

Identification of MIMO Wiener-type Koopman Models for Data-Driven Model Reduction using Deep Learning

Jan C. Schulze¹, Danimir T. Doncevic² and Alexander Mitsos^{1,3,2,*}

¹ Process Systems Engineering (AVT.SVT), RWTH Aachen University, 52074 Aachen, Germany.

² Institute of Energy and Climate Research: Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

³ JARA-ENERGY, 52056 Aachen, Germany.

Abstract: We use Koopman theory to develop a data-driven nonlinear model reduction and identification strategy for multiple-input multiple-output (MIMO) input-affine dynamical systems. While the present literature has focused on linear and bilinear Koopman models, we derive and use a Wiener-type Koopman formulation. We discuss that the Wiener structure is particularly suitable for model reduction, and can be naturally derived from Koopman theory. Moreover, the Wiener block-structure unifies the mathematical simplicity of linear dynamical blocks and the accuracy of bilinear dynamics. We present a Koopman deep-learning strategy combining autoencoders and linear dynamics that generates low-order surrogate models of MIMO Wiener type. In three case studies, we apply our framework for identification and reduction of a system with input multiplicity, a chemical reactor and a high-purity distillation column. We compare the prediction performance of the identified Wiener models to linear and bilinear Koopman models. We observe the highest accuracy and strongest model reduction capabilities of low-order Wiener-type Koopman models, making them promising for control.

Keywords: Nonlinear model reduction, Nonlinear system identification, Koopman theory, MIMO Wiener model, Deep learning

Published in *Computers and Chemical Engineering*: DOI:10.1016/j.compchemeng.2022.107781

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

1 Introduction

The identification of low-order nonlinear dynamical models is one of the main challenges in control engineering. Practical control models do not have to be highly accurate and broadly applicable, but should be simple and reliable predictors of the control-relevant dynamical response. Traditionally, low-order control models are obtained either by combining mechanistic modeling and model reduction [1,2], or via black-box identification of generic input-output models from measured data [3–5]. Nonlinear model reduction remains an active field of research, as no universal approach exists and the application of many existing methods can be mathematically involved [2,6]. On the other hand, black-box approaches such as NARMAX, Volterra series, and Hammerstein or Wiener models are nowadays established methods [7,8]. However, successful identification is strongly system-specific as it depends on a suitable choice of model type, data set, algorithm as well as user experience [4]. Recently, data-driven strategies combining elements of the two traditional approaches have been proposed, e.g., dynamic mode decomposition [9–11], Koopman-based modeling [12], lift-and-learn methods [13], and deep-learning autoencoder approaches [14,15]. Overall, generating low-order control models remains challenging in practice due to difficulties in the specification and interpretation of reduced model structures, and involved reduction or identification algorithms.

*Corresponding author: amitsos@alum.mit.edu

Here, we perform nonlinear model reduction through data-driven identification of low-order surrogate models. For this purpose, we apply Koopman theory and develop a broadly applicable reduction strategy that identifies low-order MIMO Wiener control models. These reduced models are simple and relatively minimalistic due to their linear dynamical block. Our approach employs deep-learning methods and targets a minimal need of expert knowledge of model reduction.

Koopman theory postulates that nonlinear dynamical systems can be globally exactly reformulated as linear dynamics when represented on an infinite-dimensional state space by means of nonlinear coordinate transformation [16, 17]. Recent works have aimed at identifying finite-dimensional, possibly low-order approximates of the infinite-dimensional Koopman operator [18]. Further, Koopman theory has been adopted for nonlinear system analysis, e.g., global stability analysis [19] and the extension of the local Hartman-Grobman theorem to the entire basin of attraction of an equilibrium point or periodic orbit [20]. While finding an exact low-dimensional Koopman model is often difficult or impossible, using data-driven methods, e.g., deep learning, to identify approximate Koopman models for monitoring or control is feasible in many cases. Examples for the successful applications of Koopman theory in various fields include fluid dynamics [21], molecular thermodynamics [22] and disease modeling [23]. Due to its recent extension to systems with external inputs [24–27], Koopman theory is also promising for control.

We present a data-driven non-intrusive reduction strategy for control motivated by Koopman theory. Therein, we simultaneously identify a nonlinear transformation and low-order latent dynamics approximating the nonlinear system evolution. While the existing literature on Koopman modeling for control has focused on linear [28–36] and bilinear [37–40] forms, we consider a MIMO Wiener structure. As we show, when combined with Koopman theory, the nonlinear Wiener structure provides strong model reduction capabilities beyond those of linear and bilinear models. The Wiener structure has already been used implicitly in many works applying Koopman theory to autonomous systems, e.g., [41–43], without an explicit classification as such. However, to our knowledge, Wiener-type Koopman models with controls have not yet been used in the literature. Herein, we derive the Wiener-type Koopman form for input-affine systems, showing that the MIMO Wiener structure is naturally related to Koopman theory.

Our model identification and reduction procedure extends the Koopman deep-learning framework for autonomous systems by Lusch et al. [42] to MIMO models. To this end, we combine autoencoders with non-autonomous linear dynamical blocks. This extension is mathematically justified, as shown by our derivation of the Wiener-type Koopman form. As opposed to existing MIMO Wiener identification methods [44–51], our Koopman strategy targets data-driven model reduction, and therefore uses full-state information (trajectory snapshots) instead of input-output data. The method is simple to use thanks to few requirements on model properties, identification procedure and data set, e.g., step responses are sufficient.

We apply our model reduction and identification framework to three case studies: a system with input multiplicity, a chemical reactor and a distillation column. We demonstrate that the low-order MIMO Wiener Koopman models reproduce the dynamics of the nonlinear input-affine systems precisely. Further, we compare to bilinear and linear models. In particular, we show that low-order Wiener-type models reach similar or better accuracy than low-order bilinear models, while only involving a linear dynamical block. Thus, Wiener-type models enable a higher degree of reduction and simplification.

The manuscript is structured as follows: In Section 2, we provide the theoretical background on Koopman theory and block-structured modeling. In Section 3, we derive a Koopman model of Wiener type. In Section 4, we present a deep-learning identification and model reduction strategy using Wiener-type Koopman or MIMO Wiener models. In Section 5, we demonstrate our model identification framework in three case studies. We conclude the manuscript with summary and outlook in Section 6.

2 Theoretical Background

We provide a brief review of system identification using block-structured modeling, and subsequently summarize Koopman theory and its extensions to control systems. Our presentation and notation of Koopman theory follows the comprehensive reviews in [10, 12, 17, 52].

2.1 Block-structured modeling

Block-structured modeling is a data-driven system identification approach combining nonlinear static and linear dynamical blocks [53]. Block-structured models have been used extensively in control applications for several decades, and are prominent due to their intuitive structure and sophisticated identification strategies available [8, 53]. Moreover, there exist efficient methods to embed serial block-structured models in MPC applications, e.g., [54].

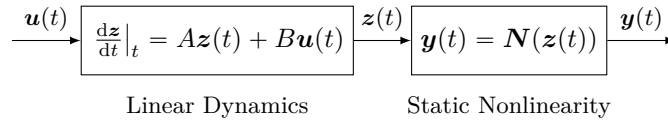


Fig. 1: Block-structure of MIMO Wiener model. Linear dynamics involve inputs $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ and states $\mathbf{z}(t) \in \mathbb{R}^{n_z}$. Output nonlinearity $\mathbf{N} : \mathbb{R}^{n_z} \rightarrow \mathbb{R}^{n_y}$ yields outputs $\mathbf{y}(t) \in \mathbb{R}^{n_y}$.

Wiener models are one common type of serial block-structured models, cascading a linear dynamical block with a nonlinear static block, Fig. 1. Using MIMO blocks leads to MIMO Wiener models, which is regarded as an extension of the single-input single-output (SISO) parallel Wiener [55] structure. However, compared to SISO models, the mapping of MIMO blocks is mathematically more complex and usually associated with more parameters. In many applications, Wiener structures are assumed to be purely empirical. However, some approaches assign the model blocks a certain physical meaning (gray-box modeling), e.g., the nonlinear static block representing the nonlinear stationary states [56] or measurement nonlinearity. Furthermore, depending on the formulation, Wiener structures are regarded as a restriction or a variant of Volterra series [57]. In Section 3, we show that the Wiener structure can also be mathematically motivated by Koopman theory.

Within the Wiener model class, there exists a variety of options for formulating the nonlinearity and linear dynamical block. Besides using, e.g., polynomial or piecewise linear functions for the static nonlinear block, several works have investigated neural networks (NNs) to represent the nonlinear block in SISO [58–62] and MIMO [46, 49] Wiener models. Combining discrete-time linear blocks and NNs promotes deep-learning environments as an alternative to standard model identification toolboxes. In addition, identification algorithms for MIMO Wiener systems have been proposed, which can be grouped into those relying on special input signals [44, 45, 50, 51], in particular Gaussian excitation, and those presupposing invertibility of the static nonlinear output block [47, 48]. Overall, MIMO Wiener identification methods rely on specific mathematical properties of the model blocks, expert knowledge or special characteristics of the identification data set.

2.2 Koopman Theory

Koopman theory facilitates the representation of nonlinear dynamics in a linear framework [12, 16]. In essence, Koopman theory applies nonlinear coordinate transformation to lift the system states into an infinite-dimensional space, where the dynamics behave linearly. The Koopman operator acts on a function space $\mathcal{F}$, e.g., Hilbert space, of nonlinear observable functions $g \in \mathcal{F} : X \rightarrow \mathbb{C}$ of the states $\mathbf{x}(t) \in X \subseteq \mathbb{R}^{n_x}$. Here, the Koopman operator advances these nonlinear observables in time in a linear fashion.

Consider a nonlinear continuous-time autonomous system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}(t)$ is the differential state at time t , and $\mathbf{f} : X \rightarrow \mathbb{R}^{n_x}$ are the continuously differentiable nonlinear dynamics. The generator $\mathcal{L}_f : \mathcal{F} \rightarrow \mathcal{F}$, also referred to as Lie operator, belongs to the Koopman operator family and induces linear dynamics:

$$\frac{dg}{dt} = \mathcal{L}_f g, \quad (2)$$

where:

$$\mathcal{L}_f g(\mathbf{x}) \equiv \nabla_{\mathbf{x}} g(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}). \quad (3)$$

This means that the nonlinear dynamics become globally linear, but the lifted states, i.e., all possible nonlinear observables, are infinite dimensional. Since the Koopman operator is linear, it has eigenfunctions $\varphi_i : X \rightarrow \mathbb{C}$ with corresponding eigenvalues $\lambda_i \in \mathbb{C}$ evolving according to the characteristic eigendynamics:

$$\frac{d\varphi_i}{dt} = \lambda_i \varphi_i. \quad (4)$$

If an observable g_j can be represented in the coordinates spanned by these eigenfunctions, its dynamic evolution is fully described by the eigendynamics.

For the practical application of Koopman theory, a finite-dimensional invariant subspace Z is sought, spanned by a finite subset of observables or eigenfunctions. Combining a finite-dimensional matrix approximation $A : Z \rightarrow Z$ (projection) of the Lie operator on this subspace Z and the respective nonlinear coordinate lifting $\mathbf{T} : X \rightarrow Z$, the continuous-time Koopman model with linear dynamics becomes:

$$\frac{dz}{dt} = Az, \quad (5a)$$

$$\mathbf{z}(t) = \mathbf{T}(\mathbf{x}(t)), \quad (5b)$$

in the observables $\mathbf{z}(t) \in Z$. Similarly, the corresponding eigenfunctions $\varphi(\mathbf{x})$ satisfy:

$$\frac{d\varphi}{dt} = \Lambda \varphi. \quad (6)$$

We limit ourselves to systems with point spectrum [63] and assume A and Λ to be constant, thus linear time-invariant (LTI) dynamics. In practical implementations, a ‘‘Koopman canonical transform’’ [64] is often useful, where Eqs. (5) and (6) are formulated in terms of real-valued observables and eigenfunctions, respectively. We adopt this formulation hereinafter and assume that Λ has block-diagonal form.

For nonlinear systems exhibiting a single globally attracting equilibrium point, a finite-dimensional Koopman invariant subspace can (theoretically) always be constructed [20]. Moreover, if the original states are included in the subspace Z , then $\mathbf{x}$ can be reconstructed from $\mathbf{z}$ by linear recombination, i.e., $\mathbf{x} = C\mathbf{z}$. Conversely, for any system exhibiting multiple fixed points or more general attractors, a finite-dimensional invariant subspace including the original states $\mathbf{x}$ cannot exist [18]. In practice, determining a truly invariant low-dimensional subspace Z with associated coordinate transformation $\mathbf{T}(\mathbf{x})$ is impractical for most problems. Instead, local approximations to invariant subspaces of the Koopman operator are constructed while trying to avoid closure issues. These approximations may also be interpreted as a truncated Koopman mode decomposition [10]. Often, they can provide sufficiently accurate predictions of the nonlinear dynamics in short-term predictions on an application-relevant part of the state space.

The so-called dynamic mode decomposition (DMD) [9, 65] and extended DMD [10] are two data-driven model identification methods closely related to Koopman theory. While extended DMD essentially performs linear regression and modal decomposition on a dictionary of nonlinear observables $\mathbf{z} = \Theta(\mathbf{x})$ generated from time-series data $\mathbf{x}(t)$, the original DMD uses the more restrictive identity dictionary $\mathbf{z} = \mathbf{x}$, trying to approximate Eq. (5) or its discrete-time counterpart.

2.3 Deep-learning strategies for Koopman models

Several works have investigated deep learning to discover Koopman observables or EDMD dictionaries for finite-dimensional representations of autonomous dynamical systems [41–43, 66–69]. Given the unknown form of the observables $\mathbf{z}(t) = \mathbf{T}(\mathbf{x}(t))$, using artificial neural networks is expedient due to their universal approximation capabilities [66]. While some approaches include the original states in the learned transformation [66, 67], most works use autoencoder networks to allow for a nonlinear state reconstruction [41–43, 68, 69], see Fig. 2. Moreover, restricting the dimension of the learned Koopman subspace to a latent variable space, combines Koopman theory and model reduction [42, 43]. In this respect, Koopman deep-learning architectures are related to other data-driven reduction strategies based on latent dynamics and nonlinear projection through autoencoders, e.g., [14, 15, 70, 71]. However, Koopman methods distinguish from these methods by simultaneously generating nonlinear transformation and linear latent dynamics.

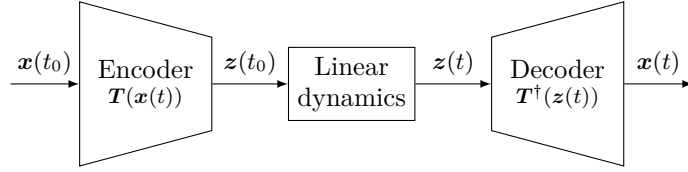


Fig. 2: Schematic illustration of deep-learning Koopman architectures for autonomous dynamical systems. Adapted from [42].

A deep-learning framework as depicted in Fig. 2 readily provides approximations of the nonlinear transformation $\mathbf{T}$ (encoder) along with the approximate inverse $\mathbf{T}^\dagger$ (decoder). While the autoencoders are typically trained on all data sampled from the relevant state space, the encoder mainly serves to provide transformed initial conditions for a simulation task, whereas the decoder recovers the original states along the dynamic trajectory. Many works use uniformly sampled data sets and prefer to identify the mathematically equivalent discrete-time linear dynamics to Eq. (5a):

$$\mathbf{z}_{k+1} = \bar{\mathbf{A}}\mathbf{z}_k, \quad (7)$$

where $k = 0, 1, \dots$ are the sampling instants. In the present literature, a simultaneous identification strategy for autoencoder and linear dynamics is achieved by summing loss functions for the nonlinear reconstruction error as well as linear or nonlinear prediction errors over multiple time steps, e.g., [42]. Regularization strategies, such as Tikhonov or ℓ_1 regularization, are frequently used in the training to prevent overfitting and stability issues, while closure issues of the subspace can be detected through validation. In addition, sparsity-promoting training for a block-diagonal structure of the linear dynamics or expert knowledge for presetting the structure of $\bar{\mathbf{A}}$ can foster the specific discovery of the eigendynamics, Eq. (6), and supports dynamical decoupling and uniqueness of the encoding and decoding [42, 69]. For example, Lusch et al. [42] combine deep autoencoders with a linear block-diagonal state-space model as depicted in Fig. 2.

2.4 Koopman theory for controlled systems

Classical Koopman theory was proposed for autonomous systems and is not directly extendable to causal nonlinear systems with external forcing of the general form:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}'(\mathbf{x}, \mathbf{u}). \quad (8)$$

Herein, $\mathbf{f}' : X \times U \rightarrow \mathbb{R}^{n_x}$ additionally depends on the inputs $\mathbf{u}(t) \in U \subseteq \mathbb{R}^{n_u}$.

Empirical extension of the linear Koopman representation of an autonomous system, Eq. (5), by external inputs can corrupt the linear dynamics [17, 18]. While efforts have been made to cast representations of dynamical systems with inputs into a Koopman-operable form [26, 28, 40, 72–74], no canonical form of Eq. (8) with Koopman dynamics explicit in the external inputs has been

presented in the literature yet. However, Surana [24] and Goswami & Paley [25] showed that a universal Koopman form exists for the continuous-time input-affine nonlinear systems class:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \sum_{i=1}^{n_u} \mathbf{h}_i(\mathbf{x}) u_i, \quad (9)$$

where $\mathbf{f}(\mathbf{x})$ and $\mathbf{h}_i(\mathbf{x})$ act on the state space. This class represents many physical systems and engineering problems. In addition, more general systems, Eq. (8), can be brought into this form when embedding additional input dynamics for $\mathbf{u}(t)$. Since our derivation of a Wiener-type Koopman model in Section 3 builds on the works of Surana [24], we discuss their derivation and follow-up works in detail below.

Assuming that all states $\mathbf{x}$ can be linearly reconstructed from a finite number of Koopman observables or eigenfunctions, i.e., $\mathbf{x} = C^\varphi \boldsymbol{\varphi}(\mathbf{x})$, Surana [24] demonstrated that the Koopman dynamics for systems of the form (9) are also input-affine:

$$\frac{d\boldsymbol{\varphi}}{dt} = \Lambda \boldsymbol{\varphi} + \sum_{i=1}^{n_u} \nabla_x \boldsymbol{\varphi}^T \mathbf{h}_i(\mathbf{x}) u_i \big|_{\mathbf{x}=C^\varphi \boldsymbol{\varphi}}. \quad (10)$$

where, noticeably, u_i are still the original inputs rather than transformed ones. Here, the eigenfunctions spanning the Koopman-invariant subspace take the form $\boldsymbol{\varphi}^c(\mathbf{x}, \mathbf{u}) = \boldsymbol{\varphi}(\mathbf{x})$, which are eigenfunctions associated with the autonomous dynamical part, $\mathbf{f}(\mathbf{x})$. Equation (10) underlines that a prerequisite for obtaining a Koopman-invariant subspace, i.e., an exact finite-order representation of the dynamics, is to find eigenfunction coordinates $\boldsymbol{\varphi}(\mathbf{x})$ such that $\nabla_x \boldsymbol{\varphi}(\mathbf{x})^T \mathbf{h}_i(\mathbf{x})$ lie in the span of $\boldsymbol{\varphi}(\mathbf{x})$ [24, 25]:

$$\nabla_x \boldsymbol{\varphi}(\mathbf{x})^T \mathbf{h}_i(\mathbf{x}) \equiv \sum_{j=1}^N \varphi_j(\mathbf{x}) \mathbf{v}_j^{(h_i)}, \quad (11)$$

where $\mathbf{v}_j^{(h_i)}$ are the vector-valued Koopman modes from a decomposition associated with $\mathbf{h}_i(\mathbf{x})$. In practice, this assumption can be hard to satisfy, and determining a closed representation may require a significantly higher number of Koopman eigenfunctions than needed for the autonomous counterpart, e.g., [17]. Finally, Eq. (10) can be brought into a form useful for control [24, 25]:

$$\frac{d\boldsymbol{\varphi}}{dt} = \Lambda \boldsymbol{\varphi} + \sum_{i=1}^{n_u} \Xi^{(i)} \boldsymbol{\varphi} u_i, \quad (12)$$

where the last term is bilinear in u_i and $\boldsymbol{\varphi}_j$, and Λ and $\Xi^{(i)}$ are matrices. Alternatively, the more intuitive Koopman model in terms of the nonlinear observables $\mathbf{z}$ reads:

$$\begin{aligned} \frac{d\mathbf{z}}{dt} &= A\mathbf{z} + \sum_{i=1}^{n_u} B^{(i)} \mathbf{z} u_i, \\ \mathbf{x} &= C\mathbf{z}, \\ \mathbf{z}_0 &= \mathbf{T}(\mathbf{x}_0). \end{aligned} \quad (13)$$

Herein, for the sake of clarity, we explicitly state the linear reconstruction, the initial condition, where $\mathbf{x}_0 = \mathbf{x}(0)$, and denote the nonlinear coordinate lifting by $\mathbf{T}(\mathbf{x}(t))$. This Koopman form employs a bilinear dynamical model and linear state reconstruction.

Surana [24] notices that the Koopman decomposition may require infinitely many observables for an arbitrarily accurate representation of the input-affine dynamics. Indeed, finite-order bilinear and linear models are known to be incapable of exhibiting chaotic behavior or output multiplicity [4, 18]. While the general limits of linear and bilinear models are known [4], a rigorous classification of systems that can be arbitrarily exactly approximated by means of a finite-order bilinear Koopman form is missing. However, in many cases, a finite-order Koopman model may represent a truncated

or approximate decomposition. Note further that the bilinearity of the Koopman model, Eq. (13), only holds for the continuous-time generator $\mathcal{L}_f$ but not generally in discrete time [37]. While an exact transformation between continuous-time and discrete-time bilinear models is possible for (pairwise) separable systems [75, 76], a discrete-time bilinear model may otherwise only be used for sufficiently fine sampling or weak nonlinearity.

According to Eq. (10), the bilinear dynamics reduce to linear dynamics, if there exist eigencoordinates satisfying [17]:

$$\nabla_x \boldsymbol{\varphi}(\mathbf{x})^T \mathbf{h}_i(\mathbf{x}) \equiv \text{const.} \quad (14)$$

If Condition (14) holds, we obtain a Koopman representation with linear dynamics:

$$\begin{aligned} \frac{d\mathbf{z}}{dt} &= A\mathbf{z} + B\mathbf{u}, \\ \mathbf{x} &= C\mathbf{z}, \\ \mathbf{z}_0 &= \mathbf{T}(\mathbf{x}_0). \end{aligned} \quad (15)$$

Note that Eq. (15) is closely related to the DMD with controls [77] approach.

Korda and Mezić [28] proposed to empirically restrict the Koopman model to a linear form, Eq. (15). Several works employed a linear Koopman approach in identification and control studies and observed a higher performance over conventional linear identification methods [28–35]. However, when significant nonlinearity is present, Eq. (15) was found to lead to prediction errors and tracking losses [37–39]. The works [37–39] demonstrated the higher predictive ability of bilinear models, Eq. (13), a result known from standard system identification literature [4]. Deep learning for Koopman models with inputs was applied in [31, 33, 35]. However, these works identified linear models, Eq. (15), without a nonlinear decoder.

3 Koopman control model with invertible coordinate transformation

In this section, we derive a Koopman model with Wiener-type block structure. In the derivation of the global bilinearization, Eq. (13), Surana [24] assumes that $\boldsymbol{\varphi}(\mathbf{x})$ is injective and that $\mathbf{x}$ can be fully and uniquely recovered by linear recombination of Koopman eigenfunctions. However, this assumption can become restrictive when aiming at low-dimensional, possibly linear, latent dynamical models [24, 40]. The identification of underlying dominant nonlinear patterns, as targeted by nonlinear model reduction, may be difficult or impossible when requiring linear state reconstruction, cf. also [14]. On the other hand, due to the exact convertibility of LTI dynamics with zeroth-order hold between discrete-time and continuous-time as well as the rich control and system theory, linear dynamics are practically favorable over bilinear dynamics.

We aim to unify the simplicity of linear dynamics, the accuracy of bilinear Koopman representations, and the model-reduction capabilities by nonlinear state reconstruction. To this end, we drop the linear state recovery assumption of [24] and assume instead that the coordinate transformations $\boldsymbol{\varphi}(\mathbf{x})$ and $\mathbf{T}(\mathbf{x})$ are continuously invertible, i.e., continuously bijective. Instead of Eq. (10), we formulate:

$$\frac{d\boldsymbol{\varphi}}{dt} = \Lambda \boldsymbol{\varphi} + \sum_{i=1}^{n_u} \nabla_x \boldsymbol{\varphi}^T \mathbf{h}_i(\mathbf{x}) u_i \big|_{\mathbf{x}=\boldsymbol{\varphi}^{-1}(\boldsymbol{\varphi})}. \quad (16)$$

Furthermore, the requirement described by Eq. (11) still applies. Analogue to Eq. (13), we then obtain a new Koopman model with nonlinear state reconstruction:

$$\begin{aligned} \frac{d\mathbf{z}}{dt} &= A\mathbf{z} + \sum_{i=1}^{n_u} B^{(i)} \mathbf{z} u_i, \\ \mathbf{x} &= \mathbf{T}^{-1}(\mathbf{z}), \\ \mathbf{z}_0 &= \mathbf{T}(\mathbf{x}_0). \end{aligned} \quad (17)$$

While Eq. (13) and Eq. (17) appear similar, the matrices and coordinate transformations may be fundamentally different. Under Condition (14), this model reduces to the linear-dynamics form:

$$\begin{aligned}\frac{dz}{dt} &= Az + Bu, \\ \mathbf{x} &= \mathbf{T}^{-1}(\mathbf{z}), \\ \mathbf{z}_0 &= \mathbf{T}(\mathbf{x}_0).\end{aligned}\tag{18}$$

In the following, we will use this model form with latent linear dynamics. The Koopman model Eq. (18) is a block-structured MIMO Wiener model, cf. Fig. 1. Notice that this model form extends to the MIMO Hammerstein-Wiener structure when adding an input nonlinearity block, i.e., when enriching the Koopman observables by nonlinear functions of the inputs.

The mathematical derivation of the Koopman model, Eq. (18), also supports the block-structure of the MIMO Wiener and parallel Wiener classes and provides a mathematical interpretation of these model architectures. On the other hand, the known limitations of Wiener modeling [4, 57] can be applied to analyze the model. In particular, Boyd and Chua [57] find that the Wiener-type structure is able to capture the behavior of systems with “fading memory” to arbitrary accuracy. In fading memory systems, the current state only depends on recent input signals but not on the distant past, which is related to a unique steady state [57]. In practice, the fading memory property may be easier to judge than the existence of a finite-dimensional subspace satisfying Condition (14).

4 Identification Strategy

In this section, we present a deep-learning identification procedure for the Wiener-type Koopman model type with external inputs that we derived in the previous section. The proposed method can also be regarded as a MIMO Wiener identification strategy. Our method is motivated by the Koopman deep-learning framework for autonomous systems by Lusch et al. [42]. However, the proposed procedure differs from the work of [42] in that we consider systems with external inputs, Fig. 3. The derivation given in Section 3 provides a mathematical foundation for this extension.

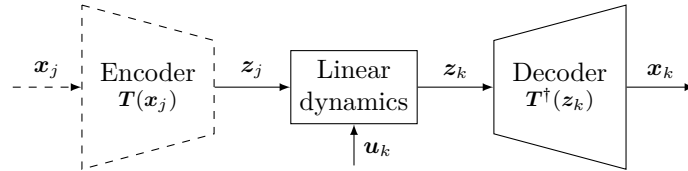


Fig. 3: Deep-learning strategy for MIMO Wiener-type models. Original states $\mathbf{x}$, transformed states $\mathbf{z}$, time instants $j, k = 0, 1, \dots$, and $k \geq j$.

Fig. 3 illustrates our NN block-structured model, which implements a discrete-time dynamical block sandwiched by a deep autoencoder, cf. Eq. (18). We use the autoencoder to construct the nonlinear state transformation, $\mathbf{T}(\cdot)$, along with its approximate inverse, $\mathbf{T}^\dagger(\cdot)$. The encoder is used in the identification procedure but is not necessarily part of a control model. The number and arrangement of neurons as well as the order n_z are model hyperparameters and affect the accuracy and degree of reduction. Similar to [42], we choose to identify discrete-time LTI dynamics because they do not need a computation of continuous time-derivatives in the training, and can be directly embedded as algebraic equations in an optimization problem. If desired, the transformation into continuous-time representation may be conducted after the training. Our procedure can be easily implemented within a deep-learning environment, e.g., Tensorflow or PyTorch. The identification does not rely on special input signals and enables a simultaneous identification of all model blocks. Importantly, although the nonlinear state transformation blocks are invertible, the strategy can be used to identify systems with input multiplicity in individual outputs as long as the steady state is unique, cf. Section 5.1.

To train the model, we use snapshots of step-response trajectories of all states with sufficiently fine sampling. Since we target data-driven model reduction, we assume that noise-free full-state information is available. We subdivide this data into shorter trajectories containing p snapshots and randomly group multiple trajectories in training batches. The trajectory length p and the batch size are training hyperparameters. Therefore, the provided identification data can be a collection of short samples. Finally, we split the batched data set into training and validation data and use an independent test data set. In the training, we encode the initial state snapshot $\mathbf{x}_0$ of a trajectory and decode single and multi-step predictions generated by the dynamical block. To this end, we sum loss terms that we compute for every p -samples trajectory in a training batch using the mean squared error (MSE):

$$L_1 = \frac{1}{p} \sum_{k=0}^p \|\mathbf{x}_k - \mathbf{T}^\dagger(\mathbf{T}(\mathbf{x}_k))\|_{\text{MSE}}, \quad (19a)$$

$$L_2 = \frac{1}{p-1} \sum_{k=0}^{p-1} \|\mathbf{x}_{k+1} - \mathbf{T}^\dagger(\mathbf{z}_{k+1}(\mathbf{x}_k))\|_{\text{MSE}}, \quad (19b)$$

$$L_3 = \frac{1}{p-1} \sum_{k=0}^{p-1} \|\mathbf{x}_{k+1} - \mathbf{T}^\dagger(\mathbf{z}_{k+1}(\mathbf{x}_0))\|_{\text{MSE}}, \quad (19c)$$

where single and multi-step predictions $\mathbf{z}_{k+1}(\mathbf{x}_j)$ are computed as:

$$\begin{aligned} \mathbf{z}_{k+1} &= A\mathbf{z}_k + B\mathbf{u}_k, \quad k = j, j+1, \dots \\ \mathbf{z}_j &= \mathbf{T}(\mathbf{x}_j). \end{aligned} \quad (20)$$

We aim at discovering the eigendynamics and thus restrict the matrix $A \in \mathbb{R}^{n_z \times n_z}$ to a block-diagonal structure. The term L_1 evaluates the nonlinear reconstruction accuracy of the autoencoder. The second term, L_2 , considers single-step predictions starting from various points along each trajectory. The third term, L_3 , evaluates multi-step predictions of up to p time steps to account for slow dynamics, all starting from the initial state $\mathbf{x}_0$ of a trajectory. We scale all losses by the mean squared sum of the data points, yielding normalized mean squared error (NMSE) loss. Additionally, we add ℓ_1 regularization on all trainable parameters. We train on a weighted linear combination of all loss terms. Details on the implementation are provided with the case studies.

5 Case Studies

We demonstrate our identification strategy in three case studies with identification problems from the literature. The first case study investigates a nonlinear system with input multiplicity [18]. Next, we consider the chemical reactor investigated by [30]. Finally, the third case study performs model identification of a distillation column [56].

In all case studies, we train Wiener-type Koopman models using our identification strategy from Section 4. We implement the identification framework with Python 3.9 using the Tensorflow 2.5 [78] API. We provide all code of our identification framework online¹. To identify suitable hyperparameters, we perform an a-priori systematic parameter search by training a variety of model structures including shallow and deep autoencoder structures with varying order n_z of the latent linear dynamics. For the sake of simplicity, we use a symmetric autoencoder design for the block-structured Koopman model, i.e., the arrangement and number of neurons in the hidden layers of the decoder is reversed as in the encoder. Further details on the training procedure and hyperparameters are provided in the Appendix.

We identify linear as well as bilinear dynamical models in all case studies as a benchmark and use the Koopman deep-learning framework described here. To this end, we remove the nonlinear hidden

¹See: <https://git.rwth-aachen.de/avt-svt/public/koopman-wiener>

layers of the decoder while keeping the nonlinear encoding, resulting in a linear recombination of the latent coordinates. These benchmark models are discrete-time versions of Eq. (13) and Eq. (15). The bilinear dynamics include an explicit term linear in the controls, as this was found to be beneficial in the training. We use sufficiently fine sampling to ensure the validity of discrete-time bilinear dynamics [37].

In all case studies, the training data are generated from numerical simulations. To use single-platform Python code throughout this work, all dynamic models are implemented in Pyomo.DAE [79] and simulated with CasADi [80]. We assume noise-free snapshot data. We evaluate the identified models on an independent test data set comprising dynamic step-response trajectories that were not seen in the training. In the test, we perform a complete forward simulation given only the initial state $\mathbf{x}_0$ and a sequence of inputs $\mathbf{u}_k$ over the full simulation horizon. We compare the predictions of the identified models to the true output profiles from the mechanistic model, and judge the performance by means of the NMSE over all scaled model outputs at all sampling instants. As a benchmark, we also identify MIMO Wiener models using the MATLAB 2021a System Identification Toolbox [81] using the same model architectures as described in the case studies, default settings and trained on the scaled data set. Since the MATLAB toolbox failed to generate low-order models of acceptable accuracy, we show these results only in the Appendix.

5.1 Case Study I: System with input multiplicity

We investigate a variant of the input-affine nonlinear system studied by Brunton et al. [18]:

$$\begin{aligned}\frac{dx_1}{dt} &= -0.1x_1 + u, \\ \frac{dx_2}{dt} &= x_1^2 - x_2, \\ y &= x_2.\end{aligned}\tag{21}$$

The system is stable and exhibits input multiplicity, since $y_s = 100u^2$. However, it has a unique steady state, $\mathbf{x}_s = [10u, 100u^2]^T$. As shown in [18], an analytic bilinear Koopman model, Eq. (13), can be constructed for this example when specifying $\mathbf{z} := [x_1, x_2, x_1^2]^T$. Here, we demonstrate that block-structured Wiener-type Koopman models with even fewer differential states $\mathbf{z}$ can be identified using our deep-learning framework.

5.1.1 Data sampling and training

The data used for the identification is generated in a dynamical simulation and comprises the trajectories of all states resulting from a sequence of 100 random uniform input steps with $u \in [-1, 1]$. Each input step has a duration of $t_{\text{step}} = 200$, and the time grid is sampled with $\Delta t = 1$. In the same way, we create a test data set comprising 20 random uniform input steps and $t_{\text{step}} = 100$. We identify full-state models, i.e., $\mathbf{y} = \mathbf{x}$. The hyperparameter search indicated that low-order latent dynamics, $n_z = 2$, can reproduce the dynamics sufficiently accurately. Moreover, shallow encoders/decoders with a single hidden layer of 20 neurons are a valid choice.

5.1.2 Results and Discussion

Figs. 4a and 4b compare the block-structured Koopman model, linear and bilinear models for $n_z = 2$. Table 1 summarizes the model hyperparameters and prediction errors for all models. All models reproduce the x_1 dynamics precisely, Fig. 4a. Furthermore, both the Wiener-type and bilinear model provide an accurate prediction of x_2 , whereas the linear model fails at reproducing the respective dynamics, Fig. 4b. As also evident from Table 1, increasing the linear order to $n_z = 10$ degrades the predictions rather than improving the model.

Since we target model reduction, we further reduce the latent space to $n_z = 1$ while keeping the autoencoder complexity, and investigate the performance of the identified models, Figs. 4c and 4d. Now, both linear and bilinear model fail to capture the system response, while the Wiener-type model continues to predict the dynamic and stationary behavior successfully. This indicates that the system dynamics allow for a joint encoding-decoding of both states, i.e., nonlinear projection, while linear reconstruction is not applicable. These results underline the strong model reduction capabilities of the Wiener structure.

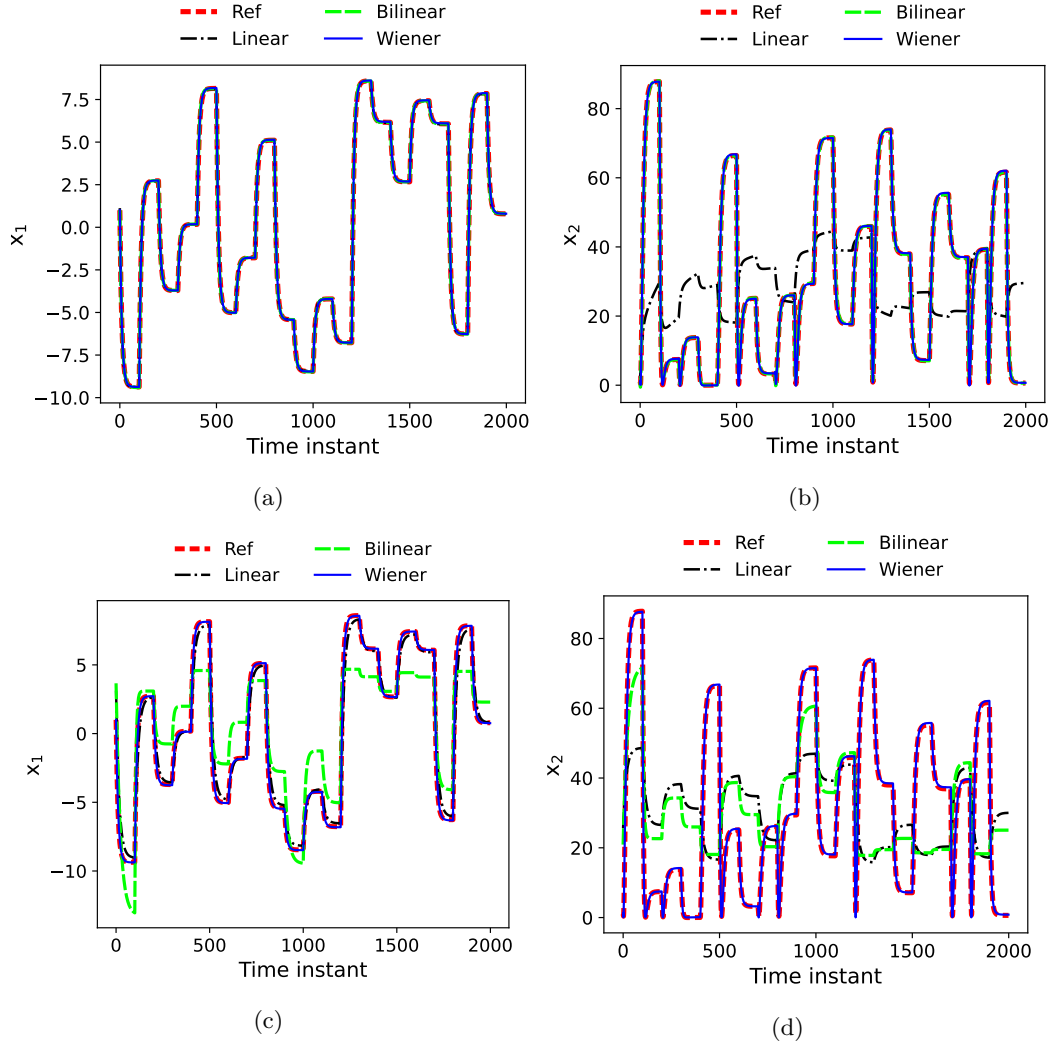


Fig. 4: Case study 1: Performance of the identified full-state models with $n_z = 2$ (a, b) and $n_z = 1$ (c, d) on the test set.

5.2 Case Study II: Chemical reactor

In this case study, we investigate the continuously operated chemical reactor for the exothermic reaction $A \rightarrow B$ presented in [30, 82]:

$$\begin{aligned} \frac{dc_A}{dt} &= \frac{F}{V}(c_{A,\text{in}} - c_A) - k_0 \exp\left(-\frac{E}{RT}\right) c_A^2, \\ \frac{dT}{dt} &= \frac{F}{V}(T_{\text{in}} - T) - \frac{k_0 \Delta H}{\rho c_p} \exp\left(-\frac{E}{RT}\right) c_A^2 + \frac{\dot{Q}}{\rho c_p V}. \end{aligned} \quad (22)$$

Herein, the molar concentration c_A and the reactor temperature T are the differential states $\mathbf{x}$. Similar to [30], we consider the heat duty $\dot{Q}$ as the varying system input u , and assume all other parameters as time-invariant. Further details on the model and parametric values can be found in [30, 82]. Most importantly, the system exhibits three steady states for $\dot{Q} = 0$. Narasingam & Kwon [30] identified a linear Koopman model with $n_z = 7$ for the high-yield operating region of product B, and presented a stabilizing MPC to control the process. However, the authors required expert knowledge to specify a dictionary of Koopman observables rather than employing a universal learning strategy. While linear, bilinear and Wiener-type models are known to be incapable of representing systems with steady-state multiplicity globally exactly [4], we aim at a local model for the high-yield operating region.

Table 1: Case study 1: Hyperparameters and prediction errors of the MIMO models: Subspace dimension n_z , no. of hidden neurons of encoder $n_{n,E}$ and decoder $n_{n,D}$, no. of original states $n_x = 2$.

Model	n_z	$n_{n,E}$	$n_{n,D}$	NMSE
Wiener	2	20	20	$2 \cdot 10^{-5}$
Bilinear	2	20	—	$2 \cdot 10^{-4}$
Linear	2	20	—	0.15
	10	50	—	0.77

5.2.1 Data sampling and training

We perform a dynamical simulation subject to a random sequence of 500 steps uniformly distributed over $\dot{Q} \in [-2000, 10\,000] \text{ kJ h}^{-1}$ at a sampling rate of 1 min^{-1} . Each input steps has a duration of $t_{\text{step}} = 4 \text{ h}$. Similarly, we create a test data set comprising 15 random uniform input steps with 2 h duration. We identify a model describing both states. A systematic parameter screening suggested that accurate predictions can be attained with extremely parsimonious models using low-dimensional static and dynamic blocks. Since we aim for low-order models rather than the most accurate predictor, we set $n_z = 1$ and use a shallow encoder/decoder each with a single hidden layer with 20 neurons.

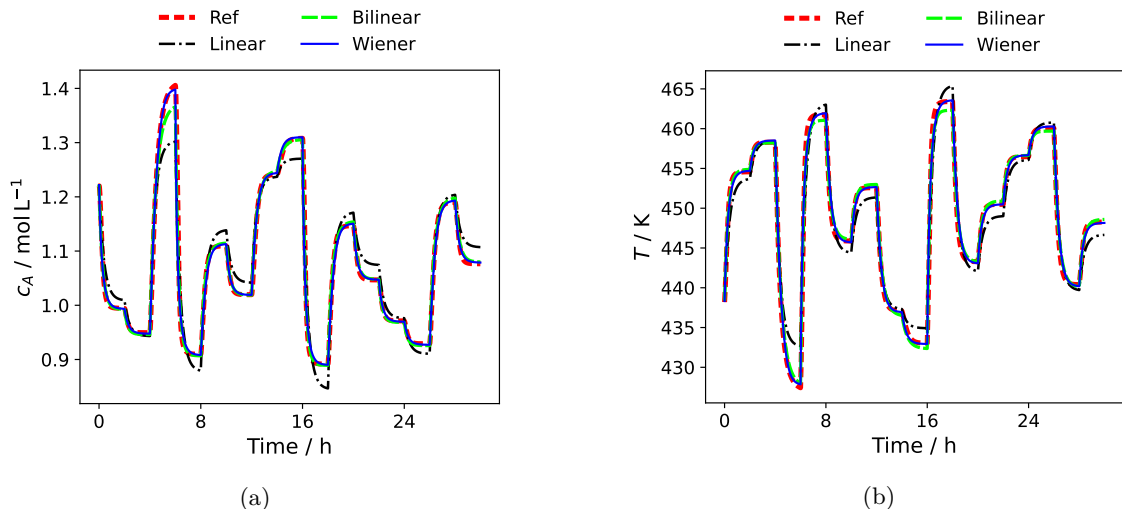


Fig. 5: Case study 2: Performance of the identified models on the test set. (a) Concentration of the reactant, (b) reactor temperature.

5.2.2 Results and Discussion

Fig. 5 displays the model performance on the test set. Table 2 collects the model parameters and the NMSE on the test set. We find that the Koopman deep-learning framework enables the identification of low-order models of all model types. However, despite the relatively narrow operating region, the prediction error of the linear model is noticeable and deviations occur as both over- and underestimation, Fig. 5. Besides steady-state offset of the linear model, there exists a pronounced error in the dynamical response. Increasing the order of the linear model to $n_z = 10$ slightly reduces the mismatch, but the prediction error is still present, see Table 2. Especially, the steady-state offset does not vanish when increasing n_z . In contrast, both Wiener and bilinear model successfully generate a precise forecast over the entire 80 h horizon. While the low-order bilinear model provides a slightly better dynamical response, the Wiener model is the only model with negligible steady-state offset over the entire operating range. We conclude that by the simple means of introducing the shallow NN decoder block with few neurons for nonlinear reconstruction,

a data-driven model with high accuracy is obtained. Noticeably, although its linear dynamical block is significantly more simple than the original nonlinear dynamics, the low-order Wiener-type model reproduces the dynamics almost exactly. The model could thus be used for highly precise tracking of quality requirements, allowing for smaller safety margins than a linear model.

Table 2: Case study 2: Hyperparameters and prediction errors of the MIMO models: Subspace dimension n_z , no. of hidden neurons of encoder $n_{n,E}$ and decoder $n_{n,D}$, no. of original states $n_x = 2$.

Model	n_z	$n_{n,E}$	$n_{n,D}$	NMSE
Wiener	1	20	20	0.002
Bilinear	1	20	—	0.002
Linear	1	20	—	0.009
	10	50	—	0.007

5.3 Case Study III: Distillation Column

High-purity distillation columns are one of the standard unit operations in chemical engineering. The transient concentration profiles inside a distillation column exhibit a nonlinear traveling wave character, i.e., an underlying coherent structure moving inside the column [83]. Block-structured SISO Wiener modeling of distillation columns has been the subject of several studies [56, 59, 84–86]. Further, the identification of linear models using log-scaled compositions is common practice in control engineering [87], which is regarded as another variant of Wiener modeling. In contrast, we aim at the identification of a low-order MIMO control model describing the evolution of all column states subject to multiple inputs.

In this case study, we investigate the high-purity methanol-propanol distillation column studied by Pearson & Pottmann [56]. The column has $N = 8$ equilibrium trays plus a total condenser and reboiler at the column top and bottom, respectively. The binary feed mixture enters the column on the fourth stage and is purified in the rectifying section above and the stripping section below. Details on the physical model and system parameters can be found in [56]. We employ our deep-learning framework for model reduction by identifying low-order MIMO dynamical models from sampled step-response state trajectories. The two input variables are the feed composition of the heavy-boiling component ($u_1 = x_f$) and the liquid molar reflux from the condenser into the column ($u_2 = L$). Similar to [56], we assume a constant vapor flow rate V , constant feed flow rate F , and perfect control of the condenser holdup through the distillate flow rate D in the model identification. Hence, we do not consider these variables as model inputs. The model has input-affine structure and $n_x = 10$ dynamical states.

5.3.1 Data sampling and training

We consider the high-purity operating region of the column, characterized by small molar fractions, $1 - x_{10}$, of the heavy-boiling component in the distillate stream. The data used for the identification is generated in a dynamical simulation, and comprises the trajectories of all state responses to a random sequence of 400 input steps. The respective input combinations (u_1, u_2) lie on a uniform grid within the bounded set $x_f \in [0.5, 0.6]$ and $L \in [0.0155, 0.0175]$ kmol/min. Each input step has 2 h duration at a data sampling rate of 1 min^{-1} , providing a balanced amount of transient and steady-state data. The simulation is initialized at the steady state corresponding to $\mathbf{u} = [0.6, 0.0175]^T$.

Since the molar fractions on the column trays vary over different orders of magnitude, we adopt the common practice [56] to train on log-transformed molar fractions of the light-boiler in the stripping section and heavy-boiler in the rectifying section. Due to the known nonlinear wave propagation phenomenon, we expect that a very low-order model is capable of providing satisfactory predictions. This was confirmed by the parameter screening. Hence, we identify MIMO models with a differential order $n_z = 2$ and autoencoder networks with a single hidden layer, Table 3.

Table 3: Case study 3: Hyperparameters and prediction error of the MIMO models: Subspace dimension n_z , no. of hidden neurons of encoder $n_{n,E}$ and decoder $n_{n,D}$, no. of original states $n_x = 10$.

Model	n_z	$n_{n,E}$	$n_{n,D}$	NMSE
Wiener	2	10	10	0.0003
Bilinear	2	20	—	0.0012
Linear	2	20	—	0.0025
	10	50	—	0.0013

5.3.2 Results and Discussion

Table 3 collects the values of the model hyperparameters along with the NMSE on the test set. Moreover, Fig. 6 visualizes the predictions by the model structures. The figure depicts the com-

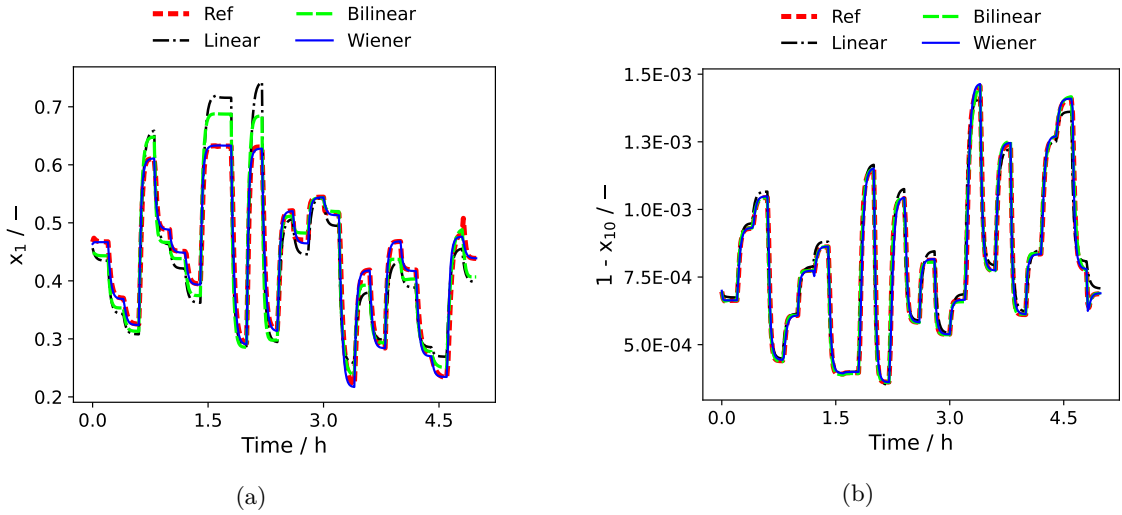


Fig. 6: Case study 3: Performance of the identified MIMO models on the test set. Predictions of the composition in the reboiler (a), and in the condenser (b).

position at the column bottom and top, as they characterize the exiting streams and are regarded as most critical with respect to prediction errors. Moreover, the prediction errors throughout the column are of the same magnitude. As visible from the figure, the Koopman deep-learning strategy enables the identification of accurate low-order control models. All models, including the linear Koopman model trained on log-transformed data, are able to capture the dominant trends of the distillation column, and especially of the distillate composition. However, the MIMO Wiener model still outperforms the linear and bilinear models. In particular, Fig. 6a displays transient and steady-state offset in the predictions of the bottom composition by the linear and bilinear models. Notice that increasing the complexity of the linear model results in a small improvement in the prediction quality but also a significantly higher number of trainable parameters, Table 3. Moreover, notice that bilinear models provided increasingly accurate predictions when increasing the dynamic order, i.e., model complexity. Conversely, even the low-order MIMO Wiener model yields highly precise long-term predictions for all column trays, which is also reflected in the NMSE performance index.

6 Conclusions and Outlook

We have presented a data-driven nonlinear model reduction and identification strategy for low-order modeling for control. To this end, we have employed a class of MIMO Wiener-type Koopman models with external inputs. While other works have focused on linear and bilinear Koopman models for input-affine systems, we have derived the Wiener-type Koopman variant here. Our derivation also demonstrates that the MIMO Wiener model structure is naturally related to Koopman theory. The Wiener structure is particularly suitable as a general-purpose model type for data-driven non-intrusive reduction, as it allows to identify dominant nonlinear dynamical patterns.

To obtain low-order MIMO control models, we have used a conceptually simple deep-learning strategy. Therein, we pair autoencoders with linear dynamics, resulting in a Wiener-type Koopman model to learn the nonlinear dynamical behavior. By restricting the latent space, i.e., linear dynamic block, to a low dimension, the deep-learning procedure is able to automatically generate strongly reduced models. Further, employing a machine learning approach reduces the required expert knowledge significantly compared to a manual selection of Koopman observables or the application of intrusive nonlinear model reduction methods. In contrast to other MIMO Wiener identification strategies in the literature, our framework uses state trajectory snapshots, implements a simultaneous identification of all model blocks and does not rely on special excitation signals.

We have discussed the advantages of the Wiener type over linear and bilinear Koopman models. The Wiener block-structure unifies the simplicity of linear dynamical models and the accuracy of bilinear forms, and provides strong model reduction capabilities. Due to the nonlinear reconstruction block, Wiener-type models can facilitate a more extensive reduction at higher accuracy than low-order bilinear models, although (high-order) bilinear models constitute a more universal model type. Moreover, employing an LTI block in a Wiener structure enables the application of standard methods for system analysis, allows for the exact transformation between discrete-time and continuous-time dynamics, and can be favorable in control and optimization procedures. While a substitution of the linear dynamical block of the Wiener structure by more universal descriptors, e.g., [15, 70], is generally possible, such extensions lead to a higher number of trainable parameters and their numerical treatment and analysis is more difficult. Moreover, our results suggest that using a MIMO Wiener model with linear dynamical block is already sufficient in many applications and yields a very minimalistic structure.

In three case studies, we have successfully demonstrated the identification and reduction capabilities of our deep-learning framework. Throughout all case studies, low-order Wiener-type models outperformed the linear and bilinear Koopman models in terms of accuracy and feasible degree of reduction. In addition, a standard identification toolbox failed to identify MIMO Wiener models for the systems and degree of reduction investigated here. These results demonstrate that, in contrast to the common practice of identifying multiple MISO models, a suitable identification and reduction method can provide a single and extremely low-order MIMO model describing all states.

We conclude that data-driven model reduction by means of applied Koopman theory provides an effective and accurate framework for low-order modeling of moderately nonlinear control systems. Therein, the MIMO Wiener structure constitutes a well suited and broadly applicable generic model form. Future work should investigate Koopman-based reduction frameworks for strongly nonlinear, e.g., chaotic, systems with controls. Moreover, our autoencoder framework could be extended to Koopman identification from noisy measurement data [19] and the identification in cases of missing state information [28, 88].

Acknowledgements We gratefully acknowledge the financial support of the Kopernikus project SynErgie 2 by the Federal Ministry of Education and Research (BMBF) and the project supervision by the project management organization Projektträger Jülich, as well as from the Helmholtz Association of German Research Centers as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE). The authors thank Johannes M. M. Faust and Pascal Schäfer for proofreading and valuable feedback.

References

- [1] W. Marquardt, “Nonlinear model reduction for optimization based control of transient chemical processes,” *AIChE Symposium Series*, 2002.
- [2] A. C. Antoulas, “An overview of approximation methods for large-scale dynamical systems,” *Annual Reviews in Control*, vol. 29, no. 2, pp. 181–190, 2005.
- [3] L. Ljung, “System identification,” *Wiley encyclopedia of electrical and electronics engineering*, pp. 1–19, 1999.
- [4] R. K. Pearson, “Selecting nonlinear model structures for computer control,” *Journal of Process Control*, vol. 13, no. 1, pp. 1–26, 2003.
- [5] M. Schoukens and K. Tiels, “Identification of block-oriented nonlinear systems starting from linear approximations: A survey,” *Automatica*, vol. 85, pp. 272–292, 2017.
- [6] U. Baur, P. Benner, and L. Feng, “Model order reduction for linear and nonlinear systems: a system-theoretic perspective,” *Archives of Computational Methods in Engineering*, vol. 21, no. 4, pp. 331–358, 2014.
- [7] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.
- [8] J. Schoukens and L. Ljung, “Nonlinear system identification: A user-oriented road map,” *IEEE control systems magazine*, vol. 39, no. 6, pp. 28–99, 2019.
- [9] P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *Journal of fluid mechanics*, vol. 656, pp. 5–28, 2010.
- [10] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, “A data-driven approximation of the koopman operator: Extending dynamic mode decomposition,” *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307–1346, 2015.
- [11] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé, “Data-driven model reduction and transfer operator approximation,” *Journal of Nonlinear Science*, vol. 28, no. 3, pp. 985–1010, 2018.
- [12] A. Mauroy, Y. Susuki, and I. Mezić, eds., *The Koopman Operator in Systems and Control*. Springer, 2020.
- [13] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, “Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems,” *Physica D: Nonlinear Phenomena*, vol. 406, p. 132401, 2020.
- [14] K. Lee and K. T. Carlberg, “Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders,” *Journal of Computational Physics*, vol. 404, p. 108973, 2020.
- [15] D. Masti and A. Bemporad, “Learning nonlinear state-space models using autoencoders,” *Automatica*, vol. 129, p. 109666, 2021.
- [16] I. Mezić, “Spectral properties of dynamical systems, model reduction and decompositions,” *Nonlinear Dynamics*, vol. 41, no. 1, pp. 309–325, 2005.
- [17] S. E. Otto and C. W. Rowley, “Koopman operators for estimation and control of dynamical systems,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 59–87, 2021.
- [18] S. Brunton, B. W. Brunton, J. L. Proctor, and J. N. Kutz, “Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control,” *PLOS ONE*, vol. 11, no. 2, p. e0150171, 2016.
- [19] A. Mauroy and J. Goncalves, “Linear identification of nonlinear systems: A lifting technique based on the Koopman operator,” *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6500–6505, 2016.
- [20] Y. Lan and I. Mezić, “Linearization in the large of nonlinear systems and Koopman operator spectrum,” *Physica D: Nonlinear Phenomena*, vol. 242, no. 1, pp. 42–53, 2013.
- [21] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, “Spectral analysis of nonlinear flows,” *Journal of fluid mechanics*, vol. 641, pp. 115–127, 2009.
- [22] C. Wehmeyer and F. Noé, “Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics,” *The Journal of chemical physics*, vol. 148, no. 24, p. 241703, 2018.
- [23] J. L. Proctor and P. A. Eckhoff, “Discovering dynamic patterns from infectious disease data using dynamic mode decomposition,” *International health*, vol. 7, no. 2, pp. 139–145, 2015.
- [24] A. Surana, “Koopman operator based observer synthesis for control-affine nonlinear systems,” *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6492–6499, 2016.
- [25] D. Goswami and D. A. Paley, “Global bilinearization and controllability of control-affine nonlinear systems: A Koopman spectral approach,” *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017.
- [26] J. L. Proctor, S. L. Brunton, and J. N. Kutz, “Generalizing Koopman theory to allow for inputs and control,” *SIAM Journal on Applied Dynamical Systems*, vol. 17, no. 1, pp. 909–930, 2018.
- [27] M. E. Villanueva, C. N. Jones, and B. Houska, “Towards global optimal control via Koopman lifts,” *Automatica*, vol. 132, p. 109610, 2021.
- [28] M. Korda and I. Mezić, “Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control,” *Automatica*, vol. 93, no. 1, pp. 149–160, 2018.
- [29] H. Arbabi, M. Korda, and I. Mezić, “A data-driven koopman model predictive control framework for nonlinear partial differential equations,” *2018 IEEE Conference on Decision and Control (CDC)*, pp. 6409–6414, 2018.

- [30] A. Narasingam and J. S.-I. Kwon, "Koopman Lyapunov-based model predictive control of nonlinear chemical process systems," *AIChE Journal*, vol. 65, no. 11, p. e16743, 2019.
- [31] I. Abraham and T. D. Murphey, "Active learning of dynamics for data-driven control using Koopman operators," *IEEE Transactions on Robotics*, vol. 35, no. 5, pp. 1071–1083, 2019.
- [32] C. Folkestad, D. Pastor, I. Mezic, R. Mohr, M. Fonoberova, and J. Burdick, "Extended dynamic mode decomposition with learned koopman eigenfunctions for prediction and control," *2020 Annual American Control Conference (ACC)*, 2020.
- [33] Y. Han, W. Hao, and U. Vaidya, "Deep Learning of Koopman Representation for Control," *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 1890–1895, 2020.
- [34] S. H. Son, H.-K. Choi, and J. S.-I. Kwon, "Application of offset-free Koopman-based model predictive control to a batch pulp digester," *AIChE Journal*, 2021.
- [35] Z. Ping, Z. Yin, X. Li, Y. Liu, and T. Yang, "Deep Koopman model predictive control for enhancing transient stability in power grids," *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, vol. 31, no. 6, pp. 1964–1978, 2021.
- [36] K. K. Lin and F. Lu, "Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism," *Journal of Computational Physics*, vol. 424, p. 109864, 2021.
- [37] S. Peitz, S. E. Otto, and C. W. Rowley, "Data-driven model predictive control using interpolated Koopman generators," *SIAM Journal on Applied Dynamical Systems*, vol. 19, no. 3, pp. 2162–2193, 2020.
- [38] D. Bruder, X. Fu, and R. Vasudevan, "Advantages of bilinear koopman realizations for the modeling and control of systems with unknown dynamics," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4369–4376, 2021.
- [39] C. Folkestad and J. W. Burdick, "Koopman NMPC: Koopman-based Learning and Nonlinear Model Predictive Control of Control-affine Systems," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [40] E. Kaiser, J. N. Kutz, and S. Brunton, "Data-driven discovery of Koopman eigenfunctions for control," *Machine Learning: Science and Technology*, 2021.
- [41] N. Takeishi, Y. Kawahara, and T. Yairi, "Learning Koopman invariant subspaces for dynamic mode decomposition," *Advances in Neural Information Processing Systems*, pp. 1131–1141, 2017.
- [42] B. Lusch, J. N. Kutz, and S. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nature Communications*, vol. 9, no. 1, pp. 1–10, 2018.
- [43] S. E. Otto and C. W. Rowley, "Linearly Recurrent Autoencoder Networks for Learning Dynamics," *SIAM Journal on Applied Dynamical Systems*, vol. 18, no. 1, pp. 558–593, 2019.
- [44] D. Westwick and M. Verhaegen, "Identifying MIMO Wiener systems using subspace model identification methods," *Signal Processing*, vol. 52, no. 2, pp. 235–258, 1996.
- [45] M. Lovera, T. Gustafsson, and M. Verhaegen, "Recursive subspace identification of linear and non-linear Wiener state-space models," *Automatica*, vol. 36, no. 11, pp. 1639–1650, 2000.
- [46] V. Arto, P. Hannu, and A. Halme, "Modeling of chromatographic separation process with Wiener-MLP representation," *Journal of Process Control*, vol. 11, no. 5, pp. 443–458, 2001.
- [47] J. C. Gómez and E. Baeyens, "Subspace-based identification algorithms for Hammerstein and Wiener models," *European Journal of Control*, vol. 11, no. 2, pp. 127–136, 2005.
- [48] A. Janczak, "Instrumental variables approach to identification of a class of MIMO Wiener systems," *Nonlinear Dynamics*, vol. 48, no. 3, pp. 275–284, 2007.
- [49] Y.-L. Hsu and J.-S. Wang, "A Wiener-type recurrent neural network and its control strategy for nonlinear dynamic applications," *Journal of Process Control*, vol. 19, no. 6, pp. 942–953, 2009.
- [50] J. Schoukens and Y. Rolain, "Parametric MIMO parallel Wiener identification," *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 2011.
- [51] B.-Q. Mu and H.-F. Chen, "Recursive identification of MIMO Wiener systems," *IEEE Transaction on automatic control*, vol. 58, no. 3, pp. 802–808, 2012.
- [52] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern Koopman theory for dynamical systems," *arXiv preprint arXiv:2102.12086*, 2021.
- [53] F. Giri and E.-W. Bai, *Block-oriented nonlinear system identification*. Springer, 2010.
- [54] S. J. Norquay, A. Palazoglu, and J. Romagnoli, "Model predictive control based on Wiener models," *Chemical Engineering Science*, vol. 53, no. 1, pp. 75–84, 1998.
- [55] M. Schoukens and Y. Rolain, "Parametric identification of parallel Wiener systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 10, pp. 2825–2832, 2012.
- [56] R. K. Pearson and M. Pottmann, "Gray-box identification of block-oriented nonlinear models," *Journal of Process Control*, vol. 10, no. 4, pp. 301–315, 2000.
- [57] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Transactions on circuits and systems*, vol. 32, no. 11, pp. 1150–1161, 1985.

- [58] H. Al-Duwaish, M. N. Karim, and V. Chandrasekar, "Use of multilayer feedforward neural networks in identification and control of Wiener model," *IEE Proceedings-Control Theory and Applications*, vol. 143, no. 3, pp. 255–258, 1996.
- [59] G. Schram, M. Verhaegen, and A. J. Krijgsman, "System Identification with Orthogonal Basis Functions and Neural Networks," *IFAC Proceedings Volumes*, vol. 29, no. 1, pp. 4150–4155, 1996.
- [60] P. Aadaleesan, N. Miglan, R. Sharma, and P. Saha, "Nonlinear system identification using Wiener type Laguerre–Wavelet network model," *Chemical Engineering Science*, vol. 63, no. 15, pp. 3932–3941, 2008.
- [61] A. Janczak, "Comparison of four gradient-learning algorithms for neural network Wiener models," *International Journal of Systems Science*, vol. 34, no. 1, pp. 21–35, 2003.
- [62] J. Peng, R. Dubay, J. M. Hernandez, and M. Abu-Ayyad, "A Wiener neural network-based identification and adaptive generalized predictive control for nonlinear SISO systems," *Industrial & engineering chemistry research*, vol. 50, no. 12, pp. 7388–7397, 2011.
- [63] I. Mezić, "Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry," *Journal of Nonlinear Science*, vol. 30, no. 5, pp. 2091–2145, 2020.
- [64] A. Surana and A. Banaszuk, "Linear observer synthesis for nonlinear systems using Koopman operator framework," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 716–723, 2016.
- [65] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *Journal of Computational Dynamics*, pp. 391–421, 2014.
- [66] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis, "Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 10, p. 103111, 2017.
- [67] E. Yeung, S. Kundu, and N. Hodas, "Learning deep neural network representations for Koopman operators of nonlinear dynamical systems," *2019 American Control Conference (ACC)*, 2019.
- [68] N. B. Erichson, M. Muehlebach, and M. W. Mahoney, "Physics-informed Autoencoders for Lyapunov-stable Fluid Flow Prediction," *arXiv preprint arXiv:1905.10866*, 2019.
- [69] S. Pan and K. Duraisamy, "Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability," *SIAM Journal on Applied Dynamical Systems*, vol. 19, no. 1, pp. 480–509, 2020.
- [70] D. Gedon, N. Wahlström, T. B. Schön, and L. Ljung, "Deep State Space Models for Nonlinear System Identification," *IFAC-PapersOnLine*, no. 54, pp. 481–486, 2021.
- [71] C. Tsay and M. Baldea, "Integrating production scheduling and process control using latent variable dynamic models," *Control Engineering Practice*, vol. 94, p. 104201, 2020.
- [72] M. O. Williams, M. S. Hemati, S. T. Dawson, I. G. Kevrekidis, and C. W. Rowley, "Extending Data-Driven Koopman Analysis to Actuated Systems," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 704–709, 2016.
- [73] S. Peitz and S. Klus, "Koopman operator-based model reduction for switched-system control of PDEs," *Automatica*, vol. 106, pp. 184–191, 2019.
- [74] S. Klus, F. Nüske, S. Peitz, J.-H. Niemann, C. Clementi, and C. Schütte, "Data-driven approximation of the Koopman generator: Model reduction, system identification, and control," *Physica D: Nonlinear Phenomena*, vol. 406, p. 132416, 2020.
- [75] A. Dunoyer, L. Balmer, K. J. Burnham, and D. J. James, "On the discretization of single-input single-output bilinear systems," *International journal of control*, vol. 68, no. 2, pp. 361–372, 1997.
- [76] P. Rumschinski, D. S. Laila, and R. Findeisen, "Discrete-to-continuous dynamics reconstruction for bilinear systems," *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 2012.
- [77] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.
- [78] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and others, "Tensorflow: A system for large-scale machine learning," *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016.
- [79] Bethany Nicholson, John D. Sirola, Jean-Paul Watson, Victor M. Zavala, and Lorenz T. Biegler, "pyomo.dae : a modeling and automatic discretization framework for optimization with differential and algebraic equations," *Mathematical Programming Computation*, vol. 10, no. 2, pp. 187–223, 2018.
- [80] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "CasADi: a software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 1–36, 2019.
- [81] The MathWorks Inc, "MATLAB: www.mathworks.com/products/matlab."
- [82] A. Alanqar, M. Ellis, and P. D. Christofides, "Economic model predictive control of nonlinear process systems using empirical models," *AIChE Journal*, vol. 61, no. 3, pp. 816–830, 2015.
- [83] W. Marquardt, "Traveling waves in chemical process," *International chemical engineering*, vol. 30, no. 4, pp. 585–606, 1990.
- [84] Y. Zhu, "Distillation column identification for control using Wiener model," *Proceedings of the 1999 American Control Conference*, 1999.

- [85] H. H. Bloemen, C. T. Chou, T. J. van den Boom, V. Verdult, M. Verhaegen, and T. C. Backx, “Wiener model identification and predictive control for dual composition control of a distillation column,” *Journal of Process Control*, vol. 11, no. 6, pp. 601–620, 2001.
- [86] S. J. Norquay, A. Palazoglu, and J. A. Romagnoli, “Application of Wiener model predictive control (WMPC) to an industrial C2-splitter,” *Journal of Process Control*, vol. 9, no. 6, pp. 461–473, 1999.
- [87] S. Skogestad and M. Morari, “Understanding the dynamic behavior of distillation columns,” *Industrial & engineering chemistry research*, vol. 27, no. 10, pp. 1848–1862, 1988.
- [88] M. Kamb, E. Kaiser, S. L. Brunton, and J. N. Kutz, “Time-delay observables for Koopman: Theory and applications,” *SIAM Journal on Applied Dynamical Systems*, vol. 19, no. 2, pp. 886–917, 2020.
- [89] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

Appendix

Additional information on the training

We formulate the training problem according to Section 4. Since we do not expect oscillatory open-loop behavior of the systems investigated, we restrict the system matrix A , Eq. (20), to a purely diagonal structure. We use the ELU activation function and a linear output layer for the autoencoder networks. All weights are initialized employing variance scaling with untruncated normal distribution, and all biases with zeros. We scale the training data from zero to one individually for each differential state. We specify the weights of the loss terms such that the three losses L_1 , L_2 and L_3 are of the same order of magnitude in the later epochs. The models are trained using batches of 32 trajectories, and 10,000 epochs to prevent premature termination. We randomly divide the training data set into 80 % training data, 20 % validation data. We execute all computations on a desktop computer with CPU at 3.0 GHz. We do not pre-train the encoder as we find no beneficial effect on the training result. We employ the Adam [89] training algorithm. The model parameters with the smallest validation loss are saved and used afterwards. Subsequently, we restore the model parameters from the epoch with the smallest validation loss. To reduce stochastic effects, we train all structures multiple times using different seeds. Table 4 collects the training hyperparameters. Training on trajectories of length $s = 50$ was found to yield good results, whereas shorter trajectories occasionally caused offset in both transients and stationary points.

Table 4: Hyperparameters of the training problem.

Parameter	Symbol	Value
Weight reconstruction	ω_1	0.1
Weight single-step prediction	ω_2	1.0
Weight multi-step prediction	ω_3	1.0
Weight ℓ_1 regularization	ω_r	1e-9
Learning rate		0.001
Trajectory length	p	50
Batch size		32
No. epochs		10 000

Comparison to MATLAB System Identification Toolbox

We compare the performance of the MIMO Wiener models identified with our deep-learning framework to MIMO Wiener models generated by the MATLAB 2021a System Identification Toolbox [81]. To this end, we use similar model hyperparameters in both frameworks, Table 5. Further, we use a piecewise-linear nonlinear block with MATLAB, corresponding to ReLU rather than ELU activation since the latter is not provided. We use the default settings of the MATLAB toolbox. However, we increase the number of major iterations of the MATLAB algorithm to 100 and use $\lambda = 0.001$ regularization, as we find that these adjustments improve the models considerably. Both identification frameworks receive the same scaled training data sets. Notice that the MATLAB toolbox jointly identifies a set of MISO models used as MIMO model, where n_z and n_n parameterize each MISO model. Hence, the MATLAB models actually involve n_y times as many parameters and equations as our Wiener-type Koopman models. For case study 3, we identify a MIMO model only for the target states x_1 and x_{10} with the MATLAB toolbox, as the generated full-state MIMO Wiener model performs significantly worse (NMSE = 0.1). The comparison is visualized in Fig. 7. In all case studies, the NMSE of the MIMO Wiener models generated with MATLAB is at least one order of magnitude greater. Moreover, in case studies 1 and 2, the obtained MATLAB models completely fail to predict one of the states. The results also underpin that a single MIMO model identified with a Koopman deep-learning approach can provide a more accurate prediction with fewer trainable parameters than a set of MISO models.

Table 5: Hyperparameters and prediction error of the MIMO Wiener models: n_z is subspace dimension, n_n is no. of hidden neurons of single-hidden-layer encoder/decoder or no. of linear segments.

MIMO Wiener model	n_z	n_n	NMSE	
			MATLAB	This work
Case Study 1	2	20	0.2	$3 \cdot 10^{-5}$
Case Study 2	1	20	0.025	0.0019
Case Study 3	2	10	0.0024	0.0003

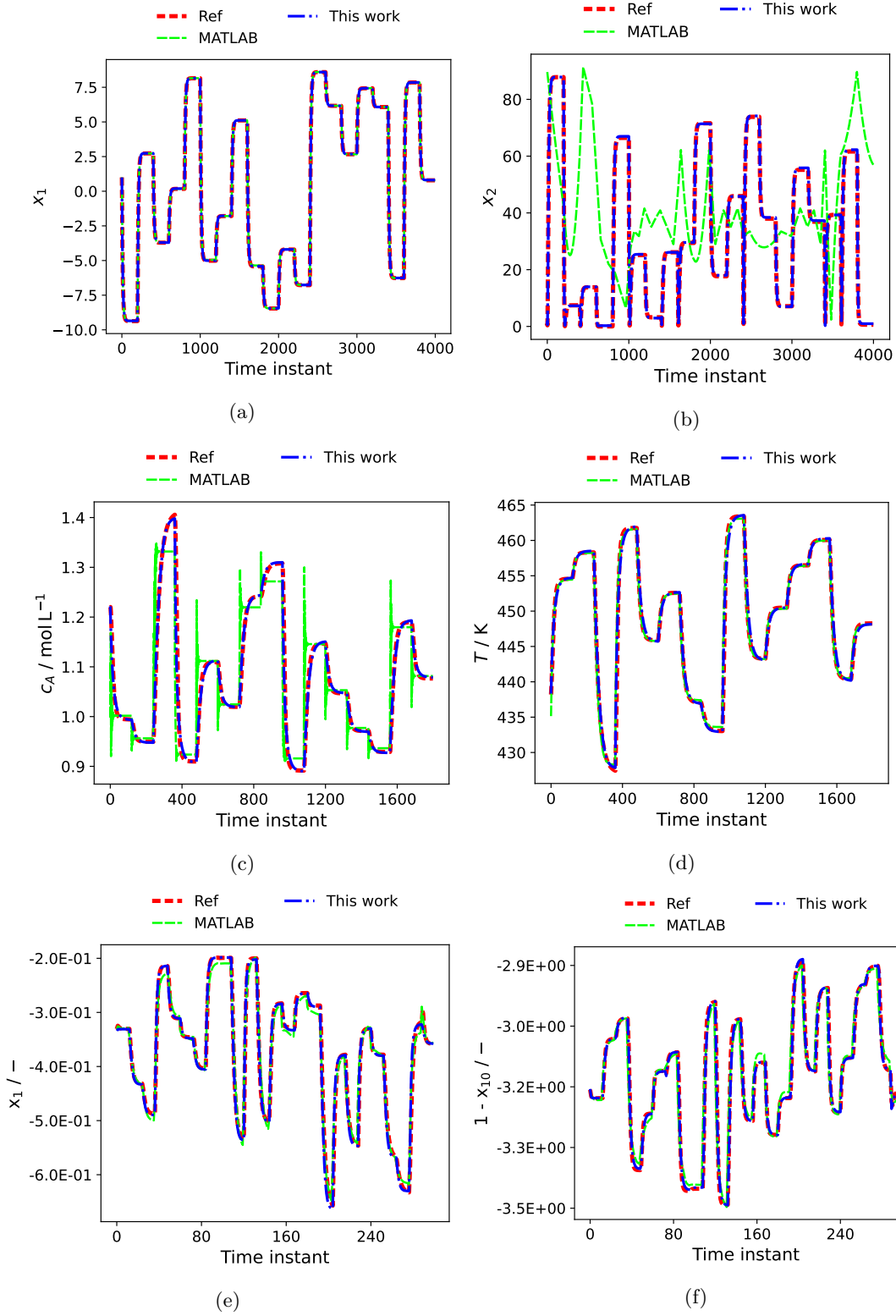


Fig. 7: Comparison of MIMO Wiener models from MATLAB Toolbox and our deep-learning framework. (a) + (b): Case study 1. (c) + (d): Case study 2. (e) + (f): Case study 3.