PAPER • OPEN ACCESS

Extending the extended dynamic mode decomposition with latent observables: the latent EDMD framework

To cite this article: Said Ouala *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 025018

View the article online for updates and enhancements.

MACHINE
LEARNING
Science and Technology

PAPER

OPEN ACCESS

Extending the extended dynamic mode decomposition with latent observables: the latent EDMD framework

Said Ouala[1,*] , Bertrand Chapron[2], Fabrice Collard[3], Lucile Gaultier[3] and Ronan Fablet[1]

[1] IMT Atlantique; Lab-STICC, 29200 Brest, France
[2] Ifremer, LOPS, 29200 Brest, France
[3] ODL, 29200 Brest, France
* Author to whom any correspondence should be addressed.

E-mail: said.ouala@imt-atlantique.fr

Abstract

Bernard O Koopman proposed an alternative view of dynamical systems based on linear operator theory, in which the time evolution of a dynamical system is analogous to the linear propagation of an infinite-dimensional vector of observables. In the last few years, several works have shown that finite-dimensional approximations of this operator can be extremely useful for several applications, such as prediction, control, and data assimilation. In particular, a Koopman representation of a dynamical system with a finite number of dimensions will avoid all the problems caused by nonlinearity in classical state-space models. In this work, the identification of finite-dimensional approximations of the Koopman operator and its associated observables is expressed through the inversion of an unknown augmented linear dynamical system. The proposed framework can be regarded as an extended dynamic mode decomposition that uses a collection of latent observables. The use of a latent dictionary applies to a large class of dynamical regimes, and it provides new means for deriving appropriate finite-dimensional linear approximations to high-dimensional nonlinear systems.

1. Introduction

Koopman operator theory (Koopman 1931) states that any nonlinear dynamical system can be lifted by a time-invariant nonlinear representation into a space where the time evolution of the system can be described by linear methods. In the last years, a combination of theoretical (Mezić 2005, 2013, 2015, Budišić *et al* 2012) and numerical (Schmid 2010, Williams *et al* 2015) efforts, as well as the increasing availability of data, promoted this formalism into one of the leading data-driven identification techniques. Modern Koopman theory aims at finding, from data, some special measurements of the state space that provide global linear models of non-linear, and potentially high dimensional dynamical systems. These representations may, however, be hard to obtain on complex systems. In practice, data-driven approximations such as the extended dynamic mode decomposition, (EDMD) (Williams *et al* 2015), assume some finite-dimensional basis of functions that span an (approximately) invariant Koopman subspace.

In this work, we wish to further expand the EDMD methodology to account for latent observables of the state variables. Specifically, we assume the finite-dimensional Koopman subspace to possibly be spanned by a known dictionary plus a collection of latent functions that form the latent dictionary. The evaluation of these latent functions is performed on a given realization of the observations. We discuss two methods for this proposed latent EDMD (LEDMD) framework, which can be used to derive Koopman approximations for both deterministic and stochastic dynamical systems. The latter relies on a state space model formulation with Gaussian uncertainties.

2. Deterministic dynamical systems and Koopman operator theory

2.1. Background on Koopman operator theory for deterministic systems

Let us assume a continuous, autonomous s-dimensional time-varying ODE, governing the state variable $\mathbf{z}_t$. This dynamical system also generates measurements $\mathbf{x}_t \in \mathbb{R}^n$ in the following state space model:

$$\dot{\mathbf{z}}_t = f(\mathbf{z}_t) \tag{1a}$$

$$\mathbf{x}_t = \mathcal{H}(\mathbf{z}_t). \tag{1b}$$

When considering the dynamical equation (1a) and given an initial condition $\mathbf{z}_{t_0}$, the solution of this equation for an interval $t \in [t_0, t_f]$ can be written as

$$\Phi_t(\mathbf{z}_{t_0}) = \mathbf{z}_{t_0} + \int_{t_0}^{t} f(\mathbf{z}_w)\mathrm{d}w \tag{2}$$

where $\Phi_t(\mathbf{z}_{t_0}) \in L$ with $L \subset \mathbb{R}^s$. We may also define a discrete state space model as follows:

$$\mathbf{z}_{k+1} = \mathcal{T}(\mathbf{z}_k) \tag{3a}$$

$$\mathbf{x}_k = \mathcal{H}(\mathbf{z}_k) \tag{3b}$$

where $\mathcal{T}$ can be the application of a given integration scheme to (1a).

Koopman (1931) introduced a new operator-based formalism, where the evolution of a dynamical system can be determined by following a set of observables of the state variable $\mathbf{z}_t$. Koopman proved that when considering an infinite-dimensional Hilbert space of observables $\mathcal{F}$, the time evolution of the dynamics is invariant to a linear Koopman operator. Formally, let $g \in \mathcal{F} : L \longrightarrow \mathbb{C}$ be a complex-valued observable of the dynamical system (3a). The collection of all these observables forms a linear vector space, on which the Koopman operator $\mathcal{K}$ is defined as follows:

$$\mathcal{K}g(\mathbf{z}_{t_k}) = g \circ \mathcal{T}(\mathbf{z}_{t_k}) \tag{4}$$

where $\circ$ is the composition operator. For time-continuous dynamical systems, we define a one-parameter semi-group of Koopman operators $\{\mathcal{K}^t\}_{t>0}$ as follows:

$$\mathcal{K}^t g(\mathbf{z}_{t_0}) = g \circ \Phi_t(\mathbf{z}_{t_0}). \tag{5}$$

We also denote by $\mathcal{A} = \lim_{t \to 0} \frac{\mathcal{K}^t - I}{t}$ the infinitesimal generator of the Koopman semi-group, which satisfies, assuming uniform continuity on a suitable banash space (Engel *et al* 2000):

$$\mathcal{K}^t = e^{\mathcal{A}t}. \tag{6a}$$

The Koopman operator $\mathcal{K}$ is a linear operator that advances observables in time. From a modeling perspective, going from an ODE to a Koopman operator formulation can be seen to trade the nonlinear complexity of the dynamical operator f for a linear operator representation based on an infinite-dimensional and non-linear set of observables.

2.2. Data-driven approximation of Koopman operator

Finding a data-driven approximation of Koopman representations consists in the definition, from a sequence of $N + 1$ measurements $\{\mathbf{x}_k\}_{k=1}^{N+1}$, of a finite-dimensional collection of observables, that can be propagated linearly in time. It is important to note that in this work, the measurements $\mathbf{x}_k$ are distinguished from the states $\mathbf{z}_k$. Indeed, in real applications, we are not guaranteed to observe the full state vector. A data-driven identification of Koopman representations thus brings another layer of complexity: the finite-dimensional set of observables is defined as a function of the measurements $\mathbf{x}_k$ and not from the state $\mathbf{z}_k$. In this situation, we need to ensue the projection (1b) does not impact the data-driven Koopman representation.

The literature on data-driven approximation of the Koopman operator mainly follows two paths. A first path is written in the language of the dynamic mode decomposition (DMD) (Schmid 2010). The DMD was first introduced to find low rank spatio-temporal coherent structures of complex dynamics. In the language

of Koopman, the DMD computes a Koopman approximation when considering the observables as the measurements $\mathbf{x}_k$ of the state space (Rowley *et al* 2009). The EDMD (Williams *et al* 2015) was then introduced to generalize the DMD algorithm to some (non-linear) functions of the measurements. Numerous basis of functions were explored in the literature, ranging from polynomial representations (Brunton *et al* 2016a) up to multi-layer perceptron (Takeishi *et al* 2017a). DMD-type algorithms were also shown to converge to a Galerkin approximation of the Koopman operator (Williams *et al* 2015). The combination of these theoretical results as well as the simplicity of the DMD algorithm motivated several developments of the method. For instance, the piDMD method (Baddoo *et al* 2023) was developed to enforce some known symmetries on the DMD approximation, and the mpEDMD (Colbrook 2022) guarantees that the EDMD approximation is measure preserving. Furthermore, in order to avoid the computation of spurious pairs of eigenvalues/eigenvectors, the ResDMD method (Colbrook *et al* 2023) was developed to assess and validate the accuracy of the pairs of eigenvalues/eigenvectors outputted by a DMD procedure. Several works also aimed at approximating the Koopman generator (Mauroy and Goncalves 2016, 2019, Klus *et al* 2020) using DMD-type algorithms, such methods can include sparsity priors on the infinitesimal generator even when the corresponding Koopman operator is not.

However, this technique can be subject to closure issues, especially for complex systems with no prior knowledge about the dynamics (and their non-linearities). The measurement noise may also have a strong influence on EDMD estimates. A number of works then proposed efficient estimates of the DMD output under certain noisy conditions. Combining the DMD with the Kalman filter (Jiang and Liu 2022), methods can both provide an efficient separation of the signal and noise, and an approximation of the stochastic Koopman operator.

A second path explores deep learning models to identify the necessary non-linear transformations of the measurements that may lead to suitable Koopman observables (Lusch *et al* 2018, Yeung *et al* 2019, Azencot *et al* 2020, Rice *et al* 2020).

Further note, considering partial observations of the state space variable, delay embedding coordinates offer a simple class of observables that can unfold (under some conditions on the parameters of the delay embedding) the structure of the underlying dynamics. These delay embedding observables were shown to be extremely efficient in the linearization of periodic and quasi-periodic dynamical systems (Arbabi and Mezic 2017). Their exploitation was also demonstrated in the decomposition of chaotic dynamics (Brunton *et al* 2017), further considering an additional forcing term. These types of representations have led to significant advances in the field, including convergence results under asymptotic regimes (Arbabi and Mezic 2017, Zhen *et al* 2022). However, in practice, the definition of their parameterization, especially the delay embedding (Kamb *et al* 2020) as well as their exploitation in computing dissipative eigenvalues of the Koopman operator (Arbabi and Mezic 2017) remains an active research topic.

3. Inverse problem formulation

We begin by outlining the EDMD method (Williams *et al* 2015) before introducing the proposed LEDMD framework.

3.1. EDMD
The main idea behind the EDMD is to estimate a finite-dimensional approximation of the Koopman operator $\mathcal{K}$ given a dictionary of functions and a dataset of snapshot pairs. Formally, we start by choosing a dictionary:

$$\mathcal{D}_M = \{\psi_1, \psi_2, \ldots, \psi_M\} \tag{7}$$

where $\psi_i \in \mathcal{F} : L \longrightarrow \mathbb{C}$ for $i = 1, 2, \ldots, M$. We also define the vector valued function $\Psi_M : L \longrightarrow \mathbb{C}^{M \times 1}$:

$$\Psi_M(\mathbf{z}_t) = [\psi_1(\mathbf{z}_t)\ \psi_2(\mathbf{z}_t) \ldots\ \psi_M(\mathbf{z}_t)]^{\mathrm{T}}. \tag{8}$$

The span of $\mathcal{D}_M$ is a subspace of $\mathcal{F}$, and can be written as follows:

$$\mathcal{F}_{\mathcal{D}_M} = \mathrm{Span}\{\psi_1, \psi_2, \ldots, \psi_M\} = \{\mathbf{a}_M^{\mathrm{T}} \Psi_M : \mathbf{a}_M \in \mathbb{C}^{M \times 1}\} \subset \mathcal{F}. \tag{9}$$

By definition, we can write any function $\phi \in \mathcal{F}_{\mathcal{D}_M}$ as:

$$\phi = \sum_{i=1}^{i=M} a_i \psi_i = \mathbf{a}_M^T \Psi_M. \tag{10}$$

If we apply the Koopman operator on the observable ϕ we have:

$$\mathcal{K}\phi = \mathbf{a}_M^T \mathcal{K}\Psi_M = \mathbf{a}_M^T \Psi_M \circ \mathcal{T}. \tag{11}$$

If we assume that the subspace spanned by $\mathcal{D}_M$ is invariant under the action of the Koopman operator we can write:

$$\mathcal{K}(\mathcal{F}_{\mathcal{D}_M}) = \mathcal{F}_{\mathcal{D}_M} \tag{12}$$

then $\mathcal{K}\phi = \mathbf{b}_M^T \Psi_M$ for some $\mathbf{b}_M \in \mathbb{C}^{M \times 1}$. Under this assumption, we can define an exact finite-dimensional Koopman operator $\mathbf{K}_M \in \mathbb{R}^{M \times M}$ that verifies $\mathbf{b}_M^T = \mathbf{a}_M^T \mathbf{K}_M$. The action of this finite-dimensional Koopman operator is defined as:

$$\mathcal{K}\phi = \mathbf{a}_M^T \mathcal{K}\Psi_M = \mathbf{a}_M^T \mathbf{K}_M \Psi_M = \mathbf{a}_M^T \Psi_M \circ \mathcal{T}. \tag{13}$$

In practice, the subspace spanned by $\mathcal{D}_M$ is not invariant to the Koopman operator. It implies that the action of the Koopman operator on some element of $\mathcal{F}_{\mathcal{D}_M}$ will not lie exactly on $\mathcal{F}_{\mathcal{D}_M}$ i.e. $\mathcal{K}\phi = \mathbf{b}_M^T \Psi_M + r$ with $r \in \mathcal{F}$. This yields to the following relation:

$$\mathcal{K}\phi = \mathbf{a}_M^T \mathcal{K}\Psi_M = \mathbf{a}_M^T \mathbf{K}_M \Psi_M + r = \mathbf{a}_M^T \Psi_M \circ \mathcal{T}. \tag{14}$$

To determine $\mathbf{K}_M$, the EDMD minimizes the following cost function given a dataset of snapshot pairs $\{(\mathbf{u}_k, \mathbf{v}_k)\}_{k=1}^N$ where $\mathbf{v}_k = \mathcal{T}(\mathbf{u}_k)$:

$$
\begin{aligned}
\mathcal{J} &= \frac{1}{2} \sum_{k=1}^N |\mathbf{r}(\mathbf{u}_k)|^2 \\
&= \frac{1}{2} \sum_{k=1}^N |\mathbf{a}^T(\Psi_M(\mathbf{v}_k) - \mathbf{K}_M \Psi_M(\mathbf{u}_k))|^2
\end{aligned}
\tag{15}
$$

where $|\cdot|$ denotes the Frobenius norm. Equation (15) is a least squares problem and the solution that minimises (15) is:

$$\mathbf{K}_M = \Psi(\mathbf{V})\Psi(\mathbf{U})^+ \tag{16}$$

where

$$\Psi(\mathbf{U}) = [\Psi(\mathbf{u}_1) \ \Psi(\mathbf{u}_2) \ \dots \ \Psi(\mathbf{u}_M)] \tag{17}$$

$$\Psi(\mathbf{V}) = [\Psi(\mathbf{v}_1) \ \Psi(\mathbf{v}_2) \ \dots \ \Psi(\mathbf{v}_M)] \tag{18}$$

and $\Psi(\mathbf{U})^+$ represents the pseudoinverse of $\Psi(\mathbf{U})$.

Overall, EDMD requires a data set of snapshot pairs, $\{(\mathbf{u}_k, \mathbf{v}_k)\}_{k=1}^N$, as well as a dictionary of observables, $\mathcal{D}_M$. Furthermore, it assumes that the subspace spanned by $\mathcal{D}_M$ is (nearly) invariant under the action of $\mathcal{K}$. Theoretical guarantees on the convergence of $\mathbf{K}_M$ to the true Koopman operator were studied in the limit of infinite data and an infinite number of observables. Specifically, under some technical assumptions on the sampling of the measurements and the form of the Hilbert space $\mathcal{F}$, Klus *et al* (2015), Williams *et al* (2015), Korda and Mezić (2018)showed that the matrix $\mathbf{K}_M$ converges to $\mathcal{K}_{\mathcal{D}_M}$, the Galerkin projection of the Koopman operator on the subspace $\mathcal{F}_{\mathcal{D}_M}$. Furthermore, Korda and Mezić (2018) studied the convergence of $\mathbf{K}_M$ to $\mathcal{K}$ as M goes to infinity in the strong operator topology. In practice, both N and M are finite, and in this setup, the choice of a relevant dictionary is still an open question. To further generalize the EDMD

algorithm, we thus propose to consider a latent dictionary. This latent dictionary will help generate observables that make the Koopman approximation more accurate.

3.2. EDMD with a latent dictionary

3.2.1. Proposed framework

Within an LEDMD framework, the data-driven derivation of numerical approximations of the Koopman operator are stated as an EDMD approximation with latent observables. These latent observables are used to account for both i) situations where the observation operator is not an embedding and ii) closure issues related to the choice of the dictionary in EDMD framework. In the LEDMD approach, we do not assume to access a dataset of snapshot pairs $\{(\mathbf{u}_k, \mathbf{v}_k)\}_{k=1}^{N}$ that satisfy $\mathbf{v}_k = \mathcal{T}(\mathbf{u}_k)$. We only assume observations of the state space model (3), that we conveniently put in the following dataset: $\{(\mathbf{x}_k, \mathbf{x}_{k+1})\}_{k=1}^{N}$.

Similarly to the EDMD, we start by choosing a dictionary $\mathcal{D}_M$ (7). This dictionary operates on measurements of the state space (3). We also assume the existence of some latent dictionary $\mathcal{D}_W$, that can be written similarly to $\mathcal{D}_M$ as follows:

$$\mathcal{D}_W = \{\psi_1^l, \psi_2^l, \ldots, \psi_W^l\} \tag{19}$$

where $\psi_i^l \in \mathcal{F} : L \longrightarrow \mathbb{C}$ for $i = 1, 2, \ldots, W$. Recall $\mathbf{x}_k = \mathcal{H}(\mathbf{z}_k)$, we also define the following vector valued functions:

We start by Ψ_M, stated as the EDMD observables. $\Psi_M : L \longrightarrow \mathbb{C}^{M \times 1}$:

$$\Psi_M(\mathcal{H}(\mathbf{z}_k)) = [\psi_1(\mathcal{H}(\mathbf{z}_k)) \; \psi_2(\mathcal{H}(\mathbf{z}_k)) \; \ldots \; \psi_M(\mathcal{H}(\mathbf{z}_k))]^{\mathrm{T}}. \tag{20}$$

We also define Ψ_W, stated as the latent observables. $\Psi_W : L \longrightarrow \mathbb{C}^{M \times 1}$:

$$\Psi_W(\mathbf{z}_k) = [\psi_1^l(\mathbf{z}_k) \; \psi_2^l(\mathbf{z}_k) \; \ldots \; \psi_W^l(\mathbf{z}_k)]^{\mathrm{T}}. \tag{21}$$

Finally, we define Ψ_{d_E} the vector valued LEDMD observables. $\Psi_{d_E} : L \longrightarrow \mathbb{C}^{d_E \times 1}$:

$$\Psi_{d_E}(\mathbf{z}_k) = [\psi_1(\mathcal{H}(\mathbf{z}_k)) \; \psi_2(\mathcal{H}(\mathbf{z}_k)) \; \ldots \; \psi_M(\mathbf{z}_k) \; \psi_1^l(\mathbf{z}_k) \; \psi_2^l(\mathbf{z}_k) \; \ldots \; \psi_W^l(\mathbf{z}_k)]^{\mathrm{T}} \tag{22}$$

where $d_E = M + W$.

We also define the matrix $\mathbf{G} \in \mathbb{R}^{M \times d_E}$ as:

$$\Psi_M(\cdot) = \mathbf{G} \Psi_{d_E}(\cdot). \tag{23}$$

If we assume that the observation operator $\mathcal{H}$ is a vector valued function with elements in $\mathcal{F}$, then the span of $\mathcal{D}_{d_E} = \mathcal{D}_M \cup \mathcal{D}_W$ is a subspace of $\mathcal{F}$, and can be written as follows:

$$\mathcal{F}_{\mathcal{D}_{d_E}} = \mathrm{Span}\{\psi_1, \psi_2, \ldots, \psi_M, \psi_1^l, \psi_2^l, \ldots, \psi_W^l\}$$
$$= \{\mathbf{a}_{d_E}^{\mathrm{T}} \Psi_{d_E} : \mathbf{a}_{d_E} \in \mathbb{C}^{d_E \times 1}\} \subset \mathcal{F}. \tag{24}$$

Similarly to the EDMD, $\mathbf{K}_{d_E}$, an approximation of the Koopman operator verifies:

$$\mathcal{K}\phi = \mathbf{a}_{d_E}^{\mathrm{T}} \mathcal{K} \Psi_{d_E} = \mathbf{a}_{d_E}^{\mathrm{T}} \mathbf{K}_{d_E} \Psi_{d_E} + r = \mathbf{a}_{d_E}^{\mathrm{T}} \Psi_{d_E} \circ \mathcal{T} \tag{25}$$

where $\phi \in \mathcal{F}_{\mathcal{D}_{d_E}}$ and $\mathbf{a}_{d_E} \in \mathbb{C}^{d_E \times 1}$. In the EDMD framework, the approximate Koopman operator $\mathbf{K}_{d_E}$ is defined as the minimizer of the following cost:

$$\mathcal{J} = \frac{1}{2} \sum_{k=1}^{N} |\mathbf{r}(\mathbf{z}_{k+1})|^2$$
$$= \frac{1}{2} \sum_{k=1}^{N} |\mathbf{a}_{d_E}^{\mathrm{T}} (\Psi_{d_E}(\mathbf{z}_{k+1}) - \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k))|^2. \tag{26}$$

The above EDMD cost function assumes direct measurements of the state variables $\mathbf{z}_t$. Furthermore, it assumes that both the dictionaries of non-linear functions $\mathcal{D}_M$ and $\mathcal{D}_W$ are known and span an (approximately) invariant Koopman subspace. In more realistic settings, we more likely have solely access to measurements $\mathbf{x}_t$ of a dynamical system. Even though we define our basis of observables on an embedding of the measurements, like for the case with delay embedding approaches (Kamb *et al* 2020), an arbitrary choice of a dictionary of observables does not systematically generate a good approximation of the Koopman operator. We thus propose to minimize cost function (26), solely given measurements of a dynamical system

and without any knowledge of the latent dictionary of observables $\mathcal{D}_W$. For this purpose, we reformulate the action of the Koopman operator on a function $\phi_M \in \mathcal{F}_{\mathcal{D}_M}$. By definition, we have:

$$\phi_M = \sum_{i=1}^{i=M} a_i \psi_i = \mathbf{a}_M^{\mathrm{T}} \Psi_M = \mathbf{a}_M^{\mathrm{T}} \mathbf{G} \Psi_{d_E} \tag{27}$$

with $\mathbf{a}_M \in \mathbb{C}^{M \times 1}$. The action of the Koopman operator on ϕ_M verifies:

$$\mathcal{K} \phi_M = \mathbf{a}_M^{\mathrm{T}} \mathbf{G} \mathcal{K} \Psi_{d_E} = \mathbf{a}_M^{\mathrm{T}} \mathbf{G} \Psi_{d_E} \circ \mathcal{T}. \tag{28}$$

Since the subspace spanned by $\mathcal{F}_{\mathcal{D}_{d_E}}$ is not invariant to the Koopman operator, the approximate Koopman operator $\mathbf{K}_{d_E}$ includes a residual term r as follows:

$$\mathcal{K} \phi_M = \mathbf{a}_M^{\mathrm{T}} \mathbf{G} \mathcal{K} \Psi_{d_E} = \mathbf{a}_M^{\mathrm{T}} \mathbf{G} \Psi_{d_E} \circ \mathcal{T} = \mathbf{a}_M^{\mathrm{T}} \mathbf{G} \mathbf{K}_{d_E} \Psi_{d_E} + r. \tag{29}$$

Given a dataset of snapshot pairs, $\{(\mathbf{x}_k, \mathbf{x}_{k+1})\}_{k=1}^N$ the LEDMD considers the following cost function:

$$\begin{aligned}
\mathcal{J}_{\mathrm{LEDMD}} &= \frac{1}{2} \sum_{k=1}^N |\mathbf{r}(\mathbf{u}_k)|^2 \\
&= \frac{1}{2} \sum_{k=1}^N |\mathbf{a}^{\mathrm{T}} (\mathbf{G} \Psi_{d_E}((\mathbf{z}_{k+1})) - \mathbf{G} \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k))|^2 \\
&= \frac{1}{2} \sum_{k=1}^N |\mathbf{a}^{\mathrm{T}} (\Psi_M(\mathcal{H}(\mathbf{z}_{k+1})) - \mathbf{G} \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k))|^2 \\
&= \frac{1}{2} \sum_{k=1}^N |\mathbf{a}^{\mathrm{T}} (\Psi_M(\mathbf{x}_{k+1})) - \mathbf{G} \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k))|^2.
\end{aligned} \tag{30}$$

If $\mathcal{D}_W$ and $\mathbf{z}_k$ are known, the LEDMD framework falls back into the standard EDMD problem described above. The solution of the minimization of $\mathcal{J}_{\mathrm{LEDMD}}$ can then be defined similarly to (15) using the pseudoinverse. However, since $\mathcal{D}_W$ and $\mathbf{z}_k$ are not known, we propose to minimize $\mathcal{J}_{\mathrm{LEDMD}}$ with respect to both $\mathbf{K}_{d_E}$ and $\{\Psi_W(\mathbf{z}_k)\}_{k=1}^{k=N}$:

$$\arg \min_{\mathbf{K}_{d_E}, \{\Psi_W(\mathbf{z}_k)\}_{k=1}^{k=N}} \mathcal{J}_{\mathrm{LEDMD}}. \tag{31}$$

3.2.2. Practical optimization problem
In this work, a regularization term is used to account for the action of $\mathbf{K}_{d_E}$ on the latent observables. Specifically, we numerically minimize the following objective function:

$$\arg \min_{\mathbf{K}_{d_E}} \min_{\{\psi_i^l(\mathbf{z}_k)\}_{i=1}^{i=W}} \sum_{k=1}^N |\Psi_M(\mathbf{x}_{k+1}) - \mathbf{G} \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k)|^2 + \beta |\Psi_{d_E}(\mathbf{z}_{k+1}) - \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k)|^2 \tag{32}$$

where β is a weighting parameter. The term $|\Psi_{d_E}(\mathbf{z}_{k+1}) - \mathbf{K}_{d_E} \Psi_{d_E}(\mathbf{z}_k)|^2$ may be regarded as a regularization term such that the inference of the latent observables $\{\Psi_W(\mathbf{z}_k)\}_{k=1}^{k=N}$ of the vector of observable Ψ_{d_E} is not solved independently for each time-step. The overall idea of the proposed framework are sketched in figure 1.

Interpretation of the LEDMD: This formulation can be interpreted as searching for the most relevant subset of *latent* observables Ψ_W, that makes the set of observables Ψ_{d_E} to evolve linearly in time. It is worth noting that this formulation naturally extends to cases where the measurements $\mathbf{x}_t$ do not form an embedding of the state of the dynamics $\mathbf{z}_t$. Indeed, the set of *latent* observables is written as a solution to an optimization problem, and not as an explicit mapping of the observations. The LEDMD can be easily extended to account for physical constraints, similarly to DMD-type algorithms (Colbrook 2022, Baddoo *et al* 2023), by including these constraints as regularization terms in (30).

Benefits of the numerical optimization: Since the optimization is carried out numerically, we can further benefit from the numerical computation of the approximate Koopman operator as well as the corresponding latent observables. It helps to formulate the Koopman operator as the matrix exponential of an infinitesimal generator. Such a formulation makes it simple to constrain the eigenvalues of the approximate Koopman operator to be unitary (when relevant, for example, in conservative systems). This formulation is simply derived by considering $\mathbf{K} = e^{\Delta t \mathbf{A}}$ where $\mathbf{A}$ is the approximate infinitesimal generator of the Koopman operator and Δt is the time step. Figure 1 shows an example where the infinitesimal generator is optimized.
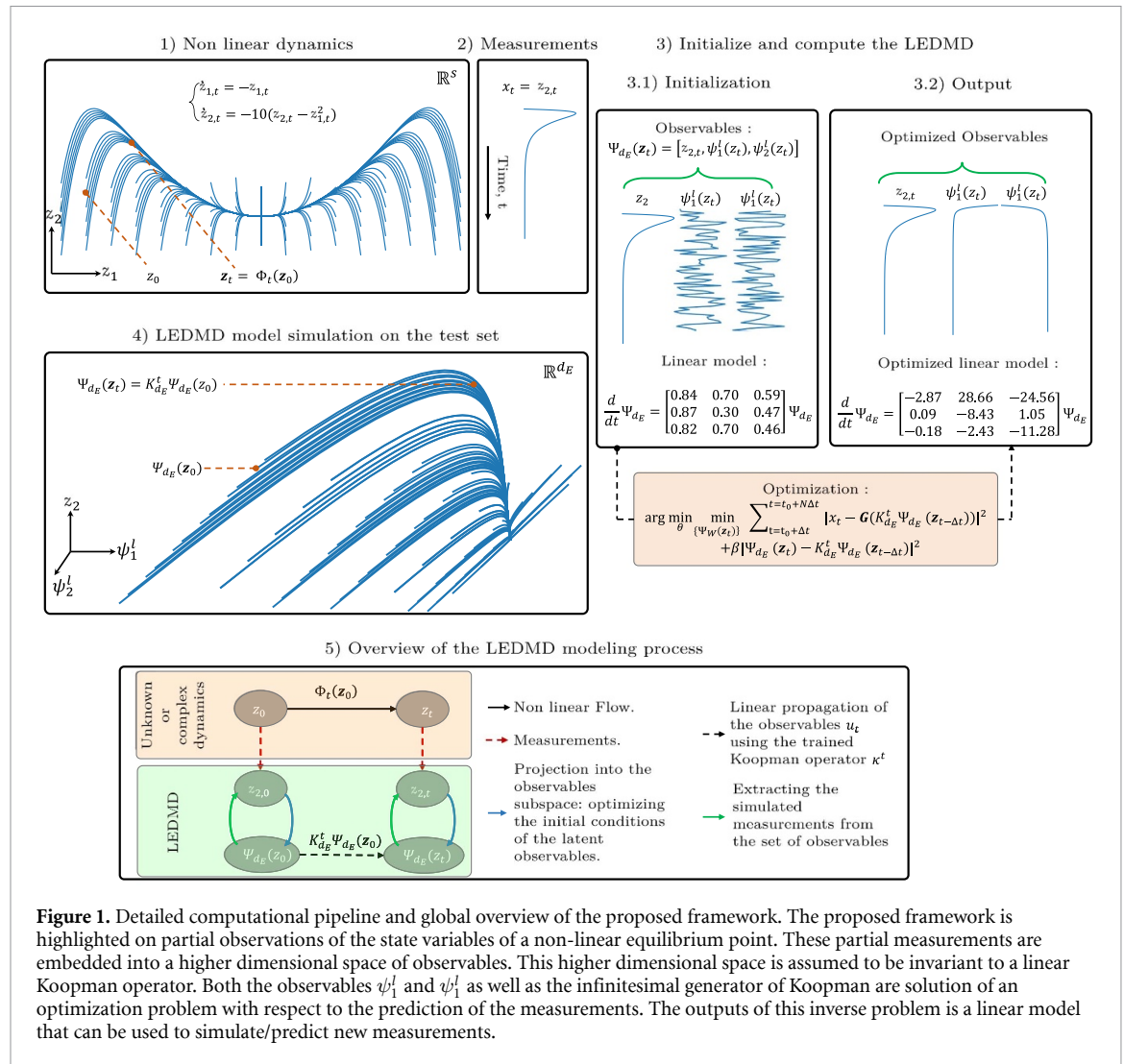
Figure 1. Detailed computational pipeline and global overview of the proposed framework. The proposed framework is highlighted on partial observations of the state variables of a non-linear equilibrium point. These partial measurements are embedded into a higher dimensional space of observables. This higher dimensional space is assumed to be invariant to a linear Koopman operator. Both the observables ψ_1^l and ψ_1^l as well as the infinitesimal generator of Koopman are solution of an optimization problem with respect to the prediction of the measurements. The outputs of this inverse problem is a linear model that can be used to simulate/predict new measurements.

Forecasting applications B the latent observables Ψ_W of the vector of observables Ψ_{d_E} are outputs of an optimization problem, forecasting applications must be approached with caution to account for a relevant choice of Ψ_W initial conditions. In practice, given a trained LEDMD model, forecasting measurements $\mathbf{x}_k$ relies on the forecasting of the entire set of observables Ψ_{d_E}. The latter amounts to finding an initial condition for the latent observables, Ψ_W and then propagating these observables by the action of the approximate Koopman operator. This issue can be addressed by finding the most relevant latent observables that minimize the forecasting cost of a predefined sequence of EDMD observables Ψ_M. Specifically, given a new sequence of measurements $\mathbf{x}_k, k \in \{N_1, \ldots, N_2\}$, forecasting these new series of measurements for $k > N_2$ is carried by following the optimization problem (32), where we infer the initial condition of Ψ_W using the following minimization:

$$\arg \min_{\{\psi_i^l(\mathbf{z}_k)\}_{i=N_1}^{i=N_2}} \sum_{k=N_1}^{N_2} |\Psi_M(\mathbf{x}_{k+1}) - \mathbf{G}\mathbf{K}_{d_E}\Psi_{d_E}(\mathbf{z}_k)|^2 + \beta|\Psi_{d_E}(\mathbf{z}_{k+1}) - \mathbf{K}_{d_E}\Psi_{d_E}(\mathbf{z}_k)|^2. \tag{33}$$

Here, we only minimize w.r.t. latent observables, Ψ_W given the trained Koopman operator $\mathbf{K}_{d_E}$. This minimization relates to a variational assimilation issue with partially-observed states and known dynamical and observation operators (Lynch and Huang 2010).

4. LEDMD and the stochastic Koopman operator

Hereafter, we describe the extension of the proposed framework for the derivation of finite-dimensional approximations of the stochastic Koopman operator.

4.1. Stochastic Koopman operator

Let us consider a state space model driven by a Random Dynamical System (RDS):

$$\mathbf{z}_{k+1} = \mathcal{T}_\Omega(\mathbf{z}_k, w_k); \ \mathbf{z}_k \in L; \ w_k \in \Omega \tag{34}$$

$$\mathbf{x}_k = \mathcal{H}(\mathbf{z}_k) \tag{35}$$

with (L, Σ_L, μ_L) and $(\Omega, \Sigma_\Omega, \mu_\Omega, \theta)$ the probability spaces associated with L and Ω respectively. $\theta : \Omega \longrightarrow \Omega$ is a measure preserving base flow. The noise process w_k is assumed to be independent from $\mathbf{z}_0 \ \mathbf{z}_0 \ \ldots \ \mathbf{z}_k$.

We can define a one-step stochastic Koopman operator $\mathcal{K}_\Omega$, associated with the RDS (34) as:

$$\mathcal{K}_\Omega g(\mathbf{z}_t) = \mathbb{E}_\Omega[g(\mathcal{T}_\Omega(\mathbf{z}_k, w_k))] \tag{36}$$

with $g \in \mathcal{F} : L \longrightarrow \mathbb{C}$ is a complex-valued observable of the dynamical system. Similarly to the deterministic case, if a finite dictionary $\mathcal{D}_{d_E}$ of elements of $\mathcal{F}$ spans a subspace $\mathcal{F}_{d_E} \subset \mathcal{F}$ that is (approximately) invariant to the stochastic Koopman operator, we can define a finite-dimensional (approximation) $\mathbf{K}_\Omega$ of $\mathcal{K}_\Omega$ as the projection of $\mathcal{K}_\Omega$ into $\mathcal{F}_{d_E}$.

4.2. Approximation of the stochastic Koopman operator

When considering the approximation of the stochastic Koopman operator from data, the standard EDMD (with a correct choice of the dictionary) has been explored in various works. Specifically, when in addition to ergodicity, the noise process is assumed to be white, Takeishi *et al* (2017b) showed the eigenvalues produced by the standard EDMD algorithms converging to the ones of the stochastic Koopman operator. However, for given noisy observables, the output of the EDMD algorithm is biased (Dawson *et al* 2016). Several works proposed variants of the DMD algorithm to correct the influence of noise on the estimation of the approximate Koopman operator in both the deterministic (Dawson *et al* 2016, Jiang and Liu 2022) and stochastic (Takeishi *et al* 2017b, Wanner 2020, Jiang and Liu 2022) settings. Yet, a fundamental issue regarding the choice of the dictionary of observables still persists.

We will show that the proposed LEDMD framework is relevant for handling RDS. Assuming a Gausian noise, e_k, we combine LEDMD with a Kalman filter, following Jiang and Liu (2022), to produce an estimate of the stochastic Koopman operator and estimates of the process and observation noises. Furthermore, using the LEDMD in a Kalman filter allows for highly irregular observations that can not be handled by simple DMD inversions.

4.3. LEDMD for the stochastic Koopman operator

The proposed LEDMD model can be used to derive the stochastic Koopman operator, given partial observations of the state of the system, and assuming an incomplete basis of observables. Similarly to the LEDMD in the deterministic case, two dictionaries of functions $\mathcal{D}_M$ and $\mathcal{D}_W$ are considered. The dictionary $\mathcal{D}_M$ acts on measurements of the RDS, while $\mathcal{D}_W$ spans the latent observables. We assume that the span of $\mathcal{D}_{d_E} = \mathcal{D}_M \cup \mathcal{D}_W$ is (approximately) invariant to the action of the stochastic Koopman operator which gives:

$$\Psi_{d_E}(\mathbf{z}_{k+1}) = \mathbf{K}_\Omega \Psi_{d_E}(\mathbf{z}_k) + \mathbf{e}_k. \tag{37}$$

The derivation of $\mathbf{K}_\Omega$ and the latent observables Ψ_W are to be considered with care to account for both the process noise and some potential measurements noise. The learning criterion proposed in (32) is likely to fail. To best circumvent this issue, similarly to Jiang and Liu (2022), we propose to formulate the LEDMD as a state space model. In this model, the measurement equation can be easily defined by writing the restriction of the full state observable to the measurements $\mathbf{x}_k$: $\Psi_O : L \longrightarrow \mathbb{C}^{n \times 1}$:

$$\begin{aligned}
\Psi_O(\mathcal{H}(\mathbf{z}_k)) &= [\psi_1^o(\mathcal{H}(\mathbf{z}_k)) \ \psi_2^o(\mathcal{H}(\mathbf{z}_k)) \ \ldots \ \psi_n^o(\mathcal{H}(\mathbf{z}_k))]^{\mathrm{T}} \\
&= [\mathbf{e}_1^*(\mathcal{H}(\mathbf{z}_k)) \ \mathbf{e}_2^*(\mathcal{H}(\mathbf{z}_k)) \ \ldots \ \mathbf{e}_n^*(\mathcal{H}(\mathbf{z}_k))]^{\mathrm{T}} \\
&= \mathbf{x}_k
\end{aligned} \tag{38}$$

with $\mathbf{e}_i^*$, $i = 1, \ldots, n$ the ith unit vector in $\mathbb{R}^n$. In this work, we assume that $\psi_i^o \in \mathcal{F}_{\mathcal{D}_{d_E}}$ for all $i = 1, \ldots, n$ so that we can write:

$$\mathbf{x}_t = \mathbf{H}\Psi_{d_E}(\mathbf{z}_k) + \boldsymbol{\nu}_k \tag{39}$$

where $\mathbf{H} = \mathbf{BG}$. Here, $\mathbf{B} \in \mathbb{R}^{n \times M}$ is some appropriate matrix of weight that satisfies $\psi_i^o = \mathbf{b}_i \Psi_M$ where $\mathbf{b}_i$ is the i-th line of $\mathbf{B}$. $\boldsymbol{\nu}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ with $\mathbf{R}_k \in \mathbb{R}^{n \times n}$ is the covariance matrix of the measurements error. The vector $\boldsymbol{\nu}$ accounts for the measurements error.

Equations (39) and (37) form a linear state space model. Assuming in this setup that the uncertainties are gaussian i.e. $\mathbf{e}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$ and $\boldsymbol{\nu}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ and the span of $\mathcal{D}_{d_E}$ is invariant to the action of the stochastic Koopman operator, the Kalman filter then represents an optimal estimator, in the least square sense, of the sequence of vector valued observables $\Psi_{d_E}(\mathbf{z}_k)$, given the measurements $\mathbf{x}_t$. Standard Kalman filtering parameterization can then be used to infer the approximate stochastic Koopman operator, and the covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$.

4.4. Kalman filter for the estimation of the stochastic LEDMD

The Kalman filter is used on equations (39) and (37) to derive an estimate of the distribution $p(\Psi_{d_E}(\mathbf{z}_k)|\{\mathbf{x}_i\}_{i=1}^{i=k})$, $k \leqslant N$. Formally, given the initial moments $\Psi_{d_E}(\mathbf{z}_k)_1^a$ and $\mathbf{P}_1^a$, the mean $\Psi_{d_E}(\mathbf{z}_k)^a$ and covariance $\mathbf{P}_k^a$ of $p(\Psi_{d_E}(\mathbf{z}_i)|\{\mathbf{x}_i\}_{i=1}^{i=k})$ can be computed as follows:

$$\Psi_{d_E}(\mathbf{z}_{k+1})^f = \mathbf{K}_\Omega \Psi_{d_E}(\mathbf{z}_k)^a \tag{40a}$$

$$\mathbf{P}_{k+1}^f = \mathbf{K}_\Omega \mathbf{P}_k^a \mathbf{K}_\Omega^{\mathrm{T}} + \mathbf{Q}_k \tag{40b}$$

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^f \mathbf{H}^{\mathrm{T}} [\mathbf{H}\mathbf{P}_{k+1}^f (\mathbf{H})^{\mathrm{T}} + \mathbf{R}_k]^{-1} \tag{40c}$$

$$\Psi_{d_E}(\mathbf{z}_{k+1})^a = \Psi_{d_E}(\mathbf{z}_{k+1})^f + \mathbf{K}_{k+1} \left[\mathbf{x}_{k+1} - \mathbf{H}\Psi_{d_E}(\mathbf{z}_{k+1})^f \right] \tag{40d}$$

$$\mathbf{P}_{k+1}^a = \mathbf{P}_{k+1}^f - \mathbf{K}_{k+1}\mathbf{H}\mathbf{P}_{k+1}^f. \tag{40e}$$

The superscripts f and a refer to the forecasting and filtering phases of the Kalman filter, respectively.

4.5. Optimization of the parameters of the stochastic LEDMD:

The use of the Kalman filter to estimate parameters of a given state space model (here Ψ_{d_E}, $\mathbf{K}_\Omega$, $\mathbf{Q}_k$ and $\mathbf{R}_k$) is classical in state space modeling literature (Tandeo *et al* 2018). The optimization of $\mathbf{K}_\Omega$, $\mathbf{Q}_k$ and $\mathbf{R}_k$ must be considered with care to account for a correct derivation of the covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, and the stochastic Koopman operator $\mathbf{K}_\Omega$. We consider the maximization of the likelihood of the observations. It provides a direct objective function to be optimized using gradient descent. Formally, and under the assumption of the state space model of equations (39) and (37), the innovation likelihood of a given observation at time k, $p(\mathbf{x}_t|\mathbf{x}_{k-1}, \mathbf{K}_\Omega)$ follows a Gaussian distribution $\mathcal{N}(\mathbf{H}\Psi_{d_E}(\mathbf{z}_k), \mathbf{H}\mathbf{P}_k^f H^{\mathrm{T}} + \mathbf{R}_k)$ (Carrassi *et al* 2017). The total likelihood of the observations under (39) and (37) then writes as a product of the likelihood of the innovations (Carrassi *et al* 2017):

$$p(\mathbf{x}_{1:N}|\mathbf{K}_\Omega) = \prod_{k=2}^{k=N} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{K}_\Omega). \tag{41}$$

Equation (41) includes crucial information regarding the parameters of the filtering scheme. It can be maximized numerically to derive the correct values for Ψ_{d_E}, $\mathbf{K}_\Omega$, $\mathbf{Q}_k$ and $\mathbf{R}_k$. Furthermore, and under some assumptions on the covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, Jiang and Liu (2022) showed that the maximization of the likelihood of the observations converges to the standard DMD formulation in the deterministic case. It also converges to the Noise-corrected DMD (Dawson *et al* 2016) when considering deterministic dynamics with a small measurement noise level. For arbitrary noises, the maximization of (41) does not have a known analytical solution. We thus propose to maximize (41) numerically using gradient descent. Specifically, the observable $\Psi_{d_E}(z_k)$ and the $\mathbf{K}_\Omega$, $\mathbf{Q}_k$ and $\mathbf{R}_k$ are estimated as follows:

- Step 1: Initialize $\mathbf{K}_\Omega$, $\mathbf{Q}_k$ and $\mathbf{R}_k$.
- Step 2: Given measurements $\{\mathbf{x}_k\}_{k=1}^{k=N}$, estimate the posterior distribution $p(\Psi_{d_E}(\mathbf{z}_k)|\{\mathbf{x}_i\}_{i=1}^{i=k})$ using the Kalman filter.

- Step 3: Minimize the negative log likelihood of (41) with respect to $\mathbf{K}_\Omega$, $\mathbf{Q}_k$ and $\mathbf{R}_k$.
- Step 4: Repeat steps 2 and 3 until convergence.

The likelihood (41) measures how good a model fits the observations. It is known in the state-of-the-art data assimilation literature as model evidence. It is also often used in model selection (Carson *et al* 2018) as it provides a direct measure of the accuracy of a given state-space model. In this work, and under the hypothesis of the Gaussian noise uncertainties, this likelihood can be used to evaluate the effectiveness of the stochastic LEDMD approximation under different dictionaries of observables. In the next experimental section, focus is given on the evaluation of the dimension of the approximate Koopman operator. We show that the likelihood is maximized for approximations of Koopman that have a sufficient dimension.

5. Application examples

The LEDMD is tested on deternimistic and stochastic dynamical systems in the following section. In the appendix, we also present further experiments on nonlinear oscillations and passengers airline dataset.

5.1. Equilibrium points
Let us consider the following system of differential equations

$$\begin{cases} \dot{z}_{1,t} = \mu z_{1,t} \\ \dot{z}_{2,t} = \alpha(z_{2,t} - z_{1,t}^2) \end{cases}. \tag{42}$$

Equation (42) is a nonlinear ODE with an equilibrium point at the origin. Studying the derivation of a linear conjugate of this equation is relevant. The equation (42) admits a three-dimensional closed-form linear Koopman representation (Brunton *et al* 2016a), by choosing as set of observables the variables z_1, z_2 and z_1^2. Formally, and considering $z_3 = z_1^2$, we may rewrite (42) as:

$$\begin{cases} \dot{z}_{1,t} &= \mu z_{1,t} \\ \dot{z}_{2,t} &= \alpha(z_{2,t} - z_{3,t}) \\ \dot{z}_{3,t} &= 2\mu z_{3,t} \end{cases}. \tag{43}$$

Full observations of the state space. In this first experiment, we consider as measurements the full state vector of (42) i.e. $\mathcal{H} = I_2$, $\mathbf{x}_t^{\mathrm{T}} = [z_{1,t}, z_{2,t}]$ (please refer to appendix B for a description of the numerical simulation and sampling of (42)). Figure 2 illustrates the performance of LEDMD framework in the identification of a linear model that perfectly matches the non-linear dynamics. Our framework is tested here with $\mathcal{D}_\mathcal{M} = \{\mathbf{e}_1^* \ \mathbf{e}_2^*\}$ and $\mathcal{D}_\mathcal{W} = \{\psi_1^l\}$. This setup generates observables $\Psi_{d_E} = [z_1 \ z_2 \ \psi_1^l(\mathbf{z}_t)]^{\mathrm{T}} \in \mathbb{R}^{3 \times 1}$ ($d_E = 3$) *i.e* a single latent observable is concatenated to the observations. It is worth noting that neither the dynamical model, nor its non-linearities are known by the proposed framework as the observable $\psi_1^l(\mathbf{z}_t)$ and the approximate Koopman operator $\mathbf{K}_{d_E}$ are solutions of an optimization problem, that minimizes the forecasting of the observations. The EDMD algorithm using, in addition to the observed states, an additional observable z_1^2 leads to a closed form Koopman translation of the non-linear differential equation (42). Indeed, as shown by equation (43), this ODE can be analytically linearized with this set of observables. However, selecting a bad observable, for instance z_2^2, drastically changes the EDMD performance as illustrated in figure 2. Choosing the right finite set of observables is key in data-driven Koopman representations, and this experiment highlights this aspect.

Partial observations of the state space We now only consider partial observations of the state of (42) i.e. $\mathcal{H} = [0, 1]$, $\mathbf{x}_t^{\mathrm{T}} = [z_{2,t}]$. Similarly to the previous experiment, the LEDMD framework is tested with $\mathcal{D}_\mathcal{M} = \{\mathbf{e}_2^*\}$ and $\mathcal{D}_\mathcal{W} = \{\psi_1^l, \psi_2^l\}$. In this configuration, two latent variables are concatenated to the observation. This application example is depicted in figure 1.

Figure 3(a) illustrates the DMD algorithm, applied to measurements of $\mathbf{z}_2$. Including high order polynomial non-linearities tends to make worst the DMD approximation as illustrated in figure 3(b). This is due to the fact that the non-linearity is present in the unobserved component $\mathbf{z}_1$. The LEDMD model is able to derive the most relevant latent observables that linearize the system, only given partial measurements. Figure 4 provides a simulation example of the LEDMD framework.
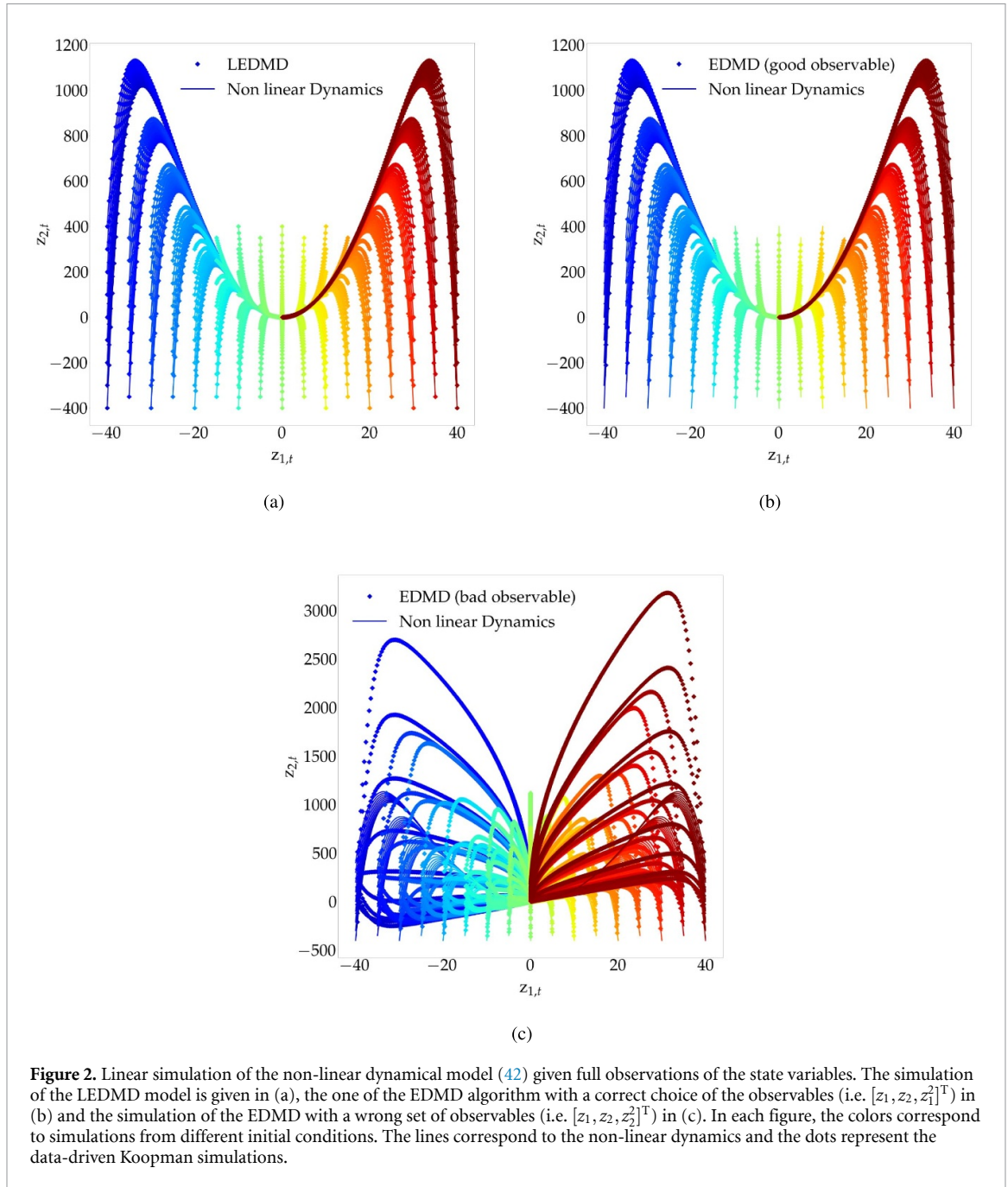
Figure 2. Linear simulation of the non-linear dynamical model (42) given full observations of the state variables. The simulation of the LEDMD model is given in (a), the one of the EDMD algorithm with a correct choice of the observables (i.e. $[z_1, z_2, z_1^2]^{\mathrm{T}}$) in (b) and the simulation of the EDMD with a wrong set of observables (i.e. $[z_1, z_2, z_2^2]^{\mathrm{T}}$) in (c). In each figure, the colors correspond to simulations from different initial conditions. The lines correspond to the non-linear dynamics and the dots represent the data-driven Koopman simulations.

5.2. LEDMD on the shallow water equation (SWE)

Chaotic systems are typical examples where the observables space of the Koopman operator is infinite-dimensional. However, numerous chaotic systems admit, in addition to a chaotic signature, several periodic and quasi-periodic modes, making suitably chosen linear models relevant for short-term forecast applications.

We consider modeling chaotic dynamics governed by the SWE. Observations of the sea surface elevation η_k are considered as measurements *i.e.* $\mathbf{x}_k = \eta_k$ (please refer to appendix B for a detailed description of the numerical simulation of the SWE). In this experiment, the LEDMD framework is built as follows. We construct Ψ_M from the empirical orthogonal function (EOF) decomposition of the measurements, i.e. $\mathcal{D}_{\mathcal{M}}$ is built with the $M = 100$ principal eigenvectors of the covariance of the measurements and $\mathcal{D}_{\mathcal{W}} = \{\psi_1^l, \ldots, \psi_W^l\}$ with $W = 600$. This setup generates observables $\Psi_{d_E}(\mathbf{z}_t) = [\mathbf{x}_t^{\mathrm{T}} \mathbf{E}^{\mathrm{T}} \; \psi_1^l(\mathbf{z}_t) \; \ldots \; \psi_W^l(\mathbf{z}_t)]^{\mathrm{T}} \in \mathbb{R}^{700 \times 1}$ ($d_E = 700$) with $\mathbf{E}$ the matrix formed by the eigenvectors of the covariance matrix of the data.

Figure 3. Linear simulation of the non-linear dynamical model (42) using the (E)DMD algorithm given partial observations of the state variables. (a) DMD approximation. The colors correspond to simulations from different initial conditions. The lines correspond to the non-linear dynamics of $z_{2,t}$ and the dots represent the DMD simulation. (b) (E)DMD Root Mean Squared Error (RMSE) forecasting performance. This error is averaged across all the test set simulations (starting from different initial conditions) for every model depicted here (DMD and EDMD with polynomial observables up to order 3).

Figure 4. Simulation of the non-linear dynamical model (42) using the LEDMD model constructed on partial observations of the state variables. The colors correspond to simulations from different initial conditions. The lines correspond to the non-linear dynamics of $z_{2,t}$ and the dots represent the LEDMD simulation. Even with partial measurements of the non-linear dynamics (42), the proposed framework perfectly matches the ground truth.
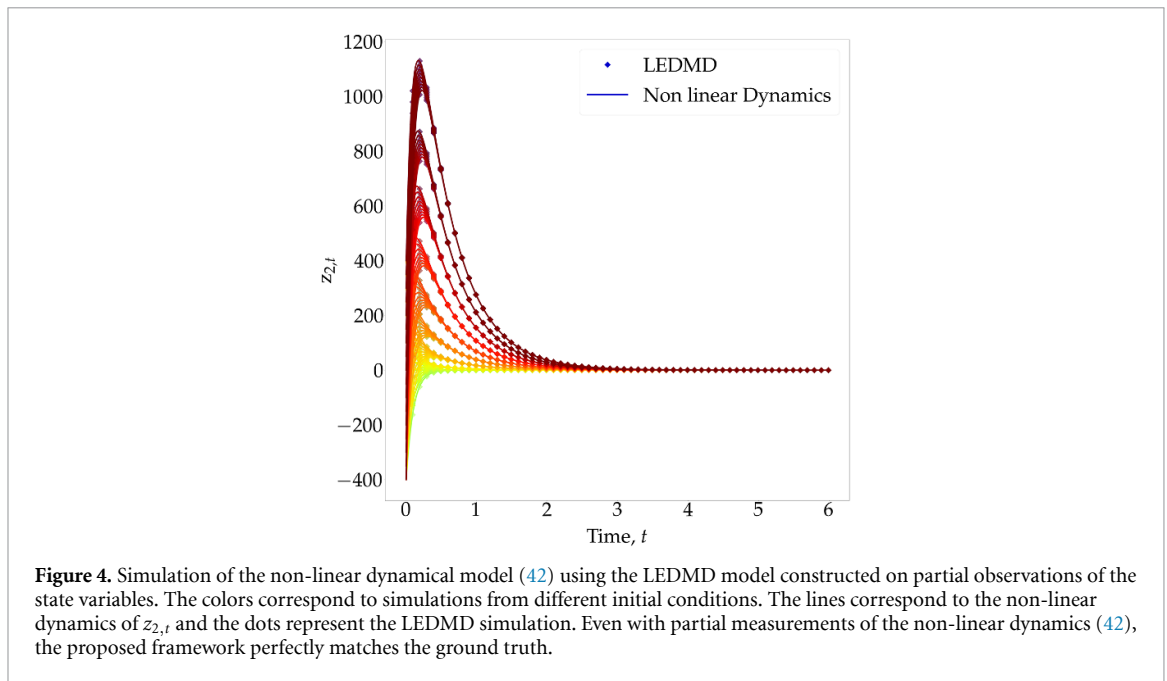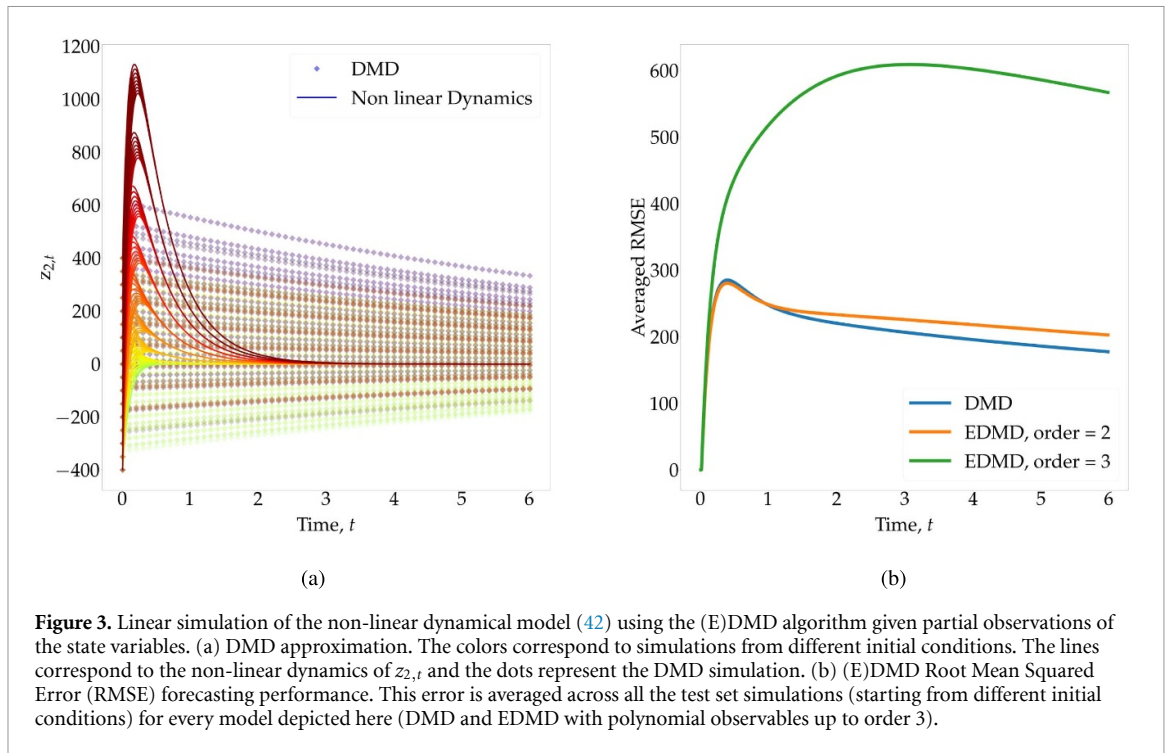
Figure 5 illustrates the forecasting performance with respect to the true state, the projection of the true state from the EOF basis, and the Hankel-DMD-based algorithm. This state-of-the-art model is built on a singular value decomposition (SVD) of delay embedding coordinates of the EOF (we use the same EOF decomposition as the on described above for the LEDMD) components (Kamb *et al* 2020). The delay embedding is computed, for every EOF component, using a lag embedding equal to one time step. We test 3 different embedding dimensions of $d_{E1} = 700$, $d_{E2} = 10\,000$ and $d_{E3} = 20\,000$. These embedding configurations generate three different Hankel DMD models that we refere to HDMD$_{d_{E1}}$, HDMD$_{d_{E2}}$ and HDMD$_{d_{E3}}$. In all these Hankel DMD models, the dimension of the SVD is set to 150, which accounts for over 95% of the total variance of the delay embedding representations. The qualitative analysis of figure 5 shows that the proposed architecture outperforms all Hankel-DMD based models by generating eddies that are closer to the ones of the true state. These observations are validated through the computation of the RMSE of each prediction time step (figure 6). Interestingly, even though increasing the embedding dimension of the
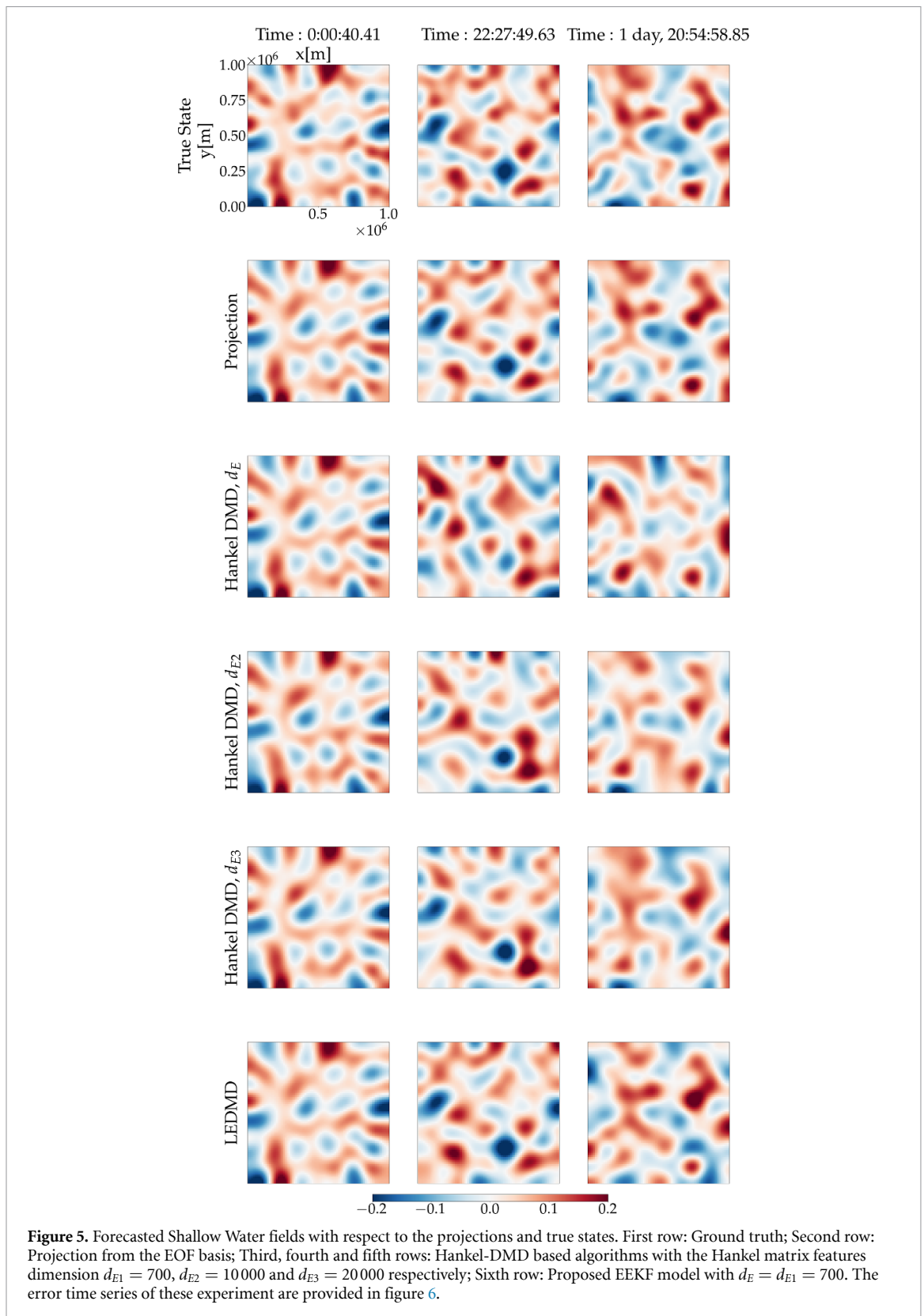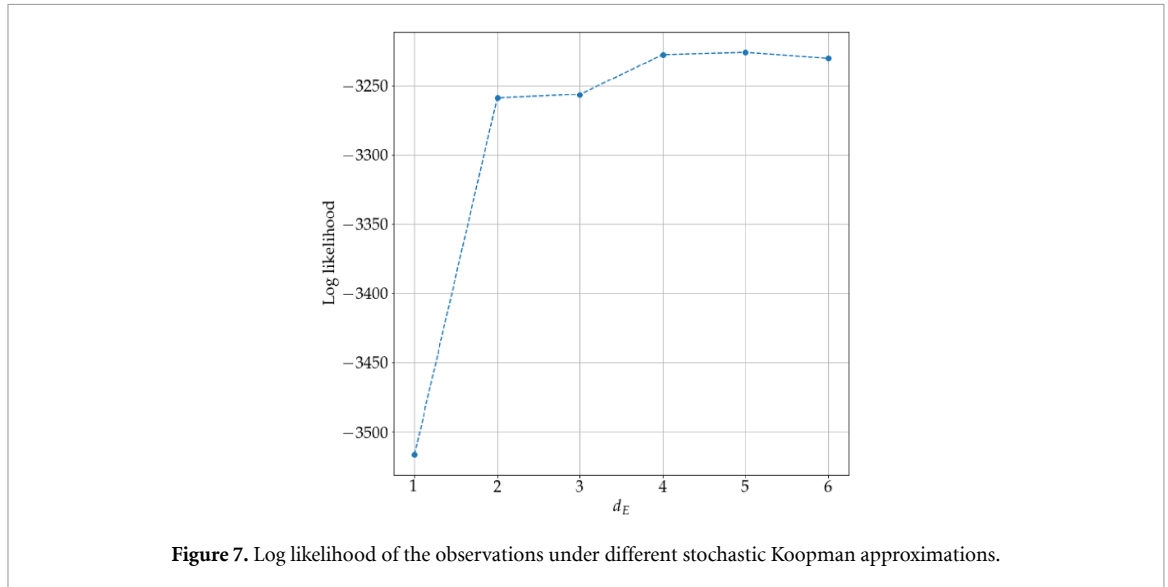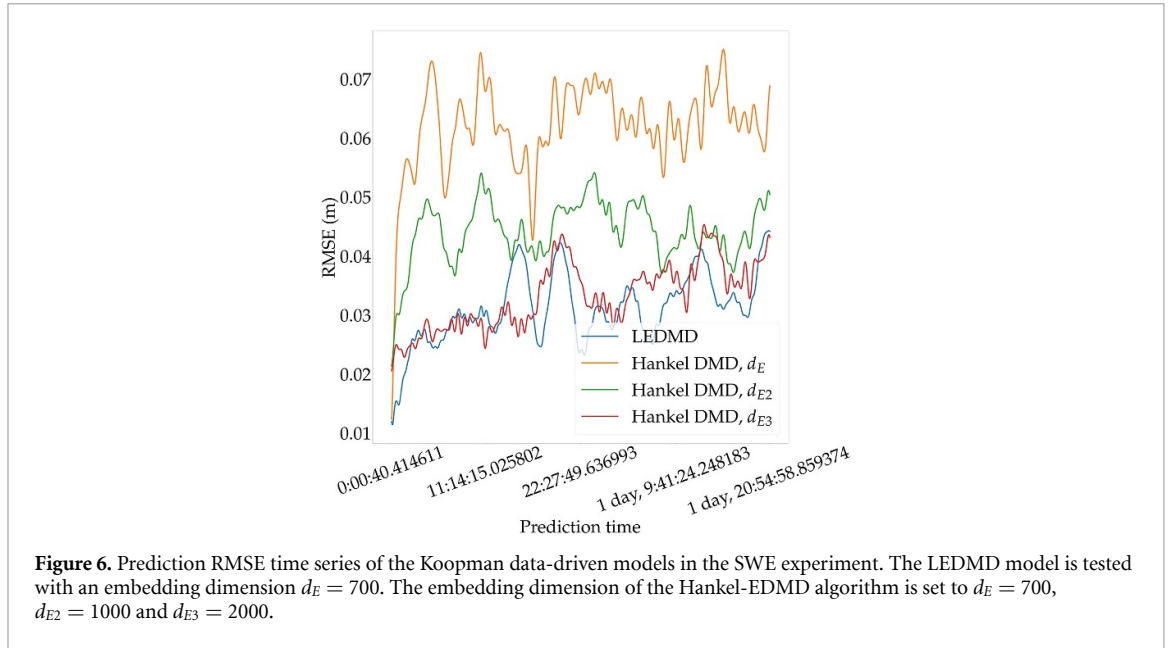
Figure 5. Forecasted Shallow Water fields with respect to the projections and true states. First row: Ground truth; Second row: Projection from the EOF basis; Third, fourth and fifth rows: Hankel-DMD based algorithms with the Hankel matrix features dimension $d_{E1} = 700$, $d_{E2} = 10\,000$ and $d_{E3} = 20\,000$ respectively; Sixth row: Proposed EEKF model with $d_E = d_{E1} = 700$. The error time series of these experiment are provided in figure 6.

Hankel matrix improves the DMD representations based on these observables, the proposed architecture leads to better results based on a smaller dimension of the set of observables.

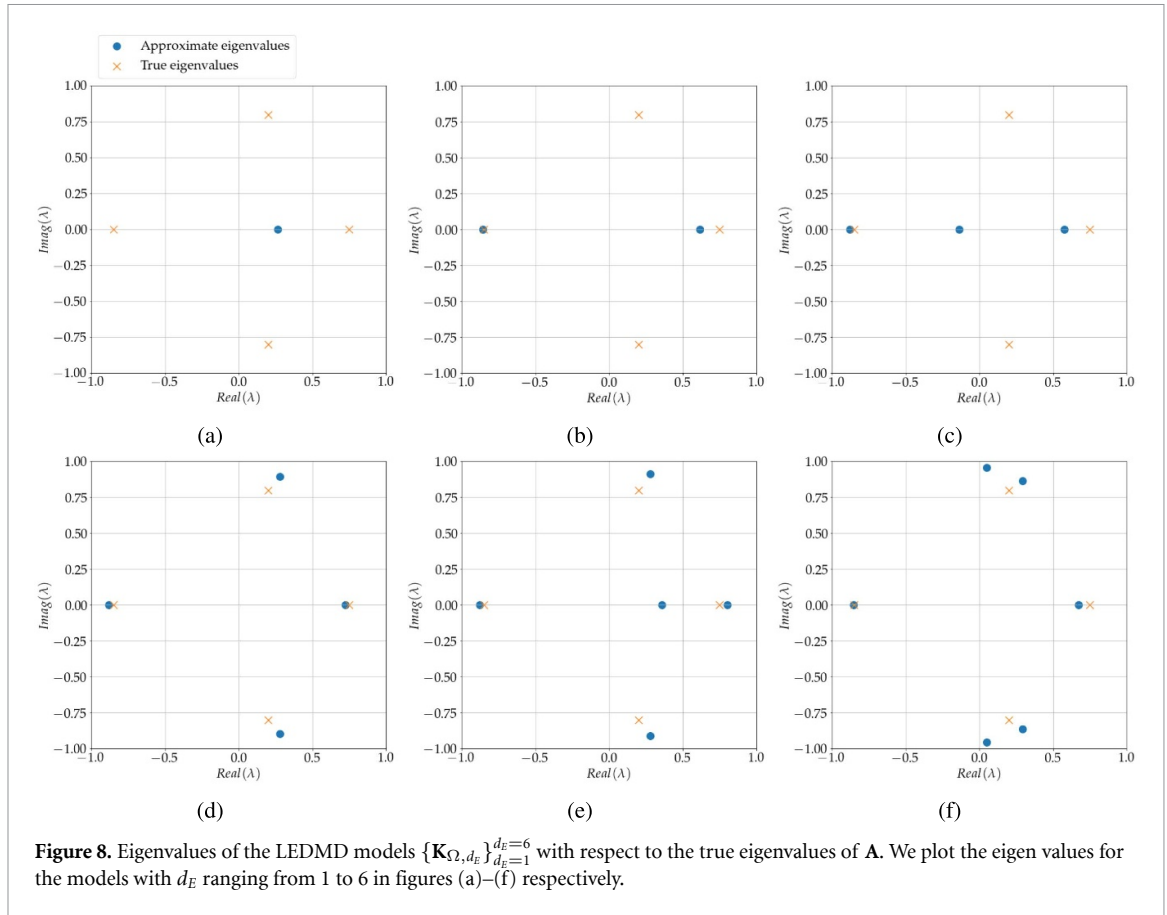## 5.3. Stochastic LEDMD representations
We now shift our attention to the stochastic version of the LEDMD framework. Let us consider the following linear system in $\mathbb{R}^4$:

Figure 6. Prediction RMSE time series of the Koopman data-driven models in the SWE experiment. The LEDMD model is tested with an embedding dimension $d_E = 700$. The embedding dimension of the Hankel-EDMD algorithm is set to $d_E = 700$, $d_{E2} = 1000$ and $d_{E3} = 2000$.

Figure 7. Log likelihood of the observations under different stochastic Koopman approximations.

$$\mathbf{z}_{k+1} = \mathbf{A}\mathbf{z}_k + \mathbf{e}_k = \begin{bmatrix} 0.75 & 0.5 & 0.1 & 2 \\ 0 & 0.2 & 0.8 & 1 \\ 0 & -0.8 & 0.2 & 0.5 \\ 0 & 0 & 0 & -0.85 \end{bmatrix} \mathbf{z}_k + \mathbf{e}_k \qquad (44)$$

with $\mathbf{e}_k \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ and $\sigma_{\mathbf{z}}^2 = 1$. We consider as observations 2000 samples of the first component of the state vector $\mathbf{z}_k$ with additional Gaussian noise i.e. $x_{k+1} = \mathbf{B}\mathbf{z}_{k+1} + b_k$ with $\mathbf{B} = [1, 0, 0, 0]^T$ and $b_k \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ and $\sigma_{\mathbf{x}}^2 = 1.5$. We consider six different stochastic Koopman operators with increasing d_E i.e. $\{\mathbf{K}_{\Omega, d_E}\}_{d_E=1}^{d_E=6}$. These operators share the same EDMD observables $\mathcal{D}_{\mathcal{M}} = \{\mathbf{e}_1^*\}$ and have an increasing number of latent observables (from no latent observables to 5). $\mathcal{D}_{\mathcal{W}} = \{\psi_1^l\}$.

We show in figure 7 the evolution of the log-likelihood of the observations for each tested stochastic Koopman approximation. Overall, this score converges for the models $\{\mathbf{K}_{\Omega, d_E}\}_{d_E>3}$ which suggests that a dimension of 4 is enough for modeling the observations. We also illustrate in figure 8 the estimated eigenvalues of each model with respect to the true eigenvalues of $\mathbf{A}$. When $d_E < 4$, the stochastic Koopman approximation is only able to match eigenvalues $\mathbf{A}$ with no imaginary parts which explain the low

Figure 8. Eigenvalues of the LEDMD models $\{\mathbf{K}_{\Omega,d_E}\}_{d_E=1}^{d_E=6}$ with respect to the true eigenvalues of $\mathbf{A}$. We plot the eigen values for the models with d_E ranging from 1 to 6 in figures (a)–(f) respectively.

log-likelihood score of figure 7 when compared to higher dimensions. At $d_E = 4$ the associated Koopman approximation has enough eigenvalues to correctly match the ones of the true RDS which explains the increase in the log-likelihood score. At higher dimensions, $\mathbf{K}_{\Omega,d_E=5}\}$ and $\mathbf{K}_{\Omega,d_E=6}\}$ have similar likelihood scores but with extra spurious eigenvalues.

6. Discussion

Introduced about a century ago, the Koopman operator formalism helps describe the evolution of a state sequence through the linear propagation of an infinite-dimensional set of observables. Finding finite rank approximations of this operator motivated tremendous amounts of works, centered around several questions, one of these interrogations concerns: which observables to use, in order to avoid loosely trading the complexity of a potentially non-linear system for a higher dimensional linear one ?

In this context, several dictionary-based families of observables have been investigated, ranging from non-linear polynomial expansions to autoencoders and deep learning. In the present work, a different perspective is considered. Instead of fixing a dictionary (or a family of dictionaries) and solving for the finite-dimensional approximation of Koopman, we write some of the Koopman observables as solutions to an optimization problem. In this setting, it helps avoid any unnecessary constraint over the observables space other than the ones constrained by the optimization of the forecasting cost of the observations. Through different numerical experiments, the proposed framework appears very efficient for the data-driven derivation of a finite-dimensional approximation of Koopman representations of dynamical systems. Whereas most state-of-the-art algorithms heavily rely on the selection of a family of basis functions, the proposed architecture can tackle several dynamical regimes, both given full or partial measurements of the state space.

When compared to standard non-linear model identification techniques (Brunton *et al* 2016b, Chen *et al* 2018, Ouala *et al* 2020, 2023), point (or discrete) spectrum approximations of Koopman, such as the proposed LEDMD framework and most of the state-of-the-art literature, can not represent the asymptotic behavior of chaotic dynamical systems. However, from an application perspective, finding linear models of non-chaotic dynamics as well as having simple predictive models of chaotic ones is highly valuable. In this

context, investigating the relevance of the LEDMD in applications such as control and data assimilation is a promising perspective. Building end-to-end trainable control/data assimilation algorithms based on this architecture should allow learning dynamical priors and linear embeddings, based on a direct application-oriented cost (such as a data assimilation cost in state reconstruction, an energy/performance based cost in the context of control, ...etc). Promoting sparsity in the proposed framework is also an important perspective. As shown in figure 1, the optimized approximation of the infinitesimal generator of Koopman contains terms that are close to zero, which emphasizes that encoding a sparsity prior to LEDMD may help to promote generalizability.

When the EDMD dictionary $\mathcal{D}_M$ spans an (approximately) invariant subspace of the Koopman operator, the latent observables Ψ_W can be omitted and the LEDMD becomes a standard EDMD method. From the numerical experiments, we show that stacking the latent observables to the vector of EDMD observables directs the LEDMD approximation towards an invariant subspace, especially when considering unknown dynamics and partial observations of the state variables. However, applying a numerical optimization instead of the standard inversion used in the EDMD misses the convergence results (for an infinite amount of data samples) of the EDMD to the Galerkin projection of the Koopman operator. This motivates studying the convergence properties and conditions of the LEDMD.

The approximation of the stochastic Koopman operator, treated in this work, further reveals an extremely important aspect, implicit to the proposed representation. Learning an LEDMD model in a stochastic fashion allows for two distinct levels of approximation within the proposed framework, namely i) the deterministic parameters, i.e. the approximate Koopman operator $\mathbf{K}_\Omega$, and ii) the stochastic components encoded through the process and observation noise covariances. When jointly learnt, those two components can trade the complexity of some given measurements and may dissociate stochastic and deterministic behaviors within a signal. In this context, investigating such aspects on more real world problems may require a way of getting rid of the Gaussianity assumption of the noises.

Data availability statement

Acknowledgment

Code availability

The code developed for this work is written using the Pytorch framework in Python. All the experiments and data used in this work are available at this link.

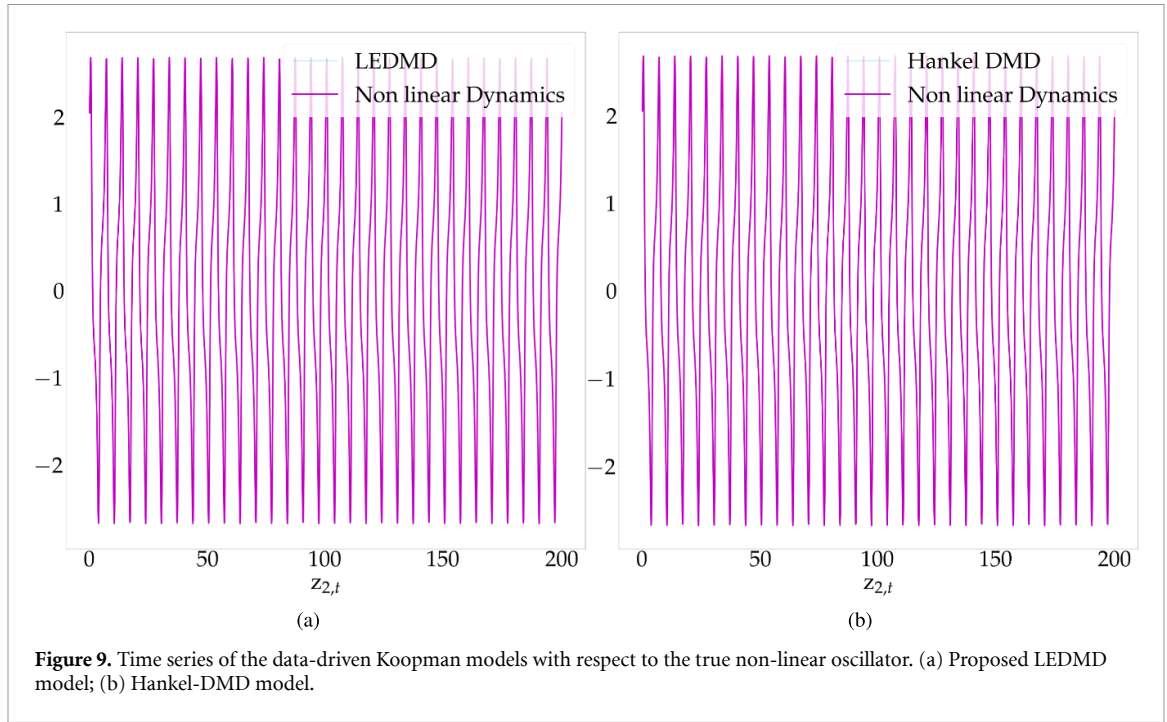## Appendix A. Additional application examples

A.1. Periodic orbits

Hankel matrix EDMD based representations (Kamb *et al* 2020) show great performance in modeling, linearly, periodic dynamics. Considering Delay embeddings as observables appears to decompose the non-linear periodic signal into a combination of linear oscillations, making the inversion of these sinusoidal modes into a linear dynamical model trivial.

While our framework can benefit from known basis of observables that can be used for instance in Ψ_M, the restriction of Ψ_M to direct measurements of the state variables (or generally, to measurements of state variables) leads to a Fourier decomposition of the measurements. Specifically, in such situations the LEDMD writes the non-linear oscillation as a composition of a finite number of linear ones with trainable frequencies encoded, for instance in the eigenvalues of the Koopman operator (or more precisely, in the eigenvalues of the infinitesimal generator).

In order to highlight this aspect, the Van der Pol oscillator, governed by the following equation, is considered.

$$\begin{cases} \dot{z}_{1,t} = z_{2,t} \\ \dot{z}_{2,t} = \mu(1 - z_{1,t})z_{2,t} - z_{1,t} \end{cases}. \tag{45}$$

Figure 9. Time series of the data-driven Koopman models with respect to the true non-linear oscillator. (a) Proposed LEDMD model; (b) Hankel-DMD model.

This dynamical system is widely used in state-of-the-art data-driven Koopman representations and can be translated into a Koopman linear model using several state-of-the-art algorithms. We show in this experiment that our proposed framework is also relevant in this context. Considering the true states as observations i.e. with $\mathcal{H} = \mathcal{M} = I_2$, we trained the proposed framework with $d_E = 100$ on a simulated trajectory of size 5000 (the trajectory was computed using the LOSDA ODE solver (Hindmarsh 1983) with a sampling rate $h = 0.1$). The Hankel-EDMD algorithm was tested with a lag embedding of a single time step and an embedding dimension $d_E = 100$. The dimension of the SVD is set to 16, which accounts for over 99% of the total variance of the delay embedding representation. The forecasting performance of the proposed model is shown in figure 9.

A.2. Air passenger time series

The previous experiment motivated the evaluation of the proposed model in forecasting real quasi-periodic signals. In this context, we consider the international Airline Passengers prediction problem. The data ranges from January 1949 to December 1960 with 144 observations in units of 1000. The first 100 data points were used as training data and we tested our approach on the remaining 44 observations. Figure 10 illustrates the forcasting performance of the proposed framework with respect to the Hankel DMD framework.

Appendix B. Generation of the data and parameterization of the non-linear equations

The data we use in our experiments are simulated by solving the systems of differential equations as follows:

- The parameters of the dynamical system (42) are set in the experiments to $\mu = -1$ and $\alpha = -10$. The training set is a single trajectory, simulated using the LOSDA ODE solver (Hindmarsh 1983) from $t = 0$ up to $t = 6$ and sampled at $h = 0.01$. The initial condition of this training set is $\mathbf{z}_0^{\mathrm{T}} = [12.0, -1.0]$. The test set consists on a collection of 289 trajectories, simulated and sampled similarly to the training sequence, but starting at different initial conditions with $\mathbf{z}_{1,0} = -40, -35, \ldots, 40$ and $\mathbf{z}_{2,0} = -400, -350, \ldots, 400$.
- The SWE dynamics used in this work is geverned by the following set of equations:

$$
\begin{cases}
\dfrac{\partial v_x}{\partial t} - F_{v_x} = -g_r \dfrac{\partial \eta}{\partial x} \\[2mm]
\dfrac{\partial v_y}{\partial t} - F_{v_y} = -g_r \dfrac{\partial \eta}{\partial y} \\[2mm]
\dfrac{\partial \eta}{\partial t} + \dfrac{\partial (\eta + H) v_x}{\partial x} + \dfrac{\partial (\eta + H) v_y}{\partial y} = 0
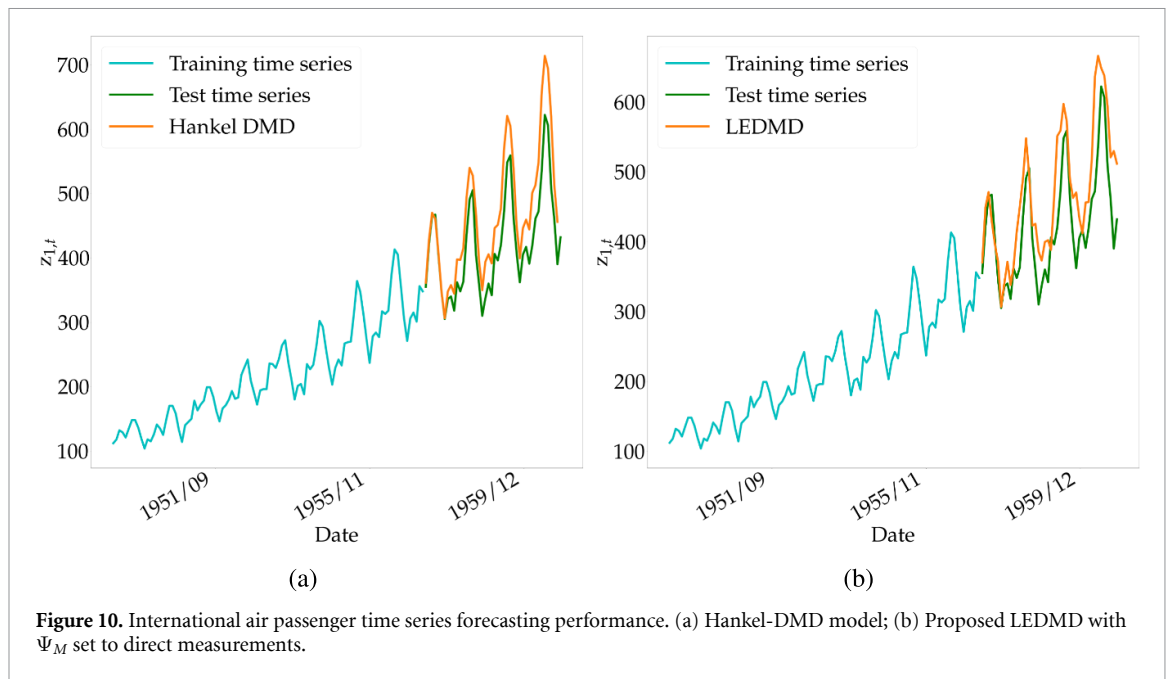\end{cases}
\tag{46}
$$

Figure 10. International air passenger time series forecasting performance. (a) Hankel-DMD model; (b) Proposed LEDMD with Ψ_M set to direct measurements.

where x and y represent the 2D directions of the fluid. η is the fluid surface elevation, and (v_x, v_y) is the fluid's horizontal flow velocity. g_r is the acceleration due to gravity which is taken to be equal to 9.81 m s^{-2}. H corresponds to the total depth (here $H = 100$ m) of the fluid and $f = f_0 + \lambda y$ is the Coriolis forcing where $f_0 = 1E - 4s^{-1}$ and $\lambda = 2E^{-11}(m \times s)^{-1}$.

The direct numerical simulation of the two-dimensional shallow-water equation is carried using an ordinary, forward in time, centered in space finite difference scheme for the momentum equation and for the linear term of the continuity equation. The non-linear term of the continuity equation is in the other hand discredited using an upwind scheme. The length of the domain is set to 1000 km $\times$ 1000 km with a corresponding regular discretization of 80×80. The temporal step size is set to satisfy the Courant–Friedrichs–Lewy condition ($h = 40.41$ s). 100 000 simulation time steps are generated from this configuration. The transient first 2500 time steps are omitted and we use the post-transient first 49 701 time-steps as training data. The remaining sequence is used as a test set.

Appendix C. Training

The trainable parameters of the LEDMD models i.e. the approximate Koopman operator and the latent observables Ψ_W are initially sampled from a uniform distribution. Furthermore, in the SWE and the VDP experiments, the eigenvalues of the $\mathbf{K}_{d_E}$ are constrained to be on the unit circle by imposing $\mathbf{K}_{d_E} = e^{\mathbf{A}_{d_E}}$ with $\mathbf{A}_{d_E} = \frac{1}{2}(\mathbf{B}_{d_E} - \mathbf{B}_{d_E}^{\mathsf{T}})$. The training of all models is carried using the Adam optimizer. We use a varying learning rate (from 0.9 to 0.001) in the experiment 5.1, a fixed learning rate of 0.001 and 0.01 in the experiments 5.2 and 5.3 respectively.

ORCID iD

Said Ouala ⬤ https://orcid.org/0000-0003-0554-3971

References

Arbabi H and Mezic I 2017 Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator *SIAM J. Appl. Dyn. Syst.* **16** 2096–126

Azencot O, Benjamin Erichson N, Lin V and Mahoney M 2020 Forecasting sequential data using consistent Koopman autoencoders *Int. Conf. on Machine Learning* (PMLR) pp 475–85

Baddoo P J, Herrmann B, McKeon B J, Kutz J N and Brunton S L 2023 Physics-informed dynamic mode decomposition *Proc. R. Soc.* A **479** 20220576

Brunton S L, Brunton B W, Proctor J L and Kaiser E 2017 and Nathan Kutz J Chaos as an intermittently forced linear system *Nat. Commun.* **8** 1–9

Brunton S L, Brunton B W, Proctor J L and Kutz J N 2016a Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control *PLoS One* **11** e0150171

Brunton S L, Proctor J L and Nathan Kutz J 2016b Discovering governing equations from data by sparse identification of nonlinear dynamical systems *Proc. Natl Acad. Sci.* **113** 3932–7

Budišić M, Mohr R and Mezić I 2012 Applied Koopmanism *Chaos* **22** 047510

Carrassi A, Bocquet M, Hannart A and Ghil M 2017 Estimating model evidence using data assimilation *Q. J. R. Meteorol. Soc.* **143** 866–80

Carson J, Crucifix M, Preston S and Wilkinson R D 2018 Bayesian model selection for the glacial–interglacial cycle *J. R. Stat. Soc.:* C (*Appl. Stat.*) **67** 25–54

Chen T Q, Rubanova Y, Bettencourt J and Duvenaud D K 2018 Neural ordinary differential equations *Advances in Neural Information Processing Systems* pp 6571–83

Colbrook M J 2022 The mpedmd algorithm for data-driven computations of measure-preserving dynamical systems (arXiv:2209.02244)

Colbrook M J, Ayton L J and Szőke Maté 2023 Residual dynamic mode decomposition: robust and verified Koopmanism *J. Fluid Mech.* **955** A21

Dawson S, Hemati M S, Williams M O and Rowley C W 2016 Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition *Exp. Fluids* **57** 1–19

Engel K-J, Nagel R and Brendle S 2000 *One-Parameter Semigroups for Linear Evolution Equations* vol 194 (Berlin: Springer)

Hindmarsh A C 1983 ODEPACK, a systematized collection of ODE solvers *IMACS Trans. on Scientific Computation* **1** 55–64

Jiang L and Liu N 2022 Correcting noisy dynamic mode decomposition with kalman filters *J. Comput. Phys.* **461** 111175

Kamb M, Kaiser E, Brunton S L and Kutz J N 2020 Time-delay observables for Koopman: theory and applications *SIAM J. Appl. Dyn. Syst.* **19** 886–917

Klus S, Koltai P and Schütte C 2015 On the numerical approximation of the Perron–Frobenius and Koopman operator (arXiv:1512.05997)

Klus S, Nüske F, Peitz S, Niemann J-H, Clementi C and Schütte C 2020 Data-driven approximation of the Koopman generator: model reduction, system identification and control *Physica* D **406** 132416

Koopman B O 1931 Hamiltonian systems and transformation in hilbert space *Proc. Natl Acad. Sci. USA* **17** 315

Korda M and Mezić I 2018 On convergence of extended dynamic mode decomposition to the Koopman operator *J. Nonlinear Sci.* **28** 687–710

Lusch B, Nathan Kutz J and Brunton S L 2018 Deep learning for universal linear embeddings of nonlinear dynamics *Nat. Commun.* **9** 1–10

Lynch P and Huang X-Y 2010 *Initialization* (Berlin: Springer) pp 241–60

Mauroy A and Goncalves J 2016 Linear identification of nonlinear systems: a lifting technique based on the Koopman operator *2016 IEEE 55th Conf. on Decision and Control* (*CDC*) (IEEE) pp 6500–5

Mauroy A and Goncalves J 2019 Koopman-based lifting techniques for nonlinear systems identification *IEEE Trans. Autom. Control* **65** 2550–65

Mezić I 2005 Spectral properties of dynamical systems, model reduction and decompositions *Nonlinear Dyn.* **41** 309–25

Mezić I 2013 Analysis of fluid flows via spectral properties of the Koopman operator *Annu. Rev. Fluid Mech.* **45** 357–78

Mezić I 2015 On applications of the spectral theory of the Koopman operator in dynamical systems and control theory *2015 54th IEEE Conf. on Decision and Control* (*CDC*) (IEEE) pp 7034–41

Ouala S, Brunton S L, Pascual A, Chapron B, Collard F, Gaultier L and Fablet R 2023 Bounded nonlinear forecasts of partially observed geophysical systems with physics-constrained deep learning *Phys. D: Nonlinear Phenom.* **446** 133630

Ouala S, Nguyen D, Drumetz L, Chapron B, Pascual A, Collard F, Gaultier L and Fablet R 2020 Learning latent dynamics for partially observed chaotic systems *Chaos* **30** 103121

Rice J, Wenwei X and August A 2020 Analyzing Koopman approaches to physics-informed machine learning for long-term sea-surface temperature forecasting (arXiv:2010.00399)

Rowley C W, Shervin Bagheri I M, Schlatter P and Henningson D S 2009 Spectral analysis of nonlinear flows *J. Fluid Mech.* **641** 115–27

Schmid P J 2010 Dynamic mode decomposition of numerical and experimental data *J. Fluid Mech.* **656** 5–28

Takeishi N, Kawahara Y and Yairi T 2017b Subspace dynamic mode decomposition for stochastic Koopman analysis *Phys. Rev. E* **96** 033310

Takeishi N, Kawahara Y and Yairi T 2017a Learning Koopman invariant subspaces for dynamic mode decomposition (arXiv:1710.04340)

Tandeo P, Ailliot P, Bocquet M, Carrassi A, Miyoshi T, Pulido M and Zhen Y 2018 Joint estimation of model and observation error covariance matrices in data assimilation: a review (arXiv:1807.11221)

Wanner M T 2020 *Robust Approximation of the Stochastic Koopman Operator* (Santa Barbara: University of California)

Williams M O, Kevrekidis I G and Rowley C W 2015 A data–driven approximation of the Koopman operator: extending dynamic mode decomposition *J. Nonlinear Sci.* **25** 1307–46

Yeung E, Kundu S and Hodas N 2019 Learning deep neural network representations for Koopman operators of nonlinear dynamical systems *2019 American Control Conf.* (*ACC*) (IEEE) pp 4832–9

Zhen Y, Chapron B, Mémin E and Peng L 2022 Eigenvalues of autocovariance matrix: a practical method to identify the Koopman eigenfrequencies *Phys. Rev. E* **105** 034205