MODEL ORDER REDUCTION OF NONLINEAR PARABOLIC PDE SYSTEMS WITH MOVING BOUNDARIES USING SPARSE PROPER ORTHOGONAL DECOMPOSITION: APPLICATION TO HYDRAULIC FRACTURING

A Thesis

by

HARWINDER SINGH SIDHU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Joseph Sang-il Kwon
Co-Chair of Committee,	Costas Kravaris
Committee Member,	Yalchin Efendiev
Head of Department,	M. Nazmul Karim

August 2018

Major Subject: Chemical Engineering

ABSTRACT

Developing reduced-order models for nonlinear parabolic partial differential equation (PDE) systems with time-varying spatial domains remains a key challenge as the dominant spatial patterns of the system change with time. To address this issue, there have been several studies where the time-varying spatial domain is transformed to the time-invariant spatial domain by using an analytical expression that describes how the spatial domain changes with time. However, this information is not available in many real-world applications, and therefore, the approach is not generally applicable. This study aims to overcome this challenge by introducing sparse proper orthogonal decomposition (SPOD)-Galerkin methodology. The proposed methodology exploits the key features of ridge and lasso regularization techniques for the model order reduction of such systems. This methodology is successfully applied to a hydraulic fracturing process, and a series of simulation results indicates that it is more accurate in approximating the original nonlinear system than the standard POD-Galerkin methodology.

ACKNOWLEDGEMENTS

I take this opportunity to convey my sincere gratitude to my principal investigator Dr. Joseph Sang-il Kwon for his continuous support and guidance throughout the course of my research. Without his timely help and guidance, it would not have been possible to accomplish the work I have done so far. His friendly approaches in clearing the doubts, insightful comments, innovative ideas, and availability at any time are the remarkable qualities which are very much appreciated. I would like to extend my gratitude towards my committee members Dr. Costas Kravaris, and Dr. Yalchin Efendiev for their insightful comments and inputs.

I would like to thank Mr. Abhinav Narasingam**,** and Mr. Prashanth Siddhamshetty, my colleagues and mentors, for their constant support and encouragement. I am grateful for all their contributions and creative ideas.

Thanks to all my colleagues from the research group for all the stimulating discussions and the fun we had. A special thanks to all my friends for keeping me sane and creating the overall experience that Texas A&M was.

This section cannot be complete without me thanking my parents and sister for carrying me so long and for being an inspiration to me. I cannot thank them enough for their constant support and encouragement throughout these two years. None of this would have been possible without them.

.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Dr. Joseph Sang-il Kwon [principal advisor] of the Department of Chemical Engineering and Texas A&M Energy Institute, Dr. Costas Kravaris of the Department of Chemical Engineering, and Dr. Yalchin Efendiev of the Department of Mathematics and Institute of Scientific Computation. All work for the thesis was completed independently by the student.

Funding Sources

Financial support from the Artie McFerrin Department of Chemical Engineering and the Texas A&M Energy Institute are gratefully acknowledged.

NOMENCLATURE

PDE	Partial Differential Equation
POD	Proper Orthogonal Decomposition
SPOD	Sparse Proper Orthogonal Decomposition
LPOD	Local Proper Orthogonal Decomposition
MOR	Model Order Reduction
ODE	Ordinary Differential Equation
ROM	Reduced Order Model
GOS	Global Optimum Search
PCA	Principal Component Analysis
SPCA	Sparse Principal Component Analysis
OLS	Ordinary Least Squares

TABLE OF CONTENTS

Page

LIST OF FIGURES

1. INTRODUCTION

A large number of industrial control problems involve highly nonlinear parabolic PDE systems with time-varying spatial domains such as hydraulic fracturing, crystal growth, and metal casting. Developing high-fidelity models from first-principles has been one of the most important research areas to achieve a fundamental understanding of these systems. However, it is not practical to employ these computationally expensive high-fidelity models for the design of real-time model-based feedback control systems. Motivated by this, model order reduction (MOR) has become an active research area and efforts in this field to develop new MOR techniques for significant CPU time reductions at the expense of model accuracy are progressing at a surprising pace (Benner et al., 2015, Rowley and Dawson, 2017).

MOR techniques are based on an observation that very often the solution of a large-scale complex system resides on a subspace whose dimension is lower than that of the original system. Many MOR techniques have been introduced and implemented in a variety of applications. For example, Nagy (1979) described modal representation of geometrically nonlinear behavior by the finite element method. Noor et al. (1981) and Noor and Peters (1981) presented reduced basis techniques for collapse analysis of shells and for predicting the post-limit-point paths of structures, respectively. Peterson (1989) introduced a reduced basis method for incompressible viscous flow calculations. Verhaegen and Dewilde (1992) presented algorithms to realize a finite dimensional, linear time-invariant state-space model from input-output data. Van Overschee and De Moor (1994) presented subspace algorithms to identify mixed deterministic-stochastic

systems. Schmid (2010) introduced dynamic mode decomposition to extract dynamic information from flow fields. Today, perhaps one of the most popular MOR techniques is based on proper orthogonal decomposition (POD) which is also known as Karhunen–Loéve analysis (Algazi and Sakrison, 1969, Graham and Kevrekidis, 1996, Glavaski et al., Rathinam and Petzold, 2002, Rathinam and Petzold, 2003, Shvartsman and Kevrekidis, 1998, Shvartsman et al., 2000, Willcox and Peraire, 2002). In POD, a set of empirical basis functions that captures dominant spatial patterns of the system is computed from spatiotemporal data obtained via experiments or large-scale high-fidelity simulations. In 1987, Sirovich introduced the method of snapshots to compute basis functions without calculating the kernel necessary for POD (Sirovich, 1987a, Sirovich, 1987b). This method is based on the assumption that each basis function can be represented by a linear combination of the snapshots. The obtained basis functions are then used in a projection method such as Galerkin's projection method to derive low-dimensional ordinary differential equation systems (ODEs) which approximate the original high-dimensional PDE systems.

The POD-based MOR techniques have been traditionally applied to various systems characterized by time-invariant spatial domains. For example, empirical basis functions computed via POD have been employed to derive accurate reduced-order models (ROMs) for dissipative PDEs arising in the modeling of reaction-diffusion systems and fluid flows (Baker and Christofides, 2000, Bangia et al., 1997, Park and Cho, 1996, Park and Jang, 2000, Park and Lee, 1998, Shvartsman and Kevrekidis, 1998). Other applications include, but are not limited to, Burgers equation (Kunisch and Volkwein, 1999), rapid thermal chemical vapor deposition processes (Baker and Christofides, 2000), batch electrochemical reactors (Zhou et al., 2001),

sheet forming processes (Arkun and Kayihan, 1998), and groundwater flow (McPhee and Yeh, 2008).

Although POD-based MOR techniques have been successfully implemented to develop accurate ROMs for parabolic PDE systems with time-invariant spatial domains, there are only few studies on parabolic PDE systems with time-varying spatial domains. Specifically, Armaou and Christofides (Armaou and Christofides, 2001a, Armaou and Christofides, 2001b, Armaou and Christofides, 2001c) employed an analytical expression that describes how the spatial domain changes with time to mathematically transform the time-varying spatial domain to the time-invariant one. Then, POD was applied to the transformed time-invariant spatial domain to compute a set of empirical basis functions for MOR of one-dimensional reaction- diffusion systems. Furthermore, a group of efforts was made to develop a new MOR technique by preserving the invariant properties of the system while transforming the time-varying spatial domain to the time-invariant one. Specifically, Fogleman et al. (2004) applied POD to the spatiotemporal data of internal combustion engine flows to compute phase-invariant basis functions. They transformed the flow velocity defined on a moving grid into a fixed grid in such a way that the divergence-free (continuity) property of the original velocity field is preserved. Izadi and Dubljevic (2013) introduced a mapping functional which relates the time-evolution of the solution of a parabolic PDE with time-varying spatial domains to the one on a fixed reference domain such that space invariant properties (e.g., thermal energy or density) of the data are preserved. They applied this method to develop ROMs of nonlinear reaction-diffusion systems and Czochralski crystal growth processes. Recently, Narasingam et al. (2017) proposed a temporally local MOR technique by partitioning the temporal domain into multiple temporal subdomains using global optimum search (GOS) framework. Then, they applied POD within

each temporal subdomain to compute a set of temporally local basis functions that captures the dominant spatial patterns of the system more effectively than the temporally global basis functions. They employed this method to develop a ROM that describes fracture propagation in a hydraulic fracturing process.

Motivated by these earlier efforts, we have adopted the idea of sparse principal component analysis (PCA) from Zou et al. (2006) and applied this idea to develop sparse proper orthogonal decomposition (SPOD) for the MOR of parabolic PDE systems with moving boundaries. Even though PCA and POD are known to be mathematically equivalent, the implementation of SPOD along with Galerkin's projection framework to develop a ROM for moving boundary problems is not a trivial task. Therefore, based on (Zou et al., 2006), we have illustrated steps required to compute basis functions via SPOD and used the basis functions in Galerkin's projection method to derive a ROM of hydraulic fracturing. Hydraulic fracturing is an important moving boundary problem in chemical and petroleum engineering and developing an accurate ROM will be beneficial for future research directions such as designing optimal pumping schedules to enhance the productivity of produced wells.

The organization of this thesis is as follows. First, a detailed procedure for transforming the time-varying spatial domain to the time-invariant one is presented and the necessity of a new model reduction technique for moving boundary problems is justified. Second, a brief introduction of the concepts and mathematical formulations for regularization techniques such as lasso, ridge and naive elastic net used in SPOD algorithm is provided. Third, the proposed methodology is addressed and a comprehensive algorithm to obtain a ROM using the basis functions obtained by SPOD is presented. Fourth, the application of the proposed methodology to develop a ROM for a highly nonlinear parabolic PDE system with moving boundaries

describing the fracture propagation in a hydraulic fracturing process is described. Lastly, a series of results is presented that demonstrates the accuracy of the proposed methodology in developing a ROM compared to the standard POD-Galerkin methodology as well as local proper orthogonal decomposition (LPOD)-Galerkin methodology (Narasingam et al., 2017).

2. PROBLEM STATEMENT

In parabolic PDE systems with moving boundaries, the main obstacle in developing an accurate ROM is that the dominant spatial patterns of the system change with time, which are not able to be captured by the standard POD. As mentioned in previous section, many researchers have tried to address this challenge by transforming the time-varying spatial domain to the time-invariant one. Based on this approach, we introduce a new methodology for MOR of a parabolic PDE system with moving boundaries describing the fracture propagation in a hydraulic fracturing process where an analytical expression describing how the spatial domain changes with time is unavailable, because the boundary of the spatial domain is a part of the solution to be determined along with other in-domain solutions. In this work, the time-varying spatial domain is viewed as a time-invariant one by leveraging the fact that the fracture width is zero when the fracture does not propagate to a specified spatial location. More specifically, the spatial domain where the fracture has not propagated is considered as a fictitious domain with zero width. This approach will result in a modified spatiotemporal data matrix as shown in Eq. (1) that has nonzero values in the upper triangular part and 0's in the lower triangular part. The upper triangular part represents the spatial locations where the fracture has propagated and the lower triangular part with 0's corresponds to the spatial locations where the fracture has not propagated yet. Please note that in the modified spatiotemporal data matrix, each row implies the temporal profile of fracture width at a particular location and each column indicates the spatial

profile of fracture width at a particular time. Incorporating zeros in the spatiotemporal data matrix will result in spurious spatial patterns that are not present in the original system. Consequently, it is not feasible to build accurate ROMs that capture the dominant process dynamics with the affordable number of global (with respect to time) basis functions.

$$\text{Modified spatiotemporal data matrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & \cdots & w_{1n} \\ 0 & w_{22} & \cdots & \cdots & w_{2n} \\ 0 & 0 & w_{33} & \cdots & w_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & w_{mn} \end{bmatrix} \quad (1)$$

Recently, there have been key developments driven by applying data science techniques to chemical engineering applications, particularly to biological and energy systems (Carothers, 2013, Carothers et al., 2009, Dubey et al., 2006, Khorshidi and Peterson, 2016, Kieslich et al., 2016a, Kieslich et al., 2016b, Lee and Lee, 2006, Lee and Wong, 2010, Lee and Lee, 2005, Qin, 2014, Wilson and Sahinidis, 2017). Motivated by these earlier efforts, we exploit the key features of regularization techniques such as lasso, ridge, and naive elastic net to deal with the spurious spatial patterns that may arise due to the addition of zeros while constructing a ROM to describe the fracture propagation in a hydraulic fracturing process.

3. THE LASSO, RIDGE AND THE NAÏVE ELASTIC NET

Consider the data set that has *m* observations and *q* predictors. Let $\mathrm{Y} = (y_1, \cdots, y_m)^T$ be the response vector and $\mathbf{X} = [X_1, \cdots, X_q]$ where for $X_j = (x_{1j}, \cdots, x_{mj})^T$ for $j = 1, \cdots, q$ be the predictors. By assuming all the predictors in **X** are normalized to have zero mean and unit variance and the response Y is normalized to have zero mean, we can write a standard linear multiple regression problem as follows:

$$y_i = \sum_{j=1}^{q} \beta_j x_{ij} + e_i\,, \qquad i = 1,2, \cdots, m \tag{2}$$

where $e_1, e_2, \cdots, e_m$ are the error terms and $\beta = \left[\beta_1, \beta_2, \cdots, \beta_q\right]^T$ are the regression coefficients. In ordinary least squares (OLS) regression analysis, these coefficients are estimated by minimizing the squared sum of residual (or error). Since the coefficients estimated by OLS have low bias but large variance, they perform poorly in prediction. This limitation of OLS can be handled by a regularization technique named ridge regression (Hoerl and Kennard, 1988) that imposes an additional L_2-penalty, $\sum_{j=1}^{q}\left|\beta_j\right|^2$, on the regression coefficients. To estimate the coefficients by ridge technique, $\hat{\beta}_{ridge}$, we minimize

$$\left\| Y - \sum_{j=1}^{q} X_j \beta_j \right\|^2 + \lambda_2 \sum_{j=1}^{q} \left|\beta_j\right|^2 \tag{3}$$

where λ_2 is a non-negative parameter. Ridge regression improves the prediction accuracy of OLS via bias-variance trade-off. However, ridge regression always keeps all the predictors in the

model because of which it is not possible to produce a parsimonious model. With an objective to remove redundant predictors from the model, a new regularization technique named lasso (Tibshirani, 1996) was introduced that imposes an additional L_1-penalty, $\sum_{j=1}^{q}|\beta_j|$, on the regression coefficients. To estimate the coefficients by lasso, $\hat{\beta}_{lasso}$, we minimize

$$\left\| Y - \sum_{j=1}^{q} X_j \beta_j \right\|^2 + \lambda_1 \sum_{j=1}^{q} |\beta_j| \tag{4}$$

where owing to the nature of L_1-penalty, some coefficients will become exact zero if λ_1 (non-negative parameter) is sufficiently large. Despite its success in many applications, lasso has some limitations (Zou and Hastie, 2005). The most relevant one to the purpose of this work, developing SPOD and applying it to moving boundary problems, is that the number of variables selected by lasso are limited by the number of observations. More specifically, if $q \gg m$, lasso can select at most *m* predictors (Efron et al., 2004). To overcome this major drawback, Zou and Hastie (2005) proposed a new technique named naive elastic net. For any non-negative λ_1 and λ_2, the coefficients can be estimated by naive elastic net, $\hat{\beta}_{en}$, as follows:

$$\hat{\beta}_{en} = \arg\min_{\beta} \left\| Y - \sum_{j=1}^{q} X_j \beta_j \right\|^2 + \lambda_2 \sum_{j=1}^{q} |\beta_j|^2 + \lambda_1 \sum_{j=1}^{q} |\beta_j| \tag{5}$$

where given a fixed λ_2, LARS-EN algorithm (Zou and Hastie, 2005) can be used to solve Eq. (5) for all λ_1. When $m \geq q$, the value of λ_2 can be zero. When $q > m$, we choose $\lambda_2 > 0$, because then the naive elastic net can potentially include all variables in the model, and thus, it can deal with the aforementioned drawback of lasso.

4. SPOD-GALERKIN METHODOLOGY

In this section, we present SPOD for MOR of nonlinear parabolic PDE systems with moving boundaries. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be the ensemble of snapshots where n and p are the number of spatial measurements and the number of snapshots, respectively. The matrix $\mathbf{X}$ is obtained after solving the high-fidelity model (or equivalently, by obtaining the experimental measurements) and incorporating zeros into the spatiotemporal data with an objective to transform the time-varying spatial domain to the time-invariant one. SPOD employs the naive elastic net technique which is a convex combination of ridge and lasso penalties as described in Section 3 to obtain the basis functions. Specifically, the ridge penalty plays a crucial role to reconstruct the basis functions for spatiotemporal data with any dimensions ($n \geq p$ or $n < p$), whereas the lasso penalty mitigates the impact of added zeros by neglecting redundant snapshots from $\mathbf{X}$. Therefore, SPOD generates basis functions that are able to capture the dominant spatial patterns of the original moving boundary system in an effective way (from the standpoint of model accuracy) as compared to the standard POD method.

We now present the methodology to obtain basis functions by employing the close connection between POD and singular value decomposition (SVD) of the spatiotemporal data matrix $\mathbf{X}$ (Kunisch and Volkwein, 1999, Kunisch and Volkwein, 2001, Pinnau, 2008). Let the SVD of $\mathbf{X}$ be

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathbf{T}} \tag{6}$$

where **U** is an $n \times n$ orthonormal matrix, **D** is an $n \times p$ diagonal matrix having only non-negative and non-increasing entries on the diagonal and **V** is an $p \times p$ orthogonal matrix. The columns $\{U_j\}_{j=1}^{n}$ of **U** are the left singular vectors of **X**, the non-zero entries of **D** are the singular values of **X** and the columns $\{V\}_{i=1}^{p}$ of **V** are the right singular vectors of **X**. It then follows that the POD basis functions may be expressed in terms of the singular values and the right singular vectors of **X** as follows:

$$\boldsymbol{\Phi_i} = \frac{1}{\sigma_i}\mathbf{X}V_i \tag{7}$$

where Φ_i is the i^{th} basis function, σ_i is the i^{th} singular value of **X**, V_i is the i^{th} column of **V**. Next, we select *k* basis functions that describe the dominant spatial patterns of the system by using the first *k* singular values (${\sigma_1}^2 \geq \cdots {\sigma_k}^2 \geq \cdots {\sigma_n}^2$) to build a ROM. Such a method of determining POD basis functions is also called the method of snapshots (Sirovich, 1987a, Sirovich, 1987b). Note that the product of the i^{th} basis function and i^{th} singular value of **X**, i.e., $Z_i = \Phi_i\sigma_i = \mathbf{X}V_i$, is a linear combination of the *p* snapshots, **X** , and thus, we can obtain the i^{th} right singular vector V_i by regressing Z_i on the *p* snapshots, **X**. For any non-negative λ_1 and λ_2, the regression coefficients estimated by naive elastic net, $\hat{\beta}_{en}$, are given by

$$\hat{\beta}_{en} = \arg\min_{\beta} \|Z_i - \mathbf{X}\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1 \tag{8}$$

Where $\beta = [\beta_1, \beta_2, \ldots, \beta_p]$, $\|\beta\|_1 = \sum_{j=1}^{p}|\beta_j|$ and $\hat{V}_i = \frac{\hat{\beta}_{en}}{\|\hat{\beta}_{en}\|}$ denotes an approximation to V_i. The L_1-penalty plays an important role by penalizing the regression coefficients (i.e., the elements of $\hat{\beta}_{en}$), to achieve a sparse approximation, i.e., $\hat{V}_i$ to the i^{th} right singular vector. This approximation will be used to alleviate the impact of added zeros by neglecting redundant snapshots from the spatiotemporal data matrix while constructing the i^{th} basis function. One

drawback of Eq. (8) is that it cannot be used to determine sparse approximations to right singular vectors without obtaining the POD basis functions from the method of snapshots. To overcome this challenge, we employed the self-contained regression developed by (Zou et al., 2006) for sparse principal component analysis (SPCA). Using the self-contained regression, SPOD can be reformulated as the following optimization problem to determine the approximations to the first *k* right singular vectors of **X**:

$$\left(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}\right) = \underset{\mathbf{A},\mathbf{B}}{\arg\min} \sum_{i=1}^{n} \|\boldsymbol{x_i} - \mathbf{A}\mathbf{B}^T \boldsymbol{x_i}\|^2 + \lambda_2 \sum_{j=1}^{k} \left\|\beta_j\right\|^2 + \sum_{j=1}^{k} \lambda_{1,j} \left\|\beta_j\right\|_1 \quad (9)$$

$$s.t.\ \mathbf{A^T A} = I_{k\times k}$$

where $\boldsymbol{x_i}$ denotes the i^{th} row vector of the matrix **X**, $\mathbf{A}_{\boldsymbol{p\times k}} = [\alpha_1, \dots, \alpha_k]$ and $\mathbf{B}_{\boldsymbol{p\times k}} = [\beta_1, \dots, \beta_k]$ are the parameters that will be solved to minimize Eq. (9) , λ_2 is the ridge penalty coefficient and $\lambda_{1,j}$ is the lasso penalty coefficient that determines the degree of sparsity of the approximation corresponding to the j^{th} right singular vector. Then, for a pair of non-negative parameters λ_2 and $\lambda_{1,j}$, $\widehat{V}_j = \frac{\widehat{\beta}_j}{\|\widehat{\beta}_j\|}$ is a sparse approximation to V_j and $\widehat{\Phi}_{\boldsymbol{j}} = \frac{1}{\sigma_j}\mathbf{X}\widehat{V}_j$ is the basis function obtained by SPOD for $j = 1, 2, \dots, k$. Furthermore, for $p > n$ data, by letting $\mathbf{B} = \mathbf{A}$ in Eq. (9) and removing the lasso penalty, the proposed SPOD reduces to the standard POD.

For $n \geq p$ data, the default choice of λ_2 can be zero. For $p > n$ data, i.e., the number of snapshots are greater than the number of spatial points, which is usually the case in moving boundary problems. Eq. (9) is valid for all $\lambda_2 > 0$, so in principle we can select any positive λ_2. In particular, Eq. (9) can be reduced to the following optimization problem if λ_2 is set to be a large value (Zou et al., 2006).

$$\left(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}\right) = \underset{\mathbf{A},\mathbf{B}}{\arg\min} -2\mathrm{tr}(\mathbf{A}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{B}) + \sum_{j=1}^{k}\left\|\beta_j\right\|^2 + \sum_{j=1}^{k}\lambda_{1,j}\left\|\beta_j\right\|_1 \quad (10)$$

$$s.t.\ \mathbf{A}^{\mathbf{T}}\mathbf{A} = I_{k\times k}$$

An iterative algorithm, which is introduced by Zou et al. (2006) can be employed to obtain the parameters $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ without computing the POD basis functions.

Let $(\Phi_i, i = 1, \cdots, k, k+1)$ be the first $k+1$ basis functions computed by the standard POD method and $\left(\widehat{\Phi}_i, i = 1, \cdots, k, k+1\right)$ be the first $k+1$ basis functions computed by the proposed SPOD method. In POD, Φ_{k+1} is not correlated with $(\Phi_i,\ i = 1, 2, \cdots, k)$, and therefore, the total energy occupied by the first $k+1$ basis functions is the sum of the energy by the first *k* basis functions and the additional energy from Φ_{k+1}. However, SPOD does not explicitly impose a constraint that enforces $\widehat{\Phi}_{k+1}$ to be uncorrelated with $\left(\widehat{\Phi}_i, i = 1, \cdots, k\right)$, because of which its energy may contain contributions from $\left(\widehat{\Phi}_i, i = 1, \cdots, k\right)$. Therefore, its energy cannot be directly added to the energy occupied by the first *k* basis functions to obtain the energy occupied by the first $k+1$ basis functions. To overcome this challenge, Zou et al. (2006) suggested a regression projection method to remove the linear dependence between the correlated basis functions. According to this method, T_j, which is the residual after removing the correlation of $\widehat{\Phi}_j$ with $\widehat{\Phi}_1, \ldots, \widehat{\Phi}_{j-1}$ can be written as follows:

$$T_j = \widehat{\Phi}_j - \mathbf{H}_{1,\cdots,j-1}\,\widehat{\Phi}_j \quad (11)$$

where $\mathbf{H}_{1,\cdots,j-1}$ is the projection matrix onto the subspace spanned by $\left\{\widehat{\Phi}_i\right\}_1^{j-1}$. Then, the energy occupied by $\widehat{\Phi}_j$ after removing the contribution from is $\widehat{\Phi}_1, \ldots, \widehat{\Phi}_{j-1}$ is $\left\|T_j\right\|^2$. Therefore, total energy occupied by the first *k* basis functions is defined as $\sum_{j=1}^{k}\left\|T_j\right\|^2$.

After normalization, the computed basis functions, $\widehat{\Phi}$, via SPOD can be used in the Galerkin's projection method to derive low- dimensional ODE systems that accurately describe the dominant dynamics of the parabolic PDE systems with moving boundaries. In Galerkin's projection method, we obtain the low-dimensional ODE system by replacing the state variable, $x(z,t)$, as follows:

$$x(z,t) = \sum_{i=1}^{k} a_i(t)\widehat{\Phi}_i(z) \tag{12}$$

where *z* is the spatial coordinate, *t* is the time coordinate, and $a_i's$ are the time-dependent coefficients. Finally, the derived system of ODEs can be numerically integrated to obtain the time-dependent coefficients, which will be used to compute a solution which approximates the full-order solution.

4.1 SPOD-Galerkin Algorithm

1. Obtain N snapshots by solving the high-fidelity model (or equivalently, by obtaining the experimental measurements) and incorporating zeros to transform the time-varying spatial domain to the time-invariant one.
2. Solve the naive elastic net problem described in Eq. (10) to obtain $\widehat{\mathbf{B}}$.
3. Obtain the sparse approximations, $\widehat{\mathbf{V}}_{p\times k} = [\hat{V}_1, \dots, \hat{V}_k]$, to the first *k* right singular vectors of **X**, by normalizing the columns of $\widehat{\mathbf{B}}$.
4. Compute the basis functions, $\{\widehat{\Phi}_i\}_{i=1}^{k}$, by multiplying the data matrix, **X**, with the obtained sparse approximation to the right singular vectors, $\widehat{\mathbf{V}}_{p\times k} = [\hat{V}_1, \dots, \hat{V}_k]$, and the corresponding singular value σ_i as follows:

$$\widehat{\Phi}_i = \frac{1}{\sigma_i}\mathbf{X}\hat{V}_i$$

5. Apply the Galerkin's projection method after normalizing the obtained basis functions to derive a low-dimensional ODE system.
6. Numerically integrate the low-dimensional ODE system to obtain the time-dependent coefficients in Galerkin's projection method, which will be used to compute a low-order solution which approximates the full-order solution.

The tuning parameters $\lambda_{1,j}$ should be selected such that each basis function obtained by SPOD algorithm has the energy similar to that of the corresponding basis function obtained from the standard POD (Zou et al., 2006). Please note that the energy occupied by the first *k* basis functions can be obtained by decomposing $\widehat{\Phi} = \left[\widehat{\Phi}_1, \ldots, \widehat{\Phi}_k\right]$ into a product of two matrices, $\widehat{\Phi} = \mathbf{QR}$ where **Q** is the orthonormal matrix and **R** is the upper triangular matrix as described by Zou et al. (2006) and Gajjar et al. (2017). It then follows that $\left\|T_j\right\|^2 = \mathbf{R}_{jj}{}^2$ and thus, the total energy occupied by the first *k* basis functions is equal to $\sum_{j=1}^{k} \mathbf{R}_{jj}{}^2$.

5. APPLICATION TO HYDRAULIC FRACTURING PROCESS

5.1 Hydraulic fracturing process

Unconventional natural-gas resources such as shale gas, coalbed (coal-seam) methane, gas hydrates are trapped in rock formations of very low porosity (2% or less) and low-permeability (0.01 to 0.0001 mD or even less) (Nikolaou, 2013). Therefore, the trapped gas cannot be extracted economically without stimulation. This challenge is addressed by the combination of directional drilling (Watters and Dunn-Norman, 1998) and hydraulic fracturing (Economides and Nolte, 1989) techniques that lead to the shale gas revolution.

A hydraulic fracturing process begins with a perforation technique, in which a well is drilled and a wire equipped with explosive charges is dropped into the well to create initial fracture channels. Then, a high-pressure clean fluid (called pad) is introduced to propagate the fractures in the rock formation at perforated sites. Subsequently, a fracturing fluid consisting of water, additives, and proppant is pumped into the wellbore at sufficiently high pressure and flow rate for further fracture propagation. Once the pumping is stopped, the fractures are closed due to the natural stress of the rock formation. During the closure process, the remaining fluid is allowed to leak off to the rock formation and the proppant is trapped inside the fracture walls. At the end of pumping, the concentration of this trapped proppant should be uniform along the fracture so that it can result in the formation of spatially uniform conductive channels to help effective extraction of the oil and gas from the reservoir (Siddhamshetty et al., 2017a,

Siddhamshetty et al., 2017b, Yang et al., 2017, Narasingam and Kwon, 2017, Narasingam et al., 2018, Siddhamshetty et al., 2018).

5.2 Modeling of fracture propagation

In this study, we apply the proposed SPOD-Galerkin projection methodology to build a ROM for a nonlinear parabolic PDE system with moving boundaries describing the fracture propagation in a hydraulic fracturing process. The dynamic model used to describe the fracture propagation is adopted from Siddhamshetty et al. (2017a) and is based on the following assumptions: (1) fracture propagation is described by Perkins, Kern, and Nordgren (PKN) model; (2) the rock properties (e.g., Young's modulus) remain constant with respect to time and space; (3) the formation layers below and above are where the fractures have sufficiently large stresses such that vertical fracture is confined within a single horizontal rock layer; and (4) the fracture length is much greater than its width, and as a result the fluid pressure across the vertical direction is constant.

A brief description of the model equations is presented below. The fluid flow rate in the horizontal direction is determined by the following equation for flow of a Newtonian fluid in an elliptical section (Nordgren, 1972, Economides and Nolte, 1989)

$$\frac{dP}{dz} = -\frac{64\mu Q}{\pi H W^3} \tag{13}$$

Where P is the net pressure, $z \in [0, L(t)]$ is the time-dependent spatial coordinate in the horizontal direction, μ is the fluid viscosity, $\boldsymbol{Q}$ is the local flow rate in the horizontal direction, H is the fracture height, and W is the fracture width**.**

For a rock under constant normal pressure, the fracture shape is elliptical as shown in Fig. 1. The relationship between the maximum fracture width (i.e., the minor axis of the ellipse) and the net fluid pressure is calculated using the following equation (Sneddon and Elliot, 1946, Gudmundsson, 1983):

$$W = \frac{2PH(1-\nu^2)}{E} \tag{14}$$

Where ν is the Poisson ratio of the formation, and E is the Young's modulus of rock formation. The volume conservation of an incompressible fluid inside the fracture is given by Nordgren (1972):

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial z} + HU = 0 \tag{15}$$

Where $A = \pi W\, H/4$ is the cross-sectional area of the elliptic fracture (Nordgren, 1972) and U is the fluid leak-off rate per unit height accounting for both walls. Eq. (15) requires two boundary equations and one initial condition as follows:

$$q_z(0,t) = Q_0, \qquad W(L(t),t) = 0 \tag{16}$$

$$W(z,0) = 0 \tag{17}$$

Where Q_0 is the water/slurry injection rate at the wellbore, q_z is the flow rate at the wellbore which is given by the following equation:

$$q_z = -\frac{\pi E W^3}{128\mu(1-\nu^2)}\frac{dW}{dz} \tag{18}$$

Plugging Eqs. (13) and (14) into Eq. (15) will generate the following nonlinear parabolic PDE

$$\frac{\pi H}{4}\frac{\partial W}{\partial t} - \frac{\pi E}{128\mu(1-\nu^2)}\left[3W^2\left(\frac{\partial W}{\partial z}\right)^2 + W^3\frac{\partial^2 W}{\partial z^2}\right] + HU = 0 \tag{19}$$

The simplest model of fluid leak-off rate per unit height during fracture propagation is given below (Howard and Fast, Economides and Nolte, 1989):

$$U = \frac{2C_{leak}}{\sqrt{t - \tau(z)}} \tag{20}$$

Where C_{leak} is the overall leak-off coefficient, *t* is the elapsed time since fracturing was initiated, and $\tau(z)$ is the time at which the fracture propagation has arrived at the location *z* for the first time.

The Eq. (19), also known as the porous medium equation has several challenges associated with it that need to be addressed from the standpoint of numerical simulation. For example, (1) an efficient coupling of governing equations with multiple nonlinear equations that describe the important physical phenomena such as rock deformation and fluid flow in hydraulic fracturing systems is essential; (2) leak-off rate has to be determined via iterations; (3) the spatial domain changes with time in hydraulic fracturing systems; and (4) the number of discretized nonlinear algebraic equations to be solved for accurate solutions grows as the fracture treatment continues, significantly increasing the computational requirements.

The values of the various process parameters used in our calculations are: $H = 10$ m, $Q_0 = 0.03$ m^3/s, $\mu = 0.56$ Pa.s, $E = 5 \times 10^3$ MPa, $\nu = 0.2$, $C_{leak} = 6.3 \times 10^{-5}$ m/s$^{1/2}$.

5.3 Numerical simulation

In this study, we did not use the method of coordinate transformation, which is one of the most widely used techniques, to deal with time-varying spatial domains, because it requires an analytical expression that describes how the spatial domain, *l(t)*, changes with time to normalize the spatial coordinate (Armaou and Christofides, 2001a). Such an analytical expression is not available a priori for hydraulic fracturing process. In such cases, the spatial (or temporal)

coordinate can be divided into equal intervals, and the resulting system of discretized algebraic equations can be solved to determine the corresponding grid size in the other coordinate system so that the boundary always remains at a grid point (Murray, 1959, Yuen and Kleinman, 1980, Bücker et al., 2009). Based on this, a numerical scheme developed by Narasingam et al. (2017) has been employed for solving the above governing equations by effectively handling the issues with the time-varying spatial domain and coupling of nonlinear equations.

Meshing Strategies: A one-dimensional (1-D) grid system is generated to represent the created fracture geometry. There are two widely used meshing strategies to deal with a time-varying spatial domain: moving meshing and periodic remeshing of a fixed domain. While the former strategy provides a less accurate solution (because of the limited number of meshes) with a reasonable computational burden, the latter provides an accurate solution at the expense of CPU time (the degree of remeshing could lead to an interpolation error in the solution). To capture the detailed process dynamics of the hydraulic fracturing system in which the boundary condition of the spatial domain keeps on changing, a fixed mesh strategy is used by additionally adapting the size of integration time step.

Numerical Solution Procedure. The steps of the numerical algorithm are described below:

1. At time step t_k, the fracture length $L(t_{k+1})$ is obtained by elongating the fracture tip by Δz.
2. The coupled equations of Eqs. 13–20 are solved to calculate the fracture width $W(z, t_{k+1})$, the net pressure $P(z, t_{k+1})$,, and the flow rate $q_z(z, t_{k+1})$, across the fracture via a finite element method.
3. Calculate $\tau(z_{k+1})$ in Eq. 20 iteratively by repeating Steps 2 and 3.

4. The time interval Δt_{k+1} is determined based on the Courant–Friedrichs–Lewy (CFL) number.
5. Set $k \rightarrow k+1$ and go to Step 1.

For 1-D case, the CFL condition has the following form for explicit numerical schemes

$$\frac{u\Delta t}{\Delta z} \leq 1 \tag{21}$$

where $u = |dW/dt|$ is the fracture width growth rate. This technique has been widely accepted to improve the computational efficiency by increasing the CFL number, (Bücker et al., 2009). In this study, Δz is fixed, and u increases with spatial domain (i.e., the fracture width changes more rapidly near the fracture tip compared to that near the wellbore), which provides room for improvement in the computational efficiency by increasing Δt near the wellbore.

6. SIMULATION RESULTS

An accurate full-order solution for a nonlinear parabolic PDE system with moving boundaries describing the fracture propagation in a hydraulic fracturing process was obtained using a numerical procedure as described in Section 5. The acquired solution led to a total of 411 and 36,144 nodes in the spatial and temporal coordinates, respectively. Fig. 1 describes the spatiotemporal evolution of the fracture width with respect to spatial and time coordinates. It can be noticed that the fracture width grows very rapidly in the beginning, and the growth rate gradually slows down with time.

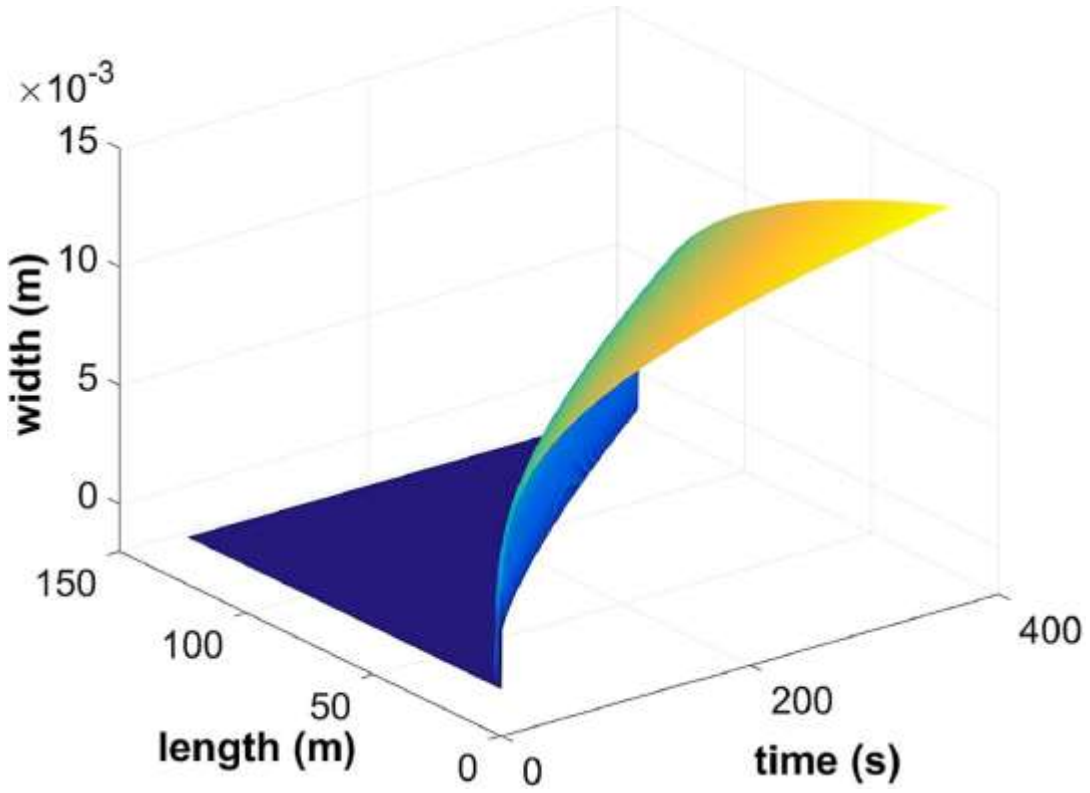

Figure 1: The evolution of fracture width obtained from the high-fidelity model. Reprinted from Sidhu et al. (2018).

We now present the computation of the ROM using the proposed SPOD-Galerkin methodology. First, a total of 1500 snapshots, out of 36,144 generated from the high-fidelity model, were selected at uniform time intervals. Second, the SPOD regression criterion described in Eq. (10) was solved to obtain a set of basis functions. From SPOD, 6 basis functions were obtained that account for 95.61% energy of the system to effectively describe the dominant spatial patterns of the original nonlinear parabolic PDE system with the moving boundaries. Specifically, $\lambda_1 = (0.01, 0.007, 0.0001, 0.0001, 0.0005,\ 0.0005)$ were selected such that each basis function obtained by SPOD algorithm has the energy similar to that of the corresponding basis function obtained from the standard POD (Zou et al., 2006). Also with these values, lasso penalty was able to diminish the impact of added zeros to the spatiotemporal data matrix. Please note that choosing larger values for the tuning parameters $\{\lambda_{1,j}\}$ will lead to highly sparse approximations to the right singular vectors of **X**, resulting in the loss of underlying dominant spatial patterns (because only a few snapshots will be used in the computation of basis functions). On the other hand, choosing $\{\lambda_{1,j}\}$ values close to zero reduces SPOD to POD. Therefore, one should select $\{\lambda_{1,j}\}$ that can effectively deal with the spurious spatial patterns that arise while transforming the time-varying spatial domain to the time-invariant one. Software in R for fitting the SPCA model (and the elastic net models) in the Comprehensive R Archive Network (CRAN) contributed package *elasticnet* was used to obtain the sparse approximations to the right singular vectors of **X**.

The computed basis functions are used in the Galerkin's projection method to obtain the ROM by replacing $W(z,t) = \sum_{i=1}^{d} a_i(t)\widehat{\Phi}_i(z)$ in Eq. (19) to get

$$\sum_{i=1}^{d}(\dot{a}_i\widehat{\Phi}_i) = C_1\left[3\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)^2\left(\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)'\right)^2 + \left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)^3\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)''\right] - \frac{C_2}{\sqrt{t-\tau(z)}} \tag{22}$$

where d is the number of basis functions, over dot and prime represent derivatives with respect to time and space, respectively, C_1 and C_2 are presented as follows:

$$C_1 = \frac{E}{32\mu H(1-\nu^2)}, \quad C_2 = \frac{8C_{leak}}{\pi} \tag{23}$$

Projecting on the basis functions $\widehat{\Phi}_j$ yields

$$\dot{a}(t) = K^{-1}[C_1\big(3u(t) + v(t)\big) - p(t)] \tag{24}$$

where

$$a(t) = [a_1(t), a_2(t), \cdots, a_d(t)]^T$$

$$u(t) = \left[\langle\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)^2\left(\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)'\right)^2, \widehat{\Phi}_j\rangle\right]$$

$$v(t) = \left[\langle\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)^3\left(\sum_{i=1}^{d}(a_i\widehat{\Phi}_i)\right)'', \widehat{\Phi}_j\rangle\right]$$

$$p(t) = \langle\frac{C_2}{\sqrt{t-\tau(z)}}, \widehat{\Phi}_j\rangle$$

$$K_{ij} = \langle\widehat{\Phi}_i, \widehat{\Phi}_j\rangle$$

Eq. (24) represents the low-dimensional ODE system of Eq. (19), which is converted into a system of algebraic equations by discretizing the time coordinate via explicit Euler method, and the algebraic equations were solved in MATLAB to calculate the time-dependent coefficients. The obtained time-dependent coefficients were used to compute a low-order solution which approximates the full-order solution. The computational time required to solve

the ROM obtained by SPOD–Galerkin methodology is 8.669 seconds whereas the time required for solving the high-fidelity model is 406.432 seconds. Therefore, a significant reduction in the CPU time has been achieved. Please note that the calculations were performed on a Dell workstation, powered by Intel(R) Core(TM) i7- 4770 CPU@3.40GHz, running the Windows 8 operating system.

Please note that the proposed approach does not provide a systematic way to determine the number of basis functions to develop a ROM. One rule of thumb to select the number of basis functions is based on their energy. In hydraulic fracturing, the first 6 basis functions obtained by SPOD takes 95.61% energy of the system and including more basis functions did not significantly increase the total energy.

For the purpose of comparison, we also computed a ROM model using the standard POD-Galerkin projection methodology with 16 basis functions that account for 99.90% energy of the system. In this work, we refer the standard POD to the following sequential approach: (a) a modified spatiotemporal data matrix is constructed and (b) POD is applied to the modified spatiotemporal data matrix to compute basis functions. In contrast to SPOD, it does not require the additional step to compute sparse approximations to right singular vectors. Fig. 2 and Fig. 3 show the fracture width profile obtained from the ROM constructed by SPOD-Galerkin and POD-Galerkin techniques, respectively. Fig. 4 shows the comparison between the full-order solution and the reduced-order solutions obtained using both of the techniques at 4 different locations within the fracture ($z = 0$ (i.e., wellbore), $z = 22.2$ m, $z = 44.7$ m and $z = 62.7$ m (i.e., the fracture center)). It can be observed from Fig. 4 that the ROM developed to approximate a nonlinear parabolic PDE system with moving boundaries using the SPOD- Galerkin methodology is more accurate than the one obtained by the standard POD-Galerkin

methodology. Please note that the proposed methodology is not able to provide a good approximate solution to the fictitious spatial domain where the fracture has not propagated. It can be observed that at $z = 0$, the approximated solutions obtained from the ROMs developed using both of the methodologies are very similar, but the difference between them increases as the distance from the wellbore increases (i.e., *z* increases). This is attributed to the fact that at $z = 0$, the fracture width is always positive for $t \geq 0$, and thus, zeros are not added to this spatial point while transforming the time-varying spatial domain to the time-invariant one. However, as *z* increases the number of added zeros also increases as shown in Eq. (1), because of which the solution obtained from the ROM developed by POD-Galerkin methodology keeps on deviating from the high-fidelity solution. On the other hand, SPOD-Galerkin methodology is able to mitigate the effect of added zeros by employing the regularization techniques, and therefore, the solution obtained from the ROM constructed using this methodology is closer to the full-order solution. Fig. 5 shows the comparison between the first four basis functions computed using the SPOD and POD method.

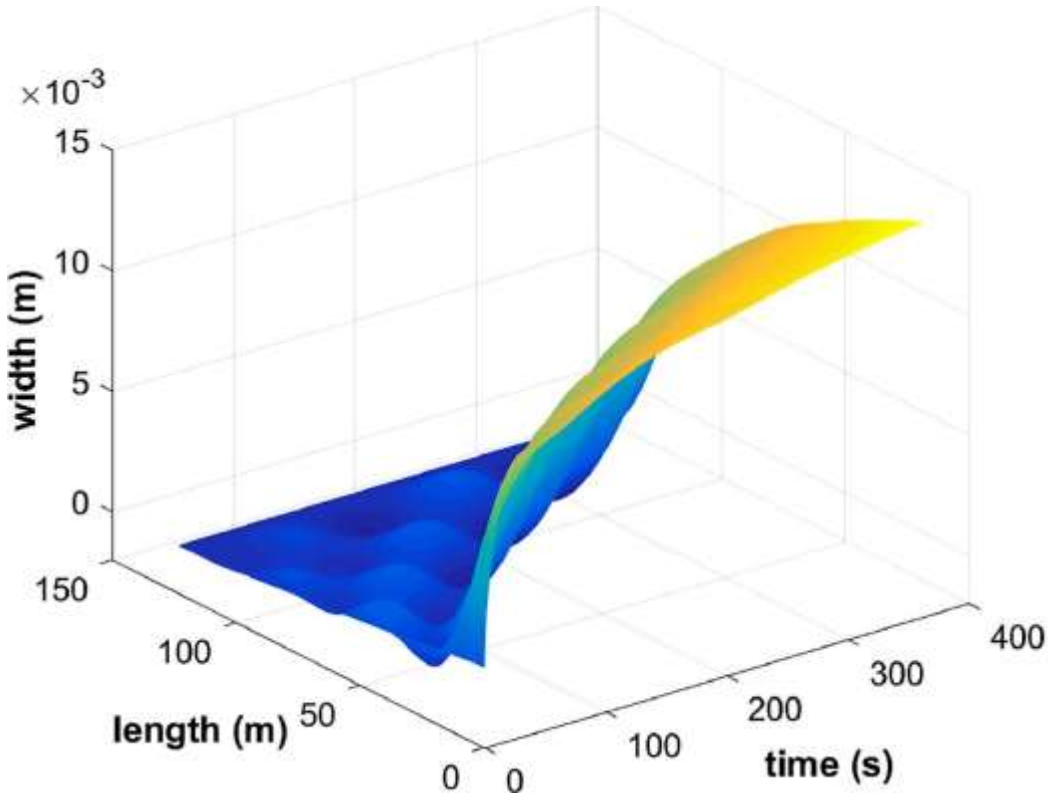

Figure 2: Approximate width profile computed from the ROM obtained by the SPOD-Galerkin methodology. Reprinted from Sidhu et al. (2018).

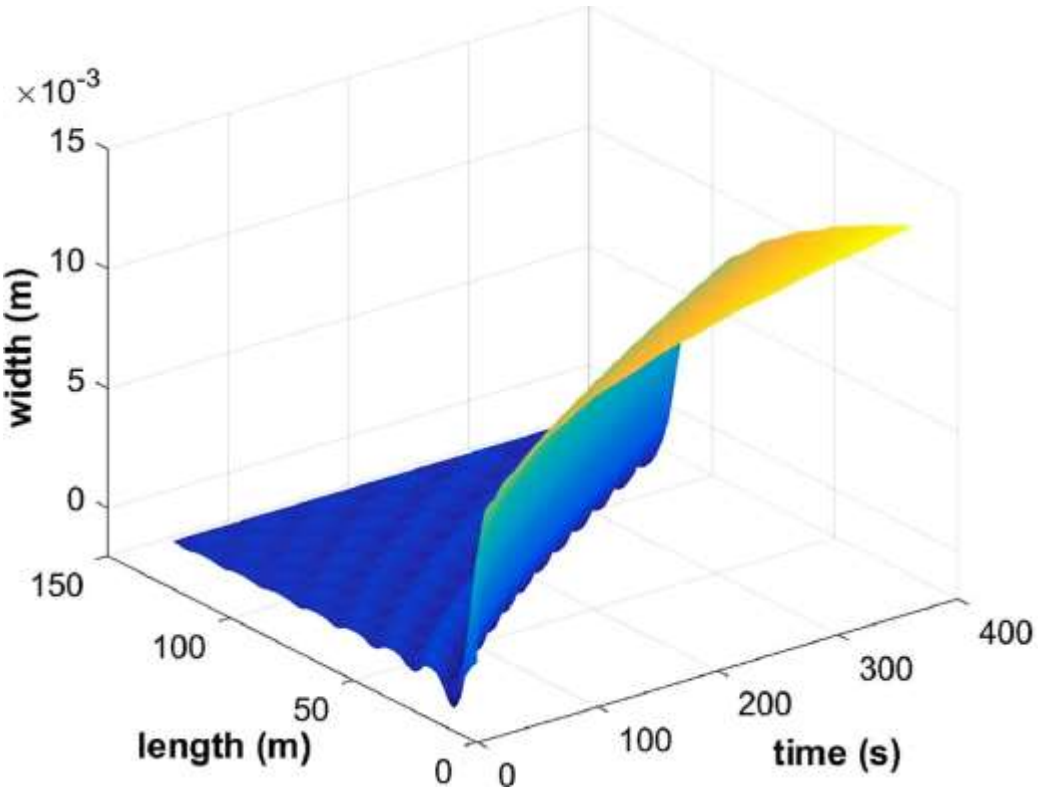

Figure 3: Approximate width profile computed from the ROM obtained by the POD-Galerkin methodology. Reprinted from Sidhu et al. (2018).

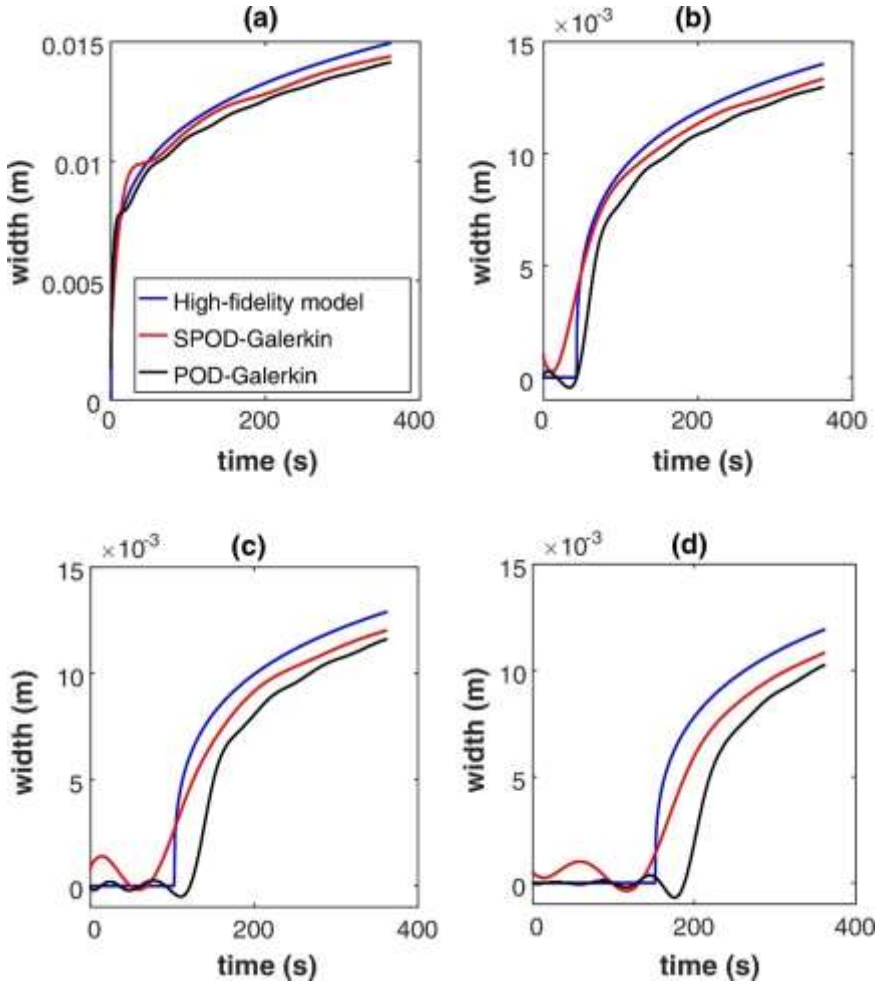

Figure 4: Comparison of width profiles obtained at four different spatial locations, (a) $z = 0$, (b) $z = 22.2$ m, (c) $z = 44.7$ m and (d) $z = 62.7$ m, from the full-order model and the ROMs obtained by the SPOD-Galerkin and POD-Galerkin methodology. Reprinted from Sidhu et al. (2018).

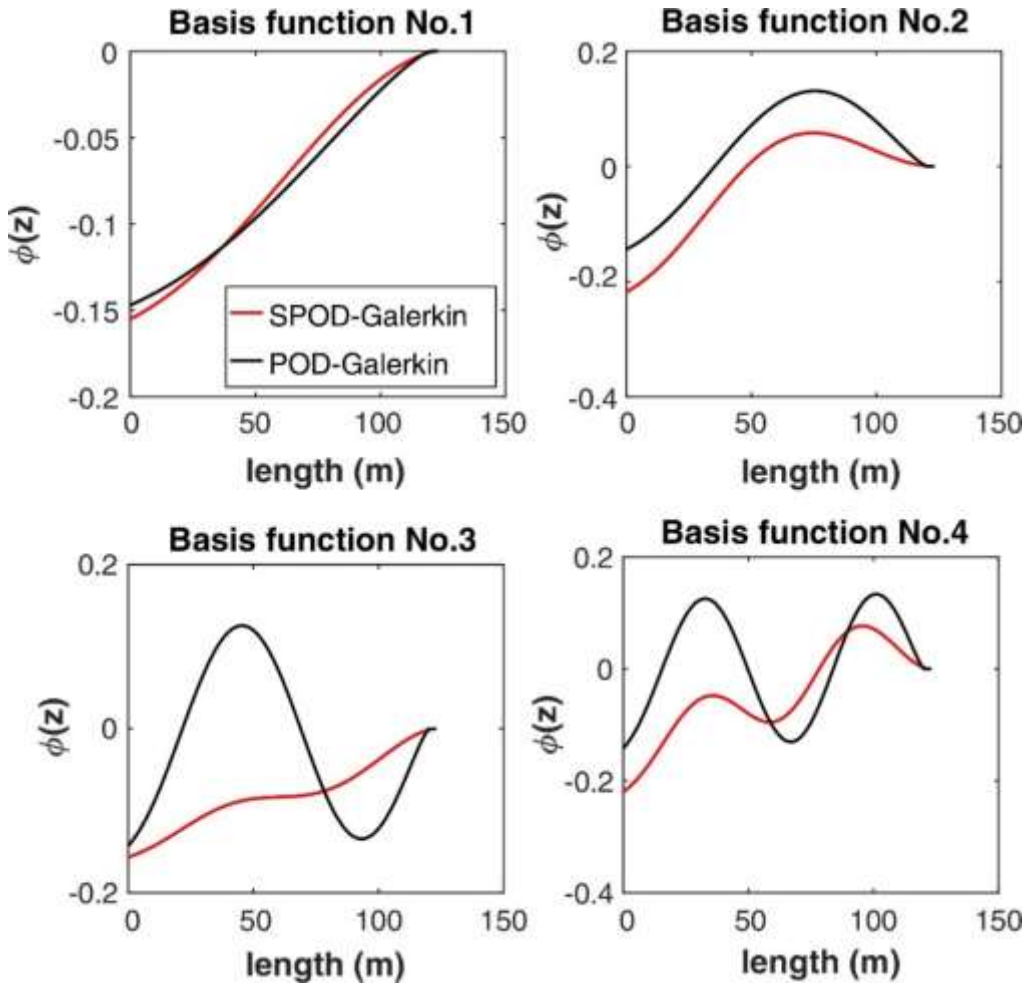

Figure 5: First four basis functions obtained by the SPOD and POD method. Reprinted from Sidhu et al. (2018).

To further demonstrate the performance of the proposed methodology, we compared the ROMs constructed using POD-Galerkin and SPOD-Galerkin methodology based on their relative error, which is calculated as follows (Armaou and Christofides, 2001b):

$$E(t) = \frac{\left\|W_{full} - W_{rom}\right\|_2}{\left\|W_{full}\right\|_2} \tag{25}$$

where $\|W_{rom}\|_2$ and $\left\|W_{full}\right\|_2$ are the l_2 norms of the width profile generated by the ROM and the high-fidelity model, respectively. Fig. 6 shows the relative error profiles E(t) for the ROMs derived using both methodologies. It can be observed that in the beginning, the relative error is high in both the cases, but it decreases with time. This can be attributed to the fact that parabolic PDE systems are characterized by the fast initial dynamics followed by the slow dynamics representing the dominant spatial patterns of the system (Balas, 1979, Chen and Chang, 1992). The computed basis functions only capture the dominant spatial patterns after the fast dynamics of the system become less significant. In practice, the fast dynamics are neglected by selecting the basis functions with large singular values (i.e., high energy) in the Galerkin's projection method, which may lead to the high initial relative error. The relative error profiles show that the ROM constructed using SPOD-Galerkin methodology provides a good approximation to the full-order solution as compared to the ROM derived using POD-Galerkin methodology. The aforementioned results clearly illustrate that the proposed SPOD-Galerkin methodology performs favorably as compared to the standard POD-Galerkin methodology in terms of both accuracy and the number of basis functions required to capture the dominant spatial patterns of a nonlinear parabolic PDE system with time-varying spatial domains. This can be attributed to the fact that the regularization techniques used in SPOD are able to circumvent the spurious spatial

patterns that may arise due to the addition of zeros while transforming the time-varying spatial domain to the time-invariant one.

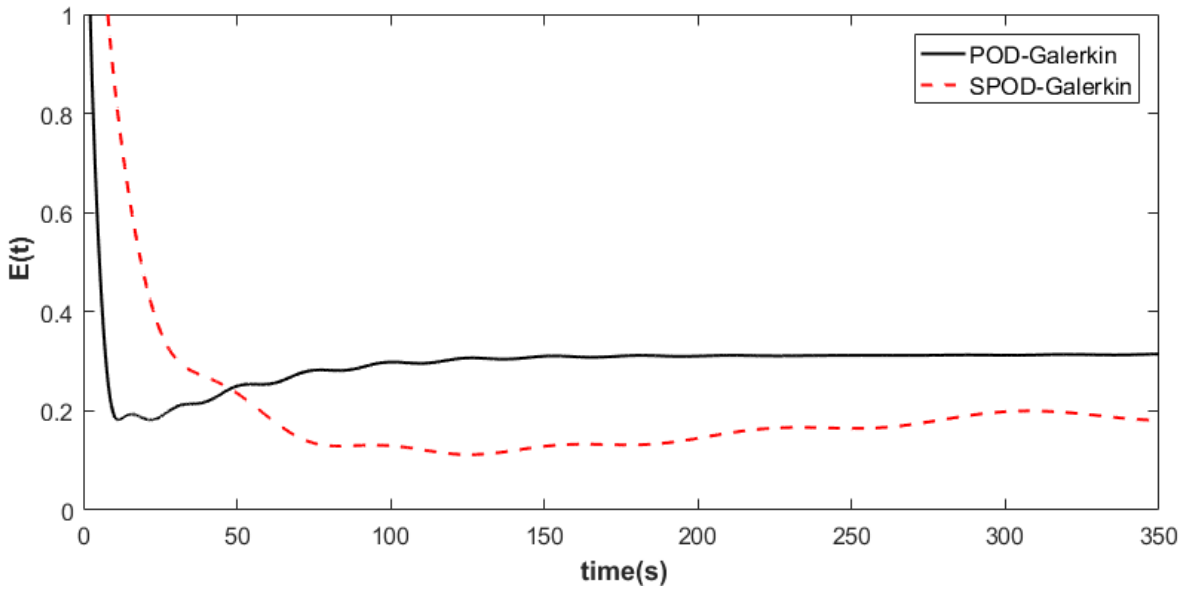

Figure 6: Profiles of the relative error with time for approximate solutions constructed from the ROM obtained by the POD-Galerkin and SPOD-Galerkin methodology.

Furthermore, we compared the performance of SPOD-Galerkin methodology with LPOD-Galerkin methodology proposed by Narasingam et al. (2017) for MOR of nonlinear parabolic PDE systems with moving boundaries. In this work, we divided the temporal domain into 16 temporal subdomains using GOS algorithm and used 6 basis functions for each temporal subdomain. Fig. 7 shows the comparison between the full-order solution and the reduced-order solutions obtained using both of the methodologies at $z = 0$ (i.e., wellbore) and $z = 62.7$ m (i.e., the fracture center). Fig. 8 shows the relative error profiles E(t) for the ROMs derived using each methodology. It can be observed from Fig. 7 and Fig. 8 that the ROMs developed using both of the methodologies are comparable with respect to their ability to approximate the full-order solution. However, we would like to highlight other key features of SPOD-Galerkin methodology. First, it does not require partitioning of the temporal domain into subdomains and the computation of local basis functions for each temporal subdomain. Second, the ROM constructed using SPOD-Galerkin methodology provides a smooth solution profile at every spatial location as shown in Fig. 4 and Fig. 7. This makes it an attractive choice for the design of model-based feedback control systems.

Please note that while solving the system of ODEs derived using LPOD-Galerkin methodology, we used the same initial condition for the high-fidelity model and the ROM for the first temporal subdomain. However, for the following temporal subdomains in LPOD-Galerkin methodology, the initial guess was obtained by using the approximated solution of the previous temporal subdomain. Also, the proposed methodology does not require an analytical expression describing how the spatial domain changes with time, which is often difficult to obtain if we have to deal with a moving boundary problem defined on nontrivial geometry. Instead, such

information is numerically obtained as a part of the solution to be determined along with other in-domain solutions from the high-fidelity simulation.

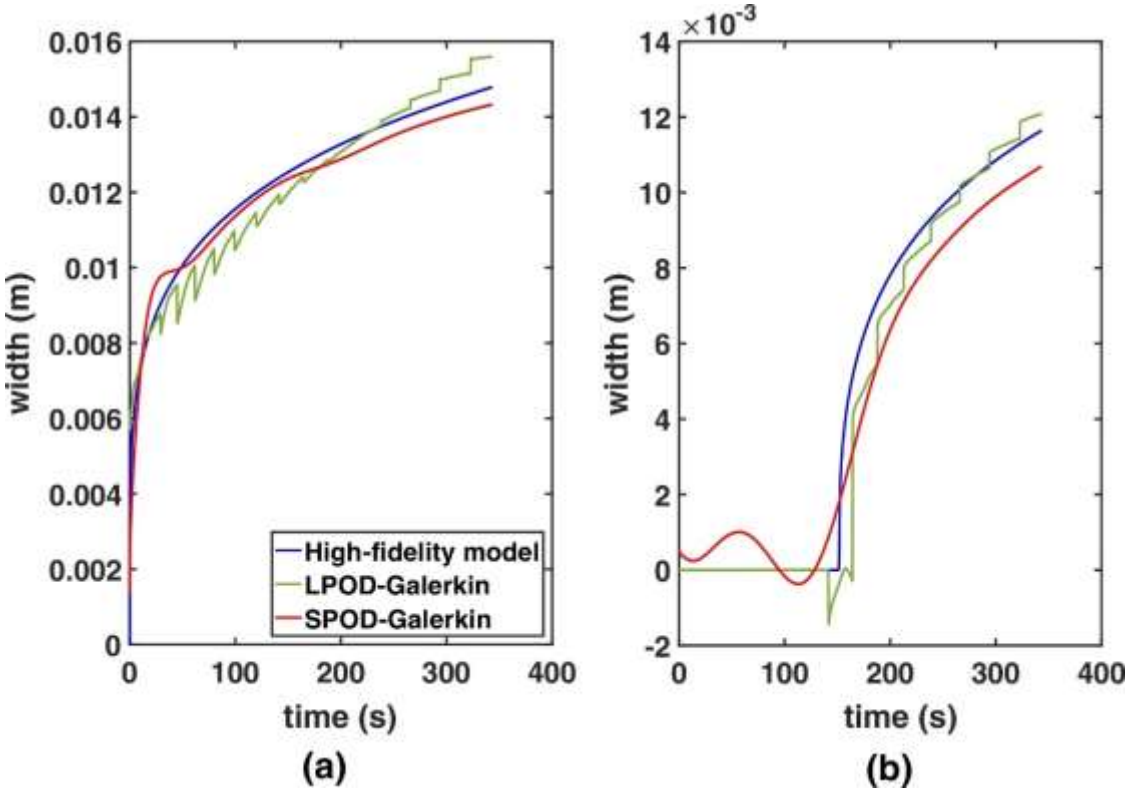

Figure 7: Comparison of width profiles obtained at two different spatial locations, (a) $z = 0$ and (b) $z = 62.7$ m, from the full-order model and ROMs obtained by the SPOD-Galerkin and LPOD-Galerkin methodology. Reprinted from Sidhu et al. (2018).

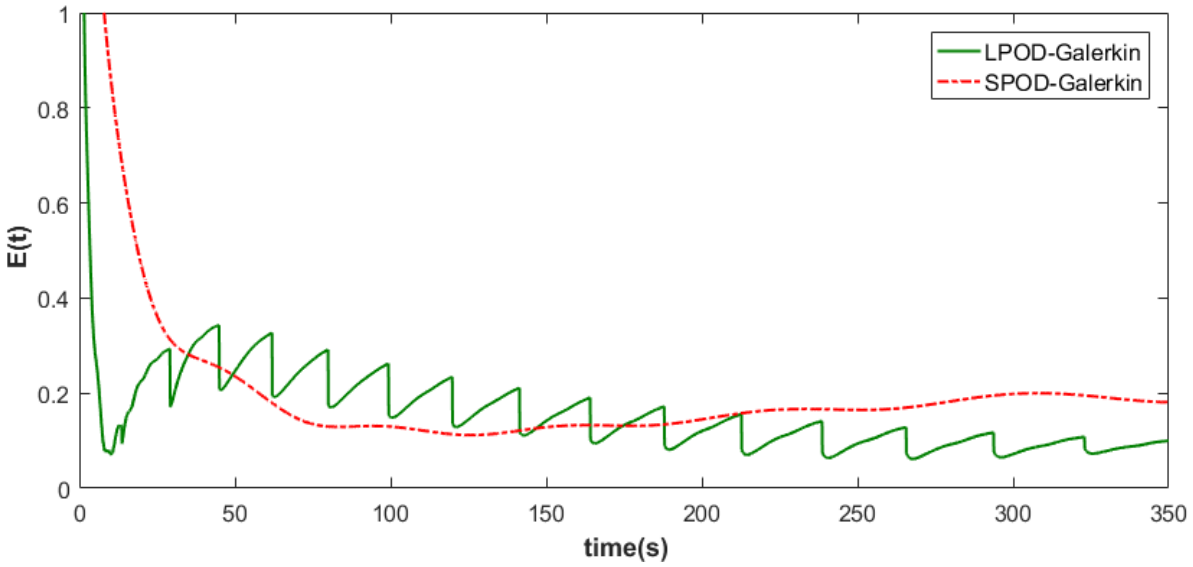

Figure 8: Profiles of the relative error with time for approximate solutions constructed from the ROM obtained by the LPOD-Galerkin and SPOD-Galerkin methodology.

7. CONCLUSIONS AND FUTURE WORK

In this study, SPOD-Galerkin projection methodology is presented to derive ROMs for nonlinear parabolic PDE systems with moving boundaries. Initially, the nonlinear system was solved to obtain the full-order solution by employing a high-order discretization scheme. Then, the proposed SPOD method was employed to generate the basis functions that effectively capture the dominant spatial patterns of the moving boundary system as compared to the basis functions generated by the standard POD method. The obtained basis functions were used in Galerkin's projection method to derive a low-dimensional ODE system, which was subsequently solved using the finite difference scheme to compute an approximate solution to the high-fidelity model. The proposed methodology was successfully applied to develop a ROM for the fracture propagation in a hydraulic fracturing process that was characterized by nonlinear parabolic PDEs with the time-varying spatial domain. In this respect, we demonstrated the proposed methodology performed favorably as compared to the standard POD-Galerkin projection method in terms of both accuracy and the number of basis functions required to capture the dominant spatial patterns of the moving boundary system. Furthermore, we have shown that the accuracy of the ROM developed using the proposed methodology to approximate the high-fidelity model is comparable with the ROM developed using LPOD-Galerkin methodology. The proposed SPOD-Galerkin methodology for developing a ROM for the parabolic PDE system describing the fracture propagation in hydraulic fracturing can be used to design optimal pumping schedules to enhance the productivity of produced wells.

REFERENCES

ALGAZI, V. & SAKRISON, D. 1969. On the optimality of the Karhunen-Loève expansion (Corresp.). *IEEE Transactions on Information Theory,* 15**,** 319-321.

ARKUN, Y. & KAYIHAN, F. 1998. A novel approach to full CD profile control of sheet-forming processes using adaptive PCA and reduced-order IMC design. *Computers & chemical engineering,* 22**,** 945-962.

ARMAOU, A. & CHRISTOFIDES, P. D. 2001a. Computation of empirical eigenfunctions and order reduction for nonlinear parabolic PDE systems with time-dependent spatial domains. *Nonlinear Analysis: Theory, Methods & Applications,* 47**,** 2869-2874.

ARMAOU, A. & CHRISTOFIDES, P. D. 2001b. Finite-dimensional control of nonlinear parabolic PDE systems with time-dependent spatial domains using empirical eigenfunctions. *International Journal of Applied Mathematics and Computer Science,* 11**,** 287-317.

ARMAOU, A. & CHRISTOFIDES, P. D. 2001c. Robust control of parabolic PDE systems with time-dependent spatial domains. *Automatica,* 37**,** 61-69.

BAKER, J. & CHRISTOFIDES, P. D. 2000. Finite-dimensional approximation and control of non-linear parabolic PDE systems. *International Journal of Control,* 73**,** 439-456.

BALAS, M. J. 1979. Feedback control of linear diffusion processes. *Information Linkage Between Applied Mathematics and Industry.* Elsevier.

BANGIA, A. K., BATCHO, P. F., KEVREKIDIS, I. G. & KARNIADAKIS, G. E. 1997. Unsteady two-dimensional flows in complex geometries: Comparative bifurcation studies

with global eigenfunction expansions. *SIAM Journal on Scientific Computing,* 18**,** 775-805.

BENNER, P., GUGERCIN, S. & WILLCOX, K. 2015. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review,* 57**,** 483-531.

BÜCKER, H. M., POLLUL, B. & RASCH, A. 2009. On CFL evolution strategies for implicit upwind methods in linearized Euler equations. *International journal for numerical methods in fluids,* 59**,** 1-18.

CAROTHERS, J. M. 2013. Design-driven, multi-use research agendas to enable applied synthetic biology for global health. *Systems and synthetic biology,* 7**,** 79-86.

CAROTHERS, J. M., GOLER, J. A. & KEASLING, J. D. 2009. Chemical synthesis using synthetic biology. *Current opinion in biotechnology,* 20**,** 498-503.

CHEN, C. C. & CHANG, H. C. 1992. Accelerated disturbance damping of an unknown distributed system by nonlinear feedback. *AIChE Journal,* 38**,** 1461-1476.

DUBEY, A., REALFF, M. J., LEE, J. H. & BOMMARIUS, A. S. 2006. Identifying the interacting positions of a protein using Boolean learning and support vector machines. *Computational biology and chemistry,* 30**,** 268-279.

ECONOMIDES, M. J. & NOLTE, K. G. 1989. *Reservoir stimulation,* Prentice Hall Englewood Cliffs, New Jersey.

EFRON, B., HASTIE, T., JOHNSTONE, I. & TIBSHIRANI, R. 2004. Least angle regression. *The Annals of statistics,* 32**,** 407-499.

FOGLEMAN, M., LUMLEY, J., REMPFER, D. & HAWORTH, D. 2004. Application of the proper orthogonal decomposition to datasets of internal combustion engine flows. *Journal of Turbulence,* 5**,** 1-3.

GAJJAR, S., KULAHCI, M. & PALAZOGLU, A. 2017. Selection of non-zero loadings in sparse principal component analysis. *Chemometrics and Intelligent Laboratory Systems,* 162**,** 160-171.

GLAVASKI, S., MARSDEN, J. E. & MURRAY, R. M. Model reduction, centering, and the Karhunen-Loeve expansion. 1998. IEEE, 2071-2076.

GRAHAM, M. D. & KEVREKIDIS, I. G. 1996. Alternative approaches to the Karhunen-Loeve decomposition for model reduction and data analysis. *Computers & chemical engineering,* 20**,** 495-506.

GUDMUNDSSON, A. 1983. Stress estimates from the length/width ratios of fractures. *Journal of structural geology,* 5**,** 623-626.

HOERL, A. & KENNARD, R. 1988. Ridge regression, in 'Encyclopedia of Statistical Sciences', Vol. 8. Wiley, New York.

HOWARD, G. C. & FAST, C. R. Optimum fluid characteristics for fracture extension. 1957. American Petroleum Institute.

IZADI, M. & DUBLJEVIC, S. 2013. Order-reduction of parabolic PDEs with time-varying domain using empirical eigenfunctions. *AIChE Journal,* 59**,** 4142-4150.

KHORSHIDI, A. & PETERSON, A. A. 2016. Amp: A modular approach to machine learning in atomistic simulations. *Computer Physics Communications,* 207**,** 310-324.

KIESLICH, C. A., SMADBECK, J., KHOURY, G. A. & FLOUDAS, C. A. 2016a. conSSert: consensus SVM model for accurate prediction of ordered secondary structure. *ACS Publications,* 56, 445-461.

KIESLICH, C. A., TAMAMIS, P., GUZMAN, Y. A., ONEL, M. & FLOUDAS, C. A. 2016b. Highly Accurate Structure-Based Prediction of HIV-1 Coreceptor Usage Suggests Intermolecular Interactions Driving Tropism. *PloS one,* 11**,** e0148974.

KUNISCH, K. & VOLKWEIN, S. 1999. Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition. *Journal of optimization theory and applications,* 102**,** 345-371.

KUNISCH, K. & VOLKWEIN, S. 2001. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische mathematik,* 90**,** 117-148.

LEE, J. H. & LEE, J. M. 2006. Approximate dynamic programming based approach to process control and scheduling. *Computers & chemical engineering,* 30**,** 1603-1618.

LEE, J. H. & WONG, W. 2010. Approximate dynamic programming approach for process control. *Journal of Process Control,* 20**,** 1038-1048.

LEE, J. M. & LEE, J. H. 2005. Approximate dynamic programming-based approaches for input–output data-driven control of nonlinear processes. *Automatica,* 41**,** 1281-1288.

MCPHEE, J. & YEH, W. W. G. 2008. Groundwater management using model reduction via empirical orthogonal functions. *Journal of Water Resources Planning and Management,* 134**,** 161-170.

MURRAY, W. D. 1959. Numerical and machine solutions of transient heat conduction problems involving phase change. *Journal of Heat Transfer,* 81**,** 106-112.

NAGY, D. A. 1979. Modal representation of geometrically nonlinear behavior by the finite element method. *Computers & Structures,* 10**,** 683-688.

NARASINGAM, A. & KWON, J. S. I. 2017. Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing. *Computers & Chemical Engineering,* 106**,** 501-511.

NARASINGAM, A., SIDDHAMSHETTY, P. & KWON, J. S. I. 2018. Handling Spatial Heterogeneity in Reservoir Parameters Using Proper Orthogonal Decomposition Based Ensemble Kalman Filter for Model-Based Feedback Control of Hydraulic Fracturing. *Industrial & Engineering Chemistry Research,* 57**,** 3977-3989.

NARASINGAM, A., SIDDHAMSHETTY, P. & KWON, J. S. I. 2017. Temporal clustering for order reduction of nonlinear parabolic PDE systems with time-dependent spatial domains: Application to a hydraulic fracturing process. *AIChE Journal,* 63**,** 3818-3831.

NIKOLAOU, M. 2013. Computer-aided process engineering in oil and gas production. *Computers & Chemical Engineering,* 51**,** 96-101.

NOOR, A. K. & PETERS, J. M. 1981. Tracing post-limit-point paths with reduced basis technique. *Computer Methods in Applied Mechanics and Engineering,* 28**,** 217-240.

NOOR, A. K., PETERS, J. M. & ANDERSEN, C. M. 1981. Reduced basis technique for collapse analysis of shells. *AIAA Journal,* 19**,** 393-397.

NORDGREN, R. P. 1972. Propagation of a vertical hydraulic fracture. *Society of Petroleum Engineers Journal,* 12**,** 306-314.

PARK, H. M. & CHO, D. H. 1996. The use of the Karhunen-Loeve decomposition for the modeling of distributed parameter systems. *Chemical Engineering Science,* 51**,** 81-98.

PARK, H. M. & JANG, Y. D. 2000. Control of Burgers equation by means of mode reduction. *International Journal of Engineering Science,* 38**,** 785-805.

PARK, H. M. & LEE, M. W. 1998. An efficient method of solving the Navier–Stokes equations for flow control. *International Journal for Numerical Methods in Engineering,* 41**,** 1133-1151.

PETERSON, J. S. 1989. The reduced basis method for incompressible viscous flow calculations. *SIAM Journal on Scientific and Statistical Computing,* 10**,** 777-786.

PINNAU, R. 2008. Model reduction via proper orthogonal decomposition. *Model Order Reduction: Theory, Research Aspects and Applications.* Springer.

QIN, S. J. 2014. Process data analytics in the era of big data. *AIChE Journal,* 60**,** 3092-3100.

RATHINAM, M. & PETZOLD, L. R. 2002. Dynamic iteration using reduced order models: a method for simulation of large scale modular systems. *SIAM journal on Numerical Analysis,* 40**,** 1446-1474.

RATHINAM, M. & PETZOLD, L. R. 2003. A new look at proper orthogonal decomposition. *SIAM Journal on Numerical Analysis,* 41**,** 1893-1925.

ROWLEY, C. W. & DAWSON, S. T. M. 2017. Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics,* 49**,** 387-417.

SCHMID, P. J. 2010. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics,* 656**,** 5-28.

SHVARTSMAN, S. Y. & KEVREKIDIS, I. G. 1998. Low-dimensional approximation and control of periodic solutions in spatially extended systems. *Physical Review E,* 58**,** 361.

SHVARTSMAN, S. Y., THEODOROPOULOS, C., RICO-MARTıNEZ, R., KEVREKIDIS, I. G., TITI, E. S. & MOUNTZIARIS, T. J. 2000. Order reduction for nonlinear dynamic models of distributed reacting systems. *Journal of Process Control,* 10**,** 177-184.

SIDDHAMSHETTY, P., KWON, J. S. I., LIU, S. & VALKÓ, P. P. 2017a. Feedback control of proppant bank heights during hydraulic fracturing for enhanced productivity in shale formations. *AIChE Journal,* 64, 1638-1650.

SIDDHAMSHETTY, P., WU, K. & KWON, J. S. I. 2018. Optimization of simultaneously propagating multiple fractures in hydraulic fracturing to achieve uniform growth using data-based model reduction. *Chemical Engineering Research & Design,* In Press.

SIDDHAMSHETTY, P., YANG, S. & KWON, J. S. I. 2017b. Modeling of hydraulic fracturing and designing of online pumping schedules to achieve uniform proppant concentration in conventional oil reservoirs. *Computers & Chemical Engineering,* 114, 306-317.

SIDHU, H. S., NARASINGAM, A., SIDDHAMSHETTY, P. & KWON, J. S. I. 2018. Model order reduction of nonlinear parabolic PDE systems with moving boundaries using sparse proper orthogonal decomposition: Application to hydraulic fracturing. *Computers & Chemical Engineering,* 112**,** 92-100.

SIROVICH, L. 1987a. Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quarterly of applied mathematics,* 45**,** 561-571.

SIROVICH, L. 1987b. Turbulence and the dynamics of coherent structures. II. Symmetries and transformations. *Quarterly of Applied mathematics,* 45**,** 573-582.

SNEDDON, I. N. & ELLIOT, H. A. 1946. The opening of a Griffith crack under internal pressure. *Quarterly of Applied Mathematics,* 4**,** 262-267.

TIBSHIRANI, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 267-288.

VAN OVERSCHEE, P. & DE MOOR, B. 1994. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica,* 30**,** 75-93.

VERHAEGEN, M. & DEWILDE, P. 1992. Subspace model identification part 1. The output-error state-space model identification class of algorithms. *International journal of control,* 56**,** 1187-1210.

WATTERS, L. T. & DUNN-NORMAN, S. 1998. *Petroleum well construction*, John Wiley & Sons.

WILLCOX, K. & PERAIRE, J. 2002. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal,* 40**,** 2323-2330.

WILSON, Z. T. & SAHINIDIS, N. V. 2017. The ALAMO approach to machine learning. *Computers & Chemical Engineering,* 106**,** 785-795.

YANG, S., SIDDHAMSHETTY, P. & KWON, J. S.-I. 2017. Optimal pumping schedule design to achieve a uniform proppant concentration level in hydraulic fracturing. *Computers & Chemical Engineering,* 101**,** 138-147.

YUEN, W. W. & KLEINMAN, A. M. 1980. Application of a variable time-step finite-difference method for the one-dimensional melting problem including the effect of subcooling. *AIChE Journal,* 26**,** 828-832.

ZHOU, X. G., ZHANG, X. S., WANG, X., DAI, Y. C. & YUAN, W. K. 2001. Optimal control of batch electrochemical reactor using K–L expansion. *Chemical engineering science,* 56**,** 1485-1490.

ZOU, H. & HASTIE, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 67**,** 301-320.

ZOU, H., HASTIE, T. & TIBSHIRANI, R. 2006. Sparse principal component analysis. *Journal of computational and graphical statistics,* 15**,** 265-286.