

Dissimilarity Analysis Based Multimode Modeling for Complex Distributed Parameter Systems

Zhi Wang, *Student Member, IEEE*, Han-Xiong Li, *Fellow, IEEE*

Abstract—For complex distributed parameter systems (DPSs) with strong nonlinearities and time-varying dynamics, the conventional spatiotemporal modeling methods become ill-suited since the elementary assumption that the process data follow a unimodal Gaussian distribution usually becomes invalid. In the paper, a multimode method is proposed for modeling of such systems. First, the original operating space is partitioned along the time dimension into several subspaces via modified dissimilarity analysis. Each subspace represents the local spatiotemporal characteristics of the original system. Second, the Karhunen-Loève decomposition (KLD) based spatiotemporal modeling approach is applied to approximate the local dynamics of each subspace. Finally, an ensemble model is obtained using the soft weighting sum of the local ones, where the corresponding weights are calculated by principal component regression. By properly decomposing the original space into several local parts, the ensemble model is capable of handling the strong nonlinearities and time-varying dynamics of the system. The validity and efficiency of the proposed method are verified on two representative applications: 1) a one-dimensional parabolic catalytic rod; 2) a two-dimensional curing thermal process. Experimental results show that the proposed method provides a superior performance regarding modeling accuracy compared to several baselines.

Index Terms—Dissimilarity analysis, distributed parameter systems, multimode modeling, principal component regression, subspace partitioning.

I. INTRODUCTION

MANY chemical or physical industrial processes (e.g., battery thermal process, fluid-flow process, and convection-diffusion-reaction process) belong to distributed parameter systems (DPSs), where the system's input/output may change along both the time and space dimensions [1–3]. Modeling and control of DPSs play a significant role in systems engineering, control theory, and industrial applications [4–6]. With recent advance in the sensor, actuator and numerical computing technologies, modeling and control of such spatiotemporal systems become feasible and practical in real applications [7].

The first-principle modeling methodology for known DPSs usually leads to the partial differential equations (PDEs) [8]. For real-time controller implementation and synthesis, the PDE is typically transformed to the low-order ordinary differential equation (ODE) to represent the dominant dynamics

of DPSs. Advanced model reduction methods are on the basis of the time-space separation framework where the process outputs are decomposed into a decoupled form consisting of several dominant spatial basis functions and temporal coefficients. Many analytical methods are applied for this transformation, such as the finite difference method (FDM) [9], the finite element method (FEM) [10], and the spectral method [11]. Nevertheless, the above model reduction approaches require knowing the precise PDE of the system in advance, which is generally not applicable in most real industrial processes. On the other hand, data-driven methods have the advantages of fewer requirements on the physical mechanism of the process, higher accuracy and better generality [12]. Karhunen-Loève decomposition (KLD), has been widely applied to derive the dominant features of DPSs with the lowest order of empirical basis functions from sampled snapshots [13].

The time-space separation methods based on KLD have achieved wide applications for industrial distributed processes [14]. Under the assumption that the process data belong to a linear space, KLD relies on the Euclidean distance as the metric to minimize, and cannot efficiently acquire the nonlinear degrees of freedom in complex systems. This intrinsic property leads to the deficiency that conventional modeling techniques are only adequate for limited types of processes whose variables are linear, unimodal, and Gaussian distributed. In reality, those restrictions can be easily violated. Snapshots generated by complex DPSs belong to manifolds for which the geodesics do not correspond in general to the Euclidean distance. Instead they have strong nonlinearity and time-varying dynamics subject to parameter variations [15], hyperbolic dynamics like traveling hump [16], changing domains [17], etc. A global reduced-order basis (ROB) often lacks robustness w.r.t. such characteristics that are part of the normal process behavior. Indeed, using only one global ROB for model reduction of nonlinear systems probably produces an expensive set of basis functions for a desired modeling accuracy, or a too small set that cannot span the different local characteristics of snapshots. Therefore, a single set of global basis functions usually provides an unsatisfactory modeling performance for approximating complex DPSs with strong nonlinearities and time-varying dynamics [18].

Snapshots associated with a nonlinear, dynamic, or parametric system could represent multiple regions of the state space. Different regions of the snapshots could correspond to divergent landscapes of the associated physical process, which can be approximated more precisely by a group of local ROB than by only one global basis. Several results were reported relevant to this concept. Dihlmann et al. [19] constructed and

This work was supported by the General Research Fund Project from Research Grant Council of Hong Kong SAR under Grant CityU: 11205615, and by the project from City University of Hong Kong 7005092. (Corresponding author: Han-Xiong Li.)

Z. Wang and H.-X. Li are with the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong (email: njuwangzhi@gmail.com, mehqli@cityu.edu.hk).

exploited ROB by using adaptive time domain partition. By using machine learning algorithms like K-means [18], global optimum search [20], or a heuristic procedure [21], the time domain is partitioned into multiple subspaces where snapshots exhibit obviously different characteristics, and a local ROB is computed for each subspace. In a similar spirit, the spatial domain is decomposed into several subspaces before the local ROBs are constructed independently [22, 23]. Additionally, subspace modeling concepts have been widely investigated for process monitoring [24], fault detection [25] and fault diagnosis [26], etc.

In this paper, we present an integral multimode method for modeling complex DPSs. To handle the strong nonlinearities and time-varying dynamics, a two-step solution is proposed: subspace decomposition via modified dissimilarity analysis and local model ensemble via principal component regression. In recent years, a series of successful theory researches and applications have demonstrated that the dissimilarity analysis method can quickly and effectively detect the change of correlations among process variables [27, 28]. We modify the conventional dissimilarity analysis method to improve its generalization ability on real-world datasets. The modified dissimilarity analysis is employed to partition the original operating space along the time dimension into several subspaces, which represent the local characteristics of the original system. Then, the local dynamics of each subspace is estimated using the spatiotemporal modeling approach, where the basis functions are extracted by KLD and the corresponding temporal coefficients are learned by extreme learning machine (ELM) due to its simplicity powerful approximation ability. Finally, an ensemble model is obtained using a soft weighting sum form of each local spatiotemporal model. To avoid multicollinearity, principal component regression (PCR) is utilized to calculate the weights of local models. By the mechanism of partitioning snapshots into subspaces with divergent local characteristics along the time dimension, the modeling framework can better approximate strong nonlinearities and time-varying dynamics of complex DPSs. With the multimode strategy, the modeling accuracy can be improved, and a larger operating range can be handled for many practical systems.

To verify the proposed method, experiments are carried out on two representative complex DPSs: a simulated infinite-dimensional parabolic catalytic rod with strong nonlinearities and a real curing thermal process with a large-scale operating domain. The proposed multimode modeling method is first compared to the conventional global modeling to demonstrate the advantage of subspace partitioning. Further, the proposed method is compared to a *naïve* subspace modeling method with blind partitioning to show the effectiveness of modified dissimilarity analysis. Experimental results show that the proposed method provides a superior performance in terms of modeling accuracy over the baselines.

The rest of the paper is organized as follows. Section II presents the preliminaries including the conventional spatiotemporal modeling and dissimilarity analysis. In Section III, the framework of the proposed method is presented, followed by the implementation details of subspace decomposition and local model ensemble. Experiments on a one-dimensional

parabolic catalytic rod and a two-dimensional curing thermal process are conducted in Section IV. Section V presents concluding remarks.

II. PRELIMINARIES

A. Spatiotemporal Modeling

In the paper, we consider a general type of parabolic DPSs with a state-space description of the following PDE:

$$\frac{\partial y(x, t)}{\partial t} = \mathcal{G} \left(y, \frac{\partial y}{\partial x}, \dots, \frac{\partial^d y}{\partial x^d} \right) + \mathbf{b}^T(x) \mathbf{u}(t), \quad (1)$$

subject to the mixed-type boundary conditions

$$\mathcal{Q} \left(y, \frac{\partial y}{\partial x}, \dots, \frac{\partial^{d-1} y}{\partial x^{d-1}} \right) \Big|_{x=x_a \text{ or } x=x_b} = 0, \quad (2)$$

and the initial condition

$$y(x, 0) = y_0(x), \quad (3)$$

where $y(x, t) \in \mathbb{R}^n$ is the state variable, $x \in [x_a, x_b] \subset \mathbb{R}$ and $t \in [0, \infty)$ are the position and time, respectively. $\mathbf{u}(t) \in \mathbb{R}^p$ denotes the temporal input vector, $\mathcal{G} \in \mathbb{R}^n$ is a complex vector function that contains a nonlinear spatial differential operator of an even order d with respect to x , $\mathbf{b}(x)$ denotes a matrix function with proper dimensions describing the distribution of inputs with respect to the space domain, \mathcal{Q} denotes a nonlinear vector function, and $y_0(x)$ denotes a smooth vector function regarding the initial output. The time-space separation framework [1] is widely investigated for modeling such unknown nonlinear DPSs. A set of orthogonal spatial basis functions, $\{\varphi_i(x)\}_{i=1}^k$, is first extracted using KLD to capture the most dominant dynamics of the distributed process. Then, a reduced-order model, $\{a_i(t)\}_{i=1}^k$, is identified in the low-dimensional temporal space using extreme learning machine [29]. Finally, combining the decoupled basis functions and the low-order temporal model can reproduce the spatiotemporal output over the entire time and space domain. More details about spatiotemporal modeling can be found in [1].

B. Dissimilarity Index

Dissimilarity analysis is a data-driven classification method for evaluating the difference between the distributions of two separate datasets [27, 28]. By utilizing eigenvalues of the covariance matrices, it explores the correlation structures between data distributions during different periods, and further detects the probable drift of operating situations.

Considering the following two data sets of snapshots: \mathbf{Y}_1 and \mathbf{Y}_2 , each column of which is an n -dimensional vector of spatial measurements from (1). $\mathbf{Y}_i, (i = 1, 2)$ contains l_i columns of snapshots. The covariance matrices are computed as $\mathbf{\Xi}_i = \mathbf{Y}_i \mathbf{Y}_i^T / l_i$. The covariance matrix of the mixture data sets is given by

$$\mathbf{\Xi} = \frac{l_1}{l_1 + l_2} \mathbf{\Xi}_1 + \frac{l_2}{l_1 + l_2} \mathbf{\Xi}_2. \quad (4)$$

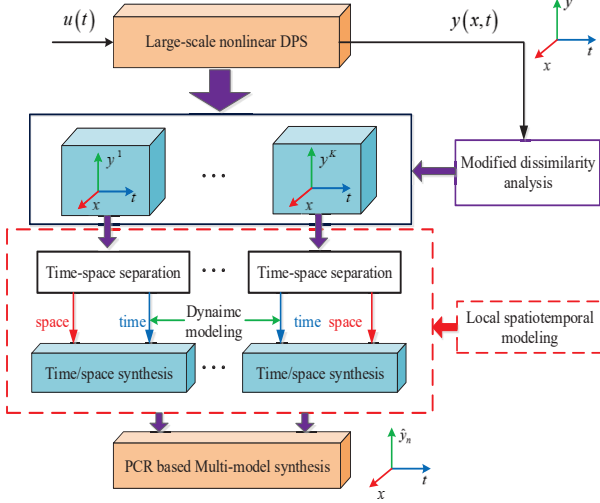


Fig. 1. The methodology framework of multimode modeling.

By using an orthogonal matrix P_0 , Ξ is diagonalized as $P_0^T \Xi P_0 = \Lambda$, then, Y_i is transformed into Z_i as

$$Z_i = \sqrt{\frac{l_i}{l_1 + l_2}} \Lambda^{-\frac{1}{2}} P_0^T Y_i = \sqrt{\frac{l_i}{l_1 + l_2}} P^T Y_i, \quad (5)$$

where $P = P_0 \Lambda^{-\frac{1}{2}}$ is the transformation matrix.

The covariance matrices of the transformed data matrices

$$\Sigma_i = \frac{1}{l_i} Z_i Z_i^T = \frac{l_i}{l_1 + l_2} P^T \Xi_i P \quad (6)$$

satisfy the property of $\Sigma_1 + \Sigma_2 = I$. The two covariance matrices can be decomposed into a set of eigenvalues and eigenvectors as

$$\Sigma_i \xi_i^j = \lambda_i^j \xi_i^j, \quad (7)$$

where λ_i^j and ξ_i^j are the eigenvalues and the corresponding eigenvectors, and the superscript j denotes the j th eigenvalue or eigenvector. The following relationships can be derived

$$\Sigma_2 \xi_i^j = (1 - \lambda_1^j) \xi_1^j, \quad 1 - \lambda_1^j = \lambda_2^j. \quad (8)$$

These two correlation equations indicate that the transformed data matrices, Z_1 and Z_2 , have the same group of principal components that are ordered in reverse.

Based on the above observation, the metric $\text{dis}(\cdot)$ is finally derived for measuring the dissimilarity between snapshots Y_1 and Y_2 as

$$\text{dis}(Y_1, Y_2) = \frac{4}{n} \sum_{j=1}^n (\lambda_j - 0.5)^2. \quad (9)$$

III. MULTIMODE MODELING

A. Framework

As illustrated in Fig. 1, the primary framework of the proposed multimode modeling methodology can be summarized as follows:

- 1) Modified dissimilarity analysis is first applied to partition the original space along the time dimension into several

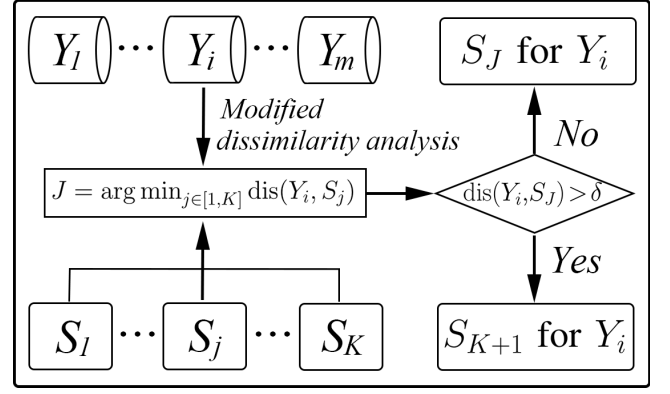


Fig. 2. Subspace decomposition via modified dissimilarity analysis.

subspaces. Each subspace represents a kind of local characteristics.

- 2) The local dynamics of each subspace is estimated using the time-space separation based modeling method.
- 3) The ensemble model is obtained using a soft weighting sum form of the local ones in the subspaces, where the corresponding weights are calculated by principal component regression.

It can be observed that, the difference between the proposed method and the conventional global modeling lies in a two-step procedure: subspace decomposition and local model ensemble. The snapshots are partitioned into multiple subspaces, where local models are estimated to capture the divergent local characteristics. With this mechanism, the modeling framework can better approximate the strong nonlinearities and time-varying dynamics, and can handle a larger operating range for complex DPSs. Moreover, since each subspace contains only a fraction of the snapshots, the computational complexity for calculating the BFs is reduced by solving a set of smaller eigenvalue problems.

B. Subspace Decomposition

For simplicity, mark $y_j = y(x, t_j)$ as the n -dimensional spatial measurement at time step t_j . Fig. 2 illustrates the idea of subspace decomposition via modified dissimilarity analysis. First, the l snapshots are collected into m small batches (Y_1, Y_2, \dots, Y_m) with the same window size of T , and the initial batch of snapshots Y_1 belongs to the first nominal subspace S_1 . Assume that there are K existing subspaces before the i th batch of snapshots $(Y_i, i=2, \dots, m)$ is available. Then, the modified dissimilarity analysis is applied for assigning the subspace which the new batch Y_i belongs to. Specifically, the dissimilarity indices, $\text{dis}(Y_i, S_j), j=1, \dots, K$, between the new batch Y_i and all existing subspaces are computed to find out the J th subspace that derives a minimal dissimilarity to the new batch. That is to say, the new batch most likely belongs to subspace S_J . Next, the minimal dissimilarity $\text{dis}(Y_i, S_J)$ is compared with a predefined threshold δ to further determine whether the new batch belongs to subspace S_J . If $\text{dis}(Y_i, S_J) > \delta$, it indicates that the new batch is sufficiently different from subspace S_J and any other existing subspace. Consequently, a new subspace S_{K+1} should

Algorithm 1: Subspace decomposition via modified dissimilarity analysis

Input: Batch size T ; number of batches m ;
dissimilarity threshold δ
Output: K subspaces (S_1, \dots, S_K)

- 1 Divide the l snapshots into m equal batches as $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$
- 2 $\mathbf{Y}_1 \leftarrow$ first batch of snapshots, $S_1 \leftarrow$ first subspace
- 3 Initialize the number of existing subspaces: $K \leftarrow 1$
- 4 Initialize the index of the pending batch: $i \leftarrow 2$
- 5 **while** $i \leq m$ **do**
- 6 $\mathbf{Y}_i \leftarrow$ snapshots during $t \in [(i-1)*T, i*T]$
- 7 $J \leftarrow \arg \min_{j \in [1, K]} \text{dis}(\mathbf{Y}_i, S_j)$
- 8 **if** $\text{dis}(\mathbf{Y}_i, S_J) > \delta$ **then**
- 9 Generate a new subspace S_{K+1}
- 10 Assign \mathbf{Y}_i to S_{K+1}
- 11 $K \leftarrow K + 1$
- 12 **else**
- 13 Assign \mathbf{Y}_i to S_J
- 14 **end**
- 15 $i \leftarrow i + 1$
- 16 **end**

be generated that the new batch belongs to. If $\text{dis}(\mathbf{Y}_i, S_J) < \delta$, then the new batch \mathbf{Y}_i is confirmed to belong to subspace S_J . Finally, all the snapshots are decomposed into K subspaces where local models are constructed to approximate the local characteristics. Together, the implementation of the proposed subspace decomposition is presented in Algorithm 1.

In Section II of the conventional dissimilarity analysis, the original data matrices \mathbf{Y}_i are transformed into \mathbf{Z}_i by the transformation matrix \mathbf{P} , where $\mathbf{P} = \mathbf{P}_0 \mathbf{\Lambda}^{-\frac{1}{2}}$. It assumes that the overall covariance matrix $\mathbf{\Xi}$ is full-rank and hence the eigenvalue matrix $\mathbf{\Lambda}$ is invertible. However, the output snapshots of distributed processes may be highly correlated, or the collected snapshots may not be representative on all the n dimensions of spatial measurements. These factors in practical applications could result that the covariance matrix $\mathbf{\Xi}$ is not full-rank (i.e., $\text{rank}(\mathbf{\Xi}) < n$), and hence $\mathbf{\Lambda}$ is not invertible. In response to this limitation, we modify the conventional dissimilarity analysis to improve its generalization ability on datasets where the covariance matrix is not full-rank.

The function of transformation matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ is to map a data set from the original n -dimensional space to a new n -dimensional one. Let r denote the rank of covariance matrix $\mathbf{\Xi}$. When $r < n$, the original data contains $n-r$ redundant dimensions and it intrinsically can be mapped into an r -dimensional space where each dimension is independent of others. Based on this insight of data statistics, we redefine the transformation matrix \mathbf{P} into a lower-dimensional form $\mathbf{P}_r \in \mathbb{R}^{n \times r}$ as $\mathbf{P}_r = \mathbf{P}_{0r} \mathbf{\Lambda}_r^{-\frac{1}{2}}$, where \mathbf{P}_{0r} is formed by the first r columns of \mathbf{P}_0 , and $\mathbf{\Lambda}_r$ denotes the r th leading principal submatrix of $\mathbf{\Lambda}$.

Through the reduced-order projection by the new transformation matrix \mathbf{P}_r , the n -dimensional snapshots \mathbf{Y}_i is mapped

to an r -dimensional data set \mathbf{Z}_i as

$$\mathbf{Z}_i = \sqrt{\frac{l_i}{l_1 + l_2}} \mathbf{P}_r^T \mathbf{Y}_i. \quad (10)$$

The covariance matrices of \mathbf{Z}_i becomes

$$\mathbf{\Sigma}_i = \frac{l_i}{l_1 + l_2} \mathbf{P}_r^T \mathbf{\Xi}_i \mathbf{P}_r. \quad (11)$$

Add the two covariance matrices together we can obtain

$$\begin{aligned} \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 &= \mathbf{P}_r^T \mathbf{\Xi} \mathbf{P}_r = \mathbf{\Lambda}_r^{-\frac{1}{2}} \mathbf{P}_{0r}^T \mathbf{P}_0 \mathbf{\Lambda} \mathbf{P}_0^T \mathbf{P}_{0r} \mathbf{\Lambda}_r^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}_r^{-\frac{1}{2}} \mathbf{P}_{0r}^T [\mathbf{P}_{0r} \ \mathbf{P}_{0(n-r)}] \begin{bmatrix} \mathbf{\Lambda}_r & \\ & \mathbf{0}_{(n-r)} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{0r}^T \\ \mathbf{P}_{0(n-r)}^T \end{bmatrix} \mathbf{\Lambda}_r^{-\frac{1}{2}} \\ &= \mathbf{I}_{r \times r}. \end{aligned} \quad (12)$$

It can be confirmed that the lower-order transformed datasets ($\mathbf{Z}_1, \mathbf{Z}_2$) also share the same group of principal components that are ordered in reverse. Finally, the modified dissimilarity index dis is defined as

$$\text{dis}(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{4}{r} \sum_{j=1}^r (\lambda_j - 0.5)^2, \quad (13)$$

where λ_j is the eigenvalues associated with the covariance matrix of the transformed dataset in the r -dimensional space.

Remark 1. The number of subspaces depends on the pre-defined “dissimilarity threshold δ ”, a fixed hyperparameter that controls the instantiation of new subspaces. Generally, a smaller threshold leads to more subspaces, which is supposed to better approximate the original system since more types of local characteristics can be separated and discovered. On the other hand, the emergence of more subspaces will increase the computational complexity of the system due to involving more spatiotemporal modeling procedures. In practice, the dissimilarity threshold can be adjusted according to the process requirements, which can be considered as a trade-off between modeling accuracy and computational efficiency.

C. Local Model Ensemble

After subspace decomposition, each subspace is estimated using the time-space separation modeling method [1], where the local spatial basis functions are learned by KLD and the corresponding time coefficients are learned by extreme learning machine (ELM) [29]. As shown in Fig. 3, once all the local spatiotemporal models are constructed, the global spatiotemporal model can be aggregated using soft weighting sum form as the following:

$$\hat{y}_k(x_i, t) = w_{i,1} \hat{y}_k^1(x_i, t) + w_{i,2} \hat{y}_k^2(x_i, t) + \dots + w_{i,K} \hat{y}_k^K(x_i, t). \quad (14)$$

where the $w_{i,j}$, ($j=1, \dots, K$) denotes the weight of j -th local spatiotemporal model $\hat{y}_k^j(x, t)$ on sensor i . The popularly used method to estimate the weight $w_{i,j}$ is the least square method with the solution:

$$\mathbf{W}_i = (\mathbf{H}_i^T \mathbf{H}_i)^{-1} \mathbf{H}_i^T \mathbf{O}_i, \quad i = 1, \dots, n. \quad (15)$$

where $\mathbf{W}_i = [w_{i,1}, w_{i,2}, \dots, w_{i,K}]^T$ represents the weight vector on sensor i , $\mathbf{H}_i = [\mathbf{O}_i^{1T}, \mathbf{O}_i^{2T}, \dots, \mathbf{O}_i^{KT}]$ is the local spatiotemporal model output matrix on sensor i , $\mathbf{O}_i^j =$

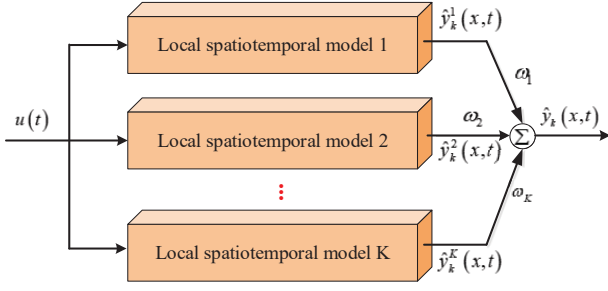


Fig. 3. Ensemble of local spatiotemporal models.

$[\hat{y}_k^j(x_i, t_1), \hat{y}_k^j(x_i, t_2), \dots, \hat{y}_k^j(x_i, t_{n_{tr}})]$ represents the output vector of j -th local spatiotemporal model on sensor i , n_{tr} denotes the number of training samples, and $\mathbf{O}_i = [y(x_i, t_1), y(x_i, t_2), \dots, y(x_i, t_{n_{tr}})]^T$ represents the actual output vector on sensor i .

The downside of this least square method is that each local spatiotemporal model represents the same system, and they are heavy correlational with each other. Therefore, the weight calculated using (15) often has the characteristic of multicollinearity, which will decrease the model accuracy. To address this problem, principal component regression can be applied here to derive a robust weight. For simplicity, let $\mathbf{H}_i = \bar{\mathbf{H}}$, $\mathbf{O}_i = \bar{\mathbf{O}}$, $\mathbf{W}_i = \bar{\mathbf{W}}$. With PCR, $\bar{\mathbf{H}}$ can be described as the following:

$$\bar{\mathbf{H}} = \mathbf{c}_1 \mathbf{d}_1^T + \mathbf{c}_2 \mathbf{d}_2^T + \dots + \mathbf{c}_K \mathbf{d}_K^T, \quad (16)$$

where $\mathbf{c}_j = \sigma_j \boldsymbol{\mu}_j$ and $\mathbf{d}_j = \mathbf{v}_j$ represent the principal component and the load component of the j -th pivot element, respectively. They both are a unit orthogonal vector. In general, the first q term of (16) can represent the dominant dynamic, where the determination of q is as same as the KLD. Therefore, (16) can be rewritten as:

$$\bar{\mathbf{H}} \doteq \bar{\mathbf{H}}_q = \mathbf{C} \mathbf{D}^T, \quad (17)$$

where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_q]$, $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_q]$. Then, (14) can be described as the following:

$$\bar{\mathbf{O}} = \bar{\mathbf{H}} \bar{\mathbf{W}} = \mathbf{C} \mathbf{D}^T \bar{\mathbf{W}}. \quad (18)$$

Suppose that $\bar{\mathbf{W}}_q = \mathbf{D}^T \bar{\mathbf{W}}$, then the least square solution of $\bar{\mathbf{W}}_q$ can be calculated as:

$$\bar{\mathbf{W}}_q = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \bar{\mathbf{O}}. \quad (19)$$

Since \mathbf{D} is an orthogonal vector, $\mathbf{D}^T = \mathbf{D}^{-1}$, so the weight $\bar{\mathbf{W}}$ can be estimated as:

$$\bar{\mathbf{W}} = \mathbf{D} \bar{\mathbf{W}}_q = \mathbf{D} (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \bar{\mathbf{O}}. \quad (20)$$

IV. SIMULATION EXPERIMENTS

To comprehensively evaluate the validity and efficiency of the multimode modeling method, two complex DPSs are investigated. One is a one-dimensional parabolic catalytic rod system with strong nonlinearities, and the other is a two-dimensional curing thermal process with a large-scale operating domain. The spatiotemporal snapshots are collected

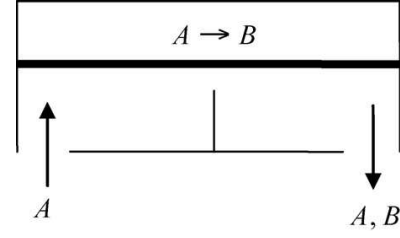


Fig. 4. The classical catalytic rod with the reaction in the form of $A \rightarrow B$.

from the simulation by using the finite difference approach and a real-time experiment, respectively.

To show the advantages of our method, we compare it to two conventional modeling approaches. One is the conventional global modeling that constructs one single spatiotemporal model for all the snapshots. The other one, denoted as “blind subspace modeling”, blindly partitions the snapshots into different subspaces and also construct a local spatiotemporal model for each subspace. Let $y(x, t)$ and $y_k(x, t)$ be the measured and the predicted system’s output, respectively. Four performance metrics are set up for modeling evaluation as follows:

- 1) Spatiotemporal error: $e(x, t) = y(x, t) - y_k(x, t)$,
- 2) Spatial normalized absolute error: $\text{SNAE}(t) = \frac{1}{n} \sum_{i=1}^n |e(x_i, t)|$,
- 3) Temporal normalized absolute error: $\text{TNAE}(x) = \frac{1}{l} \sum_{t=1}^l |e(x, t)|$.
- 4) Root of mean squared error: $\text{RMSE} = (\int \sum e(x, t)^2 dx / \int dx \sum \Delta t)^{1/2}$.

A. Case 1: Parabolic Catalytic Rod

Fig. 4 illustrates the classical transport-reaction process consisting of a long and thin rod within a reactor in the chemical industry [30]. The rod reactor is fed with pure species A , and a zeroth order exothermic catalytic reaction in the form of $A \rightarrow B$ takes place within the rod. Meantime, a cooling medium contacting the reactor is employed to cool the exothermic reaction procedure. The mathematical formulation describing the internal temperature evolution in the rod leads to the typical parabolic PDE [30]:

$$\frac{\partial y(x, t)}{\partial t} = \frac{\partial^2 y(x, t)}{\partial x^2} + \beta_T (e^{-\frac{\gamma}{1+y}} - e^{-\gamma}) + \beta_u (\mathbf{b}^T(x) \mathbf{u}(t) - y(x, t)), \quad (21)$$

subject to the Dirichlet boundary and initial conditions

$$y(0, t) = 0, \quad y(\pi, t) = 0, \quad y(x, 0) = y_0(x),$$

where $y(x, t)$ is the temperature within the rod, $\mathbf{u}(t)$ is the manipulated input or the temperature of the cooling medium, and $\mathbf{b}(x)$ is the distribution of actuators in the spatial domain. The process parameters are set as: heat of reaction $\beta_T = 50$, heat transfer coefficient $\beta_u = 2$, and activation energy $\gamma = 4$.

We employed four actuators $\mathbf{u}(t) = [u_1(t), \dots, u_4(t)]^T$ for persistently stimulating the full spectrum of the dynamics of the distributed process. These temporal inputs are set as $u_i(t) = 1.1 + (4 + 2\text{rand}) \exp(-i/5) \sin(t + 2\text{rand}) - 0.4 \exp(-i/20) \sin(5t + 2\text{rand})$, ($i = 1, \dots, 4$), where rand

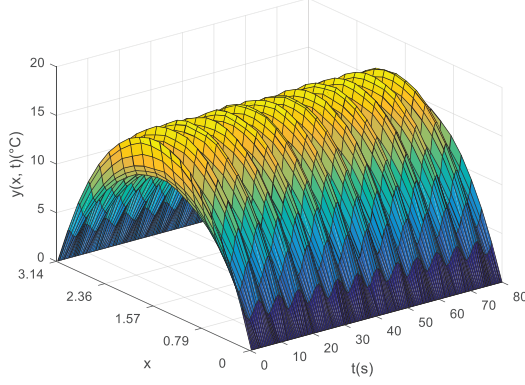


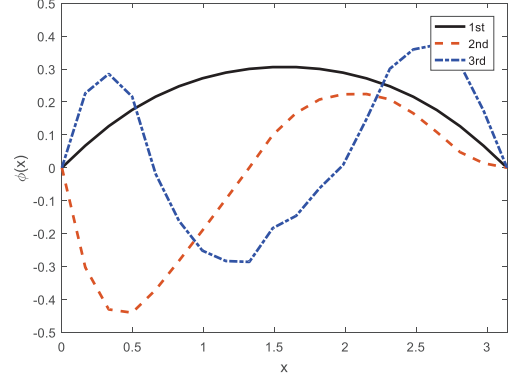
Fig. 5. Case 1: Measured output for multi-modeling.

is a uniform random distribution on the range of $[0, 1]$. The actuators are located in the spatial domain according to the distribution function $\mathbf{b}(x) = [b_1(x), \dots, b_4(x)]^T$, where $b_i(x) = H(x - (i-1)\pi/4) - H(x - i\pi/4)$, ($i = 1, \dots, 4$) and $H(\cdot)$ is the standard Heaviside function. In addition, more types of system's dynamics could be excited by setting more random phases or amplitudes.

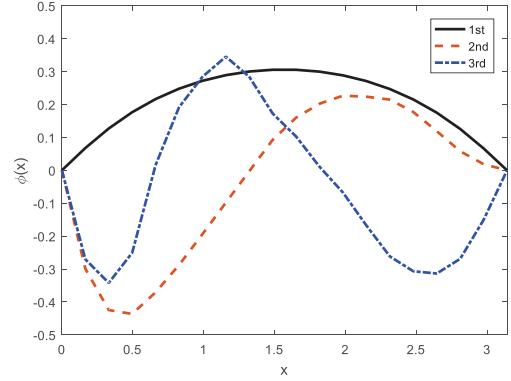
The number of distributed sensors needed for spatiotemporal modeling relies on the goal modeling accuracy and the physical complexity of the specific system. In real industrial processes, the parameters of the system are easy to change due to diverse time-varying dynamics such as catalyst deactivation, preventive maintenance, seasonal variations, sensor and process drifting, unknown disturbances, etc [27, 31]. The initial condition $y_0(x)$ is prescribed to stay steady with the temporal input $u_i(t) = 1.1$, ($i = 1, \dots, 4$). 20 sensors that are uniformly distributed in the spatial domain are employed for measuring the system's output ($n = 20$), and the sampling time interval is $\Delta t = 0.01$. The white noise with mean zero and standard deviation $\sigma(x_i) = A_d(x_i)n_d$, where $A_d(x_i) = (\max(y(x_i, t)) - \min(y(x_i, t)))/3$, ($i = 1, \dots, N$) and $n_d = 1\%$, is mixed into the noise-free data for simulating a noisy output as shown in Fig. 5. Among them, the first 4000 snapshots are used for training the model, while the last 4000 snapshots are used for testing.

Operating space division is a critical procedure to ensure the precision of the final multimode model. In the first step, modified dissimilarity analysis is applied here and the operating space can be divided from a coarse scale to a fine one progressively as the number of clusters increases. In our case, two clusters are obtained, where each cluster represents a subspace spatiotemporal dynamics. Within each subspace, the local spatiotemporal model is estimated corresponding to local snapshots. KLD is first applied for deriving the spatial basis functions. The first three dominant basis functions which are capable of capturing more than 99% of the dominant dynamics of the system are illustrated in Fig. 6. It can be observed that the two subspaces are a little different, which is reflected by the divergence of their third modes.

Based on the reduced-order subspace of the distributed



(a) Subspace S_1



(b) Subspace S_2

Fig. 6. Case 1: Three dominant spatial BFs in two subspaces of the proposed method.

process, each corresponding temporal model in the reduced-order space is approximated using ELM w.r.t. the corresponding input signals. The local spatiotemporal synthesis is reproduced by combining the derived basis functions and the estimated temporal model. In the second step, the ensemble model is obtained by aggregating the two local spatiotemporal models using (14). So far, the training process of the proposed multimode modeling method is completely finished. To test the model performance, the 4000 testing input signals are used for exciting the obtained multimode model. Fig. 7 shows the modeling performance of the proposed method regarding the predicted output and the corresponding error distribution, respectively. It can be observed that the multimode modeling method obtains a well-performing modeling accuracy in the classical catalytic rod system.

To further verify the model performance, the traditional global modeling and blind subspace modeling methods are adopted for comparison under the same configuration. These two methods are simulated at the same conditions with the proposed method. Fig. 8 presents the modeling error along the time direction (SNAE) and the space direction (TNAE), respectively. It can be observed that the proposed method improves modeling accuracy by providing smaller approximation errors than the other two methods. Moreover, Fig. 9

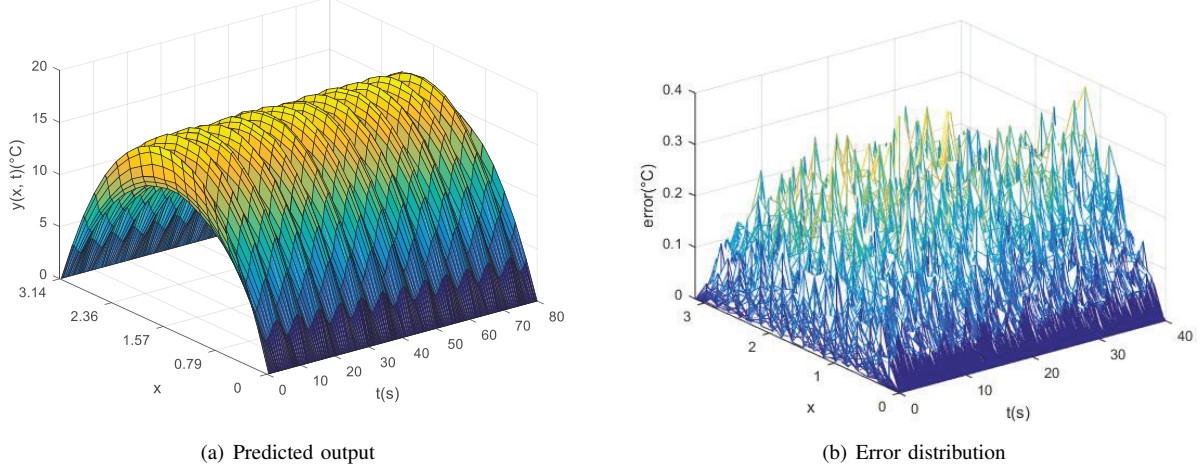


Fig. 7. Case 1: Modeling performance of the proposed method: (a) predicted output; (b) error distribution, using testing data.

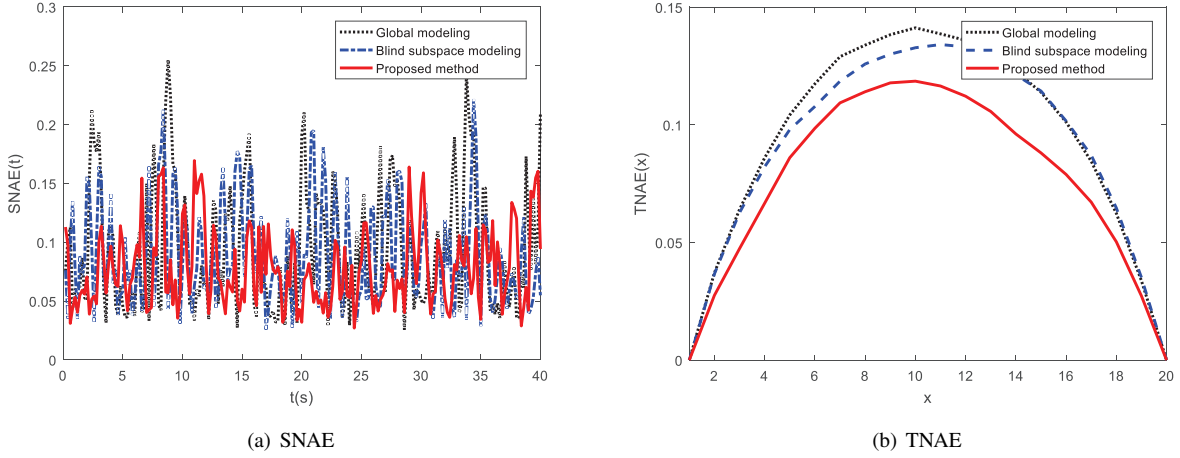


Fig. 8. Case 1: Modeling error along the time and space directions: (a) SNAE; (b) TNAE.

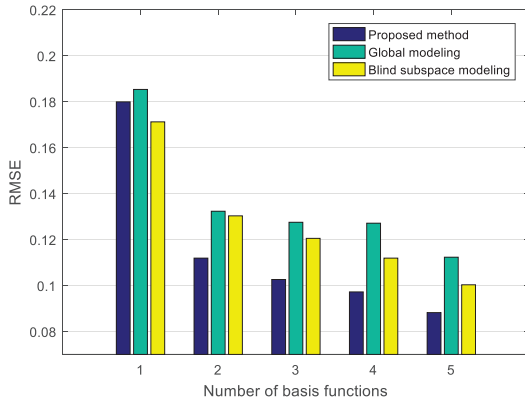


Fig. 9. Case 1: Modeling accuracy comparison under different numbers of basis functions.

illustrates the modeling RMSE regarding different numbers of dominant basis functions. Obviously, the proposed method

can achieve a smaller RMSE than the baselines when different sets of basis functions are used. The spatiotemporal model constructed using three basis functions from the proposed method provides a more precise approximation to the real system than the model reproduced using four or five basis functions from the other two methods.

B. Case 2: Curing oven

The curing oven is a typical DPS with a large-scale operating domain, which is widely utilized for the packaging process of semiconductor backend [13]. The thermal curing process also has many time-varying factors and unknown dynamics, including time-varying boundaries and parameters, which can introduce both temporal and spatial variability into its nominal situations. These aforementioned factors can impose great challenges to modeling the curing process [32]. The configuration of the curing oven is depicted in Fig. 10. There are four heaters placed at the top of the curing oven. First, by using epoxy, we are able to adhere the chip to the die pad or the lead frame cavity. Next, we employ a specific

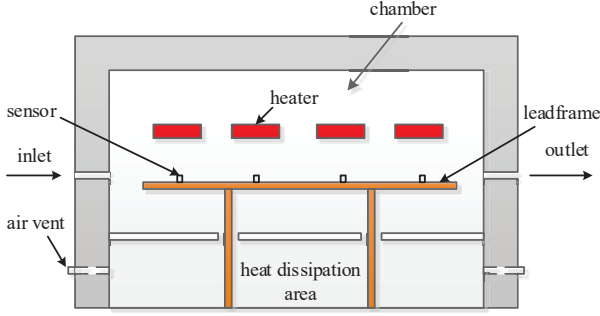


Fig. 10. Case 2: Schematic diagram of the curing oven system.

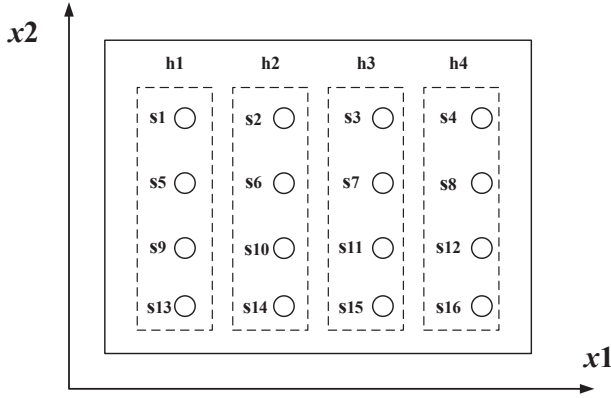


Fig. 11. Case 2: Sensors placement.

temperature for curing the epoxy in the chamber. Keeping the oven temperature uniformly distributed plays a significant role in assuring a high quality of the real curing system. Therefore, it is very important to acquire a precise spatiotemporal model of the curing oven. In our experiment, sixteen sensors are uniformly located on the lead frame as shown in Fig. 11. The four heaters were controlled by a controller. Random input signals are employed for exciting the curing system. Totally 2800 measurements are obtained at sixteen sensors with the sample interval $\Delta t = 10s$. The first 2000 measurements are employed to train the multimode modeling process, and the last 800 measurements are employed to verify the effectiveness of the developed model.

First, the above three methods are applied to derive the corresponding spatiotemporal distribution model of the snapshots with two dominant basis function selected. Three subspaces are obtained using the modified dissimilarity analysis. The spatiotemporal error distribution at the (2800-th) sample is shown in Fig. 12. Correspondingly, Fig. 13 illustrates the SNAE over the time direction and the TNAE over the space direction. The RMSEs of the three methods are: 1) Global modeling: 2.6483; 2) Blind subspace modeling: 2.4896; 3) The proposed modeling: 2.1243. It can be observed that the proposed modeling method obtains a superior modeling accuracy than the other two methods. Compared to the global method, the blind subspace modeling obtains an improved modeling accuracy by taking advantage of several sets of

basis functions. However, the system's dynamics are still not reconstructed accurately since the snapshots with largely different characteristics may be assigned to the same subspace. In contrast, the proposed method shows a promising performance with smaller approximation errors. Obviously, the reduced-order model approximated by the proposed method is much closer to the real full-order system. The improved performance benefits from the modified dissimilarity analysis that assigns snapshots with large differences to different subspaces.

V. CONCLUSIONS

In the paper, an integral multimode method is developed for spatiotemporal modeling of complex DPSs. To handle the strong nonlinearities and time-varying dynamics, a two-step solution is proposed: subspace decomposition via modified dissimilarity analysis and local model ensemble via principal component analysis. First, the original operating space is partitioned along the time dimension into several subspaces with divergent local characteristics. Second, a spatiotemporal modeling approach is applied to approximate the local dynamics of each subspace, where the basis functions are extracted by KLD and the corresponding temporal coefficients are learned by extreme learning machine. Finally, an ensemble model is aggregated using the soft weighting sum of the local ones, where the corresponding weights are learned by principal component regression. By the mechanism of partitioning snapshots into subspaces with divergent characteristics, the proposed method can provide a better approximation and prediction of the original system. Experiments on a parabolic catalytic rod and a real snap curing oven system show that the proposed method obtains a better performance regarding modeling accuracy compared to several baselines. Our future work will tackle potential multimode modeling methods from other perspectives, such as better subspace partitioning methods, partitioning the operating space along the space dimension, or along both the time and space dimensions.

REFERENCES

- [1] H.-X. Li and C. Qi, "Modeling of distributed parameter systems for applications – A synthesized review from time-space separation," *Journal of Process Control*, vol. 20, no. 8, pp. 891–901, 2010.
- [2] X.-X. Zhang, Y. Jiang, H.-X. Li, and S.-Y. Li, "SVR learning-based spatiotemporal fuzzy logic controller for nonlinear spatially distributed dynamic systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 10, pp. 1635–1647, 2013.
- [3] Y. Feng and H.-X. Li, "Dynamic spatial independent component analysis based abnormality localization for distributed parameter systems," *IEEE Transactions on Industrial Informatics*, 2019, to be published. doi: 10.1109/TII.2019.2900226.
- [4] Y. Feng and H.-X. Li, "Detection and spatial identification of fault for parabolic distributed parameter systems," *IEEE Transactions on Industrial Electronics*, 2018, to be published. doi: 10.1109/TIE.2018.2877188.

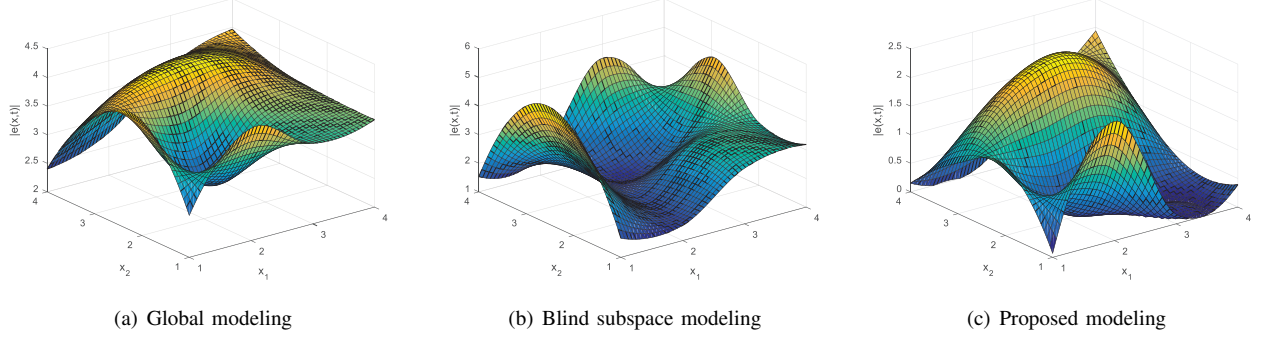


Fig. 12. Case 2: Spatiotemporal error distribution at the 2800-th test sample.

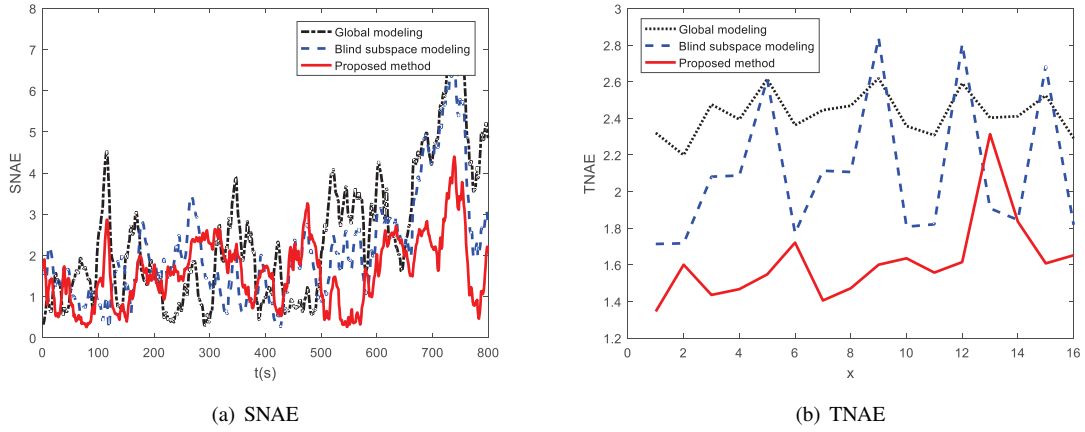


Fig. 13. Case 2: Modeling error of all tested methods along the time and space directions: (a) SNAE; (b) TNAE.

- [5] X.-X. Zhang, L.-R. Zhao, J.-J. Li, G.-T. Cao, and B. Wang, "Space-decomposition based 3D fuzzy control design for nonlinear spatially distributed systems with multiple control sources using multiple single-output SVR learning," *Applied Soft Computing*, vol. 59, pp. 378–388, 2017.
- [6] Z. Wang, H.-X. Li, and C. Chen, "Reinforcement learning based optimal sensor placement for spatiotemporal modeling," *IEEE Transactions on Cybernetics*, 2019, to be published. doi: 10.1109/TCYB.2019.2901897.
- [7] Z. Wang and H.-X. Li, "Incremental spatiotemporal learning for online modeling of distributed parameter systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, to be published. doi: 10.1109/TSMC.2018.2810447.
- [8] C. Qi and H.-X. Li, "A time/space separation-based Hammerstein modeling approach for nonlinear distributed parameter processes," *Computers & Chemical Engineering*, vol. 33, no. 7, pp. 1247–1260, 2009.
- [9] L. Guo and S. A. Billings, "State-space reconstruction and spatiotemporal prediction of lattice dynamical systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 4, pp. 622–632, 2007.
- [10] R. Rannacher and B. Vexler, "A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements," *SIAM Journal on Control and Optimization*, vol. 44, no. 5, pp. 1844–1863, 2005.
- [11] H. Deng, H.-X. Li, and G. Chen, "Spectral-approximation-based intelligent modeling for distributed thermal processes," *IEEE Transaction on Control Systems Technology*, vol. 13, no. 5, pp. 686–700, 2005.
- [12] B. Luo, T. Huang, H.-N. Wu, and X. Yang, "Data-driven H_∞ control for nonlinear distributed parameter systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 11, pp. 2949–2961, 2015.
- [13] X. Lu, T. Hu, and F. Yin, "A novel spatiotemporal fuzzy method for modeling of complex distributed parameter processes," *IEEE Transactions on Industrial Electronics*, 2018.
- [14] X.-B. Meng, H.-X. Li, and H.-D. Yang, "Evolutionary design of spatiotemporal learning model for thermal distribution in lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, 2018, to be published. doi: 10.1109/TII.2018.2866468
- [15] L. Peng and K. Mohseni, "Nonlinear model reduction via a locally weighted POD method," *International Journal for Numerical Methods in Engineering*, vol. 106, no. 5, pp. 372–396, 2016.
- [16] M. Wang, H.-X. Li, X. Chen, and Y. Chen, "Deep

- learning-based model reduction for distributed parameter systems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 12, pp. 1664–1674, 2016.
- [17] M. Izadi and S. Djuljevic, “Order-reduction of parabolic PDEs with time-varying domain using empirical eigenfunctions,” *AIChE Journal*, vol. 59, no. 11, pp. 4142–4150, 2013.
- [18] D. Amsallem, M. J. Zahr, and C. Farhat, “Nonlinear model order reduction based on local reduced-order bases,” *International Journal for Numerical Methods in Engineering*, vol. 92, no. 10, pp. 891–916, 2012.
- [19] M. Dihlmann, M. Drohmann, and B. Haasdonk, “Model reduction of parametrized evolution problems using the reduced basis method with adaptive time partitioning,” Stuttgart Research Centre for Simulation Technology, Stuttgart, Germany, Tech. Rep., 2011.
- [20] A. Narasingam, P. Siddhamshetty, and J. S. Kwon, “Temporal clustering for order reduction of nonlinear parabolic PDE systems with time-dependent spatial domains: Application to a hydraulic fracturing process,” *AIChE Journal*, vol. 63, no. 9, pp. 3818–3831, 2017.
- [21] S. Chaturantabut, “Temporal localized nonlinear model reduction with a priori error estimate,” *Applied Numerical Mathematics*, vol. 119, pp. 225–238, 2017.
- [22] S. Sahyoun and S. M. Djouadi, “Time, space, and space-time hybrid clustering POD with application to the Burgers’ equation,” in *IEEE 53rd Annual Conference on Decision and Control*, Los Angeles, CA, USA, Dec. 2014, pp. 2088–2093.
- [23] A. Corigliano, M. Dossi, and S. Mariani, “Model order reduction and domain decomposition strategies for the solution of the dynamic elastic-plastic structural problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 290, pp. 127–155, 2015.
- [24] J. Yu and S. J. Qin, “Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models,” *AIChE Journal*, vol. 54, no. 7, pp. 1811–1829, 2008.
- [25] J. Yu, J. Chen, and M. M. Rashid, “Multiway independent component analysis mixture model and mutual information based fault detection and diagnosis approach of multiphase batch processes,” *AIChE Journal*, vol. 59, no. 8, pp. 2761–2779, 2013.
- [26] K. Peng, K. Zhang, B. You, J. Dong, and Z. Wang, “A quality-based nonlinear fault diagnosis framework focusing on industrial multimode batch processes,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2615–2622, 2016.
- [27] M. Kano, S. Hasebe, I. Hashimoto, and H. Ohno, “Statistical process monitoring based on dissimilarity of process data,” *AIChE Journal*, vol. 48, no. 6, pp. 1231–1240, 2002.
- [28] C. Zhao, F. Wang, and M. Jia, “Dissimilarity analysis based batch process monitoring using moving windows,” *AIChE Journal*, vol. 53, no. 5, pp. 1267–1277, 2007.
- [29] Z. Liu and H.-X. Li, “Extreme learning machine based spatiotemporal modeling of lithium-ion battery thermal dynamics,” *Journal of Power Sources*, vol. 277, pp. 228–238, 2015.
- [30] P. D. Christofides, *Nonlinear and robust control of PDE systems: Methods and applications to transport-reaction processes*. Boston: Birkhäuser, 2001.
- [31] Y.-H. Lee, H. D. Jin, and C. Han, “On-line process state classification for adaptive monitoring,” *Industrial & Engineering Chemistry Research*, vol. 45, no. 9, pp. 3095–3107, 2006.
- [32] X. Lu, W. Zou, and M. Huang, “An adaptive modeling method for time-varying distributed parameter processes with curing process applications,” *Nonlinear Dynamics*, vol. 82, no. 1–2, pp. 865–876, 2015.



Zhi Wang (S'19) received the B.E. degree in automation from the Department of Control and Systems Engineering, Nanjing University, Nanjing, China, in 2015. He is currently pursuing the Ph.D. degree in machine learning and systems engineering with Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China.

His current research interests include reinforcement learning, system modeling, machine learning, and robotics.



Han-Xiong Li (S'94-M'97-SM'00-F'11) received his B.E. degree in aerospace engineering from the National University of Defense Technology, China in 1982, M.E. degree in electrical engineering from Delft University of Technology, The Netherlands in 1991, and Ph.D. degree in electrical engineering from the University of Auckland, New Zealand in 1997.

He is a professor in the Department of SEEM, City University of Hong Kong. He has a broad experience in both academia and industry. He has authored 2 books and about 20 patents, and published more than 200 SCI journal papers with h-index 42 (web of science). His current research interests include process modeling and control, system intelligence, distributed parameter systems, and battery management system.

Dr. Li serves as Associate Editor for IEEE Transactions on SMC: System, and was an associate editor for IEEE Transactions on Cybernetics (2002–2016) and IEEE Transactions on Industrial Electronics (2009–2015). He was awarded the Distinguished Young Scholar (overseas) by the China National Science Foundation in 2004, a Chang Jiang professorship by the Ministry of Education, China in 2006, and a national professorship in China Thousand Talents Program in 2010. He serves as a distinguished expert for Hunan Government and China Federation of Returned Overseas Chinese. He is a Fellow of the IEEE.