Missing point estimation in models described by proper orthogonal decomposition

Patricia Astrid, *Member, IEEE,* Siep Weiland, Karen Willcox, Ton Backx

Abstract

This paper presents a new method of Missing Point Estimation (MPE) to derive efficient reduced-order models for large-scale parameter-varying systems. Such systems often result from the discretization of nonlinear partial differential equations. A projection-based model reduction framework is used where projection spaces are inferred from proper orthogonal decompositions of data-dependent correlation operators. The key contribution of the MPE method is to perform online computations efficiently by computing Galerkin projections over a restricted subset of the spatial domain. Quantitative criteria for optimally selecting such a spatial subset are proposed and the resulting optimization problem is solved using an efficient heuristic method. The effectiveness of the MPE method is demonstrated by applying it to a nonlinear computational fluid dynamic model of an industrial glass furnace. For this example, the Galerkin projection can be computed using only 25% of the spatial grid points without compromising the accuracy of the reduced model.

Index Terms

model reduction, proper orthogonal decomposition, time-varying systems, parameter-varying systems

I. Introduction

This paper presents a novel reduced-order modeling strategy for large-scale parameter-varying systems. The proposed method uses selective spatial sampling to yield models of low order that can be solved efficiently in online computations. Such systems often result from the discretization of nonlinear partial differential equations (PDEs), ordinary differential equations (ODEs) and differential algebraic equations (DAEs), and have many applications of practical interest, including computational fluid dynamic (CFD) models and Electronic Design Automation models.

In recent years, simulation capabilities for systems governed by PDEs have reached a considerable level of maturity, particularly with regard to the development and use of commercial packages. For example, in the glass

This work was supported by the EET-REGLA grant EET K99002 of the Dutch Programme Economy, Ecology, Technology and the Netherlands Organizations of Scientific Research

P. Astrid is with Shell Global Solutions International B.V., P.O. Box 38000, 1030 BN Amsterdam, the Netherlands. S. Weiland is with the Department of Electrical Engineering, Control Systems Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. K. Willcox is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Rm. 37-447, Cambridge, MA 02139, U.S.A. T. Backx is with the Department of Electrical Engineering, Control Systems Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

furnace industry, such packages have served as tools for modeling physical systems, for analyzing the performance and stability of systems, and for computer aided engineering designs [6], [14], [1].

In the case of spatial-temporal systems, numerical simulation is typically achieved by spatial discretization of the governing PDEs using, for example, finite volume or finite element methods. The spatial discretization procedure leads to large-scale systems of ordinary differential equations (ODEs), typically of order $10^3 - 10^8$, depending on the complexity of the governing equations and the desired level of accuracy. The underlying governing equations are generally nonlinear and the model parameters are often functions of state variables (hence time-varying), which adds considerably to the degree of complexity [43], [2], [4], [5]. Thus, for problems of practical interest, the computational effort required to simulate these systems is substantial.

Both the large dimension of the system and the large computational requirements render such simulation models inadequate for control design and online optimization. To facilitate model-based control design, it is essential to have accurate low-order models that are significantly faster to solve than the original model. A reduced-order model can be derived using a projection-based framework, in which the system variables and governing equations are projected onto low dimensional subspaces. In the context of parameter- and time-varying systems, the resulting system is of reduced order, but is not necessarily computationally efficient to solve. This is because online simulations of the reduced models still require computations on the large scale.

The contribution of this paper is a new method – the Missing Point Estimation (MPE) approach – that achieves the goals of low model order, efficient simulation and accurate predictions using a projection-based model reduction framework combined with selective spatial sampling to efficiently perform the necessary online computations. Given a simulated or measured signal that evolves both in time and space, we first characterize a basis of spatial functions that achieves optimal approximation properties with respect to the measured signal by considering partial (finite) sums of spectral expansions. This so called ‘proper orthogonal’ (or ‘principal component’) basis has as its distinguishing features that it is data dependent, physically relevant, computable and optimal in a well defined sense. To enhance the computational speed of the resulting reduced model, we propose to sample the spatial domain in such a way that orthonormality of basis functions is preserved in the sampled domain by the introduction of a suitable bilinear form. It is shown that this bilinear form plays a crucial role in questions on exact signal reconstruction from sampled observations (missing point estimations) and on deriving expressions for alias errors in approximate signal reconstructions. An algorithm is then derived for the selection of optimal samples using a heuristic optimization method to minimize the alias error in any interpolated signal in the sampled domain. We apply the theoretical results to a model of an industrial glass feeder. The merits of the procedure for model reduction and the enhancement of computational speed by missing point estimations are demonstrated in a rather complicated heat transition mechanism in a glass feeder.

A. Previous work

The proper orthogonal decomposition (POD) technique [24], [38], [26] derives an empirical basis from a collection of simulation or experimental data. In recent years, the POD method has seen widespread, successful application to model reduction for CFD applications. For this reason, it is chosen, in combination with a glass furnace control example, as a case study for the methodology presented in this paper. However, the MPE approach is applicable to other projection-based model reduction techniques, such as balanced truncation [31], [37], [15], [22] and Krylov subspace methods [20], [18].

Improvements in the numerical efficiency of reduced-order models have been focused mostly on the development of alternative, more efficient methods to compute the approximating basis functions [30], [15], [41]. Efforts to address the problem of high computational cost for simulation of nonlinear and time-varying systems include the work of Rewienski and White, who use a trajectory piecewise-linear approximation scheme [35], [36]. In this approach, a nonlinear system is represented as a weighted combination of linear models, which are obtained by linearizing the nonlinear system at selected points along a state trajectory. This approach has been successfully applied to nonlinear analogue circuits and micromachined devices [35], [36], and to a nonlinear CFD model of a supersonic diffuser [21]. Other methods that do not rely on linearization of the system have been proposed more recently in [10], [4], and [5]. In [10], the approach for accelerating the reduced model simulation is similar to that proposed here, i.e., by constructing the nonlinear behavior using a subset of the original equations. In that work, the choice of the selected original equations is made based on *a priori* knowledge rather than using a systematic approach. The MPE approach was discussed in [4] and [5] in the context of computational fluid dynamics models, while preliminary work on MPE was presented in [2]. The mathematical foundation of MPE can be traced back to classical sampling theory [45], [33], [32] or, more specifically, to the problem of approximate signal recovery from inhomogeneously sampled multidimensional signals. Some signal recovery results from [13], [33], [23], [32] are generalized here to non-band-limited spectral expansions of multidimensional signals by arbitrary orthonormal POD bases. Quantitative criteria are introduced in [4] and [5] as a means to select suitable sample points so as to minimize alias effects in interpolations.

B. *Paper organization*

The paper is organized as follows. Preliminaries and notational issues are collected in Section II. The method of reduced-order modeling via POD is introduced in Section III. Section IV describes the construction of reduced-order models using selective spatial sampling. A heuristic optimization procedure to select the sample points is given in Section V. The concepts are implemented on a simulation model of a glass feeder, which is a section in a glass furnace. The results of the implementations are presented in Section VI. Conclusions are given in Section VII.

II. Preliminaries and notation

Let $\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times p}$ denote the field of real numbers, the sets of positive reals, real n-vectors and real $n\times p$ matrices, respectively. For $A\in\mathbb{R}^{n\times p}$, $A^\top$ is the transpose of A, $A^{-L}=(A^\top A)^{-1}A^\top$ and $A^{-R}=A^\top(AA^\top)^{-1}$ are the left and right inverses of A, respectively, assuming the inverses exist. The inner product and norm of an inner product space $\mathcal{X}$ are denoted as $(\cdot,\cdot)$ and $\|\cdot\|$, respectively, or as $(\cdot,\cdot)_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{X}}$ if the context requires indicating the underlying space. If $\mathbb{X}$ is a Lebesgue measurable set then the space of all equivalence classes (i.e. pointwise equality almost everywhere on $\mathbb{X}$) of measurable functions $f : \mathbb{X} \to \mathbb{Y}$, which are square integrable over $\mathbb{X}$ is denoted by $L_2(\mathbb{X}, \mathbb{Y})$. This is a Hilbert space when equipped with its usual inner product. The restriction of f to a subset $\mathbb{X}_0 \subset \mathbb{X}$ is the mapping $\tilde{f} : \mathbb{X}_0 \to \mathbb{Y}$ defined by $\tilde{f}(x) = f(x)$ with $x \in \mathbb{X}_0$ and is also denoted by $\tilde{f} = f|_{\mathbb{X}_0}$. The boundary of a set $\mathbb{X}_0 \subseteq \mathbb{X}$ is the set theoretic difference between its closure and interior and denoted by $\partial\mathbb{X}_0$. The function col stacks the elements in its argument as a column vector.

III. Proper orthogonal decomposition

A. The POD basis problem

In the study of dynamical systems that evolve in space and time we consider signals w that depend on a spatial variable x and on time t. That is, we consider signals $w : \mathbb{X} \times \mathbb{T} \to \mathbb{W}$ where $\mathbb{X}$ is a bounded spatial domain in a d-dimensional Euclidean space $\mathbb{R}^{\mathrm{d}}$, $\mathbb{T} \subseteq \mathbb{R}$ denotes the set of time instants of interest and $\mathbb{W}$ is a normed vector space of dimension $\dim(\mathbb{W}) = \mathrm{w}$ in which $w(x, t)$ assumes its values. For any such function w and time instant $t \in \mathbb{T}$, the map $\mathbf{w}(t) : x \mapsto w(x, t)$ is assumed to be an element of some Hilbert space $\mathcal{X}$ of functions defined on $\mathbb{X}$. We let $\mathcal{W} = L_2(\mathbb{T}, \mathcal{X})$ be the space of all functions $t \mapsto \mathbf{w}(t)$ that map $\mathbb{T}$ into $\mathcal{X}$ and that are square integrable in the sense that

$$\|\mathbf{w}\|_{\mathcal{W}} = \left(\int_{\mathbb{T}} \|\mathbf{w}(t)\|^2_{\mathcal{X}} \ \mathrm{d}t \right)^{1/2}$$

is finite. $\mathcal{W}$ becomes a Hilbert space with inner product

$$(\mathbf{v}, \mathbf{w})_{\mathcal{W}} = \int_{\mathbb{T}} (\mathbf{v}(t), \mathbf{w}(t))_{\mathcal{X}} \ \mathrm{d}t$$

where $\mathbf{v}, \mathbf{w} \in \mathcal{W}$. We will consider dynamical spatial-temporal systems $\mathcal{B}$ that are subsets of $\mathcal{W}$ and view elements of $\mathcal{B}$ as time depending functions $\mathbf{w}$ where, for fixed time $t \in \mathbb{T}$, the expression $\mathbf{w}(t)$ stands for the function $w(\cdot, t)$ in $\mathcal{X}$ that acts on the spatial domain $\mathbb{X}$.

Spectral decompositions of signals by (infinite) sequences of orthogonal functions underlie many numerical techniques of approximation. A central theme in this paper is therefore the construction of an (empirical) orthonormal basis of the Hilbert space $\mathcal{X}$ that proves useful for the representation and approximation of signals. Suppose that $\mathcal{X}$ is a separable Hilbert space so that it admits [27] a countable orthonormal basis $\{\varphi_k, k \in \mathbb{I}\}$ where the index set $\mathbb{I} = \{1, 2, \ldots\}$ has cardinality equal to the (possibly infinite) dimension of $\mathcal{X}$. Given such an orthonormal basis, we introduce for any $\mathbf{w} \in \mathcal{W}$ the spectral expansion

$$w(x, t) = \sum_{k \in \mathbb{I}} a_k(t)\varphi_k(x), \qquad x \in \mathbb{X},\ t \in \mathbb{T}, \tag{1}$$

where the expansion coefficients are given by

$$a_k(t) = (\mathbf{w}(t), \varphi_k)_{\mathcal{X}}\,, \qquad k \in \mathbb{I},\ t \in \mathbb{T}. \tag{2}$$

The approximation of $w(x, t)$ using a finite sum retaining the first n terms in (1) is denoted $w_n(x, t)$.

Definition 1: Given an observation $\mathbf{w} \in \mathcal{W}$, a *POD basis* is an orthonormal basis $\{\varphi_k, k \in \mathbb{I}\}$ of $\mathcal{X}$ with the property that the error

$$\|\mathbf{w} - \mathbf{w}_n\|^2_{\mathcal{W}} := \int_{\mathbb{T}} \|\mathbf{w}(t) - \sum_{k=1}^{n} (\mathbf{w}(t), \varphi_k)_{\mathcal{X}}\, \varphi_k\|^2_{\mathcal{X}}\ \mathrm{d}t \tag{3}$$

is minimal for all values of $n > 0$.

A POD basis therefore has the property that any truncation in the expansion (1) of w is an optimal approximation to w in the normed space $\mathcal{W}$. POD bases can be characterized and computed by means of an eigenvalue decomposition of a suitably defined correlation operator. Precisely, define, for $\mathbf{w} \in \mathcal{W}$, the *data-correlation operator* $C_w : \mathcal{X} \to \mathcal{X}$ by

$$(\psi_1, C_w\psi_2)_{\mathcal{X}} := \int_{\mathbb{T}} (\psi_1, \mathbf{w}(t))_{\mathcal{X}} \cdot (\psi_2, \mathbf{w}(t))_{\mathcal{X}}\ \mathrm{d}t \qquad \psi_1, \psi_2 \in \mathcal{X}. \tag{4}$$

It is immediate that C_w is a well defined linear, bounded, self-adjoint and non-negative operator on $\mathcal{X}$. If $\mathcal{X}$ happens to be finite dimensional, then C_w is simply a non-negative definite matrix. It can be shown [24], [8], [29] that an orthonormal basis $\{\varphi_k, k \in \mathbb{I}\}$ of $\mathbb{X}$ is a POD basis if and only if φ_k, $k \in \mathbb{I}$, are normalized eigenfunctions corresponding to the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ of C_w.

The truncation level n depends on the problem at hand and can be determined in many ways. We introduce the criterion

$$P_n = \frac{\sum_{k=1}^{n} \lambda_k}{\sum_{k\in\mathbb{I}} \lambda_k}. \tag{5}$$

The *correlation tolerance* $0 < P_{\text{tol}} < 1$ then defines the truncation level as the minimal value n for which $P_n \geq P_{\text{tol}}$. In typical applications $P_{\text{tol}} = 0.99$ [24]. Once n has been set, the reduced-order model can be constructed by conducting a Galerkin projection. For models describing diffusion phenomena in computational fluid dynamics, a number of case studies ([26], [24], [3]) show that the order can be reduced to as low as 1% of the order of the original model.

B. Approximate solutions and Galerkin projections

In most applications, spatial-temporal systems are described by partial differential equations and a typical evolution equation can be written in the form

$$D_t w(x,t) = F\left(w(x,t), \ldots, D^p_x w(x,t), \ldots\right) \tag{6}$$

subject to boundary and initial conditions on (subsets of) the (sufficiently smooth) boundary set $\partial(\mathbb{X} \times \mathbb{T})$. Here, F is some function, D_t is the partial derivative operator $D_t = \frac{\partial}{\partial t}$ and D^p_x is a compact notation for an arbitrary partial derivative in the spatial coordinate x. Precisely, D^p_x is defined by

$$D^p_x w = D^{p_1}_{x_1} D^{p_2}_{x_2} \cdots D^{p_d}_{x_d} w = \frac{\partial^{|p|} w}{\partial x_1^{p_1} \cdots \partial x_d^{p_d}},$$

where p denotes a multi-index $p = (p_1, \ldots, p_d)$ that consists of d non-negative integers p_i, and where $|p|$, the *length* of the multi-index, is defined as $|p| = \sum_{i=1}^{d} p_i$. By convention, $D^{0,\cdots,0}_x w = w$. For the important class of *linear* spatial-temporal systems, (6) simplifies to

$$D_t w(x,t) = \sum_{0 \leq |p| \leq m} c_p D_x^p w(x,t), \tag{7}$$

where $\sum_{0\leq|p|\leq m} c_p D_x^p$ is a polynomial differential operator with real coefficients c_p and degree m in the partial derivative operator D_x.

Throughout it is assumed that solutions w of (6) are continuous functions on the closure of $\mathbb{X} \times \mathbb{T}$ and are sufficiently often continuously differentiable as elements of the Hilbert space $\mathcal{W}$. The notion of an *approximate solution* of (6) will be defined in terms of projections of either the *solution space* of (6) or the *residual* associated with (6) onto a finite dimensional subspace. Specifically, let $\mathcal{S}$ be a *finite dimensional* subspace of $\mathcal{X}$, $\dim(\mathcal{S}) = n$, and suppose that $\mathtt{w} = 1$. That is, we consider scalar valued functions in (6) only.

Definition 2: Let $\mathcal{S}$ be an n dimensional subspace of $\mathcal{X}$ and let $\mathcal{W}_n = L_2(\mathbb{T}, \mathcal{S})$. An element $\mathbf{w}_n \in \mathcal{W}_n$ is an

- *approximate weak solution* of order n of (6) if

$$(D_t \mathbf{w}_n(t), \varphi)_{\mathcal{S}} = F\left((\mathbf{w}_n(t), \varphi)_{\mathcal{S}}, \ldots, (D_x^p \mathbf{w}_n(t), \varphi)_{\mathcal{S}}, \ldots\right) \tag{8}$$

  for all $\varphi \in \mathcal{S}$ and almost all $t \in \mathbb{T}$.

- *Galerkin approximate solution* of order n of (6) if

$$(D_t \mathbf{w}_n(t), \varphi)_{\mathcal{S}} = \left(F\left(\mathbf{w}_n(t), \ldots, D_x^p \mathbf{w}_n(t), \ldots\right), \varphi\right)_{\mathcal{S}} \tag{9}$$

  for all $\varphi \in \mathcal{S}$ and almost all $t \in \mathbb{T}$.

The set of all approximate weak solutions of order n is denoted $\mathcal{B}_n^{\text{weak}}$ and the set of all Galerkin approximate solutions of order n is denoted $\mathcal{B}_n^{\text{Galerkin}}$.

It is important to point out that for *linear systems* the expressions (8) and (9) coincide because of the linearity of F. This means that a projection of the solution space and a projection of the residual associated with a linear PDE (6) coincide and result in the property that $\mathcal{B}_n^{\text{weak}} = \mathcal{B}_n^{\text{Galerkin}}$. For nonlinear systems this is evidently no longer the case. We refer to [25], [40] for a rigorous treatment of solution concepts and Galerkin projections in nonlinear evolutionary PDEs.

In computational fluid dynamics, $\mathcal{S}$ typically consists of finite element approximations of functions in $\mathcal{X}$. We will be particularly interested in Galerkin approximations where $\mathcal{S} = \mathcal{X}_n = \operatorname{span}(\varphi_k, k = 1, \ldots, n)$ where $\{\varphi_k, k \in \mathbb{I}\}$ is a POD basis of $\mathcal{X}$. In that case, elements $\mathbf{w}_n \in \mathcal{W}_n$ assume the form

$$w_n(x,t) = \sum_{k=1}^{n} a_k(t) \varphi_k(x)$$

where the coefficient functions $a_k : \mathbb{T} \to \mathbb{R}$ are square integrable. Condition (8) implies that the expansion coefficients of approximate weak solutions satisfy the ordinary differential equation

$$\dot{a}_k(t) = F(a_k(t), \ldots, \sum_{\ell=1}^{n} a_\ell(t) \left(D_x^p \varphi_\ell, \varphi_k\right)_{\mathcal{S}}, \ldots) \qquad \text{for all } t \in \mathbb{T}, \quad k = 1, \ldots, n. \tag{10}$$

Similarly, (9) implies that the coefficients satisfy

$$\dot{a}_k(t) = \left(F\left(\mathbf{w}_n(t), \ldots, D_x^p \mathbf{w}_n(t), \ldots\right), \varphi \right)_{\mathcal{S}} . \tag{11}$$

In spite of a substantial reduction of model order that can be accomplished in this manner, the computational gain in computing solutions of the reduced-order model may still be modest because the evaluation of the inner products in the right-hand side of (10) or (11) is computationally intensive. Indeed, (10) and (11) amount to evaluating the inner products $(\cdot,\cdot)_{\mathcal{S}}$ over the entire spatial domain, which may become a formidable task in large-scale or high-dimensional systems, and computationally prohibitive in parameter- and time-varying models where the evaluation of these inner products needs to be performed online. This paper proposes a new methodology to accelerate these online computations. In Section IV, it is shown that under suitable conditions, the *same* reduced-order model can be obtained with a much simpler right-hand-side expression than the one in (10).

IV. MISSING POINT ESTIMATIONS AND PARTIAL OBSERVATIONS

The key contribution of this paper is a method to make the approximate models $\mathcal{B}_n^{\text{weak}}$ and $\mathcal{B}_n^{\text{Galerkin}}$ suitable for fast computations. In order to achieve this goal, we consider a sampling of the signal $w : \mathbb{X} \times \mathbb{T} \to \mathbb{W}$. The sampled points can be regarded as measurements or observations of w. We first consider the exact reconstruction of signals from sampled data or sampled measurements by means of an appropriate interpolation of the sampled signal. We then address the approximate recovery of signals from sampled or partial observations.

The methodology presented in this section extends the idea of the missing point estimation (MPE) described in [4], [5], [3], which is based upon the theory of *Gappy POD*, developed by Everson and Sirovich [17]. The gappy POD method has been applied to data reconstruction problems, such as reconstruction of facial images [17], flow structure [12], [42], and flow sensing [44]. The key idea of gappy POD is to estimate the expansion coefficients $\{a_k, k \in \mathbb{I}\}$ from incomplete (gappy) data. As such, this estimation problem belongs to the realm of signal reconstruction problems that are abundant in signal processing [45], [23], [33], [32]. Results on exact signal reconstruction that are presented here are inspired by signal reconstruction problems from inhomogeneously sampled data. Especially, the results in [13] are generalized here to non-band-limited spectral expansions of multidimensional signals in terms of an arbitrary orthonormal POD basis.

Suppose that $\mathbb{X}_0$ is a *finite* subset of N distinct points $\mathbb{X}_0 = \{x_1, \ldots, x_N\}$ in the domain $\mathbb{X}$ and suppose that a *measurement* or *partial observation* $\tilde{w}$ is available at the collection of the points in $\mathbb{X}_0$. That is, a measurement is a function $\tilde{w} : \mathbb{X}_0 \times \mathbb{T} \to \mathbb{W}$ defined on N spatial samples $\mathbb{X}_0$ and time $\mathbb{T}$ that satisfies the restriction $\tilde{w} = w|_{\mathbb{X}_0 \times \mathbb{T}}$ for some unobserved signal $w : \mathbb{X} \times \mathbb{T} \to \mathbb{W}$. Throughout, tildes and hats will be used to indicate sampled and interpolated signals, respectively. We consider here the problem to *reconstruct* the unobserved signal w from its samples $\tilde{w}$.

Suppose that $\{\varphi_k, k \in \mathbb{I}\}$ is a basis for $\mathcal{X}$ and assume that $\mathcal{X}$ is *either* finite dimensional *or* a set of continuous functions. Let $\tilde{\varphi}_k := \varphi_k|_{\mathbb{X}_0}$ denote the restriction of the basis function φ_k to the samples $\mathbb{X}_0$. Define, for $n > 0$

and a set $\{\tilde{a}_k, k = 1, \dots, n\}$ of coefficient functions $\tilde{a}_k : \mathbb{T} \to \mathbb{R}$ the expansion

$$\tilde{w}_n(x,t) := \sum_{k=1}^{n} \tilde{a}_k(t)\tilde{\varphi}_k(x), \quad x \in \mathbb{X}_0,\ t \in \mathbb{T} \tag{12}$$

together with its *interpolation*

$$\hat{w}_n(x,t) := \sum_{k=1}^{n} \tilde{a}_k(t)\varphi_k(x), \quad x \in \mathbb{X}, t \in \mathbb{T}. \tag{13}$$

The name 'interpolation' is justified since $\hat{w}_n$ coincides with $\tilde{w}_n$ on the sample points in $\mathbb{X}_0$.

Introduce the $N \times n$ real matrix $\tilde{\Phi}$ that consists of the samples of the first n basis functions φ_k, defined by

$$\tilde{\Phi} := \begin{pmatrix} \varphi_1(x_1) & \dots & \varphi_n(x_1) \\ \vdots & & \vdots \\ \varphi_1(x_N) & \dots & \varphi_n(x_N) \end{pmatrix}. \tag{14}$$

Then (12) can be written in matrix form as

$$\tilde{\mathbf{w}}_n(t) := \tilde{\Phi}\tilde{\mathbf{a}}(t), \tag{15}$$

where $\tilde{\mathbf{a}}(t) = \mathrm{col}(\tilde{a}_1(t), \dots, \tilde{a}_n(t))$ is the vector of expansion coefficients and $\tilde{\mathbf{w}}_n(t) = \mathrm{col}(\tilde{w}_n(x_1,t), \dots, \tilde{w}_n(x_N,t))$ is the vector of samples at time t.

We define for any $v, w \in \mathcal{X}$ the bilinear form:

$$(v,w)_N := \sum_{i,j=1}^{N} v(x_i)q_{i,j}w(x_j), \tag{16}$$

where $q_{i,j}$ is the (i,j)th entry of the $N \times N$ real symmetric matrix

$$Q := \tilde{\Phi}(\tilde{\Phi}^\top\tilde{\Phi})^{-1}(\tilde{\Phi}^\top\tilde{\Phi})^{-1}\tilde{\Phi}^\top,$$

and where we assume that $n \leq N$ is such that $\tilde{\Phi}$ is *injective*. Since $Q = Q^\top \geq 0$, we have that $\|w\|_N := (w,w)_N^{1/2}$ defines a semi-norm on $\mathcal{X}$. Moreover, since $(v,w)_N$ only depends on the samples $\tilde{v} = v|_{\mathbb{X}_0}$ and $\tilde{w} = w|_{\mathbb{X}_0}$ we also write, with some abuse of notation, $(\tilde{v},\tilde{w})_N$ for the right-hand side of (16) and view $(\cdot,\cdot)_N$ as a bilinear form on the *sampled* functions $\tilde{v}$ and $\tilde{w}$.

The introduction of the bilinear form (16) enables us to formulate both exact and approximate reconstruction of the signal w, as described in the following subsections.

A. Exact reconstruction

The following lemma motivates the importance of (16), relating the bilinear form on the sampled functions to the inner product in $\mathcal{X}$.

Lemma 3: If $\tilde{\Phi}$ defined in (14) is injective, then

$$(v,w)_N = (v,w)_{\mathcal{X}} \qquad \text{for all } v, w \in \mathcal{X}_n, \tag{17}$$

where $\mathcal{X}_n = \mathrm{span}(\varphi_k, k = 1, \dots, n)$.

Proof: Let $v, w \in \mathcal{X}_n$. Then, with $a_k = (v, \varphi_k)$, $b_k = (w, \varphi_k)$, $\mathbf{a} = \mathrm{col}(a_1, \ldots, a_n)$ and $\mathbf{b} = \mathrm{col}(b_1, \ldots, b_n)$, there holds $v = \sum_{k=1}^{n} a_k \varphi_k$, $w = \sum_{\ell=1}^{n} b_\ell \varphi_\ell$ and

$$(v, w)_{\mathcal{X}} = \left(\sum_{k=1}^{n} a_k \varphi_k, \sum_{\ell=1}^{n} b_\ell \varphi_\ell \right) = \sum_{k=1}^{n} a_k b_k = \mathbf{a}^\top \mathbf{b} = \tilde{\mathbf{v}}^\top \left(\tilde{\Phi}^{-L} \right)^\top \tilde{\Phi}^{-L} \tilde{\mathbf{w}} = \tilde{\mathbf{v}}^\top Q \tilde{\mathbf{w}} = (v, w)_N \,,$$

where in the fourth equality we used that (15) and the injectivity of $\tilde{\Phi}$ implies that the coefficients $\mathbf{a}$ and $\mathbf{b}$ are uniquely determined by $\tilde{\Phi}^{-L}\tilde{\mathbf{v}}$ and $\tilde{\Phi}^{-L}\tilde{\mathbf{w}}$, respectively. ■

Lemma 3 implies that (16) defines an inner product on the space $\mathcal{X}_n$ whenever $\tilde{\Phi}$ is injective. In particular, for $v, w \in \mathcal{X}_n$, this means that $(v, w)_{\mathcal{X}}$ can equivalently be evaluated on the sampled values $\tilde{v} = v|_{\mathbb{X}_0}$ and $\tilde{w} = w|_{\mathbb{X}_0}$ by employing (16). Furthermore, setting $v = \varphi_k$ and $w = \varphi_\ell$ in (17), implies that $\{\varphi_k, k = 1, \ldots, n\}$ is also an orthonormal basis of $\mathcal{X}_n$ with respect to the inner product (16).

Now define

$$\tilde{a}_k(t) = (\mathbf{w}(t), \varphi_k)_N \,, \qquad k = 1, \ldots, n, \quad t \in \mathbb{T}, \tag{18}$$

and let $\hat{w}_n$ be the corresponding interpolant (13). Using the definition in (18) and the result from Lemma 3, the following theorem provides the condition for exact reconstruction of the signal w from its partial observations.

Theorem 4: Let $\mathbb{X}_0 = \{x_1, \ldots, x_N\}$ be a set of N distinct samples, and let $\{\varphi_k, k \in \mathbb{I}\}$ be an orthonormal basis of $\mathcal{X}$. Suppose $\tilde{\Phi}$ defined in (14) has rank n. If $\mathbf{w}(t) = w(\cdot, t) \in \mathcal{X}_n = \mathrm{span}(\varphi_1, \ldots, \varphi_n)$ for $t \in \mathbb{T}$, then w can be reconstructed exactly from its partial observations $\tilde{w} = w|_{\mathbb{X}_0 \times \mathbb{T}}$ in that

$$\hat{w}_n(x, t) = w(x, t), \qquad \text{for all } x \in \mathbb{X}, \ t \in \mathbb{T}$$

by taking the expansion coefficients (18) in the interpolant (13). In particular, any signal w in the approximate models $\mathcal{B}_n^{\text{weak}}$ and $\mathcal{B}_n^{\text{Galerkin}}$ can be reconstructed exactly in this way.

Proof: If $w(\cdot, t) \in \mathcal{X}_n$ then $w(x, t) = \sum_{k=1}^{n} a_k(t)\varphi_k(x)$ so that its samples $\tilde{w}(x, t) = \sum_{k=1}^{n} a_k(t)\tilde{\varphi}_k(x)$. Hence, using vector notation, we can write $\tilde{\mathbf{w}}(t) = \tilde{\Phi}\mathbf{a}(t)$. By the injectivity of $\tilde{\Phi}$, $\mathbf{a}(t)$ is uniquely defined by the left inverse $\mathbf{a}(t) = \tilde{\Phi}^{-L}\tilde{\mathbf{w}}(t)$, where $\tilde{\Phi}^{-L} = (\tilde{\Phi}^\top\tilde{\Phi})^{-1}\tilde{\Phi}^\top$. Using the definition of Q, it follows that $\mathbf{a}(t) = \tilde{\Phi}^\top Q \tilde{\mathbf{w}}(t)$ so that, by (16), its kth entry reads $a_k(t) = (\mathbf{w}(t), \varphi_k)_N$. Consequently, with $\tilde{a}_k(t)$ defined by (18), we have $\tilde{a}_k(t) = a_k(t)$ for all $t \in \mathbb{T}$ and for all $k = 1, \ldots, n$. But then the interpolant (13) reads

$$\hat{w}_n(x, t) = \sum_{k=1}^{n} \tilde{a}_k(t)\varphi_k(x) = \sum_{k=1}^{n} a_k(t)\varphi_k(x) = w(x, t)$$

for all $x \in \mathbb{X}$ and all $t \in \mathbb{T}$, which gives the result. ■

Thus, provided that the unobserved signal $w(\cdot, t)$ belongs to $\mathcal{X}_n$ for all $t \in \mathbb{T}$, this signal can be reconstructed perfectly from its N samples $\tilde{w}$ by taking the spectral coefficients (18) in the interpolant (13). It is important to observe that, by (16), the coefficients $\tilde{a}_k$ only depend on the sampled signals $\tilde{w}$ and $\tilde{\varphi}_k$. In particular, no information of w other than its partial observations is necessary to recover w from its samples. With harmonic basis functions and equidistant samples, Theorem 4 specializes to the classical Shannon sampling theorem that has been profoundly studied in information and sampling theory [45], [33], [32]. Following standard engineering terminology, the minimum value of n for which $w \in \mathcal{X}_n$ is the *bandwidth* of w. If no such n exists, the signal is said to be non-band-limited. Theorem 4 therefore provides a signal recovery strategy for inhomogeneously sampled multidimensional signals that are represented by spectral expansions in terms of *arbitrary* orthonormal bases $\{\varphi_k\}_{k\in\mathbb{I}}$.

B. Approximate reconstruction

Of course, there are many cases where $w(\cdot,t)\notin\mathcal{X}_n$ for all time instances $t\in\mathbb{T}$. For these cases, exact signal reconstruction from samples $\tilde{w}$ will not be possible.

Using the bilinearity of (16), we have that the coefficients $\tilde{a}_k$ defined in (18) satisfy

$$\tilde{a}_k := (\mathbf{w},\varphi_k)_N = \sum_{\ell\in\mathbb{I}} a_\ell\,(\varphi_\ell,\varphi_k)_N = \sum_{\ell=1}^{n} a_\ell\,(\varphi_\ell,\varphi_k)_N + \sum_{\ell>n} a_\ell\,(\varphi_\ell,\varphi_k)_N = \\ = a_k + a_{\text{alias},k} \tag{19}$$

where

$$a_{\text{alias,k}} := \sum_{\ell>n} a_\ell\,(\varphi_\ell,\varphi_k)_N\,, \qquad k=1,\ldots,n$$

is the kth *alias coefficient*. Hence, the kth coefficient $\tilde{a}_k$ not only depends on a_k but also on the higher order expansion coefficients a_ℓ of $\mathbf{w}$ with $\ell>n$. The *alias expression* (19) is well documented for specific orthonormal bases such as bases of trigonometric functions or bases consisting of Laguerre or Chebeshev polynomials [33], [45], [13] but is hardly ever used for multidimensional signals expanded through arbitrary orthonormal bases such as the ones used here.

Due to the alias expression (19), the interpolant $\hat{w}_n$ defined in (13) with spectral coefficients (18) will in general not be equal to w and incur an interpolation error $\|w-\hat{w}\|$. Using (19), we can express this error as

$$\|w-\hat{w}_n\|_{\mathcal{W}}^2 = \|w-w_n+w_n-\hat{w}_n\|_{\mathcal{W}}^2 = \|w-w_n\|_{\mathcal{W}}^2 + \|w_n-\hat{w}_n\|_{\mathcal{W}}^2 = \\ = \sum_{k>n}\|a_k(t)\|_{L_2(\mathbb{T},\mathbb{R})}^2 + \sum_{k=1}^{n}\|a_{\text{alias},k}(t)\|_{L_2(\mathbb{T},\mathbb{R})}^2. \tag{20}$$

Here, the first summation is due to the *projection error* $w-w_n$ and the second summation is due to the *alias error*:

$$\hat{w}_n(x,t)-w_n(x,t) = \sum_{k=1}^{n} a_{\text{alias},k}(t)\varphi_k(x). \tag{21}$$

It follows that the interpolation error is never less than the projection error and never less than the alias error.

If exact signal reconstruction is not possible, one may adopt an anti-alias approach by either increasing n, or by using an anti-alias filter that forces coefficients $a_\ell=0$ for $\ell>n$. In the transform domain, the coefficients $a_{\text{alias},k}$ depend linearly on the coefficients a_ℓ, $\ell>n$. Consequently, the *alias operator* $A_n:\ell_2(\mathbb{I},\mathbb{R})\to\mathbb{R}^n$ defined by

$$A_n a := \text{col}(a_{\text{alias},k}, \quad k=1,\ldots,n) \tag{22}$$

which maps the expansion coefficients a_ℓ of a given observation to its corresponding alias error coefficients, is a linear surjective map. Its induced norm

$$\|A_n\| := \sup_{0\neq a\in\ell_2(\mathbb{I},\mathbb{R})}\frac{\|A_n a\|}{\|a\|}$$

is a suitable measure for the *alias sensitivity* and depends on the truncation level n and, by (16), on the choice of the N distinct sample points in $\mathbb{X}_0$. The following theorem characterizes the alias sensitivity. In the next section, this characterization is used to derive a quantitative metric for selecting samples.

Theorem 5: Let $\{\varphi_k,\ k \in \mathbb{I}\}$ be an orthonormal basis of $\mathcal{X}$ and let $\mathbb{X}_0 = \{x_1, \ldots, x_N\}$ be a sample set consisting of N disjoint points in $\mathbb{X}$. Then

1) The alias sensitivity $\|A_n\|$ is given by $\|A_n\| = \lambda_{\max}^{1/2}(A)$, where A is the $n \times n$ real symmetric matrix whose (k, ℓ)th entry is given by
$$A_{k,\ell} = \sum_{p>n} (\varphi_p, \varphi_k)_N \cdot (\varphi_p, \varphi_\ell)_N.$$

2) If $\mathcal{X}$ is finite dimensional and equipped with the standard Euclidean inner product, then the alias sensitivity is $\|A_n\| = \lambda_{\max}^{1/2}(A)$, where A is the matrix
$$A = (\tilde{\Phi}^\top \tilde{\Phi})^{-1} - I.$$

Proof:

1) The alias sensitivity $\|A_n\| = \lambda_{\max}^{1/2}(A)$, where $A = A_n A_n^*$. Hence, it suffices to show that $A = A_n A_n^*$. The adjoint A_n^* of $A_n : \ell_2(\mathbb{I}, \mathbb{R}) \to \mathbb{R}^n$ is the mapping $A_n^* : \mathbb{R}^n \to \ell_2(\mathbb{I}, \mathbb{R})$ defined by
$$(A_n^* b)(\ell) := \begin{cases} 0 & \text{if } 1 \leq \ell \leq n \\ \sum_{k=1}^n b_k\, (\varphi_\ell, \varphi_k)_N & \text{if } \ell > n \end{cases}.$$
Indeed, with $a \in \ell_2(\mathbb{I}, \mathbb{R})$ and $b \in \mathbb{R}^n$ there holds
$$(A_n a, b) = \sum_{k=1}^n b_k \sum_{\ell>n} a_\ell\, (\varphi_\ell, \varphi_k)_N = \sum_{\ell>n} a_\ell \sum_{k=1}^n b_k\, (\varphi_\ell, \varphi_k)_N = (a, A_n^* b)\,,$$
where the first inner product is the standard inner product in $\mathbb{R}^n$ and the last inner product is the standard inner product in $\ell_2(\mathbb{I}, \mathbb{R})$. Consequently, if e_k and e_ℓ denote the kth and ℓth unit vectors in $\mathbb{R}^n$, we have that the (k, ℓ)th entry of the $n \times n$ matrix $A_n A_n^*$ is given by
$$(e_k, A_n A_n^* e_\ell) = (A_n^* e_k, A_n^* e_\ell)_{\ell_2} = \sum_{p>n} (\varphi_p, \varphi_k)_N \cdot (\varphi_p, \varphi_\ell)_N\,.$$
Hence, $A = A_n A_n^*$ as claimed.

2) To prove the second item, suppose that $\mathcal{X}$ is finite dimensional, say of dimension K, equipped with the standard Euclidean inner product. Let $\Phi \in \mathbb{R}^{K \times K}$ be the matrix whose kth column defines the kth orthonormal basis function φ_k of $\mathcal{X}$, $k = 1, \ldots, K$. Furthermore, let $\tilde{\Phi}$ be as in (14) and define $\tilde{\Phi}_{\text{tail}}$ as the $N \times (K - n)$ matrix whose kth column is the vector of restrictions $\tilde{\varphi}_k = \varphi_k|_{\mathbb{X}_0}$, $n < k \leq K$. Then, using the orthonormality of the basis $\{\varphi_k, k = 1, \ldots, K\}$, we have that $\Phi^\top \Phi = \Phi \Phi^\top = I_K$ and
$$\begin{pmatrix} \tilde{\Phi} & \tilde{\Phi}_{\text{tail}} \end{pmatrix} \begin{pmatrix} \tilde{\Phi} & \tilde{\Phi}_{\text{tail}} \end{pmatrix}^\top = I_N. \tag{23}$$

Using the expression for A derived in the first part of this proof, (23) and (16), we infer that

$$\begin{aligned}A &= \tilde{\Phi}^\top Q \tilde{\Phi}_{\text{tail}} \tilde{\Phi}_{\text{tail}}^\top Q \tilde{\Phi} = \tilde{\Phi}^\top Q \left(I_N - \tilde{\Phi}\tilde{\Phi}^\top \right) Q \tilde{\Phi} = \\ &= \tilde{\Phi}^\top \tilde{\Phi} (\tilde{\Phi}^\top \tilde{\Phi})^{-2} \tilde{\Phi} \left(I_N - \tilde{\Phi}\tilde{\Phi}^\top \right) \tilde{\Phi} (\tilde{\Phi}^\top \tilde{\Phi})^{-2} \tilde{\Phi}^\top \tilde{\Phi} = \\ &= (\tilde{\Phi}^\top \tilde{\Phi})^{-1} \left(\tilde{\Phi}^\top \tilde{\Phi} - (\tilde{\Phi}\tilde{\Phi}^\top)^2 \right) (\tilde{\Phi}^\top \tilde{\Phi})^{-1} = \\ &= (\tilde{\Phi}^\top \tilde{\Phi})^{-1} - I_n.\end{aligned}$$

Here, we used in the second equality that (23) implies $\tilde{\Phi}_{\text{tail}} \tilde{\Phi}_{\text{tail}}^\top = I_N - \tilde{\Phi}\tilde{\Phi}^\top$. This gives the result.

■

C. Construction of MPE reduced-order models

In this subsection, the results on signal reconstruction and approximation are extended to the construction of reduced-order models using missing point estimations. This is a key enabler to derive reduced-order models that are computationally efficient to solve for nonlinear and time-varying systems. The main result of this subsection provides conditions under which the reduced-order models can equivalently be represented through function evaluations that involve the bilinear form (16).

We consider reduced-order models of a dynamic spatial-temporal system $\mathcal{B} \subset \mathcal{W}$. Let $n > 0$ and suppose that $\mathcal{B}_n^{\text{weak}}$ and $\mathcal{B}_n^{\text{Galerkin}}$ are the nth order systems specified in Definition 2 with

$$\mathcal{S} = \mathcal{X}_n = \text{span}(\varphi_1, \ldots, \varphi_n).$$

Let $\mathbb{X}_0$ consist of N distinct points in the spatial domain $\mathbb{X}$. Then define $\widetilde{B}_n^{\text{weak}}$ as the set of all functions $\mathbf{w}_n$ in $\mathcal{W}_n = L_2(\mathbb{T}, \mathcal{S})$ that satisfy

$$(D_t \mathbf{w}_n(t), \varphi)_N = F\left((\mathbf{w}_n(t), \varphi)_N, \ldots, (D_x^p \mathbf{w}_n(t), \varphi)_N, \ldots \right) \tag{24}$$

for all $\varphi \in \mathcal{S}$ and almost all $t \in \mathbb{T}$. Similarly, let $\widetilde{B}_n^{\text{Galerkin}}$ be the solution set of all $\mathbf{w}_n \in \mathcal{W}_n$ that satisfy

$$(D_t \mathbf{w}_n(t), \varphi)_N = \left(F\left(\mathbf{w}_n(t), \ldots, D_x^p \mathbf{w}_n(t), \ldots \right), \varphi \right)_N \tag{25}$$

for all $\varphi \in \mathcal{S}$ and almost all $t \in \mathbb{T}$. The evaluation of each of the arguments in the right-hand sides of (24) and (25) involves, by (16), only N function evaluations and is therefore considerably faster than the evaluation of the inner products in the right-hand-sides of (8) and (9). In particular, solutions to (24) or (25) require considerably less computational effort when compared to solving (10) and (11). Moreover, the following result shows that, under mild conditions, this computational acceleration does not incur any loss of accuracy.

Theorem 6: If $n \leq N$ is such that $\tilde{\Phi}$ defined in (14) is injective then

$$\widetilde{\mathcal{B}}_n^{\text{weak}} = \mathcal{B}_n^{\text{weak}}.$$

Moreover, if $\mathcal{B}$ is linear then

$$\widetilde{\mathcal{B}}_n^{\text{weak}} = \widetilde{\mathcal{B}}_n^{\text{Galerkin}} = \mathcal{B}_n^{\text{weak}} = \mathcal{B}_n^{\text{Galerkin}}.$$

Proof: This result is an immediate consequence of Lemma 3. Indeed, $\mathbf{w}_n \in \mathcal{B}_n^{\text{weak}}$ implies $\mathbf{w}_n(t) \in \mathcal{S} = \mathcal{X}_n$ for all t. But then, by Lemma 3, the differential equations (8) and (24) are identical, so that the solution sets $\widetilde{\mathcal{B}}_n^{\text{weak}}$ and $\mathcal{B}_n^{\text{weak}}$ coincide, provided that $\tilde{\Phi}$ is injective. The second statement is an immediate consequence of the linearity of F in (6) and linearity of the inner product in (25). ■

In other words, under a mild condition of injectivity of $\tilde{\Phi}$ the reduced-order model $\mathcal{B}_n^{\text{weak}}$ coincides with the reduced-order model $\widetilde{\mathcal{B}}_n^{\text{weak}}$ which can equivalently be obtained by (24). In addition, for the linear case the reduced-order models of Definition 2, can equivalently be obtained through (24) or (25).

The MPE method therefore yields computationally efficient reduced order models for both nonlinear and linear cases. Nonlinear reduced order models that are fast to solve are particularly attractive for a vast range of engineering applications, where nonlinear PDEs are frequently employed to describe the physical systems.

V. Choice of samples

The question how to select the N distinct samples $\mathbb{X}_0 = \{x_1, \ldots, x_N\}$ is of evident interest for the overall accuracy of the reduced-order model and has not been addressed so far. The choice of suitable sensor locations by which the system dynamics can be recovered is a prime practical motivation behind this question. This section describes selection criteria to define optimal choices of sensor locations. We propose an efficient heuristic optimization approach and two screening criteria. The criteria are independent of the original model equations, which is important for numerical tractability and simplicity of design. Indeed, for large-scale systems it is more feasible to develop selection criteria using data rather than using the model. Throughout this section it is assumed that $\mathcal{X}$ is finite dimensional, say of dimension K. $\mathcal{X}$ is identified with $\mathbb{R}^K$ and equipped with the standard Euclidean inner product. In particular, the hypothesis of item 2 of Theorem 5 applies throughout this section.

A. Optimization of the point selection

In (20) it has been shown that the estimation error $\|w - \hat{w}_n\|_{\mathcal{W}}$, obtained from the interpolation of a partial observation on the grid $\mathbb{X}_0$, can be represented as the norm of the projection error $w - w_n$ and the alias error $\hat{w}_n - w_n$. The induced norm of the alias sensitivity A_n, as characterized in Theorem 5, obviously depends on the choice of N distinct points $\mathbb{X}_0$, simply because the bilinear form (3) depends on the sample points $\mathbb{X}_0$. Let

$$e'_n(\mathbb{X}_0) := \|A_n\|^2 \tag{26}$$

express this dependence.

Using the assumptions on $\mathcal{X}$ we have, by item 2 of Theorem 5,

$$e'_n(\mathbb{X}_0) = \| \, (\tilde{\Phi}^\top \tilde{\Phi})^{-1} - I \, \| \tag{27}$$

so that the minimization of $e'_n(\mathbb{X}_0)$ over all subsets $\mathbb{X}_0 \subset \mathbb{X}$ of cardinality N amounts to selecting $\mathbb{X}_0$ in such a way that $\| \, (\tilde{\Phi}^\top \tilde{\Phi})^{-1} - I \, \|$ is minimal. The relation expressed in (27) implies that the closer $\tilde{\Phi}^\top \tilde{\Phi}$ is to the identity matrix, the smaller the sensitivity of the aliasing error, as the gain from the neglected POD modes to the alias error is small.

This result is analogous to well-known results in the literature of experimental design [11],[16],[19] where an optimal selection of N factors out of K experiments needs to be made. The search for N factors is done by maximizing a particular information matrix. Similar to what we have derived in Section IV, maximization of the information matrix also amounts to preserving the orthogonality of the information matrix when N factors are chosen. In Section IV, we introduced a bilinear form to arrive at a similar result.

If $e'_n(\mathbb{X}_0) \leq \gamma$ for some upperbound $\gamma > 0$, then $(\tilde{\Phi}^\top \tilde{\Phi})^{-1} - I \leq \gamma I$ from which we infer (after pre- and post-multiplying by $(\tilde{\Phi}^\top \tilde{\Phi})^{1/2}$ and using that $\tilde{\Phi}^\top \tilde{\Phi} \leq I$) that also

$$I - \tilde{\Phi}^\top \tilde{\Phi} \leq \gamma I.$$

To avoid computing the inverse in (27), we will instead minimize the criterion

$$e_n(\mathbb{X}_0) = \| I - \tilde{\Phi}^\top \tilde{\Phi} \| \tag{28}$$

over all subsets $\mathbb{X}_0$ of cardinality N. In particular, the above reasoning shows that $e_n(\mathbb{X}_0) \leq e'_n(\mathbb{X}_0)$. As matrix norm, we consider the Frobenius norm

$$\| X \|^2 := \sum_{i=1}^{n} \sum_{j=1}^{n} |X_{ij}|^2. \tag{29}$$

Selection of $\mathbb{X}_0$ so as to minimize $e_n(\mathbb{X}_0)$ is a combinatorial optimization problem, which is generally not a very appealing optimization strategy for large-scale systems. In this paper we employ a non-combinatorial suboptimal approach to construct $\mathbb{X}_0$, using the *greedy algorithm*. This algorithm is also implemented in [44] to characterize suitable sensor locations. Given a current subset of points $\mathbb{X}_0$, the greedy algorithm adds a point to $\mathbb{X}_0$ by looping over all possible candidate points, computing the restricted basis $\tilde{\Phi}$ that would result if the candidate point were added to $\mathbb{X}_0$, and evaluating the condition number $c(\tilde{\Phi}^\top \tilde{\Phi})$ defined by

$$c(\tilde{\Phi}^\top \tilde{\Phi}) := \frac{\lambda_{\max}(\tilde{\Phi}^\top \tilde{\Phi})}{\lambda_{\min}(\tilde{\Phi}^\top \tilde{\Phi})}. \tag{30}$$

The point that yields the lowest value of $c(\tilde{\Phi}^\top \tilde{\Phi})$ is then added to $\mathbb{X}_0$, and the process is repeated. In this way, the subset $\mathbb{X}_0$ is constructed by choosing one point at a time. The algorithm is terminated when $c(\tilde{\Phi}^\top \tilde{\Phi}) \leq c_{\text{tol}}$, where $c_{\text{tol}} > 0$ is a user defined condition number, typically chosen to be well below 100 [26], [28]. The resulting sample set $\mathbb{X}_0$ will not minimize $e_n(\mathbb{X}_0)$, but the search mechanism is efficient.

Algorithm 7: The greedy algorithm

Input a (possibly empty) set $\mathbb{X}_0^0 \subset \mathbb{X}$ of N_0 pre-selected points, a threshold $c_{\text{tol}} > 0$ for the condition number, and a set $\mathbb{Y} \subseteq \mathbb{X}$ of K_0 candidate points. Set $j = 0$.

1) Repeat the following steps until $c\left(\tilde{\Phi}^\top(\mathbb{X}_0^j)\tilde{\Phi}(\mathbb{X}_0^j)\right) \leq c_{\text{tol}}$ or until $j = K_0$.

   - Set $j = j + 1$.
   - For all $x_{k_g} \in \mathbb{Y} \backslash \mathbb{X}_0^{j-1}$, determine

     $$c_g = c\left(\tilde{\Phi}^\top(\mathbb{X}_g)\tilde{\Phi}(\mathbb{X}_g)\right),$$

     where $\mathbb{X}_g = \mathbb{X}_0^{j-1} \cup \{x_{k_g}\}$ and $g = 1, \dots, K_0 - j + 1$.

- Find the index g^* for which $c_{g^*} \leq c_g$ for all $1 \leq g \leq K_0 - j + 1$.
- Set $\mathbb{X}_0^j = \mathbb{X}_0^{j-1} \cup \{x_{k_{g^*}}\}$. Then $\mathbb{X}_0^j$ consists of $N_0 + j$ points.

2) Output $\mathbb{X}_0 = \mathbb{X}_0^j$ is a set of $N = N_0 + j$ sample points.

For very large systems, it may be computationally expensive to consider all points in the high-dimensional space $\mathbb{X}$ as candidate points. In this case, a screening criterion may be applied to first select a subset of sample points $\mathbb{Y}$, to which the greedy algorithm is applied. This criterion could also be used for selecting the initial sample set $\mathbb{X}_0^0$.

B. *Point screening criteria*

The first point screening criterion orders the points as $x_{k_1}, \ldots, x_{k_{K_0}}$ according to the quantity e_n such that

$$e_n(x_{k_1}) \leq e_n(x_{k_2}) \leq \cdots \leq e_n(x_{k_{K_0}}). \tag{31}$$

This criterion is motivated by the desire to minimize the alias sensitivity $\|A_n\|$ over all selections of N distinct rows in (14).

A second screening criterion, which incorporates the relationship with the collected snapshot data, considers the ensemble of projected signals $\mathbf{w}_n(t) \in \mathbb{R}^K$ where $t \in \mathbb{T}$ and sets

$$W_n := \begin{pmatrix} \mathbf{w}_n(t_1) & \cdots & \mathbf{w}_n(t_M) \end{pmatrix}.$$

Let $\Pi_{\mathcal{X}_0}$ denote the canonical projection from $\mathcal{X} = \mathbb{R}^K$ to $\mathcal{X}_0 = \mathbb{R}^N$, and define, for all time instants $t \in \mathbb{T}$, the projections $\bar{\mathbf{w}}_n(t) = \Pi_{\mathcal{X}_0} \mathbf{w}_n(t)$. Set

$$\overline{W}_n := \begin{pmatrix} \bar{\mathbf{w}}_n(t_1) & \cdots & \bar{\mathbf{w}}_n(t_M) \end{pmatrix}.$$

The second screening criterion measures the difference between the time correlation matrix $W_n^\top W_n$ constructed from the ensemble $\{\mathbf{w}_n(t_j), j = 1, \ldots, M\}$ and the correlation matrix $\overline{W}_n^\top \overline{W}_n$ built from the restricted ensemble $\{\bar{\mathbf{w}}_n(t_j), j = 1, \ldots, M\}$. The difference is measured by $\hat{e}_n(\mathbb{X}_0)$ defined as

$$\hat{e}_n(\mathbb{X}_0) = \| W_n^\top W_n - \overline{W}_n^\top \overline{W}_n \| . \tag{32}$$

The points in $\mathbb{X}$ are reordered as $x_{k_1}, \ldots, x_{k_K}$ such that

$$\hat{e}_n(x_{k_1}) \leq \hat{e}_n(x_{k_2}) \leq \cdots \leq \hat{e}_n(x_{k_K}). \tag{33}$$

Using either (31) or (33) the first K_0 points $\{x_{k_1}, \ldots, x_{k_{K_0}}\}$ are included in the set of candidate points $\mathbb{Y}$, which are then input to the greedy algorithm.

VI. Application

A. *Glass Melt Feeder Application*

The missing point estimation approach is implemented on a numerical model of a glass melt feeder. A glass melt feeder, shown in the schematic Figure 1, is the section of a glass furnace that is located between the refiner and the glass melt exit point. The feeder is fed by incoming glass melt from a reactor. The rate of glass melt flow is measured in tons/day and is known in the glass industry as the *pull rate*.

Glass quality is highly sensitive to variations in glass composition and energy transfer in the furnace and the feeder. The control of glass quality specifications predominantly involves the concise tracking of a non-uniform temperature distribution within a specified range. Non-concise tracking of temperature will produce defective glass products, such as irregular shapes, cracks, or bubbles [7]. The actuators in a glass feeder are the temperature distributions of the so-called *crown*, which is a combustion chamber above the glass melt. Several temperature sensors are placed in the glass melt and the measurements are fed back to controllers to adjust the crown temperature. The crown is divided into several zones; the temperature distribution in each zone is adjusted to reach the desired temperature profiles in the glass feeder.

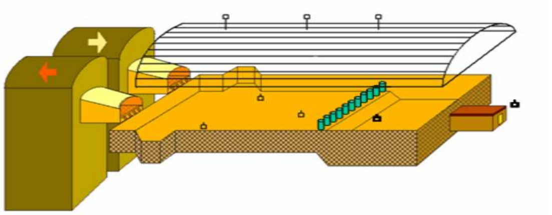

Fig. 1. Schematic view of a glass feeder, the glass melt is entering the feeder from the left side and at the right end is discharged as glass gob to the forming machine.

Until now, the glass industry has mainly used conventional PID controllers. Fast (100 to 1000 times faster than real time), accurate (absolute errors in the range of 0.2 degrees) simulation models provide an opportunity to use more sophisticated, model-based process control.

A CFD model is used for high-fidelity predictive simulations of the glass melt flow and temperature distribution in the feeder. In general, the flow can be considered to be incompressible and laminar. The flow is governed by the Navier-Stokes equations, which describe the pressure field p and the velocity field $(v_{\mathtt{x}}, v_{\mathtt{y}}, v_{\mathtt{z}})$ in the $\mathtt{x}$, $\mathtt{y}$ and $\mathtt{z}$ directions, respectively, and the energy equations for the temperature field w [9]. In this application, $w(x,t)$ refers to the *temperature* of the melt at position $x = (\mathtt{x}, \mathtt{y}, \mathtt{z}) \in \mathbb{X}$ in the feeder and at time $t \in \mathbb{T}$.

The governing equation of heat transfer in the glass feeder is given by

$$\frac{\partial \rho c_p w}{\partial t} = - \underbrace{\operatorname{div}(\rho c_p w \mathbf{v})}_{\text{convective heat transfer}} + \underbrace{\operatorname{div}(\kappa \operatorname{grad} w)}_{\text{conductive heat transfer}} + \underbrace{q,}_{\text{external energy sources}} \tag{34}$$

which is a PDE of the form (6).

Most physical parameters of the glass melt are functions of temperature. In Table I, the temperature-dependent parameters of a specific green container glass are listed (the temperature is in Kelvin).

To solve the equations numerically over the spatial domain $\mathbb{X}$ and a finite time domain $\mathbb{T}$, the feeder is discretized as shown in Figure 2, using a total of 7128 grid points. The model (34) is discretized in space using the finite

TABLE I

TEMPERATURE DEPENDENT PHYSICAL PARAMETERS.

Density ρ [kg/m^3] [39]	$\rho(w) = 2540 - 0.14w$
Viscosity $\mu(w)$ (Ns/m^2) [7]	$\mu(w) = 10^{-2.592} + \frac{4242.904}{w-541.8413}$
Specific Heat $c_p(w)$ (J/kgK) [7]	$c_p = 221 + 0.0956w$
Thermal conductivity κ (W/m.K) [39]	$\kappa(w) = 0.527 + 0.001w + 2.67 \times 10^9 w^3$

volume method [43], in which the governing PDE (34) is integrated for every grid cell. Specifically, the temperature w at a grid point P and a time instant t_k is denoted by $w_P(k)$ and given by the discrete evolution equation

$$\begin{aligned} a_P(k)w_P(k) &= a_W(k)w_W(k) + a_E(k)w_W(k) + a_S(k)w_S(k) + a_N(k)w_N(k) \\ &\quad + a_B(k)w_B(k) + a_T(k)w_T(k) + a_P^0(k)w_P(k-1) + \mathbf{S}_g(k)\mathbf{u}(k-1) \end{aligned} \tag{35}$$

where $w_W, w_E, w_N, w_S, w_T, w_B$ denote the temperature at the western, eastern, northern, southern, top, and bottom neighboring grid points, respectively. The input $\mathbf{u}(k) \in \mathbb{R}^{n_u}$ comprises n_u external sources such as the crown temperature, electrical boostings, heaters, and the terms where boundary conditions (such as inlet and outlet temperatures) are imposed. The contribution from the input to the dynamics of $w_P(k)$ is denoted as $\mathbf{S}_g(k) \in \mathbb{R}^{1\times n_u}$. The terms $a_P, a_P^0, a_W, a_E, a_S, a_N, a_T, a_S, \mathbf{S}_g \in \mathbb{R}^{1\times n_u}$ are generally time varying due to the dependencies of the physical parameters on the temperature.

Writing (35) for 7128 grid points yields the following nonlinear set of equations

$$\mathbf{A}(\mathbf{w}(k+1))\mathbf{w}(k+1) = \mathbf{A}_0(\mathbf{w}(k+1))\mathbf{w}(k) + \mathbf{B}(\mathbf{w}(k+1))\mathbf{u}(k) \tag{36}$$

where $\mathbf{w}(k+1)$ is the vector containing the unknown temperature values at time t_{k+1}.

At grid points on the domain boundary, the temperature is specified using Dirichlet or Neumann boundary conditions. The temperature at these boundary points belongs to the input terms $\mathbf{u}(k)$ in (36); they do not belong to the variables to be solved. For this application, the number of non-boundary points is 3800. In addition, we exploit symmetry in the $\mathbf{z}$ direction and only consider half of the mesh points defined in the feeder; therefore, $\mathbf{w}(k) \in \mathbb{R}^{1900}$ and $K = 1900$.

B. Complexity Analysis

In CFD and many other applications, the nonlinear system (36) is typically solved in an iterative manner. A number h_{iter} of inner iterations (in our case 100) is applied to advance from time t_k to t_{k+1}. Each inner iteration step takes the form

$$\mathbf{A}(\mathbf{w}_{k+1}^h)\mathbf{w}_{k+1}^{h+1} = \mathbf{A}_0(\mathbf{w}_{k+1}^h)\mathbf{w}(k) + \mathbf{B}(\mathbf{w}_{k+1}^h)\mathbf{u}(k), \quad h = 0, \dots, h_{\text{iter}} \tag{37}$$

where h is the inner iteration index and $\mathbf{w}_{k+1}^h$ is the approximation of $\mathbf{w}(k+1)$ at the hth iteration. Each iteration is initialized with $\mathbf{w}_{k+1}^0 = \mathbf{w}(k)$ and is terminated when either the difference $(\mathbf{w}_{k+1}^{h+1} - \mathbf{w}_{k+1}^h)$ is smaller than some specified tolerance or when $h = h_{\text{iter}}$.

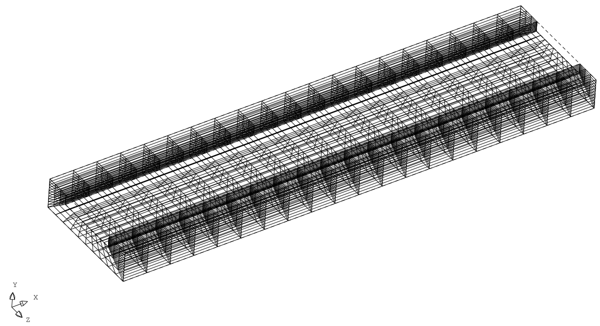

Fig. 2. Geometry and grid cells of the feeder channel. The Cartesian coordinate orientation is denoted by the $\mathtt{x}$ for length, $\mathtt{y}$ for height, and $\mathtt{z}$ for width. The entrance of the feeder (which is connected to the working end) starts from the left part and the outlet of the feeder/the spout is on the rightmost part.

Solution of the nonlinear system at each time step therefore requires solving a sequence of *linear* problems in the unknown $\mathbf{w}_{k+1}^{h+1}$ as given by (37), which can be cast as a linear parameter varying (LPV) system. Refer to [34] and [43] for more details on the implementation for this particular application.

The total computational cost of solving the full model can be classified according to the following three sources:

1) computing the coefficients of $\mathbf{A}, \mathbf{A}_0$ and $\mathbf{B}$ at every iteration within each timestep
2) solving (37) at every iteration within each timestep, using, for example, LU decomposition or conjugate gradient method;
3) other overhead cost, such as the time needed for initialization, etc.

Using a projection framework (e.g. standard POD) to derive the reduced model results in a reduced-order LPV system, which must be constructed and solved at every iteration within each timestep. As for the full model, the large-scale matrices $\mathbf{A}$, $\mathbf{A}_0$ and $\mathbf{B}$ must still be computed. In addition, the inner products $\Phi^\top \mathbf{A} \Phi$, $\Phi^\top \mathbf{A}_0 \Phi$ and $\Phi^\top \mathbf{B}$ must be computed at each iteration. The complexity of computing these inner products is, respectively, $O(p_{\mathbf{A}} K)$, $O(p_{\mathbf{A}_0} K)$ and $O(K)$, where $p_{\mathbf{A}}$ and $p_{\mathbf{A}_0}$ are the number of non-zero entries in each row of the sparse matrices $\mathbf{A}$ and $\mathbf{A}_0$. The computational cost of solving the resulting nth-order linear system is very small, since n is typically small; this is where some savings are achieved relative to the full-order system.

Implementing the MPE approach results in savings in the computation of the large-scale matrices and the necessary inner products. Specifically, if MPE is applied over $N < K$ spatial grid points, then only the corresponding N rows and columns of $\mathbf{A}$, $\mathbf{A}_0$ and $\mathbf{B}$ must be computed. In addition, the necessary inner products with the basis vectors can be computed with complexity $O(p_{\mathbf{A}} N)$, $O(p_{\mathbf{A}_0} N)$ and $O(N)$.

A comparison of the relative computational complexity of the three approaches is given in Table II. It can be seen that the MPE approach reduces the cost associated with computing the matrix coefficients and the inner products by a factor of N/K relative to the standard projection method. This means that the computational acceleration that can be achieved using the MPE approach over standard projection is directly proportional to the reduction in the dimension of $\mathbb{X}_0$. Obviously, the magnitude of this reduction that can be achieved without significant loss of accuracy is problem dependent.

TABLE II

Relative computational complexity of solving full-order, standard reduced-order, and MPE models for each iteration step. $\mathtt{C}_1$, $\mathtt{C}_2$ and $\mathtt{C}_3$ denote computational cost associated with computing the matrix coefficients, solving the linear system, and the overhead cost for the large-scale system.

Sources of computational cost	Full-order model	Standard projection method	MPE
Computing the matrix coefficients	$\mathtt{C}_1$	$\mathtt{C}_1$	$(N/K)\mathtt{C}_1$
Computing projection inner products	0	$O(p_{\mathbf{A}}K + p_{\mathbf{A}_0}K + K)$	$O(p_{\mathbf{A}}N + p_{\mathbf{A}_0}N + N)$
Solving the linear system	$\mathtt{C}_2$	$O(n^3)$	$O(n^3)$
Overhead cost	$\mathtt{C}_3$	$\mathtt{C}_3$	$\mathtt{C}_3$

However, for applications in which the dimension of the reduced basis is small, it is reasonable to expect that a significant reduction in the dimension of $\mathbb{X}_0$ can be achieved. In many applications for which model reduction is effective, the basis vectors are relatively smooth in space, which means that selective spatial sampling should be effective. In addition, as the dimension of the full-order state increases, often the required number of basis vectors remains small [26]. If this is the case, then the savings achieved using MPE will scale to very large systems.

C. Reduced-Order Models

We will simulate the process in a transition of glass color specification from green to flint (transparent) glass . This color transition is a highly nonlinear process during which the heat conductivity will change by a factor of eight. The nominal distribution of the crown temperature and the variations from the nominal temperature for every zone are depicted in Figure 3. A POD reduced-order model and corresponding acceleration by the MPE method are applied to describe the color change process in the glass feeder. In this particular example, the reduced models are derived by employing the Galerkin method.

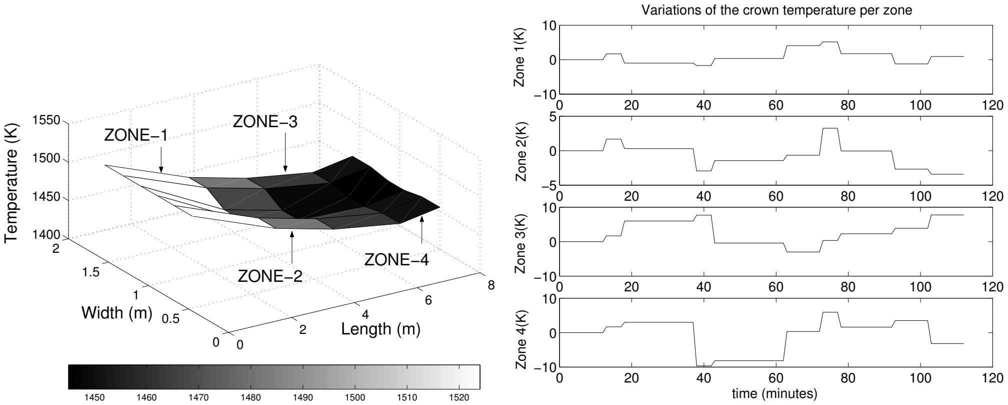

Fig. 3. The nominal crown temperature profile (left), the variations from the nominal temperature in every zone (right)

The POD basis is derived from temperature simulation data collected each minute over $M = 112$ minutes and contained in the snapshot matrix $W_{\text{snap}} = \begin{pmatrix} \mathbf{w}(t_1) & \cdots & \mathbf{w}(t_{112}) \end{pmatrix}$. The POD basis vectors, $\{\varphi_k, k = 1, \ldots, 3800\}$, are then found as the eigenvectors of the correlation matrix $C_w = \frac{1}{112} W_{\text{snap}} W_{\text{snap}}^{\top}$. The eigenspectrum of the snapshot correlation matrix is depicted in Figure 4. Eighteen POD basis functions corresponding to the $n = 18$ largest eigenvalues are chosen to construct the POD reduced-order model, $\mathcal{B}_{18}^{\text{Galerkin}}$, as defined in (10), which is calculated using $\mathcal{X}_n = \mathcal{S} = \text{span}\{\varphi_i\}_{i=1}^{18}$ as the projection space. Figure 5 shows the comparison between the results of the POD reduced-order model and the original model at two measurement points. From Figure 5, it can be seen that the reduced-order model captures the dynamics of the original model well.

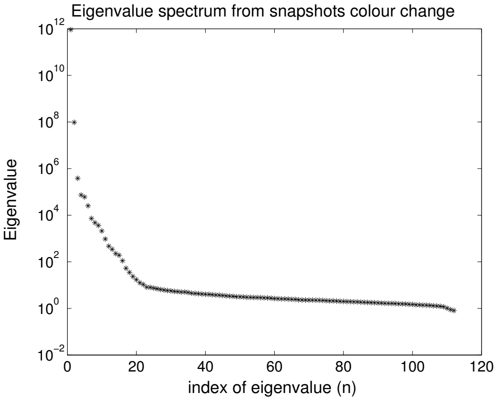

Fig. 4. The POD eigenvalue spectrum corresponding to 112 snapshots collected during simulation of a color change.

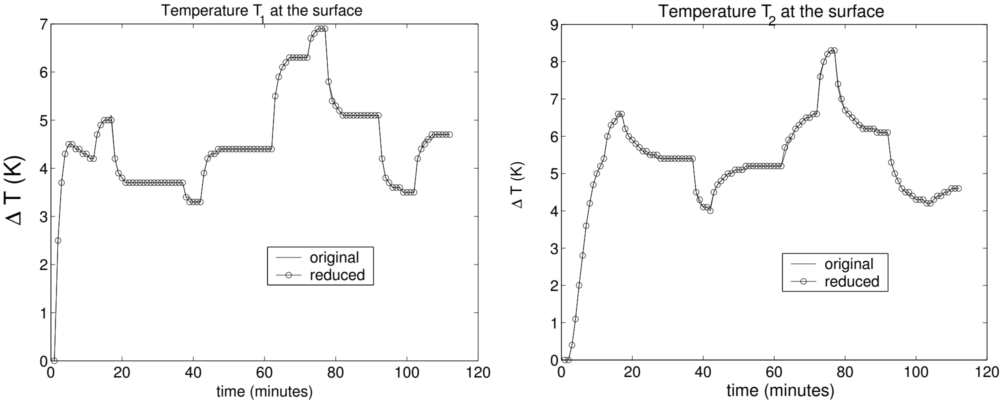

Fig. 5. POD-based reduced-order model and original temperature profiles during the color change process at two measurement points.

D. Application of MPE to the glass melt feeder

The order of the reduced model is more than 200 times lower than the original model; however, the computational time needed to solve the temperature distribution is only enhanced by a factor of 2.2. The lack of computational efficiency for this time-varying system is due to the fact that the reduced-order model requires projecting the *full* model equations onto the span of a low number of POD basis functions. The CFD matrices $\mathbf{A}(k), \mathbf{A}_0(k), \mathbf{B}(k)$ in (37) must be constantly updated to accommodate the varying physical parameters, such as the density, viscosity, and heat conductivity.

To accelerate the computation, the MPE method is implemented by selecting sample points in $\mathbb{X}_0 \subset \mathbb{X}$. The MPE method yields reduced-order models $\tilde{\mathcal{B}}_{18}^{\text{Galerkin}}$ as defined in (25), where $\mathbb{X}_0$ is determined from the selection criteria described in Section V.

An important implementation point is that, in the MPE reduced-order models, the boundary conditions must be satisfied and the set of excitation signals defined by the crown temperature must be incorporated. To achieve this, all points that are adjacent to the boundary cells have been included in $\mathbb{X}_0$. In the case of the feeder model, there are $N_0 = 265$ points that are adjacent to the boundary cells where crown temperature, inlet temperature, inlet velocity are defined. These points are considered as "obligatory points". The locations of these points define the pre-selected mask $\mathbb{X}_0^0$ in Algorithm 7.

Both screening criteria were implemented to determine a reduced set of 1635 candidate points $\mathbb{Y} = \{x_1, \ldots, x_{1635}\}$ that remain after the $N_0 = 265$ boundary points have been included in the selection. The quantities $e_{18}(x_k)$ and $\hat{e}_{18}(x_k)$ are calculated for all candidate points $x_k \in \mathbb{X}$. After $e_{18}(x_k)$ (screening criterion 1) and $\hat{e}_{18}(x_k)$(screening criterion 2) are calculated for every point, the values are ordered as in (31) and (33). Plots of the ordered $e_{18}(x_{k_j})$ and $\hat{e}_{18}(x_{k_j}), j = 1, \ldots, 1635$ are shown in Figure 6. Although the absolute differences in magnitudes of $e_{18}(x_{k_j})$ and $\hat{e}_{18}(x_{k_j})$ for the different points are small, the relative variations are important for differentiating among states. For example, suppose that we would like to construct a reduced-order model with 1000 points. The condition number of $\tilde{\Phi}^\top \tilde{\Phi}$ constructed from the 1000 points with lowest $e_{18}(x_{k_j})$ is 19.1, while the condition number of $\tilde{\Phi}^\top \tilde{\Phi}$ constructed from the 1000 points with highest $e_{18}(x_{k_j})$ is 3189.9. A low condition number of $\tilde{\Phi}^\top \tilde{\Phi}$ (less than 100) is required to use the reduced-order model for predicting different scenarios; otherwise, the prediction results can be very sensitive to any small perturbations. Inspection of Figure 6 for the second screening criterion shows

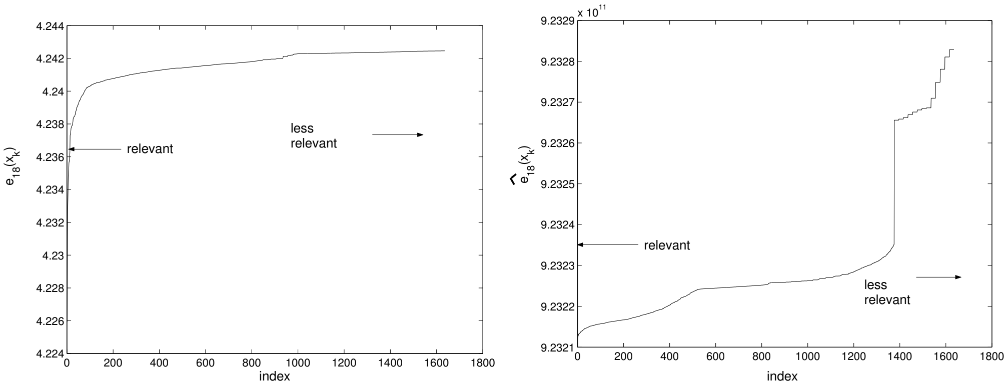

Fig. 6. The ordered e_{x_k} (left) and $\hat{e}_{x_k}$ (right) based on MPE point screening criteria.

a cut-off after 1400 points. The condition number of $\tilde{\Phi}^\top\tilde{\Phi}$ constructed by the 1400 points with lowest $\hat{e}_{18}(x_{k_j})$ is 18.52, while the condition number of $\tilde{\Phi}^\top\tilde{\Phi}$ constructed by the 1400 points with highest $\hat{e}_{18}(x_{k_j})$ is 216.3. Hence, it can be seen that both screening criteria help to separate the less relevant points from the relevant ones.

For this example, the second screening criterion tends to choose points that are spatially clustered. This is due to the fact that in this screening criterion, the POD basis is weighted by the coefficients obtained from the projection of the snapshot data. The criterion therefore tends to group states (and corresponding grid points) that have similar temperature variations, which, due to the dominant diffusive nature of the heat transfer processes in the glass melt feeder, translates directly into a grouping of points that are closely located in space. Selection of many points that are close to each other leads to a poor conditioning of the spatially restricted basis.

Two reduced-order models are constructed by the MPE method. The greedy algorithm (Algorithm 7) is applied to improve the condition number of the restricted basis inner product and reduce the number of points selected by each screening criteria to 200. After adding the obligatory boundary points, a total of 465 points are used for each MPE model. Figure 7 shows the selected spatial samples for each MPE model as grey grid cells.

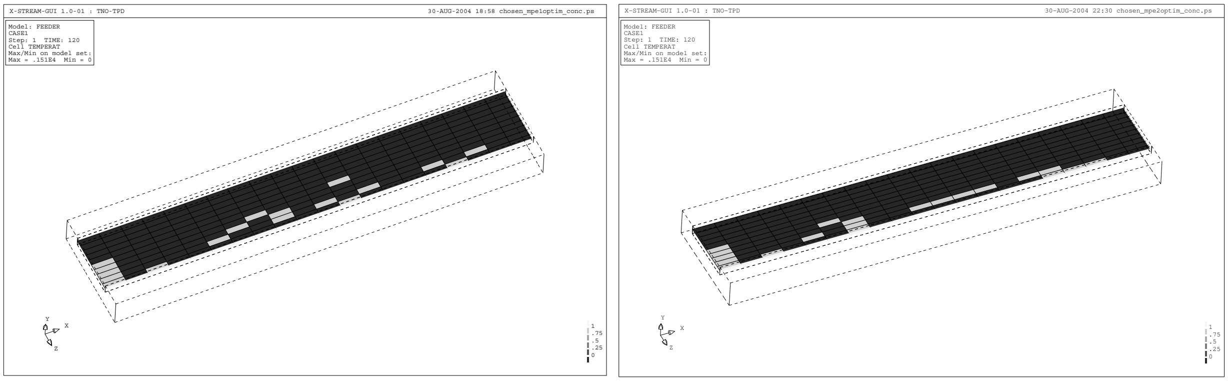

Fig. 7. Selected spatial samples in grey cells at one cross section of the feeder. The grey cells are found after implementing the greedy algorithm on points selected by screening criterion 1 (left) and screening criterion 2 (right).

Comparisons between the original and the reduced-order models constructed by the MPE method are shown in Figure 8 at two locations on the glass surface. The simulated conditions are the same as the conditions applied during the snapshot collection. From Figure 8, it is evident that the reduced models built by the MPE method adequately reconstruct the dynamics of the original model. The deviation from the original model is quantified by the maximum absolute error average ϵ_{max}, calculated as

$$\epsilon_{\text{max}} = \max_{x\in\mathbb{X}} \frac{1}{N_t} \sum_{t\in\mathbb{T}} \| \underbrace{w(x,t)}_{\text{original model}} - \underbrace{\hat{w}(x,t)}_{\text{reduced model}} \|, \tag{38}$$

where N_t is the number of time samples. The maximum absolute error for both reduced-order models constructed by the MPE method is less than 0.18K, which is about 0.1K higher than the maximum absolute error for reduced-order models constructed by the conventional POD method. The additional error is considered insignificant, as this error level is still very much below the maximum temperature variations, which is about 12K.

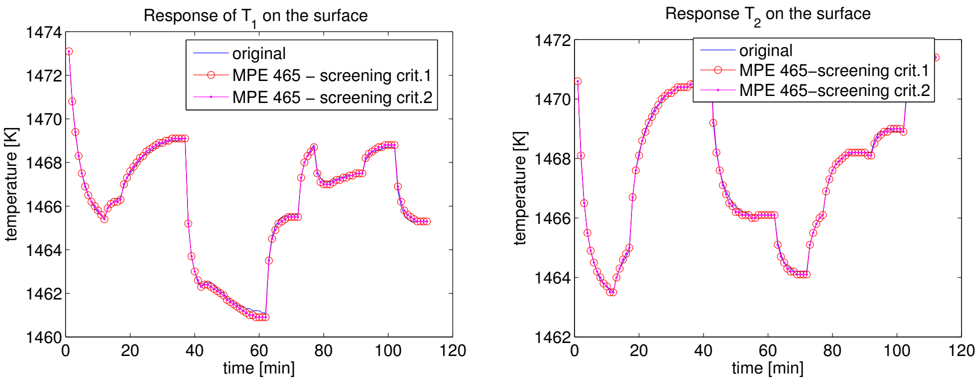

Fig. 8. Comparisons between the original model and reduced-order models constructed by the MPE method. The simulated conditions are the same as the conditions applied during the snapshot collection.

For a more challenging assessment of the POD-MPE model quality, the models are validated by exciting the system with crown temperature variations that were not considered as part of the snapshot set. In this case, all crown temperature zones are subjected to random variations from the nominal temperature distribution, distributed between 0 and 5K. The random variations are shown in Figure 9.

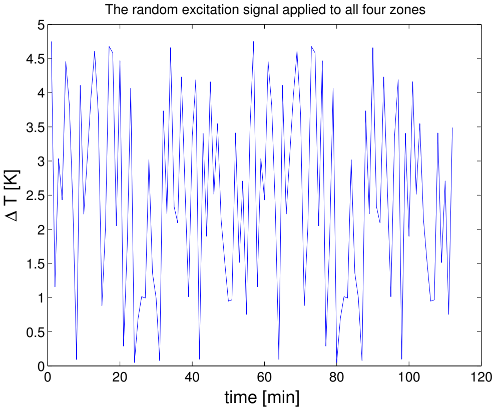

Fig. 9. The random variations from the nominal distribution of the crown temperature, applied to all four zones of the crown.

The responses of two measured locations on the glass melt surface when subjected to the random excitation signals are plotted in Figure 10. Both reduced-order models perform quite well under different excitation signals. Both reduced-order models have a maximum absolute error average of $\epsilon_{\max} < 1$K, a level of deviation that is within the 10% relative accuracy requirement, and thus is acceptable.

Figure 10 shows that the MPE model constructed from the implementation of the greedy algorithm on points screened by the first screening criterion is better for handling large temperature variations (more than 10K) while the one constructed from the points screened by the second screening criterion is more accurate when the temperature variation is small (about 4K). As explained previously, in this example the second screening criterion tends to group

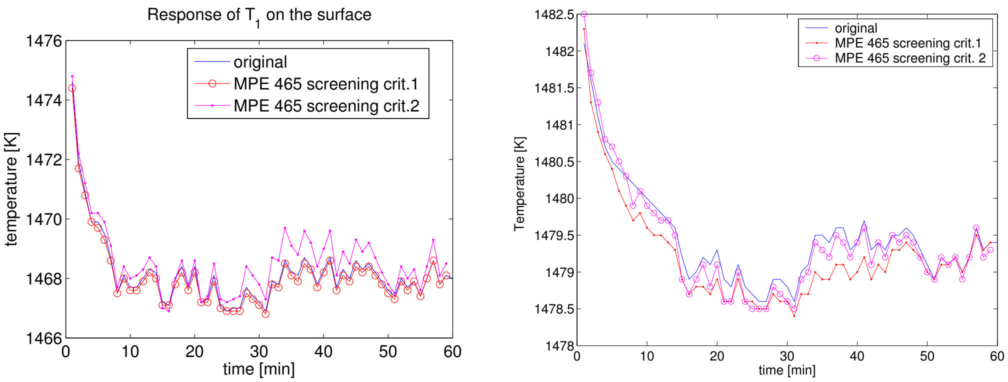

Fig. 10. The comparisons between the original model and reduced models built by the MPE method under random excitation as depicted in Figure 9. The MPE model constructed from points prescreened by screening criterion 1 handles large temperature variations better (left), while the other is more accurate for small temperature variations (right).

points that have similar temperature variations. It can be seen from Figure 7 that in some regions of the feeder, the implementation of the greedy algorithm on the points screened by this criterion result in a more clustered group of points compared to those using the first screening criterion.

Table III summarizes the simulation results using POD models and MPE models for the case of random excitations and shows the substantial computational gains that can be made using the MPE approach. The resulting average absolute errors for each screening criterion are similar, as the condition numbers of the restricted basis inner products are also similar between the two criteria. The computational gain of the reduced-order models built by MPE corresponds to 8.5 times faster than real time. The computation is performed on a 2.8 GHz processor with 512 MB RAM. The achievable computational gain depends on several factors, such as the solution methods used

TABLE III

COMPARISON BETWEEN POD AND MPE MODELS FOR RANDOM EXCITATIONS.

Model Type	**Maximum Absolute Average Error**	**Condition number** $c(H) = c(\tilde{\Phi}^\top \tilde{\Phi})$	**Computational Gain**
POD (1900 points, symmetric case) (the same excitation signals)	0.081 K	1	226%
MPE, optimized by greedy, screening criterion 1 (465 points) (the same excitation signals)	0.13 K	4.637	754%
MPE, optimized by greedy, screening criterion 1 (465 points) (validation by random excitation signals)	0.97 K	4.637	754%
MPE, optimized by greedy, screening criterion 2 (465 points) (the same excitation signals)	0.175 K	4.531	754%
MPE, optimized by greedy, screening criterion 2 (465 points) (validation by random excitation signals)	0.90 K	4.531	754%

to simulate the original model, the convergence criterion, and the algorithmic structure of the original model.

The computational gains achieved using the MPE approach may not seem sufficiently high to achieve the goal of real-time model-based control; however, in this paper, we only consider reduced-order modeling of the temperature, while the variables governing the fluid flow are still solved by the original model. If the same approach were applied to other variables, then an acceleration of 25 to 30 times faster than real time on a single processor is feasible. This would be a major breakthrough for the implementation of nonlinear, large-scale models in online control design, online tuning, and process monitoring.

VII. Conclusions

We have proposed a methodology to derive computationally efficient, reduced-order models for parameter-varying systems, such as those obtained from the discretization of nonlinear PDEs. Conventional projection-based model reduction techniques do not generally yield models that are efficient to simulate, since the original high-order model must be computed and the projection carried out at each timestep. In this paper, computational acceleration is achieved using a formal modification of the proper orthogonal decomposition method that selects a subset of the spatial domain over which to represent the dynamics of the original system. A heuristic optimization procedure, combined with two quantitative screening criteria, is proposed to select a suitable subset of grid points or state variables. The approach described in this paper is applicable to other projection-based model reduction techniques, such as balanced truncation. Demonstration of the approach on a nonlinear CFD example shows that large gains in efficiency of the reduced-order models can be obtained while retaining the nonlinear characteristics of the original system.

References

[1] J. GeBlein A.M. Lankhorst, B.D. Paarhuis. Experimental validation of the TNO forehearth model. In *7th International Seminar on Mathematical Modeling and Advanced Numerical Methods in Furnace Design and Operation*, Velke Karlovice, June 5-6 2003.

[2] P. Astrid. Fast large scale model reduction technique for large scale LTV systems. In *Proceedings of the American Control Conference*, Boston, June 2004.

[3] P. Astrid. *Reduction of process simulation models: a proper orthogonal decomposition approach*. PhD dissertation, Eindhoven University of Technology, Department of Electrical Engineering, November 2004.

[4] P. Astrid, S. Weiland, and K. Willcox. On the acceleration of the POD-based model reduction technique. In *Proceedings of the 16th International Symposium on Mathematical Theory of Network and Systems*, Leuven, July 2004.

[5] P. Astrid, S. Weiland, K. Willcox, and A.C.P.M. Backx. Missing point estimation in models described by proper orthogonal decomposition. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

[6] R.G.C. Beerkens. *Modeling of the Melting Process in Industrial Glass Furnaces*, chapter 2, pages 17–73. In Mathematical Simulation in Glass Technology, H.Loch and D.Krause (eds). Springer, Berlin, 2002.

[7] R.G.C Beerkens, H.de Waal, and F.Simonis. *Handbook for Glass Technologist.* TNO-TPD Glass Technology, Eindhoven, 1997.

[8] G. Berkooz, P. Holmes, and J.L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.

[9] B.B. Bird, W.E. Stewart, and E.N. Lightfoot. *Transport Phenomena.* John Wiley and Sons, New York, 1960.

[10] R. Bos, X. Bombois, and P. van den Hof. Accelerating large-scale nonlinear models for monitoring and control using spatial and temporal correlations. In *Proceedings of American Control Conference*, Boston, USA, 2004.

[11] G.E.P. Box and N. Draper. *Empirical Model-Building and Response Surfaces*. John Wiley and Sons, New York, 1987.

[12] T. Bui-Thanh, M. Damodaran, and K. Willcox. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. *AIAA Journal*, 42(8):1505–1516, 2004.

[13] C. Canuto, M.Y. Hussaini, A. Quarterono, and T.A. Zang. *Spectral Methods in Fluid Dynamics*. Springer Verlag, Series in computational physics, New York, 1988.

[14] M.G. Carvalho, J.Wang, and M.Nogueira. Investigation of glass melting and fining processes by means of comprehensive mathematical model. *Ceramic Transactions*, 82:143–152, 1997.

[15] Y. Chahlaoui and P. van Dooren. Estimating grammians of large scale time varying system. In *Proceedings of the 15th Triennial IFAC World Congress*, Barcelona, July 2002.

[16] Jr Dykstra, O. The Augmentation of Experimental Data to Maximize $X'X$. *Technometrics*, 13:682 –688, 1971.

[17] R. Everson and L. Sirovich. The Karhunen-Loève procedure for gappy data. *Journal Opt. Soc. Am.*, 12:1657–1664, 1995.

[18] P. Feldmann and R.W. Freund. Efficient Linear Circuit Analysis by Padé Approximation via the Lanczos Process. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 14:639–649, 1995.

[19] Z. Galil and J. Kiefer. Time- and Space-saving Computer Methods, Related to Mitchell's DETMAX. *Technometrics*, 22:301 –313, 1980.

[20] K. Gallivan, E. Grimme, and P. Van Dooren. Padé Approximation of Large-Scale Dynamic Systems with Lanczos Methods. Proceedings of the 33rd IEEE Conference on Decision and Control, December 1994.

[21] D. Gratton and K Willcox. Reduced-order, trajectory piecewise-linear models for nonlinear computational fluid dynamics. AIAA Paper 2004-2329, presented at 34th AIAA Fluid Dynamics Conference, Portland, Oregon, June 2004.

[22] S. Gugercin and A. Antoulas. A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, 77:748–766, 2004.

[23] J.R. Higgins. *Sampling Theory in Fourier and Signal Analysis*. Clarendon Press, Oxford, 1996.

[24] P. Holmes, Lumley, and G.Berkooz. *Turbulence, Coherence Structure, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge, 1996.

[25] M. Rokyta J. Málek, J. Nečas and M. Růžička. *Weak and Measure-valued Solutions to Evolutionary PDEs*. Chapman and Hall, London, 1996.

[26] M. Kirby. *Geometric Data Analysis, An Emprical Approach to Dimensionality Reduction and the Study of Patterns*. John Wiley and Sons.Inc, New York, 2001.

[27] E. Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley and Sons, Canada, 1989.

[28] E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley and Sons.Inc, New York, 1993.

[29] K. Kunisch, S. Volkwein, and L. Xie. HJB-POD based feedback design for the optimal control of evolution problems. *SIAM Journal on Applied Dynamical Systems*, to appear.

[30] S. Lall, J.E. Marsden, and S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal on Robust and Nonlinear Control*, 12(5):519–535, 2002.

[31] B. Moore. Principle component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.

[32] A. Papoulis. *Signal Analysis*. McGraw-Hill, New York, 1977.

[33] J.R. Partington. *Interpolation, Identification and Sampling*. Oxford Science Publications, Clarendon Press, London, 1997.

[34] S.V. Patankar. *Numerical Heat Transfer and Fluid Flow*. Hemisphere, London, 1980.

[35] M. Rewienski. *A Trajectory Piecewise-Linear Approach to Model Order Reduction of Nonlinear Dynamical Systems*. PhD thesis, Dept. of Electrical Engineering and Computer Science, MIT, June 2003.

[36] M. Rewienski and J. White. A Trajectory Piecewise-Linear Approach to Model Order Reduction and Fast Simulation of Nonlinear Circuits and Micromachined Devices. *Proceedings of the International Conference on Computer-Aided Design*, pages 252–7, 2001.

[37] J.M.A. Scherpen. *Balancing for Nonlinear Systems*. PhD dissertation, University of Twente, Department of Applied Mathematics, March 1994.

[38] L. Sirovich. Turbulence and the Dynamics of Coherent Structures. Part 1 : Coherent Structures. *Quarterly of Applied Mathematics*, 45(3):561–571, October 1987.

[39] J. Stanek. *Electric Melting of Glass*. Elsevier, Amsterdam, 1977.

[40] V. Thomee. *Galerkin Finite Element Methods For Parabolic Problems*, volume 25 of *Computational Mathematics*. Springer, Berlin, 1997.

[41] A. Varga. Efficient minimal realization procedure based on balancing. In *Proceedings of IMACS/IFAC Symposium on Modelling and Control of Technological Systems*, volume 2, pages 42–47, 1991.

[42] D. Venturi and G.E. Karniadakis. Gappy data and reconstruction procedures for flow past a cylinder. *Journal of Fluid Mechanics*, 519:315–336, 2004.

[43] H.K. Versteeg and W.K. Malalasekera. *An Introduction to Computational Fluid Dynamics, The Finite Volume Method*. Pearson Prentice Hall, Essex, 1995.

[44] K. Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers and Fluids*, 35(2):208–226, 2006.

[45] A.I. Zayed. *Advances in Shannon's Sampling Theory*. CRC Press, Boca Raton, 1993.