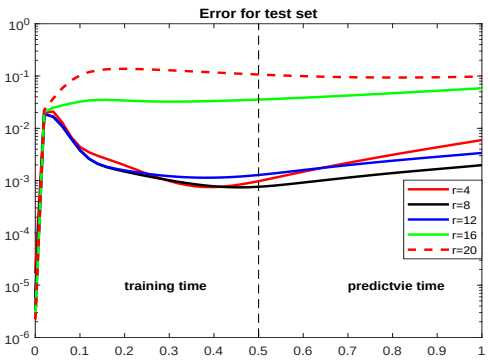

Figure 5.5: The relative errors for the test set against the different truncation DMD basis functions.

The forward model is defined on a uniform 100×100 fine grid with time step $\Delta t = 0.02$. To construct the DMD-wiNN model, we use 300 training samples with randomly generated from the uniform distribution $[0, 2] \times [0, 10]$. The relative test errors for different truncation basis functions is plotted in Fig. 5.5. The error of test set is not as expected that decreases against the number r of truncation basis functions. In contrast, the test error increases when using more basis functions. For capturing the important properties, we select $r = 12$ basis functions to ensure the effectiveness and efficiency of the optimization problem. The snapshot space with DMD-wiNN method is generated by the first 25 time levels data obtained by FEM.

Three different surrogate models (DMD-wiNN, FEm and gPC) are used for the transient nonlinear diffusion problem. For comparison, the efficiency and accuracy of estimated parameters is shown in Table 4. The cost only using FEM model is more expensive than DMD-wiNN and gPC. Although the efficiency of EnKF based on gPC (N=2) is the best, the estimated accuracy of parameters and state is lower than FEM. As the higher order gPC model is applied for solving forward model, it improves the estimated accuracy of parameter but the ADM is close to that of gPC (N=2). In terms of RS_θ, the accuracy of the proposed EnKF based on DMD-wiNN model is much better than that of FEM and gPC. Compared with gPC, it slightly slower than gPC (N=4). As expected, the computation and the estimated accuracy of parameters for the proposed method outperform that of FEM. The ADM of the proposed method is better than those of gPC. Due to the underestimation, the ADM of only using FEM is better than that of other methods.

To further investigate the accuracy of estimated parameters, we plot the prior and pos-

Table 4: CPU time, RS_θ and ADM using three different surrogates.

	FEM	DMD-wiNN	gPC (N=2)	gPC (N=4)
CPU time (s)	4530.34	163.54	52.96	117.86
RMSE_θ	0.1704	0.0287	0.1725	0.1351
spread_θ	0.0287	0.0372	0.0033	0.0121
$\lvert\mathrm{RS}_\theta\rvert$	0.0282	0.00056	0.0298	0.0181
ADM	0.0149	0.0416	0.0530	0.0579

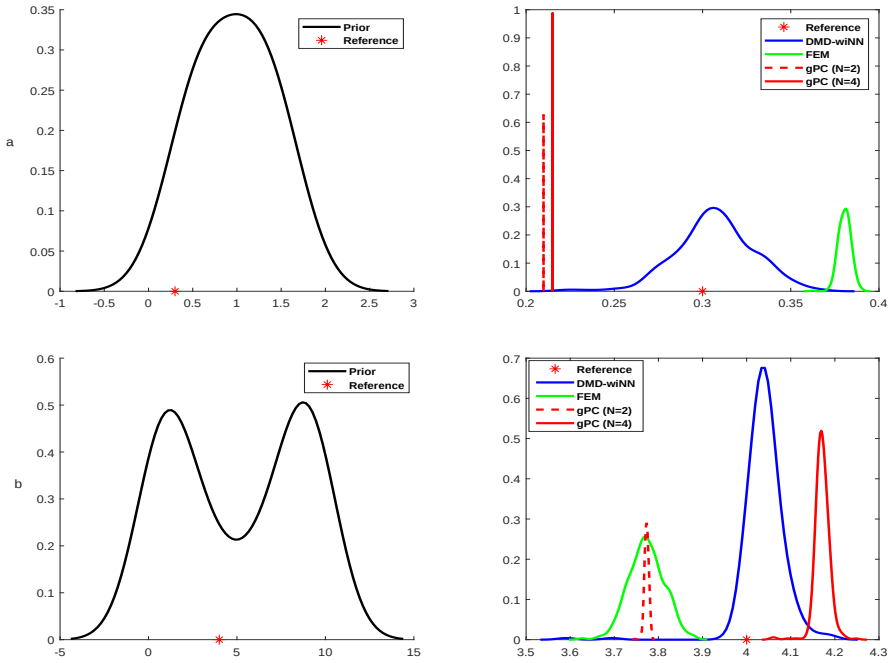

Figure 5.6: The prior and posterior distributions for parameters a (top row) and b (bottom row) using three different surrogate models (DMD-wiNN, FEM and gPC). Red star makers are the references.

terior marginal distributions using samples from DMD, FEM and gPC models in Fig. 5.6. The prior densities of a and b (Fig. 5.6 (left)) are the same for three different surrogate models. Due to the range interval of a narrower than b, the underestimation exists in EnKF using gPC models in Fig. 5.6 (upper right corner). Using FEM model slightly improve the underestimation of a, but the reference runs outside of the high probability region. The reference locates in the high probability regions constructed by samples using DMD-wiNN model. As a big range for b, the posterior density of using gPC becomes wider. For the prosetrior densities of b, the proposed method still outperforms FEM and gPC.

Figure. 5.7 shows the 95% credible and prediction intervals for model state at $u(0.5, 1; t)$ against time t. The samplers and realizations are constructed by EnKF based on the four surrogate methods (DMD-wiNN, FEM, 2 order and gPC (N=4)). We note that The 95% credible and prediction intervals from gPC are wider that from DMD-wiNN and FEM. However, measurement data and the references at most time instance are outside of the prediction using gPC (N=2) model. Using higher order basis function in gPC improves the accuracy of estimated data as shown in Fig. 5.7 (lower right corner). The measurement data is close to the prediction interval. The prediction intervals from DMD-wiNN and FEM are narrower than gPC. The proposed method slightly outperform FEM because measurement almost locate in the prediction except for the initial time instance. This is due to only a small part of measurement used for EnKF. As more data information gains in EnKF, the prediction interval becomes accuracy.

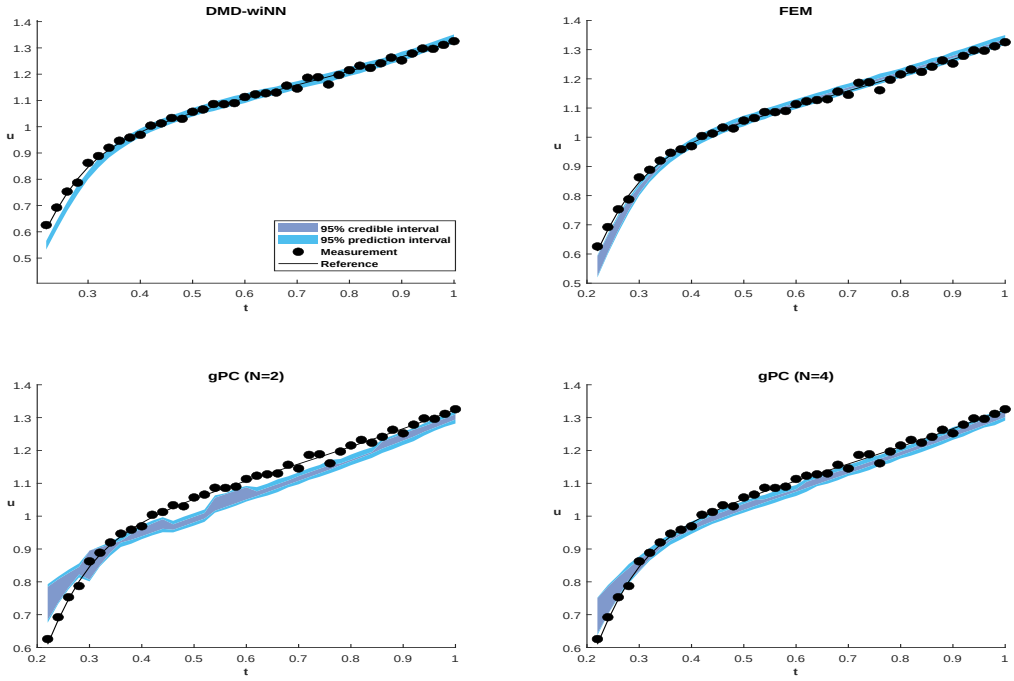

Figure 5.7: The 95% credible and prediction intervals of state $u(0.5, 1; t)$ against time t are generated from (DMD-wiNN, FEM and gPC). The black points are measurement data.

5.3. Recover the source locations and the nonlinear permeability field

In this subsection, we still consider the nonlinear advection diffusion model in Eq. (5.3) with the homogeneous Neumman boundary condition of Eq. (5.1). In this example, we prescribe an initial condition as shown in Fig. 5.8 (left) and the form is given by Eq. (5.4), where the standard deviations α_i and centers x_i, $i = 1, 2, \cdots, 5$, are given in Table 5. The form of the source term is similar to the initial condition, i.e.,

$$f(\mathbf{x}, t, \omega) = \sum_{i=1}^{2} \frac{1}{\beta_i \sqrt{2\pi}} \exp(-\frac{1}{2\beta_i^2} \|\mathbf{x} - \mathbf{x}_i(\omega)\|^2). \tag{5.7}$$

where $\beta_1 = 0.04$ and $\beta_2 = 0.07$. The source locations $\mathbf{x}_1 = (x_1, y_1)$ and $\mathbf{x}_2 = (x_2, y_2)$ are both unknown.

Figure 5.8 (right) shows the reference source. The upper left corner and lower right corner display the two reference locations for generating measurement. The reference values are set

Table 5: Standard deviations and center locations for the initial condition in Eq. (5.4).

i	1	2	3	4	5
α_i	0.07	0.05	0.07	0.05	0.05
x_i	0.2	0.25	0.35	0.45	0.55
y_i	0.15	0.15	0.2	0.2	0.12

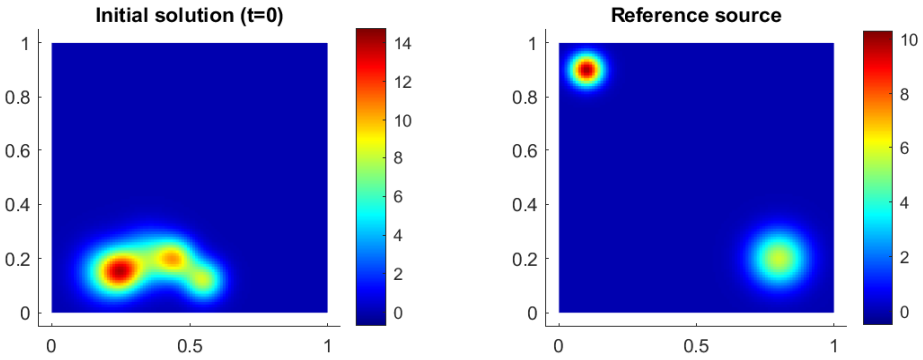

Figure 5.8: The initial condition (left) and reference source (right).

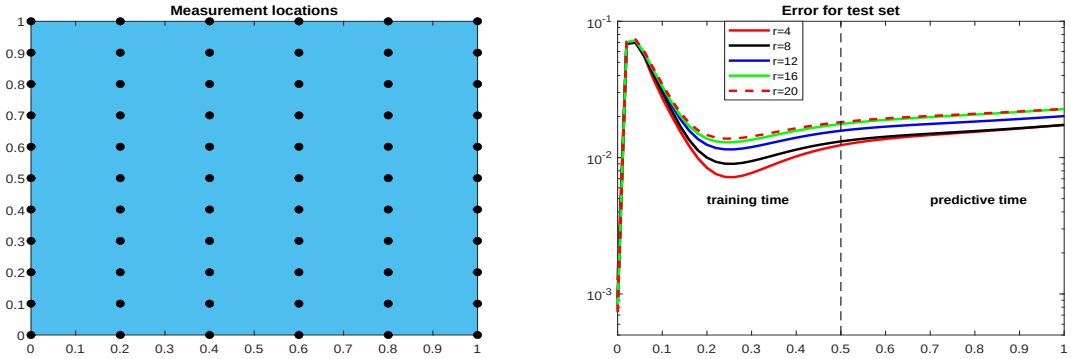

Figure 5.9: The relative errors for the test set against different truncation DMD basis functions (left) and measurement locations (right).

as $(0.1, 0.9)$ and $(0.8, 0.2)$. The advection velocity for this problem is $\mathbf{w} = 0.3\hat{i} + 0.1\hat{j}$. The diffusion coefficient κ is the nonlinear function of u given by Eq. (5.6). In this subsection, κ, $\mathbf{x}_1$ and $\mathbf{x}_2$ are all the unknowns, i.e., $\theta = (a, b, x_1, y_1, x_2, y_2)$.

For this example, the ground true parameters a and b are set as 0.3 and 4, respectively. The artificial time steps I for EnKF is set as 10. The measurement locations are distributed on the uniform 6×11 grid of the domain $[0, 1] \times [0, 1]$ as shown in Fig. 5.9 (left). Measurement data are generated from 100×100 grid with time step $\Delta t = 0.01$. For the inversion based on EnKF, ensemble size is set as 600. The forward model is defined on a uniform 100×100 fine grid with time step $\Delta t = 0.02$. 800 training samples with randomly generated from the uniform distribution $[0, 1] \times [0, 10] \times [0, 1]^4$ are used to build the DMD-wiNN model.

As the truncation basis functions of DMD-wiNN increase, the relative errors of the test set are plotted in Fig. 5.9 (right). We expect that the errors decreases as the basis functions increase. In fact, the error slightly increases when adding basis functions. To retain the main properties, we select $r = 12$ basis functions for the DMD-wiNN model. Although $r = 4$ shows a slightly smaller error, the predictive ability may decreases for the practical application. To construct the reduced order model, the snapshot space for the DMD-wiNN method is generated by the first 25 time levels data obtained by FEM.

We also investigate the efficiency and accuracy of estimated parameter and state using three different surrogate models (DMD-wiNN, FEM and gPC). The results is illustrated

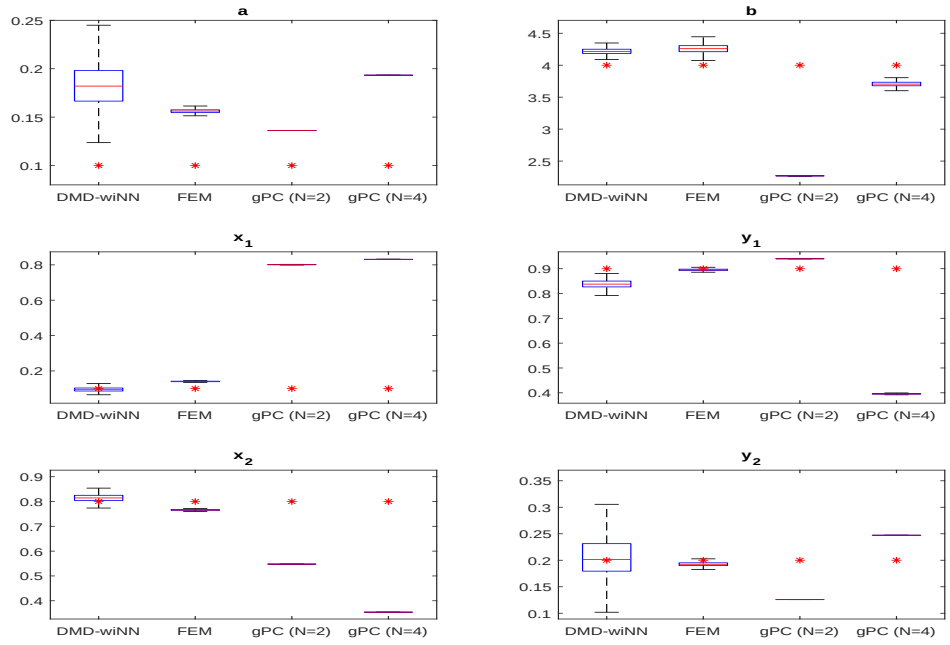

Figure 5.10: Box plot: Reference (red star), the edges of the box (unfilled rectangle) are the 25th (bottom) and 75th (top) percentiles, and 95% credible intervals (dash line) for each element of θ generated by different surrogate models (DMD-wiNN, FEM and gPC).

in Table 6. As the dimension of estimated parameter increases, the computation cost of three different surrogate models increases. This is due to the ensemble size become big for generating the accuracy estimation. The proposed method provides a more accuracy estimation both for parameter and state than that of using gPC model. Higher order gPC can improve the underestimation of estimated parameter in terms of spread. The RS_θ of using DMD-wiNN model is smaller than using FEM. It implies that DMD-wiNN model outperfroms other three surrogates for the estimated parameters. The ADM shows the accuracy of estimated state. Because no extra model error exists in FEM model, the ADM of using FEM is the smallest.

The 95% credible intervals for parameter θ by the final posterior samplers are plotted in Fig. 5.10. The X-axis of each subfigure denotes DMD-wiNN, FEM, 2 order and gPC (N=4) model and the Y-axis is the estimate value of each parameter. For EnKF using FEM and

Table 6: CPU time, RS_θ and ADM using three different surrogates.

	FEM	DMD-wiNN	gPC (N=2)	gPC (N=4)
CPU time (s)	9857.12	400.63	143.67	815.37
RMSE_θ	0.1099	0.0993	0.7704	0.4253
spread_θ	0.0286	0.0375	0.0013	0.0173
$\lvert\mathrm{RS}_\theta\rvert$	0.0113	0.0085	0.5935	0.1806
ADM	0.0147	0.0699	0.1077	0.1048

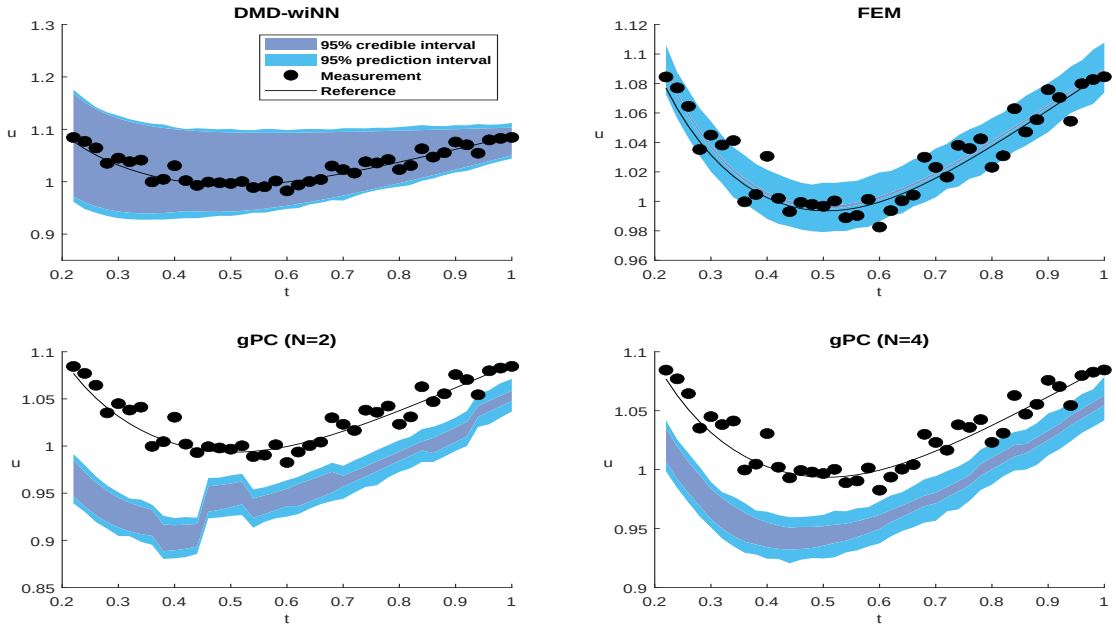

Figure 5.11: The 95% prediction and credible intervals of state $u(0.6, 0.1; t)$ with respect to t are generated from three different reduced model (DMD-wiNN, FEM and gPC). The black points are measurement data.

gPC models, the uncertainty is underestimated and leads to a narrow creadible interval. The references (red star makers) are outside the credible intervals and far from them when using gPC models. The credible intervals of x_1, x_2 and y_2 constructed by using DMD-wiNN model contain the references and other's close to the references. This implies that the proposed method can work well for the unknowns of nonlinear PDEs.

Figure 5.11 illustrates the 95% credible and prediction intervals for estimated state at $u(0.6, 0.1; t)$ with the final ensemble posterior samplers and realizations constructed by DMD-wiNN, FEM and gPC models. The black solid line is the reference values and black points are measurement data. Estimated states at different time instance are almost outside of the prediction intervals constructed by gPC models. The prediction interval is tight using FEM model and outperforms that of gPC. We note that the most reference states and measurements are both contained in this interval (upper right corner). As expected, the measurement data and the references are all within the credible interval from DMD-wiNN. Due to $t \in (0.5, 1]$ outside of the training region, it demonstrates the good predictive ability of DMD-wiNN model.

The uncertainty is from the difusion coeficient and source locations. So we also compute the correlations between these unknwons in Fig. 5.12 with samples generated from EnKF based on DMD-wiNN model. Figure 5.12 plots all of the one and two-dimensional posterior marginals of θ. The first parameter a of diffusion function and the Y-axis y_2 of second source location exist slightly positive correlation. Moreover, the negative correlations of a and y_1, a and x_2 are apparent. The X-axis x_1 and x_2 for the two source locations also exists negative correlation. The others appear uncorrelated and mutually independent based on the shape of their 2-D marginals in Fig. 5.12.

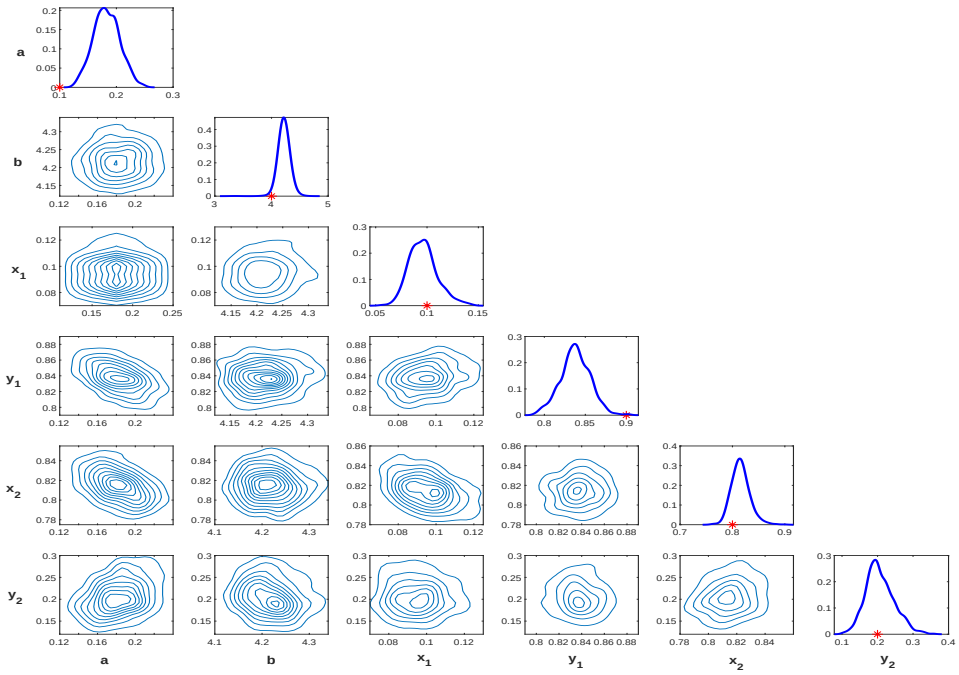

Figure 5.12: 1-D and 2-D posterior marginals of θ generated by EnKF using DMD-wiNN model.

6. Conclusions

In the paper, we develop a new low rank EnKF based on DMD-wiNN surrogate model for the parameterized time-dependent nonlinear PDEs. The proposed method can be decoupled into offline and online stages. In the offline stage, parameterized equations for each parameter in training set is first solved to generate the snapshots and get the data matrixes. Then the reduced operator matrixes is obtained by performing SVD. We adopt the wiNN to construct the efficient surrogate models of the reduced operator matrixes, which includes the reduced Koopman operator matix and SVD-modes. In the online stage, we perform EnKF updates to explore the posterior distributions of estimated parameters. For each sample, we just perform operations on low-dimensional matrixes to obtain the efficient surrogate model of the likelihood functions. For the analysis step in EnKF, we apply SVD for the ensemble deviation of simulated data. Then the second-order derivative of the forward model in the inversion process is generated by the low-rank approximation of ensemble samples. This can avoid the rank-deficient of Kalman gain when ensemble covariance is singular. For the applicability of the proposed EnKF to non-Gaussian models, we can integrate this method with normal-score transformation. The effectiveness and efficiency of the proposed method were validated by applications of Bayesian inversion in nonlinear dynamic systems.

The idea of DMD-wiNN model and low rank EnKF methods were combined together to solve the Bayesian inverse problems. For the three experiments, we only use the half of total time for training in the proposed method. Numerical results showed that the proposed method leads to a better predictive data than using FEM and gPC models. Future

works will focus on complex nonlinear dynamic systems with high-dimensional unknowns and extend the prective ability for different uncertainty quantification problems, such as data assimilation, among others.

Acknowledgments

This work was supported by National Key R&D Program of China (No.2021YFA1001300), National Natural Science Foundation of China (Nos.12271150, 12101216, 12171406), the Hunan Provincial Natural Science Foundation of China (No. 2022JJ40030), and the 2023 Guangzhou Basic and Applied Basic Research Project (No.2023A04J0035).

References

[1] H. Arbabi and I. Mezić, *Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator*, SIAM J. Appl. Dyn. Syst., 16 (2017), pp. 2096–2126.

[2] Y. Ba, L. Jiang, and N. Ou, *A two-stage ensemble Kalman filter based on multiscale model reduction for inverse problems in time fractional diffusion-wave equations*, J. Comput. Phys., 374 (2018), pp. 300–330.

[3] Y. Ba and L. Jiang, *A two-stage variable-separation Kalman filter for data assimilation*, J. Comput. Phys., 434 (2021), 110244.

[4] S. L. Brunton, J. L. Proctor, J. H. Tu, and J. N. Kutz, *Compressed sensing and dynamic mode decomposition*, J. Comput. Dyn., 2 (2016), pp. 165–191.

[5] P. Benner, S. Gugercin, and K. Willcox, *A survey of projection-based model reduction methods for parametric dynamical systems*, SIAM review, 57 (2015), pp. 483–531.

[6] M. Belkin, D. Hsu, and P. Mitra, *Overfitting or perfect fitting? risk bounds for classification and regression rules that interpolate*, Adv. Neural Inf. Process. Syst., 31 (2018), pp. 2300–2311.

[7] Y. Chen, J. S. Hesthaven, Y. Maday, et al., *Certified reduced basis methods and output bounds for the harmonic Maxwell's equations*, SIAM J. Sci. Comput., 32 (2010), pp. 970–996.

[8] P. Chen, A. Quarteroni, and G. Rozza, *A weighted reduced basis method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 51 (2013), pp. 3163–3185.

[9] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res.: Oceans, 99 (1994), pp. 10143–10162.

[10] G. Evensen, *The ensemble Kalman filter for combined state and parameter estimation*, IEEE Control Syst. Mag., 29 (2009), pp. 83–104.

[11] D. Gamerman and H. F. Lopes, *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, Chapman and Hall/CRC, 2006.

[12] Z. Gao, Y. Lin, X. Sun, and X. Zeng, *A reduced order method for nonlinear parameterized partial differential equations using dynamic mode decomposition coupled with k-nearest-neighbors regression*, J. Comput. Phys., 452 (2022), 110907.

[13] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified reduced basis methods for parametrized partial differential equations*, Berlin: Springer, 2016.

[14] Q. A. Huhn, M. E. Tano, J. C. Ragusa, et al., *Parametric dynamic mode decomposition for reduced-order modeling*, J. Comput. Phys., 475 (2023), 111852.

[15] L. Jiang and Q. Li, *Model's sparse representation based on reduced mixed GMsFE basis methods*, J. Comput. Phys., 338 (2017), pp. 285–312.

[16] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, *Sparsity-promoting dynamic mode decomposition*, Phys. Fluids, 26 (2014), 024103.

[17] R. E. Kalman, *A new approach to linear filtering and prediction problems* , J. Basic Eng., 82 (1960), pp. 35–45.

[18] J. N. Kutz, S. L. Brunton, B. W. Brunton, et al., *Dynamic mode decomposition: data-driven modeling of complex systems*, SIAM, 2016.

[19] M. Korda and I. Mezić, *On convergence of extended dynamic mode decomposition to the Koopman operator*, J. Nonlinear Sci., 28 (2018), pp. 687–710.

[20] J. Li and Y. M. Marzouk, *Adaptive construction of surrogates for the Bayesian solution of inverse problems*, SIAM J. Sci. Comput., 36 (2014), pp. A1163–A1186.

[21] J. Li and D. Xiu, *A generalized polynomial chaos based ensemble Kalman filter with high accuracy*, J. Comput. Phys., 228 (2009), pp. 5454–5469.

[22] Y. Lin, Z. Gao, Y. Chen, and X. Sun, *A dynamic mode decomposition based reduced-order model for parameterized time-dependent partial differential equations*, J. Sci. Comput., 95 (2023), 70.

[23] S. Mallat, *A wavelet tour of signal processing: the sparse way*, Academic press, 2008.

[24] J. Man, W. Li, L. Zeng, and L. Wu, *Data assimilation for unsaturated flow models with restart adaptive probabilistic collocation based Kalman filter*, Adv. Water Resour., 92 (2016), pp. 258–270.

[25] Y. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, Commun. Comput. Phys., 6 (2009), pp. 826–847.

[26] H. G. Matthies, *Stochastic finite elements: Computational approaches to stochastic partial differential equations*, ZAMM Z. Angew. Math. Mech., 88 (2008), pp. 849–873.

[27] E. Musharbash, F. Nobile, and T. Zhou, *Error analysis of the dynamically orthogonal approximation of time dependent random PDEs*, SIAM J. Sci. Comput., 37 (2015), pp. A776–A810.

[28] N. Ou, L. Jiang, and G. Lin, *A new bi-fidelity model reduction method for Bayesian inverse problems*, Internat. J. Numer. Methods Engrg., 119 (2019), pp. 941–963.

[29] O. Pajonk, B. V. Rosić, and H. G. Matthies, *Sampling-free linear Bayesian updating of model state and parameters using a square root approach*, Comput. Geosci., 55 (2013), pp. 70–83.

[30] J. L. Proctor, S. L. Brunton, and J. N. Kutz, *Generalizing Koopman theory to allow for inputs and control*, SIAM J. Appl. Dyn. Syst., 17 (2018), pp. 909–930.

[31] A. Quarteroni and G. Rozza, *Reduced order methods for modeling and computational reduction*, Berlin: Springer, 2014.

[32] B. V. Rosić, A. Litvinenko, O. Pajonk, and H. G. Matthies, *Sampling-free linear Bayesian update of polynomial chaos representations*, J. Comput. Phys., 231 (2012), pp. 5761–5787.

[33] G. Saad and R. Ghanem, *Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter*, Water Resour. Res., 45 (2009).

[34] A. Safdari-Vaighani, A. Heryudono and E. Larsson, *A radial basis function partition of unity collocation method for convection-diffusion equations arising in financial applications*, J. Sci. Comput., 64 (2015), pp. 341–367.

[35] T. Sayadi, P. J. Schmid, F. Richecoeur, et al., *Parametrized data-driven decomposition for bifurcation analysis, with application to thermo-acoustically unstable systems*, Phys. Fluids, 27 (2015), 037102.

[36] P. J. Schmid, *Dynamic mode decomposition of numerical and experimental data*, J. Fluid Mech., 656 (2010), pp. 5–28.

[37] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[38] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*, SIAM, 2005.

[39] J. H. Tu, *Dynamic mode decomposition: Theory and applications*, Princeton University, 2013.

[40] P. J. Van Leeuwen and G. Evensen, *Data assimilation and inverse methods in terms of a probabilistic formulation*, Mon. Weather Rev., 124 (1996), pp. 2898–2913.

[41] N. Wiener, *The homogeneous chaos*, Amer. J. Math., 60 (1938), pp. 897–936.

[42] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, *A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition*, J. Nonlinear Sci., 25 (2015), pp. 1307–1346.

[43] M. O. Williams,C. W. Rowley, and I. G. Kevrekidis, *A kernel-based approach to data-driven Koopman spectral analysis*, J. Comput. Dyn., 2 (2016), pp. 247–265.

[44] Y. Xing, Q. Song, and G. Cheng, *Benefit of interpolation in nearest neighbor algorithms*, SIAM J. Math. Data Sci., 4 (2022), pp. 935–956.

[45] D. Xiu and G. Karniadakis, *The Wiener-Askey polynomial chaos for the stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.

[46] L. Yan and L. Guo, *Stochastic collocation algorithms using l_1-minimization for Bayesian solution of inverse problems*, SIAM J. Sci. Comput., 37 (2015), pp. A1410-A1435.

[47] L. Yan and T. Zhou, *An adaptive multi-fidelity PC-based ensemble Kalman inversion for inverse problems*, Int. J. Uncertain. Quan., 9 (2019), pp. 205–220.

[48] L. Zeng, H. Chang, D. Zhang, et al., *A probabilistic collocation-based Kalman filter for history matching*, SPE Journal, 16 (2011), pp. 294–306.

[49] H. Zhou, J. J. Gomez-Hernandez, and H. J. H. Franssen, *An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering*, Adv. Water Resour., 34 (2011), pp. 844–864.