Nonlinear Dynamics (2026) 114:630
https://doi.org/10.1007/s11071-026-12495-8

RESEARCH

Nonlinear model order reduction of dynamical systems in process engineering: review and comparison

Jan C. Schulze[1] · **Alexander Mitsos**[1,2,3]

Received: 5 July 2025 / Revised: 16 March 2026 / Accepted: 17 March 2026 / Published online: 1 May 2026

Abstract

Computationally cheap yet accurate dynamical models are a key requirement for real-time capable nonlinear optimization and model-based control. When given a computationally expensive high-order prediction model, a reduction to a lower-order simplified model can enable such real-time applications. Herein, we review nonlinear model order reduction methods and provide a comparison of method characteristics. Additionally, we discuss both general-purpose methods and tailored approaches for chemical process systems and we identify similarities and differences between these methods. As machine learning manifold-Galerkin approaches currently do not account for inputs in the construction of the reduced state subspace, we extend these methods to dynamical systems with inputs. In a comparative case study, we apply eight established model order reduction methods to an air separation process model: POD-Galerkin, nonlinear-POD-Galerkin, manifold-Galerkin, dynamic mode decomposition, Koopman theory, manifold learning with latent predictor, compartment modeling, and model aggregation. Herein, we do not investigate hyperreduction, *i.e.*, reduction of floating point operations. Based on our findings, we discuss strengths and weaknesses of the model order reduction methods.

Keywords Model simplification · Hybrid model · Invariant manifold · Nonlinear Galerkin · Non-autonomous system

Acronyms

Acronym	Term
ANN	Artificial neural network
AEN	Autoencoder network
AGG	Aggregation method
COMP	Compartment method
DAE	Differential-algebraic system of equations
DMD	Dynamic mode decomposition
DMDc	DMD with controls (or general inputs)
FOM	Full-order model
HX	Heat exchanger
KW	Koopman-Wiener model
MFL	Manifold learning
MFL-ANN	MFL combined with latent ANN predictor
MFLu	Manifold learning with inputs
MOR	Model order reduction
ODE	Ordinary differential system of equations
POD	Proper orthogonal decomposition
POD-Res	POD residualization
ROM	Reduced order model
RMSE	Root mean squared error
SPT	Singular perturbation theory
SVD	Singular value decomposition

✉ Alexander Mitsos
amitsos@alum.mit.edu

[1] Process Systems Engineering, RWTH Aachen University, 52074 Aachen, Germany

[2] JARA-CSD, 52056 Aachen, Germany

[3] Energy Systems Engineering, Forschungszentrum Jülich, 52425 Jülich, Germany

1 Introduction

Process systems in chemical engineering are inherently nonlinear, *e.g.*, as a result of reaction kinetics, transport phenomena, and thermodynamic relationships [178]. Over the past decades, detailed mechanistic process modeling has become increasingly common in (chemical) process systems engineering. In particular, mechanistic first-principles models, accounting for conservation of mass, energy, and (less often) momentum, along with constitutive relations, possess favorable generalization capabilities. Hence, digital twins and digital shadows have been established as a valuable tool for process monitoring and operator training [141,

215]. However, in such predictive models, the high number of nonlinear differential equations is usually computationally prohibitive for online optimization and model-based control methods such as nonlinear model predictive control. Furthermore, process models are often stiff, making them even more challenging to handle [33].

Model reduction is a popular technique to reduce the computational cost of optimization by means of model simplification. In this context, a reduction in the number of differential states is called model order reduction (MOR) and the resulting model is a reduced-order model (ROM) [4, 188]. Conversely, the original model is the full-order model (FOM). MOR is particularly used for spatially distributed systems, which are commonly described by partial differential equations or modeled as networks of lumped systems. Examples of high-order process systems include computational fluid dynamics [251], distillation columns [301], heat exchangers [102], chemical reactors [81], and integrated process networks [55]. Fine-grained mechanistic modeling of such systems can result in thousands or millions of states.

Although mechanistic models can be very high-dimensional, FOMs often exhibit a certain coherence characterized by a few dominant mechanisms or patterns, *e.g.*, due to stiffness of the system [209]. Provided that a part of the transients vanish after a short period of time due to dissipative phenomena with a comparably fast exponential decay, we obtain a time-scale separation between fast stable decay and slow response [93]. As a result, the high-dimensional state evolves most of the time near a low-dimensional topological subspace, referred to as slow manifold [205]. Because this subspace is characterized by a set of invariance equations, *i.e.*, the (quasi-)steady-state conditions of the fast stable phenomena, the slow manifold is a special type of invariant manifold [68, 153, 243]. An accurate ROM is obtained when finding a low-order approximation of the FOM on the slow manifold. Such reducible systems are also called lumpable [126].

MOR methods construct low-order approximations of a FOM, *i.e.*, a ROM, either by projecting the FOM onto the slow manifold [18, 271] or by identifying an empirical ROM directly from state data [150, 259]. Projecting the FOM onto a linear subspace is called a Galerkin projection [5], a concept of high significance for this article. Below, we also discuss nonlinear variants of Galerkin methods as well as other projection methods. In many cases, a considerable reduction (removing 80% of the states or more) of a FOM is possible at a minor sacrifice of accuracy. However, the range of validity of a ROM is considerably reduced as well, *e.g.*, highly transient situations, such as process startup or shutdown, are often no longer covered.

Reduction methods can be classified according to how the FOM is utilized [253]. Specifically, intrusive methods rely on structural and parametric knowledge of the FOM. Note that the term "intrusive" as adopted here is fully independent from its use in structural mechanics [195], and as such partially inconsistent. A large subclass of intrusive methods are reduced-basis methods, which seek a low-dimensional set of basis vectors spanning a linear or affine subspace on which the system approximately evolves, followed by an (intrusive) projection of the FOM onto this subspace [208, 237]. Non-projection-based intrusive methods include simplifying assumptions and heuristic reduction. On the other hand, non-intrusive methods are fully detached from the FOM equations, *i.e.*, model-free. These methods only require time series data of the model states to generate a ROM.

Many projection-based approaches follow a two-stage procedure to derive a ROM. First, the nonlinear system is reversibly transformed into a favorable representation, *e.g.*, some canonical or normal form [197, 244, 269]. Subsequently, reduction is performed on the transformed system by systematic truncation or residualization of dynamic components. Depending on the original model and the reduction method, the resulting ROM can be anything from a white-box (all equations have a physical meaning) to an almost black-box (purely empirical) model. Examples of the first step include proper orthogonal decomposition [25], nonlinear normal modes [269, 290], input-output balancing [99, 197], singular perturbation canonical form [142], Weierstrass canonical form [247], canonical bilinear or polynomial form [66, 96, 228], and Koopman lifting [193, 300].

For some types of systems, *e.g.*, convection-dominated systems and wave-type equations, the assumption of a low-dimensional linear subspace does not hold [15, 119]. Such systems live on a nonlinear submanifold rather than on a linear subspace [133, 269]. Consequently, Galerkin projection requires a large number of basis vectors to retain high ROM accuracy [161, 271]. Instead, nonlinear subspace methods can approximate the nonlinear slow manifold more closely. Projecting the FOM onto the nonlinear submanifold is a nonlinear Galerkin [133] or manifold-Galerkin [161] projection, which is closely connected to the concept of nonlinear normal modes in mechanics [269]. Figure 1 illustrates the advantage of using a nonlinear manifold as the reduced subspace. Although the dimensionality of the 1D nonlinear subspace is lower than the 2D linear subspace, the data points generated by the 3D full-order system are well captured by either submanifold. Consequently, using a nonlinear subspace method may enable a similarly accurate ROM of lower order.

When reducing a nonlinear FOM, the CPU cost of evaluating a ROM is often not significantly decreased, despite the lower differential order [237]. Thus, to achieve a reduction in computational costs, an additional consecutive hyperreduction step is performed [101, 254], which reduces the arithmetic complexity, *i.e.*, floating point operations per model call. Hyperreduction is achieved by applying supervised approximation techniques to replace expensive terms by, *e.g.*, piecewise linear functions [17, 240], polynomials

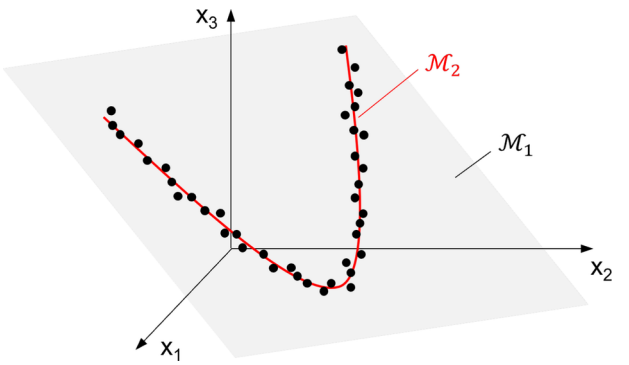

Fig. 1 Illustration of linear ($\mathcal{M}_1$, ■) and nonlinear ($\mathcal{M}_2$, —) embedded submanifolds of $\mathbb{R}^3$. The elements of $\mathcal{M}_2$ are close to $\mathcal{M}_1$ but not necessarily elements thereof. Here, the samples from the original system (•) are captured similarly well by both $\mathcal{M}_1$ and $\mathcal{M}_2$ despite their different dimensions

[30, 47, 203], sparse regression models [35], or artificial neural networks (ANNs) [101, 272]. In some cases, semi-empirical shortcut methods are an alternative to eliminate steady-state equations of a ROM [58]. Further, hyperreduction methods may also be implemented in an adaptive fashion [116, 177]. In case of a differential-algebraic ROM, hyperreduction automatically eliminates algebraic equations, so a prior reduction of algebraic variables [162, 277] is often not required. Combining data-driven and mechanistic elements, the resulting ROM can be considered a hybrid model. Hyperreduction drives the ROM further towards a black-box model but is essential for real-time capability. The overall model reduction process thus typically involves the two ingredients (i) MOR and (ii) hyperreduction.

Although non-intrusive reduction methods do not operate on the FOM equations, these methods are still based on FOM state information (and sometimes also on time derivative data). Conversely, the identification of a low-order black-box model from input-output data (*e.g.*, using subspace identification [234], Loewener framework [191, 220], or AAA algorithm [32, 200, 245]) does not involve the original FOM states and thus lacks the full information carried by the FOM states. In particular, FOM state information turns out to be highly valuable in some MOR problems, where input-output system identification may fail to extract sensible latent information [263].

Despite a long history of research, MOR remains an evolving field with major contributions made in recent years, *e.g.*, [9, 161, 259, 300]. Comprehensive reviews of MOR are found in [18, 85, 157, 185, 251] as well as textbooks on linear [5, 207] and nonlinear MOR [23, 124, 206, 282]. However, although an extensive body of literature is available, low-barrier introductory texts [4, 188] and comparative studies in the process systems field [100, 251, 258] lack a discussion of some recent developments. Furthermore, reduced modeling strategies that have been developed in the chemical engineering community are often not discussed in the general model reduction literature.

Over the years, managing the growing number of MOR options and making a suitable choice has become increasingly difficult. Therefore, this article aims to provide an overview of classical and more recent methods. We consider systems with inputs and focus on nonlinear state-space models, but we also include index-one differential-algebraic systems. We compare common methods both theoretically and in an industrially relevant case study.

As we target ROM application within numerical optimization and model predictive control, we want to preserve the knowledge of all states and retain flexibility in formulating the optimal control problem. Hence, we aim for ROMs that allow reconstruction of all original states and thus focus on input-to-state reduction methods rather than input-output methods. Consequently, we do not review input-output balancing methods for linear systems [197, 225], empirical balancing of nonlinear systems [99, 121, 158, 257], or data-driven moment matching [255]. However, we remark that for problems with fixed input-output configuration, balancing methods can provide very low-order ROMs of high accuracy and guaranteed stability preservation [140]. Further, we disregard MOR methods based on linearized FOMs, *e.g.*, applying feedback linearization before reduction [11]. Finally, we remark that model reduction of non-autonomous systems with inputs shares some conceptual similarities with MOR of stochastic dynamical systems [59, 244]. Specifically, both systems feature a-priori unknown external variables. However, we do not investigate these connections further.

The manuscript is structured as follows. First, we present a general version of the MOR problem for systems with inputs in Section 2. Therein, we characterize the reduced subspace based on the invariance equations and explicitly account for system inputs. In Section 3, we provide an overview of general-purpose MOR methods. We divide this review into intrusive methods in Section 3.1 and non-intrusive methods in Section 3.2. During the review of intrusive methods, in Section 3.1.4, we observe that current manifold learning methods do not account for external inputs. Hence, we propose an extension of the machine learning manifold-Galerkin method to non-autonomous systems with inputs in Section 3.1.5. The review of general-purpose methods concludes with a theoretical comparison in Section 3.3. In Section 4, we review problem-specific MOR approaches that have been developed in the field of process systems engineering and draw connections to methods in Section 3. Additionally, we review applications of MOR for three important unit operations, namely distillation columns, heat exchangers, and reactors. Section 5 provides a comparative case study, where we reduce the process model of an air separation unit. Therein, we compare eight MOR approaches that are selected as a rep-

resentative ensemble of the methods reviewed. Notably, we limit ourselves to MOR and do not consider hyperreduction and CPU costs of integration to avoid an extra level of methods and complexity. However, we expect that any ROM of sufficiently low order can be made real-time capable by appropriate hyperreduction. In Section 6, we summarize and discuss the strengths and weaknesses of the MOR methods. Section 7 concludes the manuscript.

Notation. Given a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_m]$, the linear space spanned by the column vectors $\boldsymbol{a}_i \in \mathbb{R}^n$ is denoted by $\mathcal{A} = \mathrm{col}(\mathbf{A})$ and called the column space of $\mathbf{A}$. For brevity, we prefer this notation over $\mathcal{A} = \mathrm{span}\{\boldsymbol{a}_1, ..., \boldsymbol{a}_m\}$. Given a real non-square matrix $\mathbf{A}$, the pseudoinverse is $\mathbf{A}^+ \triangleq (\mathbf{A}^\mathrm{T}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}$. Given some domain $\mathcal{X}$, we denote by $\mathcal{C}^r(\mathcal{X})$ the set of r-times continuously differentiable functions over $\mathcal{X}$, $r \geq 1$, and call a function f smooth if $f \in \mathcal{C}^\infty(\mathcal{D})$. We denote the identity matrix by $\mathbf{I}$ and the identity map by $\mathbf{id}$. Given a single independent coordinate $t \in \mathbb{R}$ (here time), we call the map $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ a curve. We denote the velocity of a curve $\boldsymbol{\gamma}$ at time t by $\dot{\boldsymbol{\gamma}}(t) \triangleq \frac{\mathrm{d}\gamma}{\mathrm{d}t}|_t$. Given some $\mathcal{C}^r$-map $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, where $n, m \geq 1$, we denote the linear differential map of $\boldsymbol{f}$ at $\boldsymbol{x} \in \mathbb{R}^n$ by $\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^m$. Given some point $\boldsymbol{x} \in \mathbb{R}^n$, we associate $\mathrm{D}\boldsymbol{f}(\boldsymbol{x})$ with its Jacobian matrix $\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{m\times n}$. For a function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ of two arguments, we consider $\mathrm{D}_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ and $\mathrm{D}_{\boldsymbol{y}}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ separately. An immersion is a smooth map, $\boldsymbol{h} : \mathcal{Z} \subseteq \mathbb{R}^{n_z} \to \mathcal{X} \subseteq \mathbb{R}^{n_x}$, satisfying $\mathrm{rank}(\mathrm{D}_z\boldsymbol{h}^\dagger(\cdot)) = n_z$ (injective) everywhere in $\mathcal{Z}$.

We consider embedded submanifolds of ambient space $\mathbb{R}^n$ characterized in terms of level sets. The respective definition is intuitive and sufficiently general for our purposes.

Definition 1 *(Level set,* [160]*)* Let $\boldsymbol{c} : \mathcal{X} \to \mathbb{R}^m$, $\mathcal{X} \subseteq \mathbb{R}^n$ open, $1 \leq m < n$, be a $\mathcal{C}^r$-mapping, $r \geq 1$. Denote by $\mathcal{M} \subseteq \mathbb{R}^n$ the zero level set:

$$\mathcal{M} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{c}(\boldsymbol{x}) = \mathbf{0}\}\,. \tag{1}$$

Assume that $\mathcal{M}$ is non-empty and the Jacobian $\mathrm{D}\boldsymbol{c}(\boldsymbol{x}) \in \mathbb{R}^{m\times n}$ has constant rank m at all $\boldsymbol{x} \in \mathcal{M}$. Then, $\mathcal{M}$ is a manifold of class $\mathcal{C}^r$, *i.e.*, a differentiable manifold. Moreover, $\mathcal{M}$ has dimension $k = n - m$ and is embedded in $\mathbb{R}^n$, and therefore called an embedded k-submanifold of $\mathbb{R}^n$. By the definition of continuity, the preimage $\mathcal{M} = \boldsymbol{c}^{-1}(\{\mathbf{0}\})$ is closed in $\mathbb{R}^n$ and therefore $\mathcal{M}$ is said to be properly embedded.

We remark that the embedding is not an essential part of the definition, *i.e.*, there exist more generic definitions that are independent of an ambient space [160]. Therefore, we sometimes drop the terms embedded or "sub" (manifold). Note that $\mathbb{R}^n$ is a manifold itself. By limiting ourselves to embedded submanifolds of open subsets $\mathcal{X} \subseteq \mathbb{R}^n$ of Euclidean space $\mathbb{R}^n$ defined through level sets, we circumvent geometric concepts like charts and tangent bundles. We refer the interested reader to [31, 37, 160, 199]. Below, we always consider the case $m < n$, where $n = n_x$ (FOM dimension) and $k = n_z \geq 1$ (ROM dimension). For notational and conceptual simplicity, we will focus on $\mathcal{C}^\infty$ maps and $\mathcal{C}^\infty$ manifolds. However, we remark that most methods are applicable to a less restrictive setup, such as $\mathcal{C}^2$, which is occasionally also referred to as "smooth" in the literature.

2 Model order reduction problem

In this section, we present a general formulation of the MOR problem. We are concerned with two types of non-autonomous dynamical systems with inputs. We focus on nonlinear dynamical systems with external inputs, described by a system of ordinary differential equations (ODE):

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t))\,, \tag{2}$$

where $t \in [t_0, t_\infty]$ denotes time, $\boldsymbol{x}(t) \in \mathcal{X}$ denote the differential states, $\mathcal{X} \subseteq \mathbb{R}^{n_x}$, $\boldsymbol{u}(t) \in \mathcal{U}$ are the system inputs, $\mathcal{U} \subseteq \mathbb{R}^{n_u}$, and $\boldsymbol{f} : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^{n_x}$ is smooth. We call $\mathcal{X}$ the state space, $\mathcal{U}$ the input space, and (2) a state space model. Notably, the input trajectory is generally unknown, *i.e.*, $\boldsymbol{u}(t)$ are unknown time-variant parameters. The dimension n_x is the model order.

We further consider semi-explicit differential-algebraic systems of equations (DAE) with external inputs:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{y}(t), \boldsymbol{u}(t))\,, \tag{3a}$$

$$\mathbf{0} = \boldsymbol{g}(\boldsymbol{x}(t), \boldsymbol{y}(t), \boldsymbol{u}(t))\,, \tag{3b}$$

where $\boldsymbol{x}(t) \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$, $\boldsymbol{y}(t) \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$ are the algebraic variables, and $\boldsymbol{u}(t) \in \mathcal{U} \subseteq \mathbb{R}^{n_u}$ are the external inputs. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{U}$ be open, and $\boldsymbol{f} : \mathcal{X} \times \mathcal{Y} \times \mathcal{U} \to \mathbb{R}^{n_x}$ and $\boldsymbol{g} : \mathcal{X} \times \mathcal{Y} \times \mathcal{U} \to \mathbb{R}^{n_y}$ be smooth. We say that the DAE has index one if $\mathrm{D}_{\boldsymbol{y}}\boldsymbol{g}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{u}})$ has rank n_y for any $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}) \in \hat{\mathcal{X}} \times \hat{\mathcal{U}}$ on the solution set of $\boldsymbol{g}$ [33, 217].

The class of semi-explicit DAEs covers a wide range of process models, where ODEs are often not sufficiently versatile. However, MOR methods are typically presented in the ODE framework. Fortunately, this is not a problem because an index-one DAE (3) may be regarded as an ODE that lives on a submanifold of $\hat{\mathcal{X}} \times \hat{\mathcal{Y}} \times \hat{\mathcal{U}}$ [205, 241]. Applying Definition 1 to (3b) verifies this point of view. Moreover, higher-index DAEs can usually be reduced to index-one systems [33]. Consequently, we can adopt ODE methods to index-one DAEs, *cf.* also [116, 172]. Further works that explicitly deal with DAEs are, *e.g.*, in [162, 166, 273, 277, 289]. For conciseness, we will only consider FOMs of the

form (2). However, the MOR methods can also be applied to (3) under the assumption of index-one.

The second reason for introducing both (2) and (3) is that some MOR methods reduce a high-order ODE to a low-order DAE, which we generally assume to be index-one. Hence, we will frequently encounter ROMs of the form (3). Next, we formalize the MOR problem for nonlinear systems with external inputs:

Problem 1 *(MOR)* Given a FOM (2), find a low-order index-one DAE (the prototype of a ROM):

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{f}_{\mathrm{r}}(\boldsymbol{z}(t), \boldsymbol{u}(t)), \tag{4a}$$

$$\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{z}(t), \boldsymbol{x}_{\mathrm{r}}(t), \boldsymbol{u}(t)), \tag{4b}$$

where differential states $\boldsymbol{z}(t) \in \mathcal{Z}$, $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$, $n_z \ll n_x$, algebraic variables $\boldsymbol{x}_{\mathrm{r}}(t) \in \mathcal{X}$ that approximate the FOM states $\boldsymbol{x}(t)$, and where $\boldsymbol{f}_{\mathrm{r}} : \mathcal{Z} \times \mathcal{U} \to \mathbb{R}^{n_z}$ and $\boldsymbol{h} : \mathcal{Z} \times \mathcal{X} \times \mathcal{U} \to \mathbb{R}^{n_x}$ are smooth. Let (4) be such that for any $\boldsymbol{x}(t_0) \in \mathcal{X}$, $\boldsymbol{u}(t_0) \in \mathcal{U}$, and $\boldsymbol{u}(t) = \hat{\boldsymbol{u}} \in \mathcal{U}$, $t > t_0$, the prediction error, $\varepsilon(t) \triangleq \|\boldsymbol{x}_{\mathrm{r}}(t) - \boldsymbol{x}(t)\|$, satisfies $\varepsilon(t) < \varepsilon^*$ for all $t > t_1$, where $t_1 \geq t_0$ and $\varepsilon^* \geq 0$ are sufficiently small.

We call (4a) the latent dynamics or internal dynamics of the ROM. For simplicity, we do not impose further requirements on the upper bound ε^*, *e.g.*, the existence of a contraction mapping [10, 142]. In general, a quickly decaying $\varepsilon(t)$ also depends on the ROM initialization strategy, where naïve minimization of $\|\boldsymbol{x}_{\mathrm{r}}(t_0) - \boldsymbol{x}(t_0)\|_2$ may result in a suboptimal $\boldsymbol{z}(t_0)$ [212, 242].

The dimension n_z is the reduced model order. A suitable value of n_z is fundamentally system- and application-dependent and no universal rule can be given. More specifically, the feasible degree of reduction depends on factors such as i. the type of dynamical system and its attractors (*e.g.*, multiplicity, stability/hyperbolicity), ii. the chosen reduction method, iii. the required accuracy and computational costs of the ROM, and iv. the time-scales/frequencies and subset of state-space relevant to application (*e.g.*, prediction horizon, time steps, bounded region on controllable manifold). For example, some reduction methods provide specific requirements, *e.g.*, on the slow/fast time-scale separation [106, 270], thus limiting the feasible degree of reduction.

Next, we characterize the subset of $\mathcal{X}$ on which the ROM (4) evolves. To this end, we first examine the level set:

$$\mathcal{M} \triangleq \{(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) \in \mathcal{X} \times \mathcal{Z} \times \mathcal{U} : \boldsymbol{0} = \boldsymbol{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}})\}. \tag{5}$$

Due to the index-one assumption in Definition 1, $\mathcal{M}$ constitutes a smooth embedded submanifold of $\mathcal{X} \times \mathcal{Z} \times \mathcal{U}$, where $\dim \mathcal{M} = n_z + n_u$. In addition, when fixing $\hat{\boldsymbol{u}} \in \mathcal{U}$, the level set:

$$\mathcal{M}(\hat{\boldsymbol{u}}) \triangleq \{(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}) \in \mathcal{X} \times \mathcal{Z} : \boldsymbol{0} = \boldsymbol{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}})\} \tag{6}$$

constitutes a smooth embedded n_z-submanifold of $\mathcal{X} \times \mathcal{Z}$, where $\mathcal{M}(\hat{\boldsymbol{u}}) \times \{\hat{\boldsymbol{u}}\}$ is a slice of $\mathcal{M}$. Finally, we denote the canonical projection of $\mathcal{M}(\hat{\boldsymbol{u}})$ on $\mathcal{X}$ by:

$$\mathcal{M}_{\mathcal{X}}(\hat{\boldsymbol{u}}) \triangleq \{\hat{\boldsymbol{x}} \in \mathcal{X} : \boldsymbol{0} = \boldsymbol{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}),\ \hat{\boldsymbol{z}} \in \mathcal{Z}\}. \tag{7}$$

To satisfy the requirements in Problem 1, we search for an invariant manifold $\mathcal{M}$ having the property that $\boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t))$ is tangent to $\mathcal{M}_{\mathcal{X}}(\boldsymbol{u}(t))$ if $\boldsymbol{x}(t) \in \mathcal{M}_{\mathcal{X}}(\boldsymbol{u}(t))$. Then, Eq. (4) solves the invariance equations:

$$\forall (\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) \in \mathcal{M} :\ \mathrm{D}_{\boldsymbol{x}}\boldsymbol{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}})\boldsymbol{f}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}) + \mathrm{D}_{\boldsymbol{z}}\boldsymbol{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}})\boldsymbol{f}_{\mathrm{r}}(\hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) = \boldsymbol{0}, \tag{8}$$

which can be derived by differentiating (4b) with respect to t and inserting (2) and (4a) for $\boldsymbol{u}(t) = \hat{\boldsymbol{u}}$ fixed. Finally, notice that invariance is closely connected to conservation principles, *e.g.*, the conservation of mass and energy [36, 309].

There exist various local and global approaches for determining invariant (sub)manifolds of nonlinear dynamical systems [132, 153, 283]. Within many of these approaches, the invariance equations (8) play a key role. For example, the parameterization method [40, 111, 112] and extensions [45, 210, 286, 289] have been proposed as a local intrusive method to construct invariant manifolds. If a rigorous solution of the invariance equations is impractical, the invariant manifold may be approximated based on system data (see Section 3.1.4). Finally, projecting the FOM onto an invariant manifold enables a reduced-order representation valid on this subspace.

A fundamental difference to autonomous systems is that changes in $\boldsymbol{u}(t)$ can affect $\mathcal{M}_{\mathcal{X}}$. For example, changing $\boldsymbol{u}(t)$ may rotate or shift the respective embedded $\mathcal{M}_{\mathcal{X}}$. If $\boldsymbol{u}(t)$ is piecewise constant, then $\mathcal{M}_{\mathcal{X}}$ may change its shape abruptly at the input grid points. If $\boldsymbol{u}(t)$ is continuous, then $\mathcal{M}_{\mathcal{X}}$ may gradually deform over time. When changing $\boldsymbol{u}(t)$ from $\hat{\boldsymbol{u}}^-$ to $\hat{\boldsymbol{u}}^+$ at time t_1, the prior $\boldsymbol{x}_{\mathrm{r}}^- \triangleq \lim_{t \nearrow t_1} \boldsymbol{x}_{\mathrm{r}}(t)$ may not be in $\mathcal{M}_{\mathcal{X}}(\hat{\boldsymbol{u}}^+)$. In the ROM, this offset is immediately addressed by a projection of $\boldsymbol{x}_{\mathrm{r}}^-$ onto $\mathcal{M}_{\mathcal{X}}(\hat{\boldsymbol{u}}^+)$, inherently defined by (4b). Accordingly, the ROM can exhibit a direct feedthrough from the inputs $\boldsymbol{u}(t)$ to the reconstructed states $\boldsymbol{x}_{\mathrm{r}}(t)$, as indicated by (4b). This projection corresponds to the fast-time-scale response of the FOM to changes in $\boldsymbol{u}(t)$. Furthermore, depending on $\boldsymbol{h}$, the associated projection after a change in $\boldsymbol{u}(t)$ can introduce unwanted artifacts in the ROM response (see Section 3.1.2).

A common simplification of (4) is to use a one-to-one immersion $\boldsymbol{h}^\dagger : \mathcal{Z} \to \mathcal{X}$, so that $\mathcal{M}_{\mathcal{X}} = \{\boldsymbol{h}^\dagger(\hat{\boldsymbol{z}}) : \hat{\boldsymbol{z}} \in \mathcal{Z}\}$ and $\dot{\boldsymbol{x}}_{\mathrm{r}}(t) \in \mathrm{col}(\mathrm{D}\boldsymbol{h}^\dagger(\boldsymbol{z}(t)))$ on $\mathcal{M}$. In this case, we have $\mathcal{M}(\hat{\boldsymbol{u}}) \equiv \mathcal{M}^*$, $\forall \hat{\boldsymbol{u}} \in \mathcal{U}$, and $\mathcal{M} = \mathcal{M}^* \times \mathcal{U}$ (trivial bundle), and the invariance equations (8) become:

$$\forall (\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) \in \mathcal{M} :\ \boldsymbol{f}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}) - \mathrm{D}_{\boldsymbol{z}}\boldsymbol{h}^\dagger(\hat{\boldsymbol{z}})\boldsymbol{f}_{\mathrm{r}}(\hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) = \boldsymbol{0}. \tag{9}$$

3 General-purpose reduction methods

We review state-of-the-art reduction methods that are applicable to a wide range of systems, rather than developed for a specific type of process model. Additionally, in Section 3.1.5, we develop an extension of the manifold-Galerkin method (reviewed in Section 3.1.4) to explicitly account for the dependence of the slow manifold on the inputs. For the sake of brevity, we omit equations as far as possible and rather focus on an informal discussion.

3.1 Intrusive general-purpose methods

We begin with a review of intrusive MOR methods, which involve the FOM equations in the MOR procedure. First, we discuss two classical intrusive approaches: proper orthogonal decomposition (POD) and singular perturbation theory (SPT). Afterwards, we present more recent extensions using machine learning, specifically manifold learning methods.

The POD belongs to the class of linear subspace methods, whereas SPT constructs a nonlinear subspace. Linear subspace methods are typically divided into methods based on singular values decomposition (SVD) and Krylov subspace methods [6]. Herein, we do not discuss Krylov subspace methods [12], which are major reduction techniques for linear systems, but their extension to nonlinear MIMO systems is not straightforward. We refer to [9, 228] for some recent works towards this goal. An extensive overview of MOR using Krylov subspace methods is given in [18].

3.1.1 POD-Galerkin method

The POD, also known as Karhunen-Loéve decomposition, is a method to extract a low-dimensional linear basis capturing the dominant behavior of a system [46]. Usually, the POD basis is determined through SVD of discrete state trajectory samples (snapshots) from simulation studies. Define the data matrix $\mathbf{D} \in \mathbb{R}^{n_x \times N}$, $N \geq n_x$, by stacking snapshots $\boldsymbol{x}(t_k)$, $k = 1, 2, ..., N$. Compute the mean $\bar{\boldsymbol{x}} \in \mathbb{R}^{n_x}$ and form the zero-centered matrix $\mathbf{X} = [\boldsymbol{x}(t_1) - \bar{\boldsymbol{x}}, ..., \boldsymbol{x}(t_N) - \bar{\boldsymbol{x}}]$. Then, SVD provides:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}, \tag{10}$$

where $\Sigma \in \mathbb{R}^{n_x \times N}$ is a non-square diagonal matrix of ordered entries σ_i, which are the singular values. The orthogonal matrices $\mathbf{U} \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{V} \in \mathbb{R}^{N \times N}$ collect the corresponding left and right singular vectors, respectively. The columns of $\mathbf{U}$ are also called principal components of $\mathbf{X}$, since they are ordered by their relevance to represent the data $\mathbf{X}$. As opposed to the eigendecomposition of $\mathbf{X}$, the SVD is well-conditioned and stable [5]. However, SVD is sensitive to data scaling, so improper scaling can result in a poor ROM [46].

The POD-Galerkin method consists of two main steps. First, SVD provides the POD modes $\mathbf{U}$. Next, we specify the reduced order $n_z < n_x$ and subdivide $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, where $\mathbf{U}_1 \in \mathbb{R}^{n_x \times n_z}$. Because $\mathbf{U}_1$ carries the dominant POD modes, we may truncate the modes $\mathbf{U}_2$ and use the approximation:

$$\boldsymbol{x}(t_k) - \bar{\boldsymbol{x}} \approx \mathbf{U}_1 \boldsymbol{z}(t_k), \ k = 1, 2, ..., N, \tag{11}$$

where $\boldsymbol{z}(t_k) \in \mathbb{R}^{n_z}$ are lower-dimensional coordinates and $\bar{\boldsymbol{x}} \in \mathbb{R}^{n_x}$ is the mean used in the SVD above. Eq. (11) corresponds to the optimal, *i.e.*, proper, approximation of the data $\mathbf{X}$ in the 2-norm [22]. However, the optimality of $\mathbf{U}_1$ refers only to the training data and not to ROM prediction accuracy [237]. The reduced state space is defined as the affine subspace, $\mathcal{M}_{\mathcal{X}} = \{\bar{\boldsymbol{x}} + \boldsymbol{v} : \boldsymbol{v} \in \mathrm{col}(\mathbf{U}_1)\}$. Using (11), we perform a Galerkin projection of the FOM (2) onto the basis $\mathbf{U}_1$ and obtain the ROM [237]:

$$\begin{aligned} \dot{\boldsymbol{z}}(t) &= \mathbf{U}_1^{\mathrm{T}} \boldsymbol{f}(\mathbf{U}_1 \boldsymbol{z}(t) + \bar{\boldsymbol{x}}, \boldsymbol{u}(t)), \\ \boldsymbol{x}_{\mathrm{r}}(t) &= \mathbf{U}_1 \boldsymbol{z}(t) + \bar{\boldsymbol{x}}. \end{aligned} \tag{12}$$

The POD is data-driven and therefore non-intrusive, but subsequent Galerkin projection is intrusive. Combining data-driven reduced basis and physics-based FOM, the nonlinear ROM (12) is a hybrid model. As the original map $\boldsymbol{f}$ is still contained in ROM (12), an additional hyperreduction step may be required to enable real-time applications [30]. In particular, the matrix $\mathbf{U}_1$ is usually dense, so a sparse model $\boldsymbol{f}$ may even lose its sparsity.

Alternative methods to POD exist that construct a non-orthogonal reduced basis, *e.g.*, sequential greedy approaches [29]. Projecting a FOM onto such a non-orthogonal basis is called a Petrov-Galerkin projection [5]. Greedy methods have long been considered as a remedy for the memory issues of standard SVD when applied to high-order problems. However, modern matrix decomposition methods, particularly randomized SVD [76, 104], can efficiently handle millions or billions of states and snapshots. Hence, the POD-Galerkin method has become a feasible option also for very high order.

3.1.2 Singular perturbation theory

Model reduction by SPT is a nonlinear subspace method based on time-scale separation of a FOM [67, 147]. Assume that the FOM (2) is originally in singular perturbation canonical form [142]:

$$\dot{\boldsymbol{x}}_1(t) = \boldsymbol{f}_1(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \boldsymbol{u}(t)), \tag{13a}$$

$$\varepsilon \dot{\boldsymbol{x}}_2(t) = \boldsymbol{f}_2(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \boldsymbol{u}(t)), \tag{13b}$$

where $\varepsilon \ll 1$ and $\boldsymbol{x}_1(t) \in \mathbb{R}^{n_z}$ and $\boldsymbol{x}_2(t) \in \mathbb{R}^{n_x - n_z}$. Eqs. (13a) and (13b) collect the slow and fast dynamics,

respectively. We assume that exponential stability of the fast dynamics holds uniformly.

The key idea of perturbation theory is that the system response does not change significantly when ε is perturbed. When setting $\varepsilon = 0$ in (13), the fast dynamics degenerate into quasi-stationary equations, yielding a ROM defined by:

$$\dot{\boldsymbol{x}}_1(t) = \boldsymbol{f}_1(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \boldsymbol{u}(t)), \tag{14a}$$

$$\mathbf{0} = \boldsymbol{f}_2(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \boldsymbol{u}(t)). \tag{14b}$$

Assuming that (14) is index-one, the corresponding slow manifold is given by:

$$\mathcal{M} = \{(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \hat{\boldsymbol{u}}) \in \mathbb{R}^{n_z} \times \mathbb{R}^{n_x - n_z} \times \mathbb{R}^{n_u} : \boldsymbol{f}_2(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \hat{\boldsymbol{u}}) = \mathbf{0}\}. \tag{15}$$

Overall, the projection of (13) onto $\mathcal{M}$ is inherently defined via the degeneration.

Typically, a process model is initially not in the form (13). Identifying the perturbation parameter ε is often not obvious and frequently based on human assumptions rather than system-theoretic arguments. In many cases, we seek a linear transformation, $\boldsymbol{x}(t) = \mathbf{T}_1\boldsymbol{x}_1(t) + \mathbf{T}_2\boldsymbol{x}_2(t)$. For chemical process systems, $\mathbf{T}_1$ and $\mathbf{T}_2$ are sometimes determined based on a time scale assumption on energy and mass transport [246, 295]. More rigorous approaches search for a subset of $\mathcal{X}$ or $\mathcal{X} \times \mathcal{U}$ whose defect of invariance is sufficiently small [111, 132, 142]. Further, note that implicit condensation [123, 204] and static condensation [106] in mechanics are closely connected to SPT. For example, the needed slow/fast time-scale separation for nonlinear vibrations was examined in [106, 270].

An important property of (14) is that the steady-state response remains exact. However, the degeneration of the fast dynamics occasionally introduces undesirable dynamic artifacts, especially a non-physical inverse response [20, 236]. In other words, the degenerated finite-ε dynamics have a substantial macroscopic effect when (13) does not constitute a proper normal form [244]. Also, the appearance of an inverse response depends on the degree of reduction and the determination of $\mathbf{T}_1$ and $\mathbf{T}_2$ [127]. In control, a sign-reversed model gain can degrade the closed-loop behavior of a model-based controller toward infeasibility or instability. However, a false inverse response is only problematic when appearing on the prediction time grid of the ROM application, where an extremely fast inverse response may be uncritical.

3.1.3 Residualization method

Closely related to the singular perturbation method is the POD residualization (POD-Res) or nonlinear-POD-Galerkin method [70, 93]. Instead of Galerkin projection onto a linear subspace via POD truncation, a nonlinear subspace ROM is obtained by considering the limit of infinitely fast (degenerate or slaved) dynamics. Let the POD modes $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3]$ so that $\mathbf{U}_1 \in \mathbb{R}^{n_x \times n_z}$ and $\mathbf{U}_2 \in \mathbb{R}^{n_x \times n_q}$, where $1 < n_q \leq n_x - n_z$. Accordingly, we first extend (12) and truncate $\mathbf{U}_3$ so that:

$$\dot{\boldsymbol{z}}_1(t) = \mathbf{U}_1^{\mathrm{T}} \boldsymbol{f}(\mathbf{U}_1\boldsymbol{z}_1(t) + \mathbf{U}_2\boldsymbol{z}_2(t) + \bar{\boldsymbol{x}}, \boldsymbol{u}(t)) \tag{16a}$$

$$\dot{\boldsymbol{z}}_2(t) = \mathbf{U}_2^{\mathrm{T}} \boldsymbol{f}(\mathbf{U}_1\boldsymbol{z}_1(t) + \mathbf{U}_2\boldsymbol{z}_2(t) + \bar{\boldsymbol{x}}, \boldsymbol{u}(t)), \tag{16b}$$

$$\boldsymbol{x}_{\mathrm{r}}(t) = \mathbf{U}_1\boldsymbol{z}_1(t) + \mathbf{U}_2\boldsymbol{z}_2(t) + \bar{\boldsymbol{x}}\,. \tag{16c}$$

Then, by residualization of (16b), *i.e.*, setting $\dot{\boldsymbol{z}}_2(t) = \mathbf{0}$, we obtain [271]:

$$\dot{\boldsymbol{z}}_1(t) = \mathbf{U}_1^{\mathrm{T}} \boldsymbol{f}(\mathbf{U}_1\boldsymbol{z}_1(t) + \mathbf{U}_2\boldsymbol{z}_2(t) + \bar{\boldsymbol{x}}, \boldsymbol{u}(t))\,, \tag{17a}$$

$$\mathbf{0} = \mathbf{U}_2^{\mathrm{T}} \boldsymbol{f}(\mathbf{U}_1\boldsymbol{z}_1(t) + \mathbf{U}_2\boldsymbol{z}_2(t) + \bar{\boldsymbol{x}}, \boldsymbol{u}(t))\,, \tag{17b}$$

$$\boldsymbol{x}_{\mathrm{r}}(t) = \mathbf{U}_1\boldsymbol{z}_1(t) + \mathbf{U}_2\boldsymbol{z}_2(t) + \bar{\boldsymbol{x}}\,. \tag{17c}$$

Despite the POD-based partitioning of states in the first step, the residualization provides a *nonlinear* subspace due to (17b). In [258], the authors compared POD-Galerkin and POD-Res, finding that low-order POD-Res ROMs are more accurate than POD-Galerkin ROMs but also more computationally expensive without hyperreduction.

When setting $n_q = n_x - n_z$, *i.e.*, applying an exhaustive residualization with no truncation, POD-Res can be regarded as an SPT method. This interpretation implies that the POD-based affine transformation divides the states into fast and slow dynamics. In our experience, a certain degree of POD truncation is possible without a relevant loss of precision, yet enabling a significant improvement of numerical conditioning. Therefore, we use $n_q < n_x - n_z$ below. Finally, there exist alternatives to POD in the context of nonlinear Galerkin methods [133, 186], *e.g.*, using wavelets [183]. In the next subsection, we discuss another nonlinear Galerkin variant based on machine learning, which circumvents the residualization step and is fully detached from a linear POD basis.

3.1.4 Manifold-Galerkin method

Applying machine learning to approximate a manifold from snapshot data is called manifold learning (MFL) or representation learning [181, 281]. MFL encompasses a wide array of tools and techniques that can be grouped into (i) learning methods, including autoencoder networks (AENs) [14, 152] and self-organizing maps [146], (ii) spectral methods, such as classical PCA or kernel PCA [196, 260], locally linear embeddings [249], Laplacian eigenmaps [19], and diffusion maps [52], and (iii) probabilistic methods [83]. These methods can also be combined, *e.g.*, classical PCA and quadratic decoding [15, 79] or classical PCA and AENs [229].

Not all MFL methods are suitable for MOR via Galerkin projection. The subset of viable options includes AENs, self-organizing maps, and kernel PCA [161]. Herein, we focus on MOR using AENs, which are a prominent choice, *e.g.*, [86, 138, 161]. We here consider undercomplete AENs, where $n_z < n_x$. First, the encoder network $\boldsymbol{\phi} : \mathcal{X} \to \mathbb{R}^{n_z}$, $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ open, projects the high-dimensional input vector $\boldsymbol{x} \in \mathcal{X}$ to a lower-dimensional representation $\boldsymbol{z} \in \mathbb{R}^{n_z}$. Often, $\boldsymbol{z}$ are called features or latent variables. In the second step, the decoder network $\boldsymbol{\phi}^\dagger : \mathcal{Z} \to \mathbb{R}^{n_x}$, where $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$, reconstructs the high-dimensional vector from the features. The images of the two mappings are $\boldsymbol{\phi}(\mathcal{X}) \subseteq \mathcal{Z}$ and $\boldsymbol{\phi}^\dagger(\mathcal{Z}) \subseteq \mathcal{X}$. The standard unsupervised AEN training problem minimizes the mean squared reconstruction error, $\sum_{k=1}^{N} \|\boldsymbol{x}(t_k) - \boldsymbol{\phi}^\dagger \circ \boldsymbol{\phi}(\boldsymbol{x}(t_k))\|_2^2$, given a data set $\mathcal{D} = \{\boldsymbol{x}(t_k)\}_{k=1}^{N}$ [288]. Assuming that $\boldsymbol{\phi}^\dagger$ is a one-to-one immersion, the subset $\mathcal{M}_\mathcal{X} \triangleq \{\boldsymbol{\phi}^\dagger(\hat{\boldsymbol{z}}) : \hat{\boldsymbol{z}} \in \mathcal{Z}\}$ is an n_z-dimensional embedded submanifold of $\mathcal{X}$. When using an mean squared training loss, the composition $\boldsymbol{\phi}^\dagger \circ \boldsymbol{\phi}$ approximates an orthogonal projection from $\mathcal{X}$ to $\mathcal{M}_\mathcal{X}$.

Intrusive MOR based on AENs is presented in [138, 161] and refined, *e.g.*, in [211, 229]. Here, we adopt the method of [161] and we comment on the refinements below. Training an AEN on snapshots approximates the slow manifold by $\mathcal{M}_\mathcal{X} \times \mathcal{U}$. Differentiation of the approximation $\boldsymbol{x}(t) \approx \boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))$ and insertion into (2) yields:

$$\begin{aligned} \mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))\dot{\boldsymbol{z}}(t) &= \boldsymbol{f}(\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)), \boldsymbol{u}(t)) \,, \\ \boldsymbol{x}_\mathrm{r}(t) &= \boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)) \,, \end{aligned} \tag{18}$$

where $\dot{\boldsymbol{z}}(t) \in \mathbb{R}^{n_z}$ and $\boldsymbol{x}_\mathrm{r}(t) \in \mathbb{R}^{n_x}$. Clearly, the map $\boldsymbol{\phi}^\dagger$ (and thus the activation function) need to be differentiable. To obtain a well-posed ROM, we multiply (18) by $\mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))^\mathrm{T}$:

$$\mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))^\mathrm{T}\mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))\dot{\boldsymbol{z}}(t) = \mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))^\mathrm{T}\boldsymbol{f}(\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)), \boldsymbol{u}(t)) \,, \tag{19a}$$

$$\boldsymbol{x}_\mathrm{r}(t) = \boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)) \,. \tag{19b}$$

As a result, we obtain a manifold-Galerkin projection of FOM (2) onto $\mathcal{M}_\mathcal{X}$ [161]. By applying the definition of the pseudoinverse, $\mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))^+$, Eq. (19) can be finally brought into the form of (4):

$$\dot{\boldsymbol{z}}(t) = \mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))^+ \boldsymbol{f}(\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)), \boldsymbol{u}(t)) \,, \tag{20a}$$

$$\boldsymbol{0} = \boldsymbol{x}_\mathrm{r}(t) - \boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)) \,. \tag{20b}$$

In practice, we employ (19) rather than (20) to circumvent matrix inversion or SVD associated with pseudoinversion in the model. Note that both (18) and (20) do not involve the encoding $\boldsymbol{\phi}$. While $\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))$ is an embedding of $\mathcal{Z}$ into $\mathcal{X}$, the map $\mathrm{D}\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))^+$ is a linear projection from $\mathbb{R}^{n_x}$ to $\mathbb{R}^{n_z}$ (more precisely from the tangent space to $\mathcal{X}$ at $\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t))$ onto the tangent space to $\mathcal{Z}$ at $\boldsymbol{z}(t)$). When using a linear AEN, then (20) reduces to (12). Finally, we highlight the close connection between (20) and other approaches using a single parameterization/decoder map, *e.g.*, [132, 242].

An alternative to pseudoinversion is to apply the left-inverse property, $\mathrm{D}\boldsymbol{\phi}(\boldsymbol{\phi}^\dagger(\cdot)) \circ \mathrm{D}\boldsymbol{\phi}^\dagger(\cdot) = \mathbf{I}$, to (18). This approach is also valid for non-orthogonal (oblique) projections and yields [211]:

$$\dot{\boldsymbol{z}}(t) = \mathrm{D}\boldsymbol{\phi}(\boldsymbol{\phi}^\dagger(\cdot))\boldsymbol{f}(\boldsymbol{\phi}^\dagger(\boldsymbol{z}(t)), \boldsymbol{u}(t)) \tag{21}$$

instead of (20a). However, computing $\mathrm{D}\boldsymbol{\phi}(\boldsymbol{\phi}^\dagger(\cdot))$ can be expensive and the precise satisfaction of the left-inverse property is a crucial prerequisite of this method.

Despite their intuitive structure and popularity in recent years, standard dense AENs suffer from limited applicability, poor scaling in the number of parameters and training costs, expensive model evaluation, moderate accuracy, and potential singularity issues. We discuss these challenges next. First, the simultaneous strength and weakness of AENs is their feedforward structure, which renders AENs conceptually and numerically simple to deal with. However, some manifolds cannot be globally represented by a single parameterization/decoder as provided by an AEN. Consequently, AENs may not be applicable to some problems. On the positive side, the slow manifold is represented by a feedforward structure rather than by a set of implicit equations, see (4b).

A second issue is that very high n_x and n_z cause an explosion in the parameters of dense encoder and decoder networks, resulting in intractable training and model evaluation. Additionally, embedding an AEN into a ROM increases the number of floating point operations of model evaluation considerably. For spatially distributed systems, computational improvements can be achieved by using convolutional instead of fully connected AENs [74, 248]. Moreover, there exist more parameter-efficient approaches to realize the autoencoder structure, *e.g.*, polynomials with compressed tensor coefficients have linear complexity scaling [94, 281]. On the other hand, we remark that the FOM equations, $\boldsymbol{f}$ in (20), are often already computationally expensive, necessitating a hyperreduction step in any case. Then, the AEN evaluation costs are secondary as the right-hand side of (20a) will be ultimately replaced by a cheaper function.

To improve ROM accuracy, the standard AEN training problem may be extended by a robustness loss [310] as well as a velocity loss [211], which we regard as a step towards physics-informed model reduction. Furthermore, a standard AEN often faces injection errors, *i.e.*, violating the identity $\boldsymbol{\varphi} \circ \boldsymbol{\varphi}^\dagger = \mathbf{id}$ or equivalently $\boldsymbol{\varphi}^\dagger \circ \boldsymbol{\varphi} \circ \boldsymbol{\varphi}^\dagger \circ \boldsymbol{\varphi} = \boldsymbol{\varphi}^\dagger \circ \boldsymbol{\varphi}$ (idempotence). This identity can be included into the training problem [281] or enforced structurally [211]. Moreover, learning an invariant foliation instead of a single manifold may increase accuracy of the reduced space [280, 281].

The AEN-based manifold-Galerkin method is sensitive to ill-conditioning of the Jacobian $\mathrm{D}\boldsymbol{\phi}^\dagger(\cdot)$. Unfortunately, the Jacobian from a standard AEN training may be arbitrarily ill-conditioned. However, conditioning issues can be addressed by including a rank loss [202] in the training or employing an invertible network structure possessing an exact (full-rank) Jacobian left-inverse [211].

Herein, we are particularly interested in systems with input. Although [161, 211] consider input-affine or parametric systems, these works do not account for the effect of inputs or parameters on the slow manifold, *cf.* (7). Instead, a trivial bundle $\mathcal{M} = \mathcal{M}_\mathcal{X} \times \mathcal{U}$ is constructed. However, we expect this simplification to demand an unnecessarily high-dimensional slow manifold for some systems. Consequently, we employ the concepts from Section 2 to propose a respective extension below.

3.1.5 Manifold-Galerkin method with inputs

We extend the machine learning manifold-Galerkin method [161] to account for the parametric effect of system inputs on the slow manifold $\mathcal{M}$ and $\mathcal{M}_\mathcal{X}$. This explicit parametric dependency has not been accounted for in the learning strategy proposed in [161] and related works (see Sections 3.1.4 and 3.2.4).

Restricting ourselves to AENs, we are generally interested in maps of the form $\boldsymbol{\varphi} : \mathcal{X} \times \mathcal{U} \to \mathcal{Z} \times \mathcal{U}$ and $\boldsymbol{\varphi}^\dagger : \mathcal{Z} \times \mathcal{U} \to \mathcal{X} \times \mathcal{U}$. The corresponding manifold is $\mathcal{M} = \{(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) \in \mathcal{X} \times \mathcal{Z} \times \mathcal{U} : \mathbf{0} = \hat{\boldsymbol{x}} - \boldsymbol{\varphi}^\dagger(\hat{\boldsymbol{z}}, \hat{\boldsymbol{u}})\}$, where $\boldsymbol{\varphi} \circ \boldsymbol{\varphi}^\dagger = \mathbf{id}$. For convenience, we substructure $\boldsymbol{\varphi}^\dagger$ by stacking the maps $\boldsymbol{\psi}^\dagger : \mathcal{Z} \times \mathcal{U} \to \mathcal{X}$ and $\mathbf{id} : \mathcal{U} \to \mathcal{U}$, so that $\mathcal{M}_\mathcal{X}(\hat{\boldsymbol{u}}) = \{\hat{\boldsymbol{x}} \in \mathcal{X} : \mathbf{0} = \hat{\boldsymbol{x}} - \boldsymbol{\psi}^\dagger(\hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}),\ \hat{\boldsymbol{z}} \in \mathcal{Z}\}$. Similarly, we substructure $\boldsymbol{\varphi}$ into the stacked maps $\boldsymbol{\psi} : \mathcal{X} \times \mathcal{U} \to \mathcal{Z}$ and $\mathbf{id} : \mathcal{U} \to \mathcal{U}$.

Given the above characterization of $\mathcal{M}$, Eq. (8) becomes:

$$\forall (\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) \in \mathcal{M} : \ \boldsymbol{f}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}) - \mathrm{D}_z \boldsymbol{\psi}^\dagger(\hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) \boldsymbol{f}_\mathrm{r}(\hat{\boldsymbol{z}}, \hat{\boldsymbol{u}}) = \mathbf{0}. \quad (22)$$

To guarantee that $\boldsymbol{z}(t)$ is not algebraically coupled to $\boldsymbol{u}(t)$, we simplify $\boldsymbol{\psi}$ to $\boldsymbol{\psi} : \mathcal{X} \to \mathcal{Z}$. Consequently, the encoding only depends on $\boldsymbol{x}(t)$. We apply an AEN training that minimizes the loss $\sum_{k=1}^N \|\boldsymbol{x}(t_k) - \boldsymbol{\psi}^\dagger(\boldsymbol{\psi}(\boldsymbol{x}(t_k)), \boldsymbol{u}(t_k)\|_2^2$. Similarly to Section 3.1.4, this loss may be extended to incorporate system knowledge and reduce projection errors. We call our approach manifold learning with inputs (MFLu). Next, we project $\boldsymbol{f}$ onto $\mathcal{M}_\mathcal{X}(\boldsymbol{u}(t))$:

$$\mathrm{D}_z \boldsymbol{\psi}^\dagger(\boldsymbol{z}(t), \boldsymbol{u}(t))\, \dot{\boldsymbol{z}}(t) = \boldsymbol{f}(\boldsymbol{\psi}^\dagger(\boldsymbol{z}(t), \boldsymbol{u}(t)), \boldsymbol{u}(t)). \quad (23)$$

Finally, after left-multiplication by the pseudoinverse $\mathrm{D}_z \boldsymbol{\psi}^\dagger(\boldsymbol{z}(t), \boldsymbol{u}(t))^+$, we obtain the ROM:

$$\begin{aligned} \dot{\boldsymbol{z}}(t) &= \mathrm{D}_z \boldsymbol{\psi}^\dagger(\boldsymbol{z}(t), \boldsymbol{u}(t))^+ \, \boldsymbol{f}(\boldsymbol{\psi}^\dagger(\boldsymbol{z}(t), \boldsymbol{u}(t)), \boldsymbol{u}(t)) \,, \\ \boldsymbol{x}_\mathrm{r}(t) &= \boldsymbol{\psi}^\dagger(\boldsymbol{z}(t), \boldsymbol{u}(t)) \,. \end{aligned} \quad (24)$$

This extension is related to studies on parametric model reduction; see [22, 24] for an overview. However, present works mostly deal with stationary parametric systems and use extensions of linear basis methods. In particular, [38, 179] combine a reduced POD basis and polynomial interpolation to build low-dimensional steady-state models of parametric problems. Refinements of this approach are found in [120, 184]. Reduction of parametric systems via manifold learning has also been investigated in the literature: To reduce the dimensionality of steady parametric flow problems, [72] present a strategy based on isomaps, and [57] apply kernel PCA. In [122], diffusion maps are used to learn submanifolds of the input-output space for reducing the parameter dimension. However, these works do not consider MOR of non-autonomous dynamical systems with inputs. Finally, we remark that MFLu is conceptually not restricted to AENs and may be adapted to other types of manifold learning.

3.2 Non-intrusive general-purpose methods

Besides the classical intrusive reduction methods, recent advances in machine learning and system identification have opened new avenues toward data-driven non-intrusive model reduction [34, 82, 144]. Importantly, non-intrusive reduction does not operate on the FOM equations and is therefore model-free. In addition, many non-intrusive methods combine MOR and hyperreduction in a single step. However, notice that non-intrusive reduction may still exploit some structural knowledge, *e.g.*, knowing that the FOM is a polynomial model [151]. Finally, the generic and homogeneous structure of non-intrusive ROMs can be exploited in optimization [262, 265]. Like Krylov subspace methods, we disregard non-intrusive rational approximation [191, 200, 220, 255], as these methods are not (yet) applicable to nonlinear multi-input systems.

3.2.1 Dynamic mode decomposition

Among the conceptually most simple yet effective non-intrusive methods is dynamic mode decomposition (DMD) [105, 259], a non-intrusive linear approach to construct linear ROMs, *i.e.*, linear latent dynamics on a linear subspace. DMD employs SVD to identify and truncate the principal linear modes from trajectory snapshot data. An extension of DMD to systems with inputs or controls (DMDc) is presented in [231], providing a ROM:

$$\begin{aligned} \boldsymbol{z}(k+1) &= \mathbf{A}\boldsymbol{z}(k) + \mathbf{B}\boldsymbol{u}(k), \\ \boldsymbol{x}_\mathrm{r}(k) &= \mathbf{U}_1 \boldsymbol{z}(k), \end{aligned} \quad (25)$$

where $\mathbf{U}_1 \in \mathbb{R}^{n_x \times n_z}$, $\mathbf{A} \in \mathbb{R}^{n_z \times n_z}$, and $\mathbf{B} \in \mathbb{R}^{n_z \times n_u}$. We refer to [231] for details on the computation of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{U}_1$. For a stable linear FOM, [173] presents bounds on the prediction

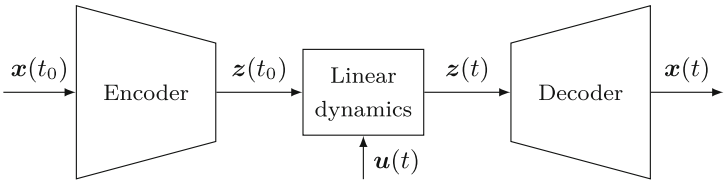

Fig. 2 Model structure of the Koopman-Wiener ROM

error, $e(k) \triangleq x_r(k) - x(k)$, for $k > N$. However, as discussed in [131], DMDc may encounter inaccuracies when applied to significantly nonlinear systems.

3.2.2 Applied Koopman theory

An extension of the DMD algorithm was proposed in [300]. Their approach still provides a linear ROM, but the mathematical procedure to construct the dynamics involves a prior nonlinear lifting step based on Koopman theory. Koopman theory [193, 214] postulates that a nonlinear dynamical system possesses a linear operator representation when lifted to a generally infinite-dimensional space of observables $g : \mathcal{X} \to \mathbb{C}$. Constructing a finite-dimensional or even low-order Koopman model by means of (E)DMD corresponds to a Galerkin projection of the Koopman operator onto a set of basis functions g_i, collected in a vector $\boldsymbol{g}$. Hence, despite being a linear method, DMD is applicable to nonlinear systems, when substituting state snapshots $\boldsymbol{x}^{(k)}$ by a sufficiently rich set of embedded snapshots $\boldsymbol{g}(\boldsymbol{x}^{(k)})$. Common choices of real-valued observables include Hermite polynomials, radial basis functions, sparse regression, and ANNs [167, 300]. More recently, kernel methods have gained attention due to their strong approximation capabilities [56, 114, 145].

Notably, a Koopman model does not necessarily predict states but only deals with observables. However, we can adapt the Koopman framework to state prediction by demanding that $\boldsymbol{x}(t)$ can be inferred from $\boldsymbol{g}(\boldsymbol{x}(t))$. To this end, a common assumption is that $\boldsymbol{x}(t)$ can be linearly reconstructed from the nonlinear observables. For example, [148] and follow-ups [109, 110, 201] use extended DMDc:

$$\begin{aligned} \dot{\boldsymbol{z}}(t) &= \mathbf{A}\boldsymbol{z}(t) + \mathbf{B}\boldsymbol{u}(t), \\ \boldsymbol{z}(t_0) &= \boldsymbol{g}(\boldsymbol{x}(t_0)), \\ \boldsymbol{x}(t) &= \mathbf{C}\boldsymbol{z}(t), \end{aligned} \tag{26}$$

where $\boldsymbol{z}(t) \in \mathbb{R}^{n_z}$ are the Koopman states, $\boldsymbol{g} : \mathcal{X} \to \mathbb{R}^{n_z}$ is the nonlinear observable map, and $\mathbf{A} \in \mathbb{R}^{n_z \times n_z}$, $\mathbf{B} \in \mathbb{R}^{n_z \times n_u}$, $\mathbf{C} \in \mathbb{R}^{n_x \times n_z}$. A ROM is obtained when setting $n_z < n_x$.

In the past years, Koopman theory has become a popular approach for both model bilinearization [223, 278, 299] and non-intrusive MOR [213, 222]. For autonomous systems, [176] relaxes the assumption of a global linear projection and proposed to combine linear dynamics and nonlinear reconstruction by means of a decoder network $\boldsymbol{g}^{\dagger}$. In [263], we have shown that this strategy is also valid for input-affine systems. Under certain assumptions on the observables $\boldsymbol{g}$, see [263], the respective ROM takes the form:

$$\begin{aligned} \dot{\boldsymbol{z}}(t) &= \mathbf{A}\boldsymbol{z}(t) + \mathbf{B}\boldsymbol{u}(t), \\ \boldsymbol{z}(t_0) &= \boldsymbol{g}(\boldsymbol{x}(t_0)), \\ \boldsymbol{x}_r(t) &= \boldsymbol{g}^{\dagger}(\boldsymbol{z}(t)), \end{aligned} \tag{27}$$

which is visualized by Fig. 2. Notice that the linear dynamics in (27) is a special case of the bilinear dynamics derived in [263]. For non-oscillating systems, we may choose a diagonal $\mathbf{A}$ and thereby reduce the trainable parameters. The combination of linear dynamics and nonlinear static map has Wiener structure and we term (27) a Koopman-Wiener (KW) model. In contrast to (20), we do not only project the nonlinear dynamics onto some subspace of $\mathcal{X}$, but we additionally demand the projection $\boldsymbol{g}$ to be linearizing. Consequently, the learning problem is more complex than standard MFL. Realizing the KW framework is most straightforward in a non-intrusive fashion, where a training problem combines reconstruction, prediction, and regularization loss terms [263].

Finally, the flexibility of machine learning enables modifications of the KW structure. For example, the encoder may serve as a state estimator [264].

3.2.3 Lift-and-project methods

Lift-and-project methods expand the FOM into an even higher-dimensional canonical form before projecting this representation to a low-dimensional subspace. The lifting step usually targets a linear, bilinear, or quadratic representation of the FOM. Notably, this extra step enables (i) the use of linear or quadratic/polynomial MOR methods, and (ii) circumvents hyperreduction due to mathematically simpler ROMs. Methods such as Koopman lifting [149], Carleman bilinearization [275], piecewise linearization [239], variational analysis via Taylor expansion [66, 228], and quadratization [39, 96, 117] lay the theoretical foundation of lift-and-project methods. An overview is given in [18] and software to automate the lifting was recently presented in [39, 65, 118].

In cases where intrusive variants of lift-and-project methods are mathematically involved or too computationally expensive, non-intrusive implementations are convenient to circumvent lifting the FOM. For example, the studies [151, 232] presented a non-intrusive lifting method for polynomial

FOMs, where snapshot data is lifted and POD-projected to learn a quadratic ROM. Therein, the model regression of the quadratic ROM is called operator inference [21, 221]. Extensions of operator inference to polynomial manifolds and ROMs are found, *e.g.*, in [78, 79]. Also, the non-intrusive (extended) DMDc and KW methods can be regarded as a variant of operator inference, where the latent dynamics are restricted to a linear or bilinear form.

3.2.4 Reduced basis and manifold learning methods

An alternative non-intrusive reduction approach is to combine reduced basis methods or manifold learning with system identification methods. For example, given POD-reduced snapshots, the latent ROM dynamics may be learned using recurrent neural networks [293], temporal convolutional ANNs [302], or feedforward ANN predictors [77]. Analogously, given AEN-compressed snapshots, an ROM can be obtained, *e.g.*, using ANNs [297] or Hammerstein-Wiener models [285]. These approaches are non-intrusive counterparts to (20). Non-intrusive operator inference for quadratic manifolds was presented in [79], where the authors infer a quadratic ROM from POD-projected snapshot data. This approach was extended to polynomial manifolds and ROMs in [78].

The above methods are two-stage procedures because the slow manifold is constructed independently of the latent dynamics. However, this separation may introduce a bottleneck, *i.e.*, the latent variables may not be suitable latent states, depending on the type of predictor. Interestingly, a similar bottleneck is also present in intrusive model reduction, as highlighted in [211, 212]. Non-intrusive single-stage approaches address this issue by learning slow manifold and latent dynamics simultaneously. Conveniently, the KW method reviewed above inherently follows this strategy [263]. In a similar but empirical setup, the works [90, 294] use AENs and linear time-invariant dynamics to learn stochastic ROMs. A combination of AEN and quadratic dynamics is used in [92]. Alternative approaches combine MFL and recurrent ANNs [87] or feedforward ANNs [8, 73, 86, 189] (termed MFL-ANN below), as well as MFL and sparse identification of latent dynamics [53].

3.3 Discussion and comparison

We present a comparison of the reviewed methods with respect to their nonlinear reduction properties in Table 1. First, we distinguish methods by their intrusive character (**Intr**). Intrusive MOR methods do not inherently perform hyperreduction (**Hr**) and thus require additional model simplification efforts for real-time applications. In contrast, non-intrusive methods use ROM structures that are typically (by design) much cheaper to evaluate so that a separate hyperreduction step is not needed.

Next, we distinguish whether the methods are snapshot-based (**Snap**). This categorization reveals another practical limitation, namely the ability to deal with a large number of inputs (**Nu**). Snapshot-based methods usually face the curse of dimensionality in n_u, since an independent excitation of inputs is commonly used to adequately cover the relevant state-space regions. As a consequence, snapshot-based methods are less straightforward to apply to systems with many inputs, *i.e.*, $n_u \gg 1$. Hence, for systems with many independent inputs $\boldsymbol{u}(t)$, snapshot-free intrusive methods are the most suitable.

Despite the curse of dimensionality in data sampling, input-response snapshots contain valuable information about controllability. As discussed in [298], when the snapshots are created via broad excitation of the system through the inputs, these data sample the controllable subspace. As a consequence, building a ROM based on input-response data inherently approximates the controllable subspace and eliminates the uncontrollable states (**Ctr**).

Next, we distinguish between methods that construct a linear (or affine) subspace versus a nonlinear manifold (**Manif**). As discussed in Section 1, using a linear subspace can limit the possible degree of reduction. On the other hand, computing a reduced basis by means of POD is numerically robust and computationally cheaper than MFL. The effort of a system-theoretic reduction, such as SPT, depends on the system knowledge and availability of heuristics. Overall, there is a trade-off between reduced order and reduction effort.

Lastly, we distinguish the reduction methods by the type of latent dynamics (**Lat**), see (4a). In general, the ROM of a nonlinear FOM is also nonlinear. However, DMDc and KW reduction methods construct linear latent dynamics, whereby the nonlinear ROM is simplified and structurally regularized. In other words, a ROM with linear latent dynamics cannot falsely exhibit strong nonlinear phenomena such as chaos or state multiplicity [219]. At the same time, these ROMs are limited to mildly nonlinear systems or short prediction horizons. However, a ROM with linear or bilinear latent dynamics is oftentimes sufficient, where more complex structures, *e.g.*, MFL-ANN, bear a higher risk of overfitting and undesired dynamical behavior.

Only a subgroup of nonlinear model reduction methods provide a-priori guarantees on stability preservation (**Stab**). These guarantees are established by system-theoretic arguments (SPT, AGG, POD-Res.) or through constrained learning problems (DMD, KW, and operator inference). In reduced basis methods, stability preservation may be enabled through variable transformation or modification of the inner product [16, 250]. Other methods require a careful a-posteriori investigation of ROM stability [22, 119]. We also mention methods developed to incorporate physics priors, particularly for Hamiltonian [2, 36, 95, 268], port-

Table 1 Comparison of nonlinear MOR methods that preserve state information. **Intr**: intrusive method, **Snap**: snapshot-based method, **Manif**: type of submanifold, **Lat**: type of latent dynamics, **Hr**: hyperreduction features, **Nu**: suitable for many inputs ($n_u \gg 1$), **Ctr**: approximates controllable subspace, **Stab**: preserves stability. ✓: yes, (✓): practical limitations, ×: no

	Method	Section	Intr	Snap	Manif	Lat	Hr	Nu	Ctr	Stab
a)	POD-Galerkin	3.1.1	✓	✓	Lin.	Nonlin.	×	(✓)	✓	(✓)
b)	SPT	3.1.2	✓	×	Nonlin.	Nonlin.	×	✓	×	✓
c)	POD-Res.	3.1.3	✓	✓	Nonlin.	Nonlin.	×	(✓)	✓	✓
d)	MFL-Galerkin	3.1.4	✓	✓	Nonlin.	Nonlin.	×	(✓)	✓	×
e)	MFLu-Galerkin	3.1.5	✓	✓	Nonlin.	Nonlin.	×	(✓)	✓	×
f)	(Extended) DMDc	3.2.1/2	×	✓	Lin.	Lin./Bilin.	✓	(✓)	✓	✓
g)	KW	3.2.2	×	✓	Nonlin.	Lin./Bilin.	✓	(✓)	✓	✓
h)	Operator inference	3.2.3	×	✓	Polyn.	Polyn.	✓	(✓)	✓	✓
i)	POD-ANN	3.2.4	×	✓	Lin.	Nonlin.	✓	(✓)	✓	×
j)	MFL-ANN	3.2.4	×	✓	Nonlin.	Nonlin.	✓	(✓)	✓	×

Hamiltonian [80, 97, 192], and Lagrangian systems [42, 75]. These formalisms support the construction of ROMs that respect energy and mass conservation and preserve stability [36, 164]. Here, the port-Hamiltonian framework is particularly suitable for dynamical systems with inputs or interconnected process subsystems.

Regarding the ROM error, $\boldsymbol{e}(t) \triangleq \boldsymbol{x}_\mathrm{r}(t) - \boldsymbol{x}(t)$, many publications on MOR techniques do not derive universal error bounds but perform a-posteriori ROM validation. Rigorous error bounds are presented, *e.g.*, in [10, 26, 140, 142, 267]. The accuracy of a ROM is directly coupled to modal closure and memory effects [3]. Hence, methods for ROM closure modeling, *e.g.*, the Mori-Zwanzig formalism [91], support tighter error bounds.

4 Reduced modeling of chemical processes

We now review specialized MOR methods developed in the process systems engineering field as well as applications of general-purpose methods to such problems. In Section 4.1, we review three specializations of MOR methods from the previous section, namely (i) compartment modeling, (ii) model aggregation, and (iii) collocation method. Afterwards, in Section 4.2, we discuss the application of MOR to three very common unit operations: (i) distillation columns, (ii) heat exchangers (HX), and (iii) reactors. Again, we omit equations as far as possible and rather focus on an informal discussion. Besides deterministic reduction, reduced modeling is often achieved based on knowledge or intuition through simplifying assumptions in the modeling process [64]. We also comment on these aspects.

4.1 Problem-specific approaches

4.1.1 Compartment modeling

The key concept in compartment modeling (COMP) is the definition of interconnected functional subsystems, termed compartments, between the local microscale and the global system scale [136]. When applied for MOR, the introduction of compartments is often motivated by a time scale assumption, where local phenomena inside a compartment are assumed to be much faster than the overall response of the compartment [20]. Given a discretized model of a conservative distributed system, a compartment is a cluster of neighboring cells, and the compartment equations are obtained by superposition of the extensive balance equations, *e.g.*, absolute enthalpy or total molar holdup, of these cells. In terms of intensive quantities, *e.g.*, molar enthalpy and molar fraction, this superposition corresponds to a mass-weighted averaging.

By applying the above time scale assumption, the local dynamics of cells within a compartment are degenerated to quasi-stationary equations, and we obtain a ROM. In addition, replacing quasi-stationary equations by shortcut methods, averaging rules, or regression models realizes a hyperreduction. COMP is a heuristic approach to transform an FOM into Eqs. (13) and (14). Hence, we regard COMP as a heuristic SPT method, inheriting properties such as the potential appearance of a non-physical inverse response.

4.1.2 Model aggregation

In order to eliminate the potential inverse response in COMP ROMs and improve model sparsity, [165] presents the aggregation method (AGG). AGG is a heuristic modification of SPT for 1D distributed systems. Instead of transforming the FOM into the canonical form (13), the authors multiply (only) the left-hand side of the FOM by a matrix $\tilde{\mathbf{H}} = \mathrm{diag}(\tilde{h}_1, \tilde{h}_2, ..., \tilde{h}_{n_x})$, $\tilde{h}_i \geq 0$. Thereby, the time constants of individual stage dynamics are artificially increased or reduced. $\tilde{\mathbf{H}}$ is a hyperparameter and commonly selected such that $\mathrm{trace}(\tilde{\mathbf{H}}) \approx n_x$. Let Ω be the $n_x \times n_x$ permutation matrix such that $\Omega\tilde{\mathbf{H}} = \left[\begin{smallmatrix}\mathbf{H} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right]$, where $\mathbf{H} = \mathrm{diag}(h_1, h_2, ..., h_{n_z})$, $h_i > 1$. The resulting ROM reads:

$$\mathbf{H}\dot{\boldsymbol{x}}_1(t) = \boldsymbol{f}_1(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \boldsymbol{u}(t)), \tag{28a}$$

$$\mathbf{0} = \boldsymbol{f}_2(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \boldsymbol{u}(t)). \tag{28b}$$

A systematic comparison of AGG to COMP is given in [170]. By contrasting (14) and (28), we notice that the manifolds on which the ROMs evolve are identical if $\boldsymbol{f}_2$ coincide. Hence, both methods may realize a different ROM living on the same manifold $\mathcal{M}$.

4.1.3 Collocation method

Motivated by the collocation method for solving ordinary and partial differential equations [71, 103], the studies [49, 301] adapt the collocation method to MOR of spatially discrete problems. The underlying assumption is a polynomial coherence of the differential variables of the FOM. Then, a ROM is obtained by combining polynomial interpolation and a number of discrete cells (collocation points), *e.g.*, finite volume elements or column stages. However, the collocation points do not necessarily correspond to cells in the FOM. Intuitively, the number and position of collocation points depend on the system, *i.e.*, whether a low-order polynomial is able to capture the underlying patterns characterizing the slow manifold. The positions of the collocation points are typically chosen a-priori as the roots of Legendre, Hahn, or Jacobi polynomials and not further adapted online [276, 279]. In other words, the spatial collocation points are determined once and for all, *i.e.*, do not change over time.

4.2 Applications

4.2.1 Distillation columns

Detailed distillation column models typically include stage-by-stage differential energy, mass, and possibly momentum balances [128]. Common simplifying assumptions are ideally mixed stages with no diffusion effects, thermodynamic vapor-liquid equilibrium, and quasi-stationary pressure and enthalpy dynamics of both phases [236, 246]. These assumptions rely on time-scale assumptions and may be regarded as a heuristic application of SPT, rendering such mechanistic models already reduced. However, these models are often too expensive for online applications and therefore require MOR [1, 261].

Early applications of MOR to columns are found in [49, 301] for COL and [84] for COMP. As mentioned above, when applying COL, the collocation points do not necessarily correspond to physical separation stages, *i.e.*, can be pseudo-stages. Moreover, constructing COL models based on a log-transformed FOM can improve model accuracy [41].

Using SPT, [20] derives a widely applicable COMP model form and shows a higher accuracy over COL. The COMP ROMs therein combine dynamic compartment balances and quasi-stationary stage models. The separation stage whose differential balances are replaced by compartment balances is termed sensitivity stage. The appearance of a non-physical inverse response depends on the number of compartments and the position of the sensitivity stages [127, 129]. However, avoiding an inverse response may limit the degree of order reduction. On the other hand, the particular structure of COMP models, featuring large sets of neighboring quasi-stationary stage models, enables hyperreduction via steady-state shortcut methods, *e.g.*, FUG shortcut [58] or group methods [60, 137]. Typically, these shortcuts have fewer parameters than universal regressors (*e.g.*, ANNs) and are less prone to overfitting. On the other hand, ANNs can reach almost arbitrary precision [256].

In steady operation, a distillation column exhibits a set of coherent composition patterns, which can be analytically derived under simplifying assumptions [130, 187]. These analytical solutions are wave-shaped profiles, and therefore called nonlinear wave propagation model. Ref. [143] proposed a semi-empirical ROM for multi-species distillation columns, by combining wave propagation equations, global column balances, and a boundary equilibrium assumption. The resulting ROM is a compartment model (a non-physical inverse response is also observed here). In [43], this ROM was extended to account for effects due to non-constant holdup.

The AGG method was originally developed for distillation columns [165]. Like COMP models, AGG ROMs involve clusters of quasi-stationary trays that can be simplified using shortcut models. Different from COMP, the AGG ROM states may be associated with actual stages in a distillation column. While avoiding a non-physical inverse response, the AGG predictions usually converge more slowly to the FOM states after a change of system inputs [169].

Besides the specialized MOR methods, general-purpose MOR methods have been applied to distillation column models, *e.g.*, using POD-Galerkin [100, 115], standard SPT [156], KW [262, 264], and DMDc [233]. However, as opposed to these general-purpose methods, the specialized methods, COMP, nonlinear wave propagation, AGG, and COL, build on the intrinsic system structure, providing generic low-order models that can be written (almost) independently of a FOM. In particular, these specialized ROMs can be pre-configured for generic columns with similar stage models and in many cases steady-state data is sufficient to customize these ROMs for a specific column. Moreover, COMP and AGG may access an extra set of hyperreduction methods in the form of steady-state shortcuts as described above.

4.2.2 Heat exchangers

HX are another type of distributed systems that are commonly associated with high-order FOMs. For industrial processes, the 3D discretized model of a multi-stream HX can reach an order of over a million states [102]. Clearly, such fine-grained models are prohibitive for real-time applications.

As an alternative to MOR, simplifying assumptions enable a heuristic degeneration or lumping of differential equations. Common assumptions are (i) a time scale separation between energy dynamics of fluids and walls and (ii) for parallel flow HX, a uniform common wall temperature in each cross section perpendicular to the principal flow axis [230]. Common examples of lumped parameter models are ε-NTU, P-NTU and LMTD [266]. However, these are typically only valid at stationary points and under simplified assumptions on geometry, heat transfer, and thermodynamics.

Under reasonable simplifying assumptions, such as constant heat capacities and linear heat transfer correlations, the HX FOMs are (quasi-)linear or bilinear [175]. In that case, linear or bilinear MOR may be applied [303], which is less involved than general nonlinear MOR. Moreover, when including dynamical HX submodels in a process model, the heat transfer phenomena often have significantly smaller time constants than mass conversion and mass transfer problems [13]. Under these assumptions, a dynamical HX model may fully degenerate into a quasi-stationary model.

Applications of general-purpose MOR methods to HX include POD-Galerkin [50, 180, 304], DMD [305], and POD combined with ANN predictors [168]. Furthermore, both classical collocation and Galerkin projection are standard methods to approximate low-order solutions of partial differential equations [108, 266]. In addition, all three specialized approaches (COMP, AGG, COL) have been applied to spatially discretized dynamical models of 1D distributed HX. Specifically, [41, 235] use the collocation method to reduce finite volume models of HXs. Lumped COMP models of HX are presented in [190, 194, 308]. Also, the reduced modeling of HX with phase change (evaporators or condensers) is often accomplished using the moving boundary model [134], which is a COMP model. The moving boundary model uses compartments with flexible boundaries to account for spatially moving phase regions and can exhibit a non-physical inverse response [287]. Finally, AGG is applied to HX in [171].

Similar to separation columns, there exist steady-state shortcut models and analytic solutions to replace steady-state equations in COMP and AGG models. Common lumped parameter models are reviewed in [266]. The work [175] presents an analytic solution to the stationary multi-stream HX equations under simplifying assumptions. Further, HX networks may be used as a surrogate model of multi-stream HX [113, 307].

4.2.3 Reactors

There are two main aspects that can result in high-order FOMs of chemical reactors. First, reaction systems can include high-dimensional multispecies kinetics with numerous reaction pathways [182, 238, 291]. Second, the spatial distribution of a reactor can prohibit lumped models, *e.g.*, for non-ideally mixed tanks or for tubular reactors [125, 271].

For kinetic systems, a reduction in the number of species corresponds to MOR, whereas a reduction in the number of modeled reactions may be regarded as hyperreduction. Analogously to the simplifying assumptions discussed for column and HX modeling, the FOM of a kinetic system is often already a simplification of the real mechanisms and so is the mechanistic model of a reactor. Naturally, for two-timescale kinetic systems with a clear partitioning between slow and fast dynamics, SPT can be directly applied [292].

A large set of works has been devoted to constructing invariant manifolds by solving the invariance equations of chemical kinetic systems with non-obvious slow and fast partitioning. These methods employ system linearization [182], computational singular perturbation [159], thermodynamic dissipativity and symmetry [89] relation graphs [174, 224], or sensitivity analysis and numerical optimization [28, 61, 227]. Extensive reviews are given in [48, 88, 274]. Recently, machine learning has been increasingly applied to reduce kinetic systems [135, 139, 218, 296]. Interestingly, the chemical reaction network formalism has also been applied as mathematical framework to facilitate quadratization of generic ODEs [65, 117].

For spatially distributed reaction systems, *e.g.*, reaction diffusion problems and tube reactors, the application of POD-Galerkin method [7, 125, 154, 216], nonlinear-Galerkin method [69, 93, 271], manifold-Galerkin method [161], and lift-and-reduce techniques [21, 151] have been reported. As shown in [161], nonlinear subspace methods can significantly outperform linear basis methods in the very-low-order region. The classical collocation method is applied, *e.g.*, in [81, 107, 163], to obtain low-order solutions of respective partial differential equations. Moreover, an overview of compartment modeling of distributed chemical reactors is found in [98].

For integrated reaction-separation systems (with recycle), the works [51, 155] performed a time scale analysis by means of SPT to derive ROMs for control. Along the same lines, [13] investigated integrated reaction-HX networks. Ref. [258] compared the POD-Galerkin and POD-Res method on an integrated reaction-separation system, finding similar ROM accuracy but higher CPU costs of POD-Res ROMs.

5 Case study

We compare the performance of MOR methods on an industrial process by studying the air separation unit presented in [44]. Clearly, comparing all the methods reviewed in Sections 3 and 4 would exceed the scope of this article. Hence, we examine a subset of methods listed in Table 1. Specifically, we compare the MOR approaches (a-g) and (j). Additionally,

we consider the two tailored methods COMP and AGG from Section 4.

We briefly comment on this choice of methods. Despite successful applications in the literature, we investigate neither COL nor nonlinear wave propagation models, because the former are known to exhibit a significant steady-state offset [20] and the latter are a special case of COMP. We refer to [41, 43] for applications of these methods to the process considered herein. As discussed in Section 3.2.3, we regard KW and DMDc as a variant of operator inference (h) and do not investigate other forms herein. Of the many possible approaches (i) and (j) reviewed in Section 3.2.4, we choose the MFL-ANN method [189] due to conceptual simplicity and low implementation efforts. Further, class (i) is theoretically included in (j) and therefore not investigated separately. In particular, we expect all single-stage methods in Section 3.2.4 to reach a comparable accuracy. Finally, recall that POD-Res and COMP represent data-driven and heuristic variants of SPT, respectively. In this regard, our case study covers SPT. As discussed in Section 1, we focus on MOR and do not investigate a potential hyperreduction and CPU times.

Figure 3 illustrates the process under investigation, which produces gaseous nitrogen and is built of a main air compressor with precooling, three multi-stream HXs, two turbines, a high-pressure rectifying column with 30 trays, and an integrated reboiler-condenser unit. The unit has four manipulated input variables, which are the molar flow rate of air F_{mac}, the molar flow fraction to the turbine ζ_{turb}, the reflux fraction at the column top ζ_{cond}, and the purge stream F_{drain}.

The system is open-loop unstable due to the integrator behavior of the liquid molar holdup M_{irc} of the reboiler tank. For unstable processes, data sampling is more challenging, especially if only a small subregion of the state space is relevant for ROM application. However, most chemical plants feature a stabilizing base layer as part of the automation hierarchy, so stability is not an issue [63]. For simplicity, we thus assume that the reboiler tank is equipped with an ideal inventory controller that manipulates F_{drain} to stabilize M_{irc} at its nominal value. Additionally, we fix ζ_{turb} to its nominal value. Consequently, we have two variable inputs $\boldsymbol{u} = [F_{\mathrm{mac}}, \zeta_{\mathrm{cond}}]^{\mathrm{T}}$.

For our purpose, the selected process features some characteristic properties of industrial process systems, such as variable coupling and nonlinear process response. Moreover, we regard the process as a good compromise between the complexity of the process system and the complexity of the MOR problem. However, the selected process does not incorporate a chemical reactor and has few inputs. The investigation of MOR for complex processes including reaction and many input variables is indispensable and should be undertaken in future studies.

5.1 Full-order model

We consider the three main species of air nitrogen, argon, and oxygen. In contrast, we do not model carbon dioxide, hydrocarbons, or other residual chemical species here. The FOM comprises material and energy balances of all process units as well as constitutive equilibrium-based thermodynamic equations. The column is modeled in a tray-to-tray fashion assuming thermal equilibrium with constant relative volatilities, constant specific heat capacities of all species, and ideal mixture (no excess properties). Both the column bottom tray and the reboiler are modeled as special column trays with an extra feed stream and fixed liquid outflow, respectively. The condenser is a total condenser with negligible material holdup and modeled using lumped quasi-stationary balance equations. We apply a logarithmic transformation of molar fractions to support the extraction of composition patterns, *i.e.*, the logarithmic molar fractions are the respective differential states. The HX models feature energy balances for the metal wall and quasi-stationary balances for the fluids, assuming a single flow chanel for each stream, 1D spatial discretization with n_{HX} cells per stream, and the common wall assumption. Following [44], we set $n_{\mathrm{HX1}} = 25$, $n_{\mathrm{HX2}} = 25$, and $n_{\mathrm{HX3}} = 1$. Lastly, first-order low-pass filters account for the lagged vapor flow rate in heat exchangers and column.

In total, the FOM comprises $n_u = 2$ inputs and $n_x = 176$ differential states, and is implemented in `Modelica`. Although n_x may appear low compared to some structural mechanic or fluid dynamics problems, the system investigated here constitutes a representative example of model reduction in the process control field, *cf.* [62, 235, 261]. The admissible input set $\mathcal{U}$ is a box defined by $F_{\mathrm{mac}} \in [30\,\mathrm{mol/s}, 50\,\mathrm{mol/s}]$ and $\zeta_{\mathrm{cond}} \in [0.51, 0.54]$. As the initial state $\boldsymbol{x}_0$, we use the nominal steady state corresponding to $F_{\mathrm{mac}} = 52.2\,\mathrm{mol/s}$, $\zeta_{\mathrm{cond}} = 0.535$, $\zeta_{\mathrm{turb}} = 0.805$, and $M_{\mathrm{irc}} = 25\,\mathrm{kmol}$. At $\boldsymbol{x}_0$, the process produces 25 mol/s nitrogen of 100 ppm impurity grade.

5.2 Data sampling

In order to guarantee a fair comparison of snapshot-based methods against each other, we always use the same training data. This data set is recorded by simulating the FOM subject to randomly ordered inputs drawn from an equidistant 5×5 grid in $\mathcal{U}$. Following [44], the sampling time is $\Delta t_{\mathrm{s}} = 15\,\mathrm{min}$. We impose the input sequence as steps (zero-order hold) of 240 h (10 days) each. Thereby, we include information from both the fast spectrum of the system response and the stationary states. We illustrate the input profile $\boldsymbol{u}(t)$ in the supplementary information.

We simulate the FOM in `Dymola 2025x` using the integrator `DASSL` [226] with integration tolerance 10^{-6}. The recorded data set comprises $N = 24\,000$ snapshots of each

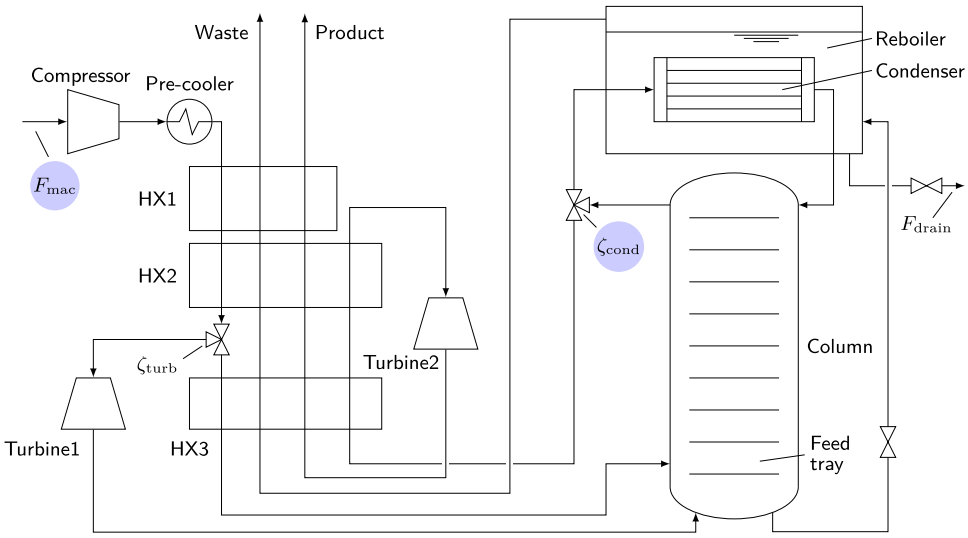

Fig. 3 Schematic process flowsheet of the investigated air separation unit. Blue circles indicate the system inputs $\boldsymbol{u}(t)$

state. The corresponding snapshot matrices have the form $\mathbf{X} \in \mathbb{R}^{n_x \times N}$ and $\mathbf{U} \in \mathbb{R}^{n_u \times N}$. These snapshots cover the operating space between 15 mol/s and 27 mol/s production rate and product quality between 10 ppm and 1000 ppm impurities.

5.3 Implementation

As previously discussed, POD and MFL are strongly affected by scaling. Herein, we scale all states and inputs individually to the range $[-1, 1]$ and apply zero-mean centering. We remark that a tailored scaling strategy, *e.g.*, scaling based on physical relevance and range, may enable more accurate ROMs (with respect to a target application). Like the FOM, we implement all intrusive ROMs in `Modelica` and perform all simulations in `Dymola` using `DASSL` with an integration tolerance of 10^{-6}. The non-intrusive ROMs (DMDc, KW, MFL-ANN) are directly evaluated in the respective machine learning environments in `Python` (see below). Next, we provide further details on the implementation.

5.3.1 POD methods and DMDc

We implement POD and DMDc using `SciPy` as a `Python` interface to the SVD implementation in `LAPACK`. All other matrix computations are performed using `Numpy`. We specify `float64` precision, which was found to be crucial for obtaining reliable SVD results. We scale the latent states $\boldsymbol{z}(t)$ heuristically, which was found to improve model conditioning. For DMDc, we calculate the coefficient matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{U}_1$ in (25) using the formulas given in [231]. To derive a POD-Res model, we use a truncated residualization, where $n_q = 50 - n_z$. The truncation improves simulation convergence without significant loss of precision.

5.3.2 MFL-Galerkin and MFLu-Galerkin methods

We train AENs in `Tensorflow` using the standard training problem [288]. To reduce the number of trainable parameters and speed up the training, we prereduce the scaled and zero-centered snapshots by projecting the data onto the first 100 POD modes [229]. For the success of MFL, it is crucial to not rescale the POD-projected snapshots again. The encoders and decoders have a single fully connected nonlinear hidden layer with 50 neurons with (smooth) `tanh` activation, respectively. Using deeper networks or additional neurons did not improve the reconstruction accuracy considerably.

We divide the pre-processed snapshots into 80% training data and 20% validation data. For every network, we run three trainings with random seeds and select the weights with the best performance on the validation data set. We train all networks for a total of 20 000 epochs and minibatch size 256, using `Adam` with initial learning rate of 10^{-4} for the first 15 000 epochs and `SGD` with learning rate of 10^{-6} for subsequent finetuning. For the implementation of MFLu, we modify the AEN structure from $(\boldsymbol{\varphi}, \boldsymbol{\varphi}^\dagger)$ to $(\boldsymbol{\psi}, \boldsymbol{\psi}^\dagger)$ as discussed in Section 3.1.5. However, the general training setup remains the same. To include the AENs in the ROMs, we implement the decoder network structure in `Modelica`.

5.3.3 KW models

We use our machine learning framework [263] to train reduced KW models. This framework is based on `Tensorflow` and implements the KW model structure along with a training procedure based on snapshot data. As in Section 5.3.2,

we pre-reduce the training data through projection onto the principal 100 POD modes. The trajectory snippets required for the prediction loss are created by sliding along the training data set in a moving horizon fashion, stopping every 10 samples and copying a series of 100 consecutive snapshots. The encoder and decoder networks have a symmetric structure with two hidden layers and linear output layer. The number of hidden neurons are determined via linear interpolation between 100 and n_z. We use `ELU` activation and train the model for 10 000 epochs and minibatch size 32, where the first 8 000 epochs use `ADAM` with learning rate of 10^{-3} and the final 2 000 epochs use `SGD` with learning rate of 10^{-6} for finetuning. All other hyperparameters have the default values from [263]. The ROM simulation employs the graph-based execution mode of `Tensorflow`.

5.3.4 MFL-ANN models

We implement the method in [189] by means of a modification of our aforementioned framework [263]. To this end, we substitute the linear dynamics in (27) by a feedforward ANN to learn the discrete-time latent predictor:

$$z_{k+1} = \boldsymbol{F}(z_k, \boldsymbol{u}_k), \tag{29}$$

where $\boldsymbol{F} : \mathcal{Z} \times \mathcal{U} \to \mathcal{Z}$ and $k \in \{0, 1, ...\}$. The ANN $\boldsymbol{F}$ features two hidden layers of each $2n_z$ neurons and `ReLU` activation, and a linear output layer. This specification is parsimonious compared to [189], who use three hidden layers of each $5n_z$ neurons. However, we found that overparameterization results in poor training convergence. All other settings, including AEN structure and training hyperparameters, are identical to (5.3.3).

5.3.5 COMP and AGG models

Both methods are applied to the FOM in a unit-by-unit fashion. However, the submodels of HX3, column feed stage, and reboiler/condenser already have a minimal form and cannot be further reduced by the COMP or AGG. Since the dynamic relevance of these units is unclear without further investigation, we leave them untouched. Moreover, preliminary experimentation showed that the HX dynamics have a noticeable effect on the overall system response. Hence, we do not replace these units by quasi-stationary models, *i.e.*, we create and insert ROM submodels for HX1, HX2, and column.

The number and size of compartments and the position of the sensitive elements (referring to column trays and HX cells simultaneously) are degrees of freedom of the COMP method and may be regarded as hyperparameters. Similarly, the hyperparameters of AGG are the number and positions of aggregation elements. We follow [256] and use compartments of a uniform size, with sensitive elements located in the center of the respective compartments. A perturbation analysis of the positions confirmed this choice. We select the same central positions for the aggregation elements. Using the same number of compartments or aggregation elements n_c for each HX1, HX2, and column, we obtain $n_z = 10 + 6 \cdot n_c$. In all AGG models, we specify $\mathbf{H}$ in (28) such that $\mathrm{diag}(\mathbf{H}) = n_x$.

5.4 Definition of the comparison criteria

We compare the selected MOR techniques in two tests. First, we only assess the accuracy of the linear or nonlinear subspaces, *i.e.*, the slow manifold approximations, constructed by the MOR methods. This assessment does not yet include ROMs but merely determines the average accuracy of projecting the FOM state snapshots of a test data set onto the respective low-dimensional subspaces. In particular, the projection error constitutes a lower bound on the ROM prediction error [237]. Second, we simulate the ROMs, *i.e.*, we examine the prediction capabilities of the ROMs. This prediction test is the most important criterion as we assess the overall ROM accuracy. We initialize all ROMs at the nominal steady state described above.

In both evaluations, we use an independent test data set that comprises a series of random uniformly distributed zero-order hold inputs (illustration see supplementary information). Furthermore, we consider different degrees of reduction, *i.e.*, we vary the reduced order n_z. As previously discussed, for COMP and AGG, the values of n_z are not freely selectable but are restricted by the subsystem-by-subsystem application of the methods, where viable options are $n_z \in \{16, 22, 28, ...\}$. To facilitate a comparison of the MOR methods, we include such values of n_z into the study.

We evaluate the accuracy in terms of the root mean squared error (RMSE), which we define in terms of the Frobenius norm:

$$\mathrm{RMSE}(\mathbf{X}, \mathbf{X}_\mathrm{r}) \triangleq \sqrt{\frac{1}{n_x N} \|\mathbf{X} - \mathbf{X}_\mathrm{r}\|_\mathrm{F}^2}. \tag{30}$$

Therein, $\mathbf{X} \in \mathbb{R}^{n_x \times N}$ is the scaled test snapshot matrix, having the snapshots $\boldsymbol{x}^{(k)}$, $k = 1, ..., N$, as matrix rows. The matrix $\mathbf{X}_\mathrm{r} \in \mathbb{R}^{n_x \times N}$ collects the projected or predicted snapshots, $\boldsymbol{x}_\mathrm{r}^{(k)} \in \mathbb{R}^{n_x}$, $k = 1, ..., N$. We scale and shift $\mathbf{X}$ and $\mathbf{X}_\mathrm{r}$ consistently with the training data.

We note, however, that the choice of the error norm and test data set can have a significant impact on the conclusions and should always be selected based on the ROM application. For example, the RMSE averages out peak deviations so that the slow and steady-state errors are prioritized over maximum errors, *e.g.*, an inverse response on a short-term scale

does not drastically increase the error. Herein, our test data set comprises a significant share of slow response and near-stationary snapshots. We consider this choice to be suitable for examining ROMs that capture the slow manifold.

Finally, we describe how the projections for test 1 are obtained. Regarding the linear basis methods POD and DMDc, both the linear subspaces and the projections coincide, respectively. The projection is given by the projection matrix, $\Pi \triangleq \mathbf{U}_1 \mathbf{U}_1^{\mathrm{T}}$, where $\mathbf{U}_1$ is from (11) or (25). To compute the projections performed by MFL, MFLu, KW, and MFL-ANN, we encode and decode the test data through the respective AENs. The projections in POD-Res, COMP and AGG are inherently defined through the degeneration/residualization. We evaluate these projections by solving the respective nonlinear systems of equations, Eqs. (17b), (14b) and (28b), on the test data in `Dymola`. Recall that for the selected hyperparameters, the slow manifolds of AGG and COMP coincide; however, the projections differ.

5.5 Results: projection accuracy

We present the results of the projection test. Table 2 compares the projection RMSE on the test data set. First, we consider small $n_z \leq 10$, *i.e.*, the very-low-order region. As expected, the nonlinear subspace methods (POD-Res, MFL, MFLu) perform significantly better than the linear basis methods (POD, DMDc). The respective difference in RMSE is approximately one order of magnitude. Overall, the smallest projection errors are reached by the MFLu approach. However, the difference in RMSE between MFL and MFLu decreases with increasing n_z. This confirms the input-dependency of the slow manifold as discussed in Section 3.1.5. Clearly, the error difference is system-dependent and may be even more substantial for a strongly parameterized slow manifold. Finally, the projection RMSEs of the non-intrusive nonlinear subspace methods KW and MFL-ANN are comparable to POD. The higher RMSE than for MFL is attributed to the additional loss terms included in the KW and MFL-ANN trainings.

Moving to higher orders, *i.e.*, $n_z \in \{16, 28\}$, the improvement by the MFL-based methods is only moderate, which was also observed in [161, 229]. We attribute this limited improvement to the challenging convergence of the respective unsupervised learning problems using local solvers, *i.e.*, training AENs by stochastic gradient descent algorithms. As a result, POD truncation outperforms the MFL methods at $n_z = 28$, facilitating the overall smallest projection error. Interestingly, the linear POD subspace projection also exceeds the accuracy of the nonlinear POD-Res method. A further investigation of the POD-Res results revealed notable deviations between $\boldsymbol{x}(t)$ and $\boldsymbol{x}_{\mathrm{r}}(t)$ at the points of abrupt changes in $\boldsymbol{u}(t)$. Finally, both COMP and AGG exhibit relatively large errors compared to the other ROM methods. Here, the COMP subspace shows significant deviations from the FOM samples near the input steps but relatively small errors otherwise. In contrast, the maximum deviations of AGG are lower, but a greater number of samples have a notable projection error.

5.6 Results: ROM predictions

We now present the results of the ROM prediction test. Table 3 collects the state prediction RMSE of the ROMs. The RMSE in Table 3 is consistently higher than in Table 2, because ROM predictions are less accurate than the mere projection of FOM snapshots onto the ROM subspace.

Both linear subspace methods, POD-Galerkin and DMDc, do not reach the desired accuracy at very low order, *i.e.*, $n_z \leq 10$. For reduced orders $11 \leq n_z \leq 21$, the simulations of the POD-Galerkin ROMs did not converge at all. As we discuss in the supplementary information, these critical values of n_z lie below the truncation threshold of $n_z = 23$, obtained from the POD singular values. At order $n_z = 28$, the POD-Galerkin ROM successfully provides an accurate prediction, whereas the DMDc prediction is still poor. Clearly, the nonlinearity of the FOM prohibits highly accurate DMDc models. Finally, we notice a significant gap between projection error (Table 2) and projection error (Table 3) for both POD-Galerkin and DMDc.

Compared to the POD-Galerkin results, the POD-Res model is more accurate throughout all orders. This advantage is particularly significant in the very-low-order region. In contrast to the projection accuracy (Table 2), POD-Res outperforms the POD-Galerkin ROMs in the prediction. Even for $n_z = 2$, we notice a fairly small prediction error of POD-Res. Since an RMSE of 10^{-2} roughly corresponds to 1% average error in the scaled states, we expect a satisfactory precision of POD-Res ROMs of order $n_z = 10$ (and possibly even $n_z = 5$) in tasks such as control. When moving to higher n_z, the accuracy improves further.

Similarly to POD-Res, the methods MFL and MFLu enable a higher precision than the linear basis methods at very low order (except for $n_z = 2$). Moreover, at $n_z = 5$, the MFLu-Galerkin ROM clearly outperforms the MFL-Galerkin ROM. On the other hand, at higher order, both models perform similarly. At $n_z = 28$, the simulation of both MFL and MFLu ROMs did not converge. Using a different activation function or adding neurons to the AENs did not resolve this issue. A possible cause is the combination of relatively high ROM order at moderate AEN accuracy (similarly to the issues with POD-Galerkin ROMs). Although we are confident that modifications based on more extensive training might facilitate convergence, we did not undertake such attempts. Instead, we highlight the potential numerical issues

Table 2 Results of projection accuracy. The RMSE indicates the average error between the states projected on the reduced subspace versus the original FOM states. We consider different ROM orders n_z

n_z	POD/DMDc	POD-Res	MFL	MFLu	KW	MFL-ANN	COMP	AGG
2	$7.8 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$5.3 \cdot 10^{-3}$	$1.5 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	–	–
5	$1.6 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$2.2 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$1.4 \cdot 10^{-2}$	$2.5 \cdot 10^{-2}$	–	–
10	$3.8 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$1.3 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$7.3 \cdot 10^{-3}$	$2.5 \cdot 10^{-2}$	–	–
16	$1.1 \cdot 10^{-3}$	$5.5 \cdot 10^{-3}$	$7.3 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$4.8 \cdot 10^{-2}$
28	$9.0 \cdot 10^{-5}$	$3.6 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$5.9 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$

Table 3 Results of ROM predictions. The RMSE indicates the average error between the state predictions by the ROMs versus the true FOM states. We consider different ROM orders n_z. Non-converged simulations are indicated by $(*)$

n_z	POD	DMDc	POD-Res	MFL	MFLu	KW	MFL-ANN	COMP	AGG
2	$1.1 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$	$3.6 \cdot 10^{-2}$	$6.1 \cdot 10^{-1}$	$2.9 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$	$1.7 \cdot 10^{-1}$	–	–
5	$6.3 \cdot 10^{-2}$	$1.5 \cdot 10^{-1}$	$2.1 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$2.2 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$2.9 \cdot 10^{-2}$	–	–
10	$4.9 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$8.8 \cdot 10^{-3}$	$2.1 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$9.9 \cdot 10^{-3}$	$3.3 \cdot 10^{-2}$	–	–
16	$(*)$	$9.4 \cdot 10^{-2}$	$7.9 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$	$2.5 \cdot 10^{-1}$	$4.9 \cdot 10^{-2}$
28	$8.7 \cdot 10^{-3}$	$7.8 \cdot 10^{-2}$	$7.5 \cdot 10^{-3}$	$(*)$	$(*)$	$1.1 \cdot 10^{-2}$	$3.3 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$

with higher-order MFL-Galerkin ROMs when working in a standard AEN training setup.

The non-intrusive KW and MFL-ANN ROMs have a prediction error comparable to the intrusive nonlinear subspace methods. For each n_z, the differences in RMSE are insignificant, where KW models perform only slightly better. Notably, the MFL-ANN models do not perform better than the KW models. A comparison to the performance on the training data set revealed a tendency of MFL-ANN to overfit the training data (the RMSEs on the training data were approximately one order of magnitude smaller than on the test data), which was less pronounced for the KW structure. From $n_z = 10$ to $n_z = 28$, there is no improvement in RMSE for both model types, *i.e.*, increasing the ROM order does not improve the accuracy. Finally, we notice that the projection and prediction errors are comparable for both models, respectively.

For COMP and AGG models with a single compartment and aggregation cell for each unit operation, respectively, we find significant prediction errors. In particular, the COMP model is far from the desired accuracy due to severe inverse response. Increasing the ROM order results in a notably smaller RMSE, where the AGG results reach the desired error level. However, the COMP response still exhibits a notable inverse response, diminishing ROM accuracy (see Fig. 4 below).

Besides comparing the numerical results, we also comment on the numerical properties of the ROMs. Performing the simulations was most straightforward using the non-intrusive ROMs (DMDc, KW, MFL-ANN), which are numerically favorable to deal with due to the sequential model structure. On the other hand, simulating the POD-Res and COMP models was computationally most expensive, and required a very careful initialization and scaling of the model equations to converge the simulations. As a result, the high accuracy of the POD-Res model is offset by the fact that it is difficult to handle numerically.

To illustrate the results, we examine the molar fraction of oxygen on the column top tray, which is directly connected to the quality of the nitrogen product. We show the predictions for $n_z = 5$ in Fig. 4a and for $n_z = 28$ in Fig. 4b. In Fig. 4a, we observe a significant offset between ROM prediction and FOM response for POD, DMDc, and MFL. This finding is consistent with Table 3. For these ROMs, the offset involves both a transient and a steady-state mismatch. Moreover, despite the relatively small RMSE of POD-Res, the graph shows significant inverse response near the input steps, whereas the steady states are captured accurately. All other ROMs (MFLu, KW, MFL-ANN) reproduce the FOM trajectory closely.

Observing the higher-order case in Fig. 4b, we find significantly better predictions by the POD and POD-Res models. At the same time, the DMDc forecast remains inaccurate. The effect of overfitting discussed above for the MFL-ANN ROMs is also visible. Compared to Fig. 4a, the MFL-ANN prediction is less accurate despite a higher-order model. Further, notable deviations from the FOM trajectory are visible for the COMP model in the form of severe inverse response. In contrast, the AGG ROM provides a close reproduction of the FOM trajectory.

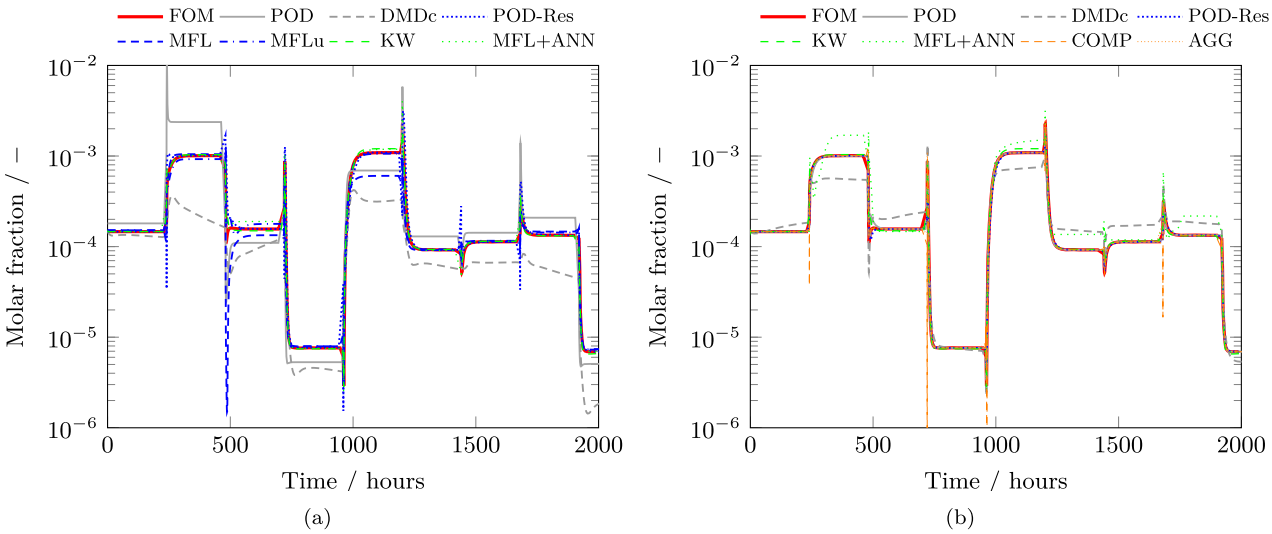

Fig. 4 Results of test 2 (prediction). Depicted is the liquid molar fraction of oxygen on the column top stage over the first 2000 h of the test. ROM order is (a) $n_z = 5$ and (b) $n_z = 28$

6 Discussion

Based on theoretical properties of the MOR methods and the findings of the case study, we summarize characteristic properties of the methods investigated. However, we emphasize that definitive statements on method ranking are delicate, because in addition to n_z, most MOR techniques possess further hyperparameters that can strongly affect ROM behavior, *e.g.*, the position of the sensitive element (COMP, AGG) or the ANN architecture (MFL, KW, MFL-ANN). Although we performed a basic validation of crucial hyperparameters within the case study, we do not claim that this screening was exhaustive. Additionally, the performance of a certain type of ROM may be system-dependent, and in that sense our experimental results may be specific to air separation units.

Comparing the results of intrusive (MFL, MFLu) to non-intrusive nonlinear subspace methods (KW, MFL-ANN) in the case study, we conclude that similar ROM accuracy is reached when targeting very low-order models. Such very-low-order ROMs are important in control applications, where CPU reduction has a high priority whereas moderate errors are acceptable. Often, a moderate prediction error can be compensated by state estimation and model adaptation, especially if the error is predominantly steady-state offset [198]. On the other hand, despite a relatively low RMSE, the POD-Res results exhibited an inverse response in the very-low-order setup, which can be problematic in control applications.

When prioritizing accuracy over low order, non-intrusive methods may face difficulties to reach arbitrary precision due to their inherently difficult learning problems. In such cases, intrusive strategies can be easier to apply. In particular, the POD-based methods were able to generate very accurate ROMs given sufficiently high order. In this context, we acknowledge the simplicity and approximation guarantees (on the training data) of the POD. In contrast, high projection and prediction accuracies were difficult to reach when using MFL strategies.

The advantage of non-intrusive over intrusive reduction methods is that non-intrusive strategies combine MOR and hyperreduction. The weakness of non-intrusive methods is that they can be prone to overfitting if not carefully regularized. One possible structural regularization strategy is to restrict the latent dynamics to a linear or bilinear form. In particular, applied Koopman theory combines MFL and linearization into a KW structure, which enables reliable ROMs of the moderately nonlinear systems.

POD-Res and COMP may be considered as data-driven and heuristic variants of SPT, respectively. However, COMP retains most of the model sparsity, whereas POD-Res ROMs are generally dense and can therefore be more challenging to solve. Here, both POD-Res and COMP were numerically the most delicate to handle. Specifically, large-scale implicit nonlinear systems of equations and non-physical inverse response made these models difficult to converge.

In terms of reduction effort, the (non-intrusive) KW and MFL-ANN strategies involved the most extensive offline computations. Achieving sufficiently accurate models may require multiple trainings and extensive hyperparameter tuning, which introduces delays into the model reduction procedure. Furthermore, non-intrusive methods are most sensible to the training data in our experience. In particular, a balanced amount of transient and stationary data is essential for the successful generation of ROMs that enable precise long-term forecasts.

The model reduction efforts in intrusive reduction methods (including tailored methods) demand a great share of model building and reformulation (debugging, robustification). On the other hand, the training problems used within intrusive methods, *e.g.*, AEN training, are typically less computationally demanding than in non-intrusive approaches. In our experience, intrusive data-driven methods are slightly more tolerant to a poorly designed data set (*e.g.*, too few steady-state data) than non-intrusive methods. For intrusive methods, further efforts have to be expected in terms of hyperreduction, often necessary to obtain a real-time capable ROM.

Lastly, we comment on the model reduction procedure for interconnected systems. Applying the tailored methods, such as COMP and AGG, is naturally done in a subsystem-by-subsystem fashion, since these methods provide reduced models of the subsystems, *e.g.*, distillation columns or HXs, of an interconnected system, *e.g.*, a process flowsheet model. In particular, when working in a structured modeling environment, such as `Unisim Design` or `gPROMS`, these ROMs can be pre-configured and included into the model library. Reducing an interconnected system model is then achieved by substituting full-order submodels by reduced submodels. The "subsystem-by-subsystem" approach preserves the model sparsity originating from the interconnection. However, the feasible degree of reduction may be limited, as shown by the case study results. Further, the reduction problem becomes a combinatorial problem due to potentially many subsystem ROMs to be configured and combined.

From a mathematical perspective, the subsystem-by-subsystem strategy falls into the port-Hamiltonian formalism. Notably, in this framework, the respective ROMs are invariant under couplings and preserve their properties. Hence, this formalism should be investigated in future work to develop more accurate alternatives to COMP and AGG.

In contrast to the subsystem-by-subsystem approach, general-purpose methods are most easily applied to the entire FOM. Then, the complete model of the interconnected system is reduced as a whole. As shown by the case study, such an approach can enable a lower ROM order. However, for systems with many inputs, the data sampling efforts grow exponentially with n_u, which can result in a curse of dimensionality. This curse may be broken by a subsystem-by-subsystem strategy if the high-dimensional inputs enter the system only locally and are evenly distributed over the submodels.

7 Conclusions and outlook

Model order reduction spans a collection of techniques to construct a low-order model approximating the evolution of a high-order system. Herein, we have reviewed and compared reduction methods for nonlinear dynamical models with inputs. In particular, we have contrasted the different strategies both theoretically and in a case study. As expected, there is no universal best choice when selecting a reduction method. Rather, a suitable choice depends on targeted ROM properties.

If very-low-order ROMs are needed, then intrusive nonlinear subspace methods, *e.g.*, the MFLu approach, as well as non-intrusive nonlinear subspace methods, *e.g.*, the KW model, have provided accurate ROMs within our case study. Notably, non-intrusive reduction methods combine MOR and hyperreduction in a single step, providing a ROM that is real-time capable. If high ROM accuracy is needed but the degree of reduction is of secondary importance, then linear subspace methods, *e.g.*, the POD-Galerkin method, are a good choice due to their (theoretically) arbitrary degree of accuracy and simplicity of application. Finally, when desiring structural preservation of a given network or flowsheet model comprising interconnected subsystems, then structure-preserving (especially port-Hamiltonian) ROMs or problem-specific ROMs are a straightforward choice. These approaches reduce the subunits individually by means of preconfigured ROM blocks.

Besides reviewing and comparing existing MOR methods, we have also proposed an extension of the manifold-Galerkin method to systems with inputs (Section 3.1.5). In particular, we stress that system inputs $\boldsymbol{u}(t)$ can deform the slow manifold $\mathcal{M}_{\mathcal{X}}$. Consequently, if the effect of inputs is disregarded, an accurate approximation of $\mathcal{M}_{\mathcal{X}}$ may necessitate a high ROM order.

Within the case study, we validated the proposed manifold-Galerkin extension for systems with inputs. In cases of very strong reduction, we found that our approach outperforms the standard method both in terms of the projection accuracy and the ROM prediction accuracy. As expected, our extended approach and the established approach performed similarly at higher orders.

In our case study, we did not investigate hyperreduction or CPU cost of integration. Instead, we expect that a sufficiently low order combined with hyperreduction enables real-time applications, *e.g.*, [256, 262]. Clearly, real-time capability depends on the application, so there is no universal CPU threshold to classify performance. In any case, accounting for hyperreduction already within MOR could be a promising step towards more holistic methods. Even though our case study represents a realistic MOR problem in the process systems control field, future work should investigate higher-dimensional problems and examine additional reduction methods, *e.g.*, [39, 78].

Despite increasing computer power, nonlinear MOR (especially for parametric and non-autonomous systems) remains a challenging discipline worth of investigation. The curse of dimensionality, when sampling systems with

many inputs or parameters, is connected to the need for robust methods applicable in the low-data limit. In this regard, structure-preserving MOR is a promising approach to enrich the non-intrusive manifold learning task by additional (physics-based) information. Moreover, the scalability of manifold learning methods to very-high-order problems remains an open challenge. Recently, promising results have been presented using quadratic/polynomial manifold learning [78, 281] and convolutional autoencoders [74, 248].

Although several methods offer stability preservation, such guarantees are not inherently provided by every method [119]. Consequently, a careful method selection or a-posteriori error and stability analysis are required. Another important topic is the development of a-priori error bounds on ROM accuracy.

Finally, we notice some limitations of current reduction methods with respect to numerical optimization and control applications. First, most numerical optimization techniques are based on derivative information, in particular, the Jacobian of the ROM variables $\boldsymbol{z}(t)$ and $\boldsymbol{x}_{\mathrm{r}}(t)$ in (4) with respect to each other and the inputs $\boldsymbol{u}(t)$. However, standard MOR methods do not explicitly account for the accuracy of such derivatives. For learning-based MOR and hyperreduction approaches, Sobolev trainings [54, 284] may improve derivative accuracy. Additionally, in our experience, sensitivity integration is often considerably more expensive than standard integration, and therefore accounting for sensitivities in the reduction process might enable fundamental improvements in CPU costs of optimization applications. Second, model reduction of non-smooth or even discontinuous models, *e.g.*, for controlling a process start-up, is a relevant but mostly unexplored topic [27, 306].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11071-026-12495-8.

Acknowledgements A pre-print version of this manuscript has been posted on https://arxiv.org/pdf/2506.12819 under the arXiv.org non-exclusive distribution license. The authors thank Danimir Doncevic and Eike Cramer for fruitful discussions. Finally, the authors thank the anonymous reviewers and Tony Roberts for their valuable feedback and suggestions.

Author Contributions Jan C. Schulze: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Visualization, Writing - original draft preparation A. Mitsos: Supervision, Resources, Funding acquisition, Writing - review and editing

Funding Open Access funding enabled and organized by Projekt DEAL. The authors gratefully acknowledge the financial support of the Kopernikus project SynErgie by the German Federal Ministry of Research, Technology and Space (BMFTR) and project supervision by Projektträger Jülich (PtJ).

Data Availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

1. Abdullah, Z., Aziz, N., Ahmad, Z.: Nonlinear Modelling Application in Distillation Column. Chem. Prod. Process. Model. **2**(3), 12 (2007). https://doi.org/10.2202/1934-2659.1082
2. Afkham, B.M., Hesthaven, J.S.: Structure Preserving Model Reduction of Parametric Hamiltonian Systems. SIAM J. Sci. Comput. **39**(6), A2616–A2644 (2017). https://doi.org/10.1137/17M1111991
3. Ahmed, S.E., Pawar, S., San, O., Rasheed, A., Iliescu, T., Noack, B.R.: On closures for reduced order models–a spectrum of first-principle to machine-learned avenues. Phys. Fluids **33**(9), 091301 (2021)
4. Antoulas, A.C.: An overview of approximation methods for large-scale dynamical systems. Annu. Rev. Control. **29**(2), 181–190 (2005)
5. Antoulas, A.C.: Approximation of Large-Scale Dynamical Systems. Siam (2005)
6. Antoulas, A.C., Sorensen, D.C., Gugercin, S.: A survey of model reduction methods for large-scale systems. Contemp. Math. **280**, 193–220 (2001)
7. Armaou, A., Christofides, P.D.: Dynamic optimization of dissipative PDE systems using nonlinear order reduction. Chem. Eng. Sci. **57**(24), 5083–5114 (2002)
8. Assael, J.A.M., Wahlström, N., Schön, T.B., Deisenroth, M.P.: Data-efficient learning of feedback policies from image pixels using deep dynamical models. arXiv preprint arXiv:1510.02173 (2015)
9. Astolfi, A.: Model reduction by moment matching for linear and nonlinear systems. IEEE Trans. on Auto. Control **55**(10), 2321–2336 (2010)
10. Aulbach, B.: A reduction principle for nonautonomous differential equations. Arch. Math. **39**(3), 217–232 (1982). https://doi.org/10.1007/BF01899528
11. Baader, F.J., Althaus, P., Bardow, A., Dahmen, M.: Demand response for flat nonlinear MIMO processes using dynamic ramping constraints. Comput. & Chem. Eng. **172**, 108171 (2023). https://doi.org/10.1016/j.compchemeng.2023.108171
12. Bai, Z.: Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. Appl. Numer. Math. **43**(1–2), 9–44 (2002). https://doi.org/10.1016/S0168-9274(02)00116-2

13. Baldea, M., Daoutidis, P.: Model reduction and control of reactor-heat exchanger networks. J. Process Control **16**(3), 265–274 (2006). https://doi.org/10.1016/j.jprocont.2005.06.007
14. Baldi, P., Hornik, K.: Neural networks and principal component analysis: learning from examples without local minima. Neural Netw. **2**(1), 53–58 (1989). https://doi.org/10.1016/0893-6080(89)90014-2
15. Barnett, J., Farhat, C.: Quadratic approximation manifold for mitigating the Kolmogorov barrier in nonlinear projection-based model order reduction. J. Comput. Phys. **464**, 111348 (2022)
16. Barone, M.F., Kalashnikova, I., Segalman, D.J., Thornquist, H.K.: Stable Galerkin reduced order models for linearized compressible flow. J. Comput. Phys. **228**(6), 1932–1946 (2009)
17. Barrault, M., Maday, Y., Nguyen, N.C., Patera, A.T.: An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. Comptes Rendus. Mathématique **339**(9), 667–672 (2004). https://doi.org/10.1016/j.crma.2004.08.006
18. Baur, U., Benner, P., Feng, L.: Model order reduction for linear and nonlinear systems: a system-theoretic perspective. Arch. of Comput. Methods in Eng. **21**(4), 331–358 (2014)
19. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 1373–1396 (2003)
20. Benallou, A., Seborg, D.E., Mellichamp, D.A.: Dynamic compartmental models for separation processes. AIChE J. **32**(7), 1067–1078 (1986)
21. Benner, P., Goyal, P., Kramer, B., Peherstorfer, B., Willcox, K.: Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms. Comput. Methods Appl. Mech. Eng. **372**, 113433 (2020)
22. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Rev. **57**(4), 483–531 (2015)
23. Benner, P., Ohlberger, M., Cohen, A., Willcox, K.: Model Reduction and Approximation: Theory and Algorithms. Siam (2017)
24. Benner, P., Ohlberger, M., Patera, A., Rozza, G., Urban, K. (eds.): Model Reduction of Parametrized Systems. MS&A. Springer International Publishing, Cham (2017). doi:https://doi.org/10.1007/978-3-319-58786-8
25. Berkooz, G., Holmes, P., Lumley, J.L.: The proper orthogonal decomposition in the analysis of turbulent flows. Annu. Rev. Fluid Mech. **25**(1), 539–575 (1993)
26. Besselink, B., van de Wouw, N., Nijmeijer, H.: Model reduction for nonlinear systems with incremental gain or passivity properties. Automatica **49**(4), 861–872 (2013). https://doi.org/10.1016/j.automatica.2013.01.004
27. Bettini, L., Cenedese, M., Haller, G.: Model reduction to spectral submanifolds in piecewise smooth dynamical systems. Int. J. Non-Linear Mech. **163**, 104753 (2024)
28. Bhattacharjee, B., Schwer, D.A., Barton, P.I., Green, W.H.: Optimally-reduced kinetic models: reaction elimination in large-scale kinetic mechanisms. Combust. Flame **135**(3), 191–208 (2003)
29. Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., Wojtaszczyk, P.: Convergence rates for greedy algorithms in reduced basis methods. SIAM J. Math. Anal. **43**(3), 1457–1472 (2011). https://doi.org/10.1137/100795772
30. Bos, R., Bombois, X., van den Hof, P.: Accelerating large-scale non-linear models for monitoring and control using spatial and temporal correlations. In: Proceedings of the 2004 American Control Conference, pp. 3705–3710 (2004). https://doi.org/10.23919/ACC.2004.1384488
31. Boumal, N.: An Introduction to Optimization on Smooth Manifolds. Cambridge University Press (2023)
32. Bradde, T., Grivet-Talocia, S., Aumann, Q., Gosea, I.V.: A modified AAA algorithm for learning stable reduced-order models from data. J. Sci. Comput. **103**(1), 14 (2025)
33. Brenan, K.E., Campbell, S.L., Petzold, L.R.: Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. Siam (1995)
34. Brunton, S.L., Noack, B.R., Koumoutsakos, P.: Machine learning for fluid mechanics. Annu. Rev. Fluid Mech. **52**(1), 477–508 (2020). https://doi.org/10.1146/annurev-fluid-010719-060214
35. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. **113**(15), 3932–3937 (2016). https://doi.org/10.1073/pnas.1517384113
36. Buchfink, P., Glas, S., Haasdonk, B.: Symplectic model reduction of Hamiltonian systems on nonlinear manifolds and approximation with weakly symplectic autoencoder. SIAM J. Sci. Comput. **45**(2), A289–A311 (2023). https://doi.org/10.1137/21M1466657
37. Buchfink, P., Glas, S., Haasdonk, B., Unger, B.: Model reduction on manifolds: a differential geometric framework. Physica D **468**(1), 134299 (2024). https://doi.org/10.1016/j.physd.2024.134299
38. Bui-Thanh, T., Damodaran, M., Willcox, K.: Proper orthogonal decomposition extensions for parametric applications in compressible aerodynamics. In: 21st AIAA applied aerodynamics conference, pp. 4213 (2003)
39. Bychkov, A., Issan, O., Pogudin, G., Kramer, B.: Exact and optimal quadratization of nonlinear finite-dimensional nonautonomous dynamical systems. SIAM J. Appl. Dyn. Syst. **23**(1), 982–1016 (2024)
40. Cabré, X., Fontich, E., de La Llave, R.: The parameterization method for invariant manifolds III: overview and applications. J. Diff. Equ. **218**(2), 444–515 (2005)
41. Cao, Y., Swartz, C.L.E., Flores-Cerrillo, J., Ma, J.: Dynamic modeling and collocation-based model reduction of cryogenic air separation units. AIChE J. **62**(5), 1602–1615 (2016)
42. Carlberg, K., Tuminaro, R., Boggs, P.: Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics. SIAM J. Sci. Comput. **37**(2), B153–B184 (2015)
43. Caspari, A., Offermanns, C., Ecker, A.M., Pottmann, M., Zapp, G., Mhamdi, A., Mitsos, A.: A wave propagation approach for reduced dynamic modeling of distillation columns: optimization and control. J. Process Control **91**, 12–24 (2020)
44. Caspari, A., Tsay, C., Mhamdi, A., Baldea, M., Mitsos, A.: The integration of scheduling and control: top-down vs. bottom-up. J. Process Control **91**, 50–62 (2020)
45. Castelli, R., Lessard, J.P., Mireles James, J.D.: Parameterization of invariant manifolds for periodic orbits i: efficient numerics via the floquet normal form. SIAM J. Appl. Dyn. Syst. **14**(1), 132–167 (2015). https://doi.org/10.1137/140960207
46. Chatterjee, A.: An introduction to the proper orthogonal decomposition. Curr. Sci. 808–817 (2000)
47. Chaturantabut, S., Sorensen, D.C.: Nonlinear model reduction via discrete empirical interpolation. SIAM J. Sci. Comput. **32**(5), 2737–2764 (2010)
48. Chiavazzo, E., Gorban, A.N., Karlin, I.V.: Comparison of invariant manifolds for model reduction in chemical kinetics. Commun. Comput. Phys **2**(5), 964–992 (2007)
49. Cho, Y.S., Joseph, B.: Reduced-order steady-state and dynamic models for separation processes. part ii. application to nonlinear multicomponent systems. AIChE J. **29**(2), 270–276 (1983)
50. Christ, P., Sattelmayer, T.: Reduced order modelling of flow and mixing in an automobile HVAC system using proper orthogonal decomposition. Appl. Therm. Eng. **133**, 211–223 (2018). https://doi.org/10.1016/j.applthermaleng.2018.01.023

51. Christofides, P.D., Daoutidis, P.: Feedback control of two-time-scale nonlinear systems. Int. J. Control **63**(5), 965–994 (1996). https://doi.org/10.1080/00207179608921879
52. Coifman, R.R., Lafon, S.: Diffusion maps. Appl. Comput. Harmon. Anal. **21**(1), 5–30 (2006). https://doi.org/10.1016/j.acha.2006.04.006
53. Conti, P., Gobat, G., Fresca, S., Manzoni, A., Frangi, A.: Reduced order modeling of parametrized systems through autoencoders and SINDy approach: continuation of periodic solutions. Comput. Methods Appl. Mech. Eng. **411**, 116072 (2023). https://doi.org/10.1016/j.cma.2023.116072
54. Czarnecki, W.M., Osindero, S., Jaderberg, M., Swirszcz, G., Pascanu, R.: Sobolev training for neural networks. Adv. Neural. Inf. Process. Syst. **30**, (2017)
55. Daoutidis, P.: DAEs in model reduction of chemical processes: an overview. In: Surveys in Differential-Algebraic Equations II, pp. 69–102. Springer (2015)
56. Das, S., Giannakis, D.: Koopman spectra in reproducing kernel Hilbert spaces. Appl. Comput. Harmon. Anal. **49**(2), 573–607 (2020). https://doi.org/10.1016/j.acha.2020.05.008
57. Díez, P., Muixí, A., Zlotnik, S., García-González, A.: Nonlinear dimensionality reduction for parametric problems: a kernel proper orthogonal decomposition. Int. J. Numer. Meth. Eng. **122**(24), 7306–7327 (2021). https://doi.org/10.1002/nme.6831
58. Diwekar, U.M.: How simple can it be? – a look at the models for batch distillation. Comput & Chem Eng **18**, S451–S457 (1994). https://doi.org/10.1016/0098-1354(94)80074-X
59. Dsilva, C.J., Talmon, R., Gear, C.W., Coifman, R.R., Kevrekidis, I.G.: Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems. SIAM J. Appl. Dyn. Syst. **15**(3), 1327–1351 (2016)
60. Ecker, A.M., Thomas, I., Häfele, M., Wunderlich, B., Obermeier, A., Ferstl, J., Klein, H., Peschel, A.: Development of a new column shortcut model and its application in process optimisation. Chem. Eng. Sci. **196**, 538–551 (2019). https://doi.org/10.1016/j.ces.2018.10.035
61. Edwards, K., Edgar, T.F., Manousiouthakis, V.I.: Kinetic model reduction using genetic algorithms. Comput & Chem Eng **22**(1–2), 239–246 (1998)
62. El Wajeh, M., Mhamdi, A., Mitsos, A.: Optimal design and flexible operation of a fully electrified biodiesel production process. Ind & Eng Chem Res **63**(3), 1487–1500 (2024). https://doi.org/10.1021/acs.iecr.3c03074
63. Engell, S., Harjunkoski, I.: Optimal operation: scheduling, advanced control and their integration. Comput & Chem Eng **47**, 121–133 (2012). https://doi.org/10.1016/j.compchemeng.2012.06.039
64. Esche, E., Müller, D., Kraus, R., Wozny, G.: Systematic approaches for model derivation for optimization purposes. Chem. Eng. Sci. **115**, 215–224 (2014). https://doi.org/10.1016/j.ces.2013.11.041
65. Fages, F., Hemery, M., Soliman, S.: On BIOCHAM symbolic computation pipeline for compiling mathematical functions into biochemistry. ACM Commun. Comput. Algebra **58**(2), 15–22 (2025)
66. Feng, L., Zeng, X., Chiang, C., Zhou, D., Fang, Q.: Direct nonlinear order reduction with variational analysis. In: Proceedings Design, Automation and Test in Europe Conference and Exhibition, pp. 1316–1321. IEEE Comput. Soc (2004). doi:https://doi.org/10.1109/DATE.2004.1269077
67. Fenichel, N.: Geometric singular perturbation theory for ordinary differential equations. J. Diff. Equ. **31**(1), 53–98 (1979)
68. Fenichel, N., Moser, J.K.: Persistence and smoothness of invariant manifolds for flows. Indiana Univ. Math. J. **21**(3), 193–226 (1971)
69. Foias, C., Jolly, M.S., Kevrekidis, I.G., Sell, G.R., Titi, E.S.: On the computation of inertial manifolds. Phys. Lett. A **131**(7–8), 433–436 (1988)
70. Foias, C., Temam, R.: The algebraic approximation of attractors: the finite dimensional case. Physica D **32**(2), 163–182 (1988). https://doi.org/10.1016/0167-2789(88)90049-8
71. Franke, C., Schaback, R.: Solving partial differential equations by collocation using radial basis functions. Appl. Math. Comput. **93**(1), 73–82 (1998)
72. Franz, T., Zimmermann, R., Görtz, S., Karcher, N.: Interpolation-based reduced-order modelling for steady transonic flows via manifold learning. Int. J. of Comput. Fluid Dyn. **28**(3–4), 106–121 (2014)
73. Fresca, S., Dede', L., Manzoni, A.: A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs. J. Sci. Comput. **87**, 1–36 (2021)
74. Fresca, S., Manzoni, A.: POD-DL-ROM: enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition. Comput. Methods Appl. Mech. Eng. **388**, 114181 (2022). https://doi.org/10.1016/j.cma.2021.114181
75. Friedl, K., Jaquier, N., Lundell, J., Asfour, T., Kragic, D.: A riemannian framework for learning reduced-order lagrangian dynamics. https://doi.org/10.48550/arXiv.2410.18868
76. Frieze, A., Kannan, R., Vempala, S.: Fast Monte-Carlo algorithms for finding low-rank approximations. J. ACM **51**(6), 1025–1041 (2004)
77. Gao, H., Wang, J.X., Zahr, M.J.: Non-intrusive model reduction of large-scale, nonlinear dynamical systems using deep learning. Physica D **412**, 132614 (2020). https://doi.org/10.1016/j.physd.2020.132614
78. Geelen, R., Balzano, L., Wright, S., Willcox, K.: Learning physics-based reduced-order models from data using nonlinear manifolds. Chaos: An Interdiscip. J. of Nonlinear Sci. **34**(3), (2024). https://doi.org/10.1063/5.0170105
79. Geelen, R., Wright, S., Willcox, K.: Operator inference for non-intrusive model reduction with quadratic manifolds. Comput. Methods Appl. Mech. Eng. **403**, 115717 (2023). https://doi.org/10.1016/j.cma.2022.115717
80. Geng, Y., Ju, L., Kramer, B., Wang, Z.: Data-driven reduced-order models for port-Hamiltonian systems with operator inference. Comput. Methods Appl. Mech. Eng. **442**, 118042 (2025)
81. Georgakis, C., Aris, R., Amundson, N.R.: Studies in the control of tubular reactors-i general considerations. Chem. Eng. Sci. **32**(11), 1359–1369 (1977). https://doi.org/10.1016/0009-2509(77)85032-X
82. Ghattas, O., Willcox, K.: Learning physics-based models from data: perspectives from inverse problems and model reduction. Acta Numer **30**, 445–554 (2021). https://doi.org/10.1017/S0962492921000064
83. Ghojogh, B., Crowley, M., Karray, F., Ghodsi, A.: Elements of Dimensionality Reduction and Manifold Learning. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-10602-6
84. Gilles, E.D., Retzbach, B.: Reduced models and control of distillation columns with sharp temperature profiles. In: 1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes, pp. 865–870 (1980)
85. Givon, D., Kupferman, R., Stuart, A.: Extracting macroscopic dynamics: model problems and algorithms. Nonlinearity **17**(6), R55–R127 (2004). https://doi.org/10.1088/0951-7715/17/6/R01
86. Gobat, G., Fresca, S., Manzoni, A., Frangi, A.: Reduced order modeling of nonlinear vibrating multiphysics microstructures with deep learning-based approaches. Sens. (Basel, Switzerland) **23**(6), (2023). https://doi.org/10.3390/s23063001

87. Gonzalez, F.J., Balajewicz, M.: Deep convolutional recurrent autoencoders for learning low-dimensional feature dynamics of fluid systems. arXiv (2018). doi:https://doi.org/10.48550/arXiv.1808.01346
88. Gorban, A.N., Karlin, I.V.: Method of invariant manifold for chemical kinetics. Chem. Eng. Sci. **58**(21), 4751–4768 (2003)
89. Gorban, A.N., Karlin, I.V., Zinovyev, A.Y.: Invariant grids for reaction kinetics. Physica A **333**, 106–154 (2004)
90. Goroshin, R., Mathieu, M.F., LeCun, Y.: Learning to linearize under uncertainty. Adv. Neural. Inf. Process. Syst. **28**, (2015)
91. Gouasmi, A., Parish, E.J., Duraisamy, K.: A priori estimation of memory effects in reduced-order models of nonlinear systems using the Mori-Zwanzig formalism. Proc. of the Royal Soc. A **473**(2205), 20170385 (2017). https://doi.org/10.1098/rspa.2017.0385
92. Goyal, P., Benner, P.: Generalized quadratic embeddings for nonlinear dynamics using deep learning. Physica D **463**(24), 134158 (2024). https://doi.org/10.1016/j.physd.2024.134158
93. Graham, M.D., Kevrekidis, I.G.: Alternative approaches to the Karhunen-Loeve decomposition for model reduction and data analysis. Comput. & Chem. Eng. **20**(5), 495–506 (1996)
94. Grasedyck, L., Kressner, D., Tobler, C.: A literature survey of low-rank tensor approximation techniques. GAMM-Mitteilungen **36**(1), 53–78 (2013)
95. Gruber, A., Tezaur, I.: Canonical and noncanonical Hamiltonian operator inference. Comput. Methods Appl. Mech. Eng. **416**(3), 116334 (2023). https://doi.org/10.1016/j.cma.2023.116334
96. Gu, C.: QLMOR: a projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **30**(9), 1307–1320 (2011). https://doi.org/10.1109/TCAD.2011.2142184
97. Gugercin, S., Polyuga, R.V., Beattie, C., van der Schaft, A.: Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. Automatica **48**(9), 1963–1974 (2012)
98. Haag, J., Gentric, C., Lemaitre, C., Leclerc, J.P.: Modelling of chemical reactors: from systemic approach to compartmental modelling. Int. J. Chem. Reactor Eng. **16**(8), 20170172 (2018). https://doi.org/10.1515/ijcre-2017-0172
99. Hahn, J., Edgar, T.F.: An improved method for nonlinear model reduction using balancing of empirical gramians. Comput. & Chem. Eng. **26**(10), 1379–1397 (2002)
100. Hahn, J., Kruger, U., Edgar, T.F.: Application of model reduction for model predictive control. IFAC Proc. Vol. **35**(1), 393–398 (2002). https://doi.org/10.3182/20020721-6-ES-1901.00634
101. Hahn, J., Lextrait, S., Edgar, T.F.: Nonlinear balanced model residualization via neural networks. AIChE J. **48**(6), 1353–1357 (2002). https://doi.org/10.1002/aic.690480621
102. Haider, P., Freko, P., Acher, T., Rehfeldt, S., Klein, H.: A transient three-dimensional model for thermo-fluid simulation of cryogenic plate-fin heat exchangers. Appl. Therm. Eng. **180**, 115791 (2020). https://doi.org/10.1016/j.applthermaleng.2020.115791
103. Hairer, E., Norsett, S.P., Wanner, G.: Solving Ordinary Differential Equations I: Nonstiff Problems. Springer (1993)
104. Halko, N., Martinsson, P.-G., Tropp, J.A.: Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. **52**(2), 217–288 (2011)
105. Haller, G., Kaszás, B.: Data-driven linearization of dynamical systems. Nonlinear Dyn. **112**(21), 18639–18663 (2024)
106. Haller, G., Ponsioen, S.: Exact model reduction by a slow-fast decomposition of nonlinear mechanical systems. Nonlinear Dyn. **90**(1), 617–647 (2017). https://doi.org/10.1007/s11071-017-3685-9
107. Hansen, K.W.: Analysis of transient models for catalytic tubular reactors by orthogonal collocation. Chem. Eng. Sci. **26**(10), 1555–1569 (1971)
108. Hansen, K.W., Demandt, K.: Dynamics of a cross-flow heat exchanger with fins. Int. J. Heat Mass Transf. **17**(9), 1029–1036 (1974). https://doi.org/10.1016/0017-9310(74)90184-7
109. Hao, W., Huang, B., Pan, W., Wu, D., Mou, S.: Deep Koopman learning of nonlinear time-varying systems. Automatica **159**, 111372 (2024). https://doi.org/10.1016/j.automatica.2023.111372
110. Hara, K., Inoue, M., Sebe, N.: Dissipativity-constrained learning of MPC with guaranteeing closed-loop stability. Automatica **157**, 111271 (2023). https://doi.org/10.1016/j.automatica.2023.111271
111. Haro, A., Canadell, M., Figueras, J.L., Luque, A., Mondelo, J.M.: The parameterization method for invariant manifolds. Appl. Math. Sci. **195**, 10743–10768 (2016)
112. Haro, A., de La Llave, R.: A parameterization method for the computation of invariant tori and their whiskers in quasi-periodic maps: rigorous results. J. Diff. Equ. **228**(2), 530–579 (2006). https://doi.org/10.1016/j.jde.2005.10.005
113. Hasan, M.M.F., Karimi, I.A., Alfadala, H.E., Grootjans, H.: Operational modeling of multistream heat exchangers with phase changes. AIChE J. **55**(1), 150–171 (2009). https://doi.org/10.1002/aic.11682
114. Héas, P., Herzet, C., Combès, B.: Nonlinear reduced modeling of dynamical systems using Kernel methods and Low-Rank approximation. J. of Nonlinear Sci. **35**(3), 51 (2025). https://doi.org/10.1007/s00332-025-10149-4
115. Hedengren, J.D., Edgar, T.F.: Order reduction of DAE models. IFAC Proc. Vol. **38**(1), 106–111 (2005)
116. Hedengren, J.D., Edgar, T.F.: Order reduction of large scale DAE models. Comput. & Chem. Eng. **29**(10), 2069–2077 (2005). https://doi.org/10.1016/j.compchemeng.2005.05.006
117. Hemery, M., Fages, F., Soliman, S.: On the complexity of Quadratization for polynomial differential equations. In: Abate, A., Petrov, T., Wolf, V. (eds.) Computational Methods in Systems Biology, Lecture Notes in Computer Science, vol. 12314, pp. 120–140. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-60327-4_7
118. Hemery, M., Fages, F., Soliman, S.: Compiling elementary mathematical functions into finite chemical reaction networks via a polynomialization algorithm for ODEs. In: International Conference on Computational Methods in Systems Biology, pp. 74–90 (2021)
119. Hesthaven, J.S., Pagliantini, C., Rozza, G.: Reduced basis methods for time-dependent problems. Acta Numer **31**, 265–345 (2022)
120. Hesthaven, J.S., Ubbiali, S.: Non-intrusive reduced order modeling of nonlinear problems using neural networks. J. Comput. Phys. **363**, 55–78 (2018). https://doi.org/10.1016/j.jcp.2018.02.037
121. Himpe, C., Ohlberger, M.: Cross-gramian-based combined state and parameter reduction for large-scale control systems. Math. Probl. Eng. **2014**, 843869 (2014)
122. Holiday, A., Kooshkbaghi, M., Bello-Rivas, J.M., Gear, C.W., Zagaris, A., Kevrekidis, I.G.: Manifold learning for parameter reduction. J. Comput. Phys. **392**, 419–431 (2019). https://doi.org/10.1016/j.jcp.2019.04.015
123. Hollkamp, J.J., Gordon, R.W.: Reduced-order models for nonlinear response prediction: implicit condensation and expansion. J. Sound Vib. **318**(4–5), 1139–1153 (2008). https://doi.org/10.1016/j.jsv.2008.04.035
124. Holmes, P.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry, 2nd edn. Cambridge University Press (2012)

125. Hoo, K.A., Zheng, D.: Low-order control-relevant models for a class of distributed parameter systems. Chem. Eng. Sci. **56**(23), 6683–6710 (2001)
126. Horstmeyer, L., Atay, F.M.: Characterization of exact lumpability for vector fields on smooth manifolds. Diff. Geom. Appl. **48**, 46–60 (2016). https://doi.org/10.1016/j.difgeo.2016.06.001
127. Horton, R.R., Bequette, B.W., Edgar, T.F.: Improvements in dynamic compartmental modeling for distillation. Comput. & Chem. Eng. **15**(3), 197–201 (1991). https://doi.org/10.1016/0098-1354(91)85006-G
128. Howard, G.M.: Unsteady state behavior of multicomponent distillation columns: part i: simulation. AIChE J. **16**(6), 1022–1029 (1970). https://doi.org/10.1002/aic.690160627
129. Huang, Z.W., Xu, Y.M., Fang, C.Z., Tang, J.: Improvements in dynamic compartmental modelling of distillation columns. J. Process Control **3**(3), 139–145 (1993). https://doi.org/10.1016/0959-1524(93)80019-8
130. Hwang, Y.L., Helfferich, F.G.: Nonlinear waves and asymmetric dynamics of countercurrent separation processes. AIChE J. **35**(4), 690–693 (1989)
131. Iacob, L.C., Tóth, R., Schoukens, M.: Koopman form of nonlinear systems with inputs. Automatica **162**, 111525 (2024). https://doi.org/10.1016/j.automatica.2024.111525
132. Jain, S., Haller, G.: How to compute invariant manifolds and their reduced dynamics in high-dimensional finite element models. Nonlinear Dyn. **107**(2), 1417–1450 (2022)
133. Jauberteau, F., Rosier, C., Temam, R.: The nonlinear Galerkin method in computational fluid dynamics. Appl. Numer. Math. **6**(5), 361–370 (1990). https://doi.org/10.1016/0168-9274(90)90026-C
134. Jensen, J.M., Tummescheit, H.: Moving boundary models for dynamic simulations of two-phase flows. In: Proceedings of the 2nd International Modelica Conference, pp. 235–244 (2002)
135. Ji, W., Qiu, W., Shi, Z., Pan, S., Deng, S.: Stiff-PINN: physics-informed neural network for stiff chemical kinetics. J. Phys. Chem. A **125**(36), 8098–8106 (2021). https://doi.org/10.1021/acs.jpca.1c05102
136. Jourdan, N., Neveux, T., Potier, O., Kanniche, M., Wicks, J., Nopens, I., Rehman, U., Le Moullec, Y.: Compartmental modelling in chemical engineering: a critical review. Chem. Eng. Sci. **210**, 115196 (2019). https://doi.org/10.1016/j.ces.2019.115196
137. Kamath, R.S., Grossmann, I.E., Biegler, L.T.: Aggregate models based on improved group methods for simulation and optimization of distillation systems. Comput. & Chem. Eng. **34**(8), 1312–1319 (2010)
138. Kashima, K.: Nonlinear model reduction by deep autoencoder of noise response data. In: IEEE 55th Conference on Decision and Control (CDC), pp. 5750–5755 (2016). doi:https://doi.org/10.1109/CDC.2016.7799153
139. Kasiraju, S., Vlachos, D.G.: LearnCK: mass conserving neural network reduction of chemistry and species of microkinetic models. Reaction Chem. & Eng. **9**(1), 119–131 (2023). https://doi.org/10.1039/D3RE00279A
140. Kawano, Y., Besselink, B., Scherpen, J.M.: Nonlinear model reduction using balancing methods. In: Encyclopedia of Systems and Control Engineering, vol. 55, pp. 280–296. Elsevier (2026). doi:https://doi.org/10.1016/B978-0-443-14081-5.00033-7
141. Kender, R., Kaufmann, F., Rößler, F., Wunderlich, B., Golubev, D., Thomas, I., Ecker, A.M., Rehfeldt, S., Klein, H.: Development of a digital twin for a flexible air separation unit using a pressure-driven simulation approach. Comput. & Chem. Eng. **151**, 107349 (2021)
142. Khalil, H.K.: Nonlinear Systems, vol. 115, 3rd edn. Patience Hall (2002)
143. Kienle, A.: Low-order dynamic models for ideal multicomponent distillation processes using nonlinear wave propagation theory. Chem. Eng. Sci. **55**(10), 1817–1828 (2000)
144. Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., Noé, F.: Data-driven model reduction and transfer operator approximation. J. of Nonlinear Sci. **28**(3), 985–1010 (2018)
145. Köhne, F., Philipp, F.M., Schaller, M., Schiela, A., Worthmann, K.: L^∞-error bounds for approximations of the Koopman operator by Kernel extended dynamic mode decomposition. SIAM J. Appl. Dyn. Syst. **24**(1), 501–529 (2025)
146. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biol. Cybern. **43**(1), 59–69 (1982). https://doi.org/10.1007/BF00337288
147. Kokotovic, P.V., O'Malley, R.E., Jr., Sannuti, P.: Singular perturbations and order reduction in control theory–an overview. Automatica **12**(2), 123–132 (1976)
148. Korda, M., Mezić, I.: Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. Automatica **93**(1), 149–160 (2018). https://doi.org/10.1016/j.automatica.2018.03.046
149. Korda, M., Mezić, I.: Optimal construction of Koopman eigenfunctions for prediction and control. IEEE Trans. on Autom. Control **65**(12), 5114–5129 (2020)
150. Kramer, B., Peherstorfer, B., Willcox, K.E.: Learning nonlinear reduced models from data with operator inference. Annu. Rev. Fluid Mech. **56**(1), 521–548 (2024). https://doi.org/10.1146/annurev-fluid-121021-025220
151. Kramer, B., Willcox, K.E.: Nonlinear model order reduction via lifting transformations and proper orthogonal decomposition. AIAA J. **57**(6), 2297–2307 (2019)
152. Kramer, M.A.: Autoassociative neural networks. Comput. & Chem. Eng. **16**(4), 313–328 (1992). https://doi.org/10.1016/0098-1354(92)80051-A
153. Krauskopf, B., Osinga, H.M., Doedel, E.J., Henderson, M.E., Guckenheimer, J., Vladimirsky, A., Dellnitz, M., Junge, O.: A survey of methods for computing (un) stable manifolds of vector fields. Int. J. of Bifurcation and Chaos **15**(03), 763–791 (2005)
154. Krischer, K., Rico-Martinez, R., Kevrekidis, I.G., Rotermund, H.H., Ertl, G., Hudson, J.L.: Model identification of a spatiotemporally varying catalytic reaction. AIChE J. **39**(1), 89–98 (1993)
155. Kumar, A., Daoutidis, P.: Nonlinear dynamics and control of process systems with recycle. J. Process Control **12**(4), 475–484 (2002). https://doi.org/10.1016/S0959-1524(01)00014-2
156. Kumar, A., Daoutidis, P.: Nonlinear model reduction and control for high-purity distillation columns. Ind. & Eng. Chem. Res. **42**(20), 4495–4505 (2003)
157. Kumar, R., Ezhilarasi, D.: A state-of-the-art survey of model order reduction techniques for large-scale coupled dynamical systems. Int. J. of Dyn. and Control **11**(2), 900–916 (2023)
158. Lall, S., Marsden, J.E., Glavaški, S.: Empirical model reduction of controlled nonlinear systems. IFAC Proc. Vol. **32**(2), 2598–2603 (1999)
159. Lam, S.H., Goussis, D.A.: The CSP method for simplifying kinetics. Int. J. Chem. Kinet. **26**(4), 461–486 (1994)
160. Lee, J.M.: Introduction to Smooth Manifolds, vol. 218. Springer, New York, New York, NY (2012). https://doi.org/10.1007/978-1-4419-9982-5
161. Lee, K., Carlberg, K.T.: Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. J. Comput. Phys. **404**, 108973 (2020)
162. Lee, M.Y.L.: Estimation of the error in the reduced basis method solution of differential algebraic equation systems. SIAM J. Numer. Anal. **28**(2), 512–528 (1991)
163. Lefèvre, L., Dochain, D., Feyo de Azevedo, S., Magnus, A.: Optimal selection of orthogonal polynomials applied to the integration of chemical reactor equations by collocation methods. Comput. &

Chem. Eng. **24**(12), 2571–2588 (2000). https://doi.org/10.1016/S0098-1354(00)00597-4
164. Lepri, M., Bacciu, D., Della Santina, C.: Neural autoencoder-based structure-preserving model order reduction and control design for high-dimensional physical systems. IEEE Control Syst. Lett. **8**, 133–138 (2023)
165. Lévine, J., Rouchon, P.: Quality control of binary distillation columns via nonlinear aggregated models. Automatica **27**(3), 463–480 (1991)
166. Li, M., Jain, S., Haller, G.: Model reduction for constrained mechanical systems via spectral submanifolds. Nonlinear Dyn. **111**(10), 8881–8911 (2023). https://doi.org/10.1007/s11071-023-08300-5
167. Li, Q., Dietrich, F., Bollt, E.M., Kevrekidis, I.G.: Extended dynamic mode decomposition with dictionary learning: a data-driven adaptive spectral decomposition of the Koopman operator. Chaos: An Interdiscip. J. of Nonlinear Sci. **27**(10), 103111 (2017)
168. Li, X., Xu, Q., Wang, S., Luo, K., Fan, J.: A novel data-driven reduced-order model for the fast prediction of gas-solid heat transfer in fluidized beds. Appl. Therm. Eng. **253**, 123670 (2024). https://doi.org/10.1016/j.applthermaleng.2024.123670
169. Linhart, A., Skogestad, S.: Computational performance of aggregated distillation models. Comput. & Chem. Eng. **33**(1), 296–308 (2009)
170. Linhart, A., Skogestad, S.: Reduced distillation models via stage aggregation. Chem. Eng. Sci. **65**(11), 3439–3456 (2010)
171. Linhart, A., Skogestad, S.: An aggregation model reduction method for one-dimensional distributed systems. AIChE J. **58**(5), 1524–1537 (2012). https://doi.org/10.1002/aic.12688
172. Löffler, H.P., Marquardt, W.: Order reduction of non-linear differential-algebraic process models. J. Process Control **1**(1), 32–40 (1991). https://doi.org/10.1016/0959-1524(91)87005-I
173. Lu, Q., Zavala, V.M.: Image-based model predictive control via dynamic mode decomposition. J. Process Control **104**, 146–157 (2021)
174. Lu, T., Law, C.K.: A directed relation graph method for mechanism reduction. Proc. Combust. Inst. **30**(1), 1333–1341 (2005)
175. Luo, X., Li, M., Roetzel, W.: A general solution for one-dimensional multistream heat exchangers and their networks. Int. J. Heat Mass Transf. **45**(13), 2695–2705 (2002). https://doi.org/10.1016/S0017-9310(02)00003-0
176. Lusch, B., Kutz, J.N., Brunton, S.: Deep learning for universal linear embeddings of nonlinear dynamics. Nat. Commun. **9**(1), 1–10 (2018). https://doi.org/10.1038/s41467-018-07210-0
177. Lüthje, J.T., Schulze, J.C., Caspari, A., Mhamdi, A., Mitsos, A., Schäfer, P.: Adaptive learning of Hybrid models for nonlinear model predictive control of distillation columns. IFAC-PapersOnLine **54**(3), 37–42 (2021). https://doi.org/10.1016/j.ifacol.2021.08.215
178. Luyben, W.L.: Chemical Reactor Design and Control. John Wiley & Sons (2007)
179. Ly, H.V., Tran, H.T.: Modeling and control of physical processes using proper orthogonal decomposition. Math. Comput. Model. **33**(1–3), 223–236 (2001). https://doi.org/10.1016/S0895-7177(00)00240-5
180. Ma, J., Kim, D., Braun, J.E.: Proper orthogonal decomposition for reduced order dynamic modeling of vapor compression systems. Int. J. Refrig **132**, 145–155 (2021). https://doi.org/10.1016/j.ijrefrig.2021.09.016
181. Ma, Y., Fu, Y.: Manifold Learning Theory and Applications. CRC Press (2011). https://doi.org/10.1201/b11431
182. Maas, U., Pope, S.B.: Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space. Combust. Flame **88**(3–4), 239–264 (1992). https://doi.org/10.1016/0010-2180(92)90034-M
183. Mahadevan, N., Hoo, K.A.: Wavelet-based model reduction of distributed parameter systems. Chem. Eng. Sci. **55**(19), 4271–4290 (2000)
184. Mainini, L., Willcox, K.: Surrogate modeling approach to support real-time structural assessment and decision making. AIAA J. **53**(6), 1612–1626 (2015)
185. Mallick, S., Mittal, M.: AI-based model order reduction techniques: a survey. Arch. of Comput. Methods in Eng. **32**(4), 1–26 (2025)
186. Marion, M., Temam, R.: Nonlinear Galerkin methods. SIAM J. Numer. Anal. **26**(5), 1139–1157 (1989). https://doi.org/10.1137/0726063
187. Marquardt, W.: Nonlinear model reduction for binary distillation. Dyn. and Control of Chem. React. and Distillation Columns 123–128 (1986)
188. Marquardt, W.: Nonlinear model reduction for optimization based control of transient chemical processes. AIChE Symposium Ser. (2002)
189. Masti, D., Bemporad, A.: Learning nonlinear state-space models using autoencoders. Automatica **129**, 109666 (2021)
190. Mathisen, K.W., Morari, M., Skogestad, S.: Dynamic models for heat exchangers and heat exchanger networks. Comput. & Chem. Eng. **18**, S459–S463 (1994)
191. Mayo, A.J., Antoulas, A.: A framework for the solution of the generalized realization problem. Linear Algebra Appl. **425**(2–3), 634–662 (2007)
192. Mehrmann, V., Unger, B.: Control of port-Hamiltonian differential-algebraic systems and applications. Acta Numer **32**, 395–515 (2023). https://doi.org/10.1017/S0962492922000083
193. Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn. **41**(1), 309–325 (2005)
194. Michel, A., Kugi, A.: Accurate low-order dynamic model of a compact plate heat exchanger. Int. J. Heat Mass Transf. **61**, 323–331 (2013). https://doi.org/10.1016/j.ijheatmasstransfer.2013.01.072
195. Mignolet, M.P., Przekop, A., Rizzi, S.A., Spottswood, S.M.: A review of indirect/non-intrusive reduced order modeling of nonlinear geometric structures. J. Sound Vib. **332**(10), 2437–2460 (2013). https://doi.org/10.1016/j.jsv.2012.10.017
196. Mika, S., Schölkopf, B., Smola, A., Müller, K.R., Scholz, M., Rätsch, G.: Kernel PCA and de-noising in feature spaces. Adv. Neural. Inf. Process. Syst. **11**, (1998)
197. Moore, B.: Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Trans. on Autom. Control **26**(1), 17–32 (1981)
198. Morari, M., Maeder, U.: Nonlinear offset-free model predictive control. Automatica **48**(9), 2059–2067 (2012)
199. Munkres, J.R.: Analysis on Manifolds. Addison-Wesley Publishing (1991)
200. Nakatsukasa, Y., Sète, O., Trefethen, L.N.: The AAA algorithm for rational approximation. SIAM J. Sci. Comput. **40**(3), A1494–A1522 (2018)
201. Narasingam, A., Kwon, J.S.I.: Koopman Lyapunov-based model predictive control of nonlinear chemical process systems. AIChE J. **65**(11), e16743 (2019)
202. Nazari, P., Damrich, S., Hamprecht, F.A.: Geometric autoencoders – what you see is what you decode. In: 40th International Conference on Machine (ICLR), (2023)
203. Nguyen, N.C., Peraire, J.: An efficient reduced-order modeling approach for non-linear parametrized partial differential equations. Int. J. Numer. Meth. Eng. **76**(1), 27–55 (2008). https://doi.org/10.1002/nme.2309
204. Nicolaidou, E., Hill, T.L., Neild, S.A.: Nonlinear mapping of non-conservative forces for reduced-order modelling. Proc. of the Royal Soc. A **478**, 2268 (2022). https://doi.org/10.1098/rspa.2022.0522

205. Nijmeijer, H., van der Schaft, A.: Nonlinear Dynamical Control Systems. Springer, New York, New York, NY (1990). https://doi.org/10.1007/978-1-4757-2101-0
206. Noack, B.R., Morzynski, M., Tadmor, G.: Reduced-Order Modelling for Flow Control, vol. 528. Springer Science & Business Media (2011)
207. Obinata, G., Anderson, B.D.O.: Model Reduction for Control System Design. Communications and Control Engineering. Springer, London and Berlin and Heidelberg (2001)
208. Ohlberger, M., Rave, S.: Reduced basis methods: success, limitations and future challenges. In: Proceedings Of The Conference Algoritmy, pp. 1–12 (2016)
209. Okeke, B.E.: Roussel: an invariant-manifold approach to lumping. Math. Modell. of Nat. Phenomena **10**(3), 149–167 (2015)
210. Opreni, A., Vizzaccaro, A., Touzé, C., Frangi, A.: High-order direct parametrisation of invariant manifolds for model order reduction of finite element structures: application to generic forcing terms and parametrically excited systems. Nonlinear Dyn. **111**(6), 5401–5447 (2023). https://doi.org/10.1007/s11071-022-07978-3
211. Otto, S.E., Macchio, G.R., Rowley, C.W.: Learning nonlinear projections for reduced-order modeling of dynamical systems using constrained autoencoders. Chaos: An Interdiscip. J. of Nonlinear Sci. **33**(11), 1 (2023). https://doi.org/10.1063/5.0169688
212. Otto, S.E., Padovan, A., Rowley, C.W.: Optimizing oblique projections for nonlinear systems using Trajectories. SIAM J. Sci. Comput. **44**(3), A1681–A1702 (2022). https://doi.org/10.1137/21M1425815
213. Otto, S.E., Rowley, C.W.: Linearly recurrent autoencoder networks for learning dynamics. SIAM J. Appl. Dyn. Syst. **18**(1), 558–593 (2019). https://doi.org/10.1137/18M1177846
214. Otto, S.E., Rowley, C.W.: Koopman operators for estimation and control of dynamical systems. Ann. Rev. of Control, Robotics, and Auton. Syst. **4**, 59–87 (2021)
215. Pantelides, C.C., Renfro, J.G.: The online use of first-principles models in process operations: review, current status and future needs. Comput. & Chem. Eng. **51**, 136–148 (2013). https://doi.org/10.1016/j.compchemeng.2012.07.008
216. Park, H.M., Cho, D.H.: Low dimensional modeling of flow reactors. Int. J. Heat Mass Transf. **39**(16), 3311–3323 (1996)
217. Parker, R., Nicholson, B., Siirola, J., Laird, C., Biegler, L.: An implicit function formulation for optimization of discretized index-1 differential algebraic systems. Comput. & Chem. Eng. **168**, 108042 (2022). https://doi.org/10.1016/j.compchemeng.2022.108042
218. Pateras, J., Zhang, C., Majumdar, S., Pal, A., Ghosh, P.: Physics-informed machine learning for automatic model reduction in chemical reaction networks. Sci. Rep. **15**(1), 7980 (2025)
219. Pearson, R.K.: Selecting nonlinear model structures for computer control. J. Process Control **13**(1), 1–26 (2003)
220. Peherstorfer, B., Gugercin, S., Willcox, K.: Data-driven reduced model construction with time-domain loewner models. SIAM J. Sci. Comput. **39**(5), A2152–A2178 (2017). https://doi.org/10.1137/16M1094750
221. Peherstorfer, B., Willcox, K.: Data-driven operator inference for nonintrusive projection-based model reduction. Comput. Methods Appl. Mech. Eng. **306**, 196–215 (2016). https://doi.org/10.1016/j.cma.2016.03.025
222. Peitz, S., Klus, S.: Koopman operator-based model reduction for switched-system control of PDEs. Automatica **106**, 184–191 (2019)
223. Peitz, S., Otto, S.E., Rowley, C.W.: Data-driven model predictive control using interpolated Koopman generators. SIAM J. Appl. Dyn. Syst. **19**(3), 2162–2193 (2020)
224. Pepiot-Desjardins, P., Pitsch, H.: An efficient error-propagation-based reduction method for large chemical kinetic mechanisms. Combust. Flame **154**(1–2), 67–81 (2008). https://doi.org/10.1016/j.combustflame.2007.10.020
225. Pernebo, L., Silverman, L.: Model reduction via balanced state space representations. IEEE Trans. Autom. Control **27**(2), 382–387 (1982). https://doi.org/10.1109/TAC.1982.1102945
226. Petzold, L.: Differential/algebraic equations are not ODE's. SIAM J. Sci. Stat. Comput. **3**(3), 367–384 (1982)
227. Petzold, L., Zhu, W.: Model reduction for chemical kinetics: an optimization approach. AIChE J. **45**(4), 869–886 (1999)
228. Phillips, J.R.: Projection-based approaches for model reduction of weakly nonlinear, time-varying systems. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **22**(2), 171–187 (2003). https://doi.org/10.1109/TCAD.2002.806605
229. Phillips, T.R.F., Heaney, C.E., Smith, P.N., Pain, C.C.: An autoencoder-based reduced-order model for eigenvalue problems with application to neutron diffusion. Int. J. Numer. Meth. Eng. **122**(15), 3780–3811 (2021)
230. Picon-Nunez, M., Polley, G.T., Medina-Flores, M.: Thermal design of multi-stream heat exchangers. Appl. Therm. Eng. **22**(14), 1643–1660 (2002)
231. Proctor, J.L., Brunton, S.L., Kutz, J.N.: Dynamic mode decomposition with control. SIAM J. Appl. Dyn. Syst. **15**(1), 142–161 (2016)
232. Qian, E., Kramer, B., Peherstorfer, B., Willcox, K.: Lift & learn: physics-informed machine learning for large-scale nonlinear dynamical systems. Physica D **406**, 132401 (2020). https://doi.org/10.1016/j.physd.2020.132401
233. Qian, X., Dang, Q., Jia, S., Yuan, Y., Huang, K., Chen, H., Zhang, L.: Operation of distillation columns using model predictive control based on dynamic mode decomposition method. Ind. & Eng. Chem. Res. **62**(50), 21721–21739 (2023)
234. Qin, S.J.: An overview of subspace identification. Comput. & Chem. Eng. **30**(10–12), 1502–1513 (2006)
235. Quarshie, A.W.K., Swartz, C.L.E., Madabhushi, P.B., Cao, Y., Wang, Y., Flores-Cerrillo, J.: Modeling, simulation, and optimization of multiproduct cryogenic air separation unit startup. AIChE J. **69**(2), e17953 (2023). https://doi.org/10.1002/aic.17953
236. Ranzi, E., Rovaglio, M., Faravelli, T., Biardi, G.: Role of energy balances in dynamic simulation of multicomponent distillation columns. Comput. & Chem. Eng. **12**(8), 783–786 (1988). https://doi.org/10.1016/0098-1354(88)80016-4
237. Rathinam, M., Petzold, L.R.: A new look at proper orthogonal decomposition. SIAM J. Numer. Anal. **41**(5), 1893–1925 (2003). https://doi.org/10.1137/S0036142901389049
238. Reizman, B.J., Jensen, K.F.: An automated continuous-flow platform for the estimation of Multistep reaction Kinetics. Organic Process Res. & Dev. **16**(11), 1770–1782 (2012). https://doi.org/10.1021/op3001838
239. Rewienski, M., White, J.: A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **22**(2), 155–170 (2003). https://doi.org/10.1109/TCAD.2002.806601
240. Rewieński, M., White, J.: Model order reduction for nonlinear dynamical systems based on trajectory piecewise-linear approximations. Linear Algebra Appl. **415**(2–3), 426–454 (2006). https://doi.org/10.1016/j.laa.2003.11.034
241. Rheinboldt, W.C.: Differential-algebraic systems as differential equations on manifolds. Math. Comput. **43**(168), 473–482 (1984). https://doi.org/10.1090/S0025-5718-1984-0758195-5
242. Roberts, A.J.: Appropriate initial conditions for asymptotic descriptions of the long term evolution of dynamical systems. The J. of the Australian Math. Soc. Ser. B. Appl. Math. **31**(1), 48–75 (1989). https://doi.org/10.1017/S0334270000006470

243. Roberts, A.J.: The utility of an invariant manifold description of the evolution of a dynamical system. SIAM J. Math. Anal. **20**(6), 1447–1458 (1989). https://doi.org/10.1137/0520094
244. Roberts, A.J.: Normal form transforms separate slow and fast modes in stochastic dynamical systems. Physica A **387**(1), 12–38 (2008). https://doi.org/10.1016/j.physa.2007.08.023
245. Rodriguez, A.C., Balicki, L., Gugercin, S.: The p-AAA algorithm for data-driven modeling of parametric dynamical systems. SIAM J. Sci. Comput. **45**(3), A1332–A1358 (2023)
246. Roffel, B., Betlem, B.H.L., de Ruijter, J.A.: First principles dynamic modeling and multivariable control of a cryogenic distillation process. Comput. & Chem. Eng. **24**(1), 111–123 (2000)
247. Romijn, R.: Empirical Model Reduction of Differential-Algebraic Equation Systems. RWTH Aachen University, Dissertation (2016)
248. Romor, F., Stabile, G., Rozza, G.: Non-linear manifold reduced-order models with convolutional autoencoders and reduced over-collocation method. J. Sci. Comput. **94**(3), 74 (2023). https://doi.org/10.1007/s10915-023-02128-2
249. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Sci. **290**(5500), 2323–2326 (2000)
250. Rowley, C.W., Colonius, T., Murray, R.M.: Model reduction for compressible flows using POD and Galerkin projection. Physica D **189**(1–2), 115–129 (2004)
251. Rowley, C.W., Dawson, S.T.: Model reduction for flow analysis and control. Annu. Rev. Fluid Mech. **49**(1), 387–417 (2017). https://doi.org/10.1146/annurev-fluid-010816-060042
252. Rozza, G., Huynh, D.B.P., Patera, A.T.: Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. Arch. of Comput. Methods in Eng. **15**(3), 229–275 (2008). https://doi.org/10.1007/s11831-008-9019-9
253. Rozza, G., Stabile, G., Ballarin, F. (eds.): Advanced Reduced Order Methods and Applications in Computational Fluid Dynamics. Society for Industrial and Applied Mathematics, Philadelphia, PA (2022). doi:https://doi.org/10.1137/1.9781611977257
254. Ryckelynck, D.: Hyper-reduction of mechanical models involving internal variables. Int. J. Numer. Meth. Eng. **77**(1), 75–89 (2009)
255. Scarciotti, G., Astolfi, A.: Data-driven model reduction by moment matching for linear and nonlinear systems. Automatica **79**, 340–351 (2017)
256. Schäfer, P., Caspari, A., Kleinhans, K., Mhamdi, A., Mitsos, A.: Reduced dynamic modeling approach for rectification columns based on compartmentalization and artificial neural networks. AIChE J. **65**(5), e16568 (2019)
257. Scherpen, J.M.A.: Balancing for nonlinear systems. Syst. & Control Lett. **21**(2), 143–153 (1993)
258. Schlegel, M., vd Berg, J., Marquardt, W., Bosgra, O.H.: Projection based model reduction for dynamic optimization. In: AIChE Annual Meeting. Indianapolis (2002)
259. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5–28 (2010)
260. Schölkopf, B., Smola, A., Müller, K.R.: Kernel principal component analysis. In: International conference on artificial neural networks, pp. 583–588 (1997)
261. Schulze, J.C., Caspari, A., Offermanns, C., Mhamdi, A., Mitsos, A.: Nonlinear model predictive control of ultra-high-purity air separation units using transient wave propagation model. Comput. & Chem. Eng. **145**, 107163 (2021)
262. Schulze, J.C., Doncevic, D.T., Erwes, N., Mitsos, A.: Data-Driven Model Reduction and Nonlinear Model Predictive Control of an Air Separation Unit by Applied Koopman Theory. Foundations of Computer Aided Process Operations / Chemical Process Control (FOCAPO) (2023)
263. Schulze, J.C., Doncevic, D.T., Mitsos, A.: Identification of MIMO Wiener-type Koopman models for data-driven model reduction using deep learning. Comput. & Chem. Eng. **161**(15), 107781 (2022). https://doi.org/10.1016/j.compchemeng.2022.107781
264. Schulze, J.C., Mitsos, A.: Data-driven nonlinear model reduction using Koopman theory: integrated control form and NMPC case study. IEEE Control Syst. Lett. (L-CSS) **6**, 2978–2983 (2022). https://doi.org/10.1109/LCSYS.2022.3181443
265. Schweidtmann, A.M., Mitsos, A.: Deterministic global optimization with artificial neural networks embedded. J. Optim. Theory Appl. **180**(3), 925–948 (2019)
266. Shah, R.K., Sekulic, D.P.: Fundamentals of Heat Exchanger Design. John Wiley & Sons (2003)
267. Shakib, M.F., Scarciotti, G., Pogromsky, A.Y., Pavlov, A., van de Wouw, N.: Model reduction by moment matching with preservation of global stability for a class of nonlinear models. Automatica **157**, 111227 (2023)
268. Sharma, H., Wang, Z., Kramer, B.: Hamiltonian operator inference: physics-preserving learning of reduced-order models for canonical Hamiltonian systems. Physica D **431**(19), 133122 (2022). https://doi.org/10.1016/j.physd.2021.133122
269. Shaw, S.W., Pierre, C.: Non-linear normal modes and invariant manifolds. J. Sound Vib. **150**(1), 170–173 (1991). https://doi.org/10.1016/0022-460X(91)90412-D
270. Shen, Y., Béreux, N., Frangi, A., Touzé, C.: Reduced order models for geometrically nonlinear structures: assessment of implicit condensation in comparison with invariant manifold approach. Eur. J. Mech. A. Solids **86**, 104165 (2021). https://doi.org/10.1016/j.euromechsol.2020.104165
271. Shvartsman, S.Y., Kevrekidis, I.G.: Nonlinear model reduction for control of distributed systems: a computer-assisted study. AIChE J. **44**(7), 1579–1595 (1998). https://doi.org/10.1002/aic.690440711
272. Shvartsman, S.Y., Theodoropoulos, C., Rico-Martinez, R., Kevrekidis, I.G., Titi, E.S., Mountziaris, T.J.: Order reduction for nonlinear dynamic models of distributed reacting systems. J. Process Control **10**(2–3), 177–184 (2000). https://doi.org/10.1016/S0959-1524(99)00029-3
273. Sjöberg, J., Fujimoto, K., Glad, T.: Model reduction of nonlinear differential-algebraic equations. IFAC Proc. Vol. **40**(12), 176–181 (2007)
274. Snowden, T.J., van der Graaf, P.H., Tindall, M.J.: Methods of model reduction for large-scale biological systems: a survey of current methods and trends. Bull. Math. Biol. **79**, 1449–1486 (2017)
275. Steeb, W.H., Wilhelm, F.: Non-linear autonomous systems of differential equations and Carleman linearization procedure. J. Math. Anal. Appl. **77**(2), 601–611 (1980). https://doi.org/10.1016/0022-247X(80)90250-4
276. Stewart, W.E., Levien, K.L., Morari, M.: Simulation of fractionation by orthogonal collocation. Chem. Eng. Sci. **40**(3), 409–421 (1985)
277. Sun, C., Hahn, J.: Reduction of stable differential-algebraic equation systems via projections and system identification. J. Process Control **15**(6), 639–650 (2005)
278. Surana, A.: Koopman operator based observer synthesis for control-affine nonlinear systems. In: 2016 IEEE 55th Conference on Decision and Control (CDC), pp. 6492–6499 (2016). doi:https://doi.org/10.1109/CDC.2016.7799268
279. Swartz, C.L.E., Stewart, W.E.: A collocation approach to distillation column design. AIChE J. **32**(11), 1832–1838 (1986)
280. Szalai, R.: Invariant spectral foliations with applications to model order reduction and synthesis. Nonlinear Dyn. **101**(4), 2645–2669 (2020)
281. Szalai, R.: Data-driven reduced order models using invariant foliations, manifolds and autoencoders. J. of Nonlinear Sci. **33**(5), 75 (2023). https://doi.org/10.1007/s00332-023-09932-y

282. Touzé, C., Frangi, A.: Model Order Reduction for Design, Analysis and Control of Nonlinear Vibratory Systems, vol. 614. Springer Nature, Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-67499-0
283. Touzé, C., Vizzaccaro, A., Thomas, O.: Model order reduction methods for geometrically nonlinear structures: a review of nonlinear techniques. Nonlinear Dyn. **105**(2), 1141–1190 (2021). https://doi.org/10.1007/s11071-021-06693-9
284. Tsay, C.: Sobolev trained neural network surrogate models for optimization. Comput. & Chem. Eng. **153**, 107419 (2021). https://doi.org/10.1016/j.compchemeng.2021.107419
285. Tsay, C., Baldea, M.: Integrating production scheduling and process control using latent variable dynamic models. Control. Eng. Pract. **94**, 104201 (2020). https://doi.org/10.1016/j.conengprac.2019.104201
286. van den Berg, J.B., Mireles James, J.D., Reinhardt, C.: Computing (Un)stable manifolds with validated error bounds: non-resonant and resonant spectra. J. of Nonlinear Sci. **26**(4), 1055–1095 (2016). https://doi.org/10.1007/s00332-016-9298-5
287. Vaupel, Y., Huster, W.R., Holtorf, F., Mhamdi, A., Mitsos, A.: Analysis and improvement of dynamic heat exchanger models for nominal and start-up operation. Energy **169**, 1191–1201 (2019). https://doi.org/10.1016/j.energy.2018.12.048
288. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning, pp. 1096–1103 (2008)
289. Vizzaccaro, A., Gobat, G., Frangi, A., Touzé, C.: Direct parametrisation of invariant manifolds for non-autonomous forced systems including superharmonic resonances. Nonlinear Dyn. **112**(8), 6255–6290 (2024). https://doi.org/10.1007/s11071-024-09333-0
290. Vizzaccaro, A., Shen, Y., Salles, L., Blahoš, J., Touzé, C.: Direct computation of nonlinear mapping via normal form for reduced-order models of finite element nonlinear structures. Comput. Methods Appl. Mech. Eng. **384**, 113957 (2021). https://doi.org/10.1016/j.cma.2021.113957
291. Vlachos, D.G.: Reduction of detailed kinetic mechanisms for ignition and extinction of premixed hydrogen/air flames. Chem. Eng. Sci. **51**(16), 3979–3993 (1996). https://doi.org/10.1016/0009-2509(96)00239-4
292. Vora, N., Daoutidis, P.: Nonlinear model reduction of chemical reaction systems. AIChE J. **47**(10), 2320–2332 (2001)
293. Wang, Z., Xiao, D., Fang, F., Govindan, R., Pain, C.C., Guo, Y.: Model identification of reduced order fluid dynamics systems using deep learning. Int. J. Numer. Meth. Fluids **86**(4), 255–268 (2018). https://doi.org/10.1002/fld.4416
294. Watter, M., Springenberg, J., Boedecker, J., Riedmiller, M.: Embed to control: A locally linear latent dynamics model for control from raw images. Adv. Neural. Inf. Process. Syst. **28**, (2015)
295. Weischedel, K., McAvoy, T.J.: Feasibility of decoupling in conventionally controlled distillation columns. Ind. & Eng. Chem. Fundamentals **19**(4), 379–384 (1980)
296. Wen, M., Spotte-Smith, E.W.C., Blau, S.M., McDermott, M.J., Krishnapriyan, A.S., Persson, K.A.: Chemical reaction networks and opportunities for machine learning. Nat. Comput. Sci. **3**(1), 12–24 (2023). https://doi.org/10.1038/s43588-022-00369-z
297. Wiewel, S., Becher, M., Thuerey, N.: Latent space physics: towards learning the temporal evolution of fluid flow. Computer Graphics Forum **38**, 71–82 (2019)
298. Willcox, K., Peraire, J.: Balanced model reduction via the proper orthogonal decomposition. AIAA J. **40**(11), 2323–2330 (2002)
299. Williams, M.O., Hemati, M.S., Dawson, S.T., Kevrekidis, I.G., Rowley, C.W.: Extending data-driven Koopman analysis to actuated systems. IFAC-PapersOnLine **49**(18), 704–709 (2016). https://doi.org/10.1016/j.ifacol.2016.10.248
300. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the koopman operator: extending dynamic mode decomposition. J. of Nonlinear Sci. **25**(6), 1307–1346 (2015)
301. Wong, K.T., Luus, R.: Model reduction of high-order multistage systems by the method of orthogonal collocation. The Canadian J. of Chem. Eng. **58**(3), 382–388 (1980)
302. Wu, P., Sun, J., Chang, X., Zhang, W., Arcucci, R., Guo, Y., Pain, C.C.: Data-driven reduced order model with temporal convolutional neural network. Comput. Methods Appl. Mech. Eng. **360**, 112766 (2020). https://doi.org/10.1016/j.cma.2019.112766
303. Xia, A.-G., Hartley: Modelling and simulation of a heat exchanger. In: IEEE International Conference on Systems Engineering, pp. 453–456. IEEE (1991). https://doi.org/10.1109/ICSYSE.1991.161174
304. Xu, B., Yebi, A., Hoffman, M., Onori, S.: A rigorous model order reduction framework for waste heat recovery systems based on proper orthogonal decomposition and galerkin projection. IEEE Trans. Control Syst. Technol. **28**(2), 635–643 (2018)
305. Xu, Y., Peng, M., Cammi, A., Introini, C., Xia, G.: Model order reduction of a once-through steam generator via dynamic mode decomposition. Ann. Nucl. Energy **201**, 110457 (2024). https://doi.org/10.1016/j.anucene.2024.110457
306. Yamaleev, N.K., Pathak, K.A.: Nonlinear model reduction for unsteady discontinuous flows. J. Comput. Phys. **245**, 1–13 (2013). https://doi.org/10.1016/j.jcp.2013.03.002
307. Yee, T.F., Grossmann, I.E., Kravanja, Z.: Simultaneous optimization models for heat integration—i. area and energy targeting and modeling of multi-stream exchangers. Comput. & Chem. Eng. **14**(10), 1151–1164 (1990). https://doi.org/10.1016/0098-1354(90)85009-Y
308. Zavala-Rio, A., Santiesteban-Cos, R.: Reliable compartmental models for double-pipe heat exchangers: an analytical study. Appl. Math. Model. **31**(9), 1739–1752 (2007). https://doi.org/10.1016/j.apm.2006.06.005
309. Zhang, Z., Zhao, J.: General numerical framework to derive structure preserving reduced order models for thermodynamically consistent reversible-irreversible PDEs. J. Comput. Phys. **521**(9), 113562 (2025). https://doi.org/10.1016/j.jcp.2024.113562
310. Zhou, C., Paffenroth, R.C.: Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 665–674 (2017)