Physica D: Nonlinear Phenomena
April 2023, Volume 446 Pages 133630 (19p.)
https://doi.org/10.1016/j.physd.2022.133630
https://archimer.ifremer.fr/doc/00815/92707/

Bounded nonlinear forecasts of partially observed geophysical systems with physics-constrained deep learning

Ouala Said [1, *], Brunton Steven L. [2], Chapron Bertrand [3], Pascual Ananda [4], Collard Fabrice [5], Gaultier Lucile [5], Fablet Ronan [1]

[1] IMT Atlantique; Lab-STICC, 29200 Brest, France
[2] University of Washington, USA
[3] Ifremer, Lops, 29200 Brest, France
[4] IMEDEA, UIB-CSIC, 07190 Esporles, Spain
[5] ODL, 29200 Brest, France

* Corresponding author : Said Ouala, email address : said.ouala@imt-atlantique.fr

sbrunton@uw.edu ; Bertrand.Chapron@ifremer.fr ; ananda.pascual@imedea.uib-csic.es ; dr.fab@oceandatalab.com ; lucile.gaultier@oceandatalab.com ; ronan.fablet@imt-atlantique.fr

Abstract :

The complexity of real-world geophysical systems is often compounded by the fact that the observed measurements depend on hidden variables. These latent variables include unresolved small scales and/or rapidly evolving processes, partially observed couplings, or forcings in coupled systems. This is the case in ocean-atmosphere dynamics, for which unknown interior dynamics can affect surface observations. The identification of computationally-relevant representations of such partially-observed and highly nonlinear systems is thus challenging and often limited to short-term forecast applications. Here, we investigate the physics-constrained learning of implicit dynamical embeddings, leveraging neural ordinary differential equation (NODE) representations. In particular, we restrict the NODE representation to linear-quadratic dynamics and enforce a global boundedness constraint, which promotes the generalization of the learned dynamics to arbitrary initial conditions. The proposed architecture is implemented within a deep learning framework, and its relevance is demonstrated with respect to state-of-the-art schemes for different case studies representative of geophysical dynamics.

Highlights

► Neural ODEs are studied for the modeling of partially observed dynamical systems. ► The models are constrained to remain bounded using the Schlegel boundedness theorem. ► This framework allows deriving dynamical models with good asymptotic behaviors. ► These models also have good generalization properties for unprecedented initial conditions.

Keywords : partially-observed systems, Embedding, Boundedness, Deep learning, Neural ODE, Forecasting

1 Introduction

Modeling the dynamics of complex systems is one of the key drivers of technological development. Dynamical systems provide interpretable models of the evolution of real-world phenomena, enabling understanding and prediction. There are a number of practical and theoretical limitations that arise when characterizing and modeling physical phenomena, such as the lack of known governing equations [1], the presence of hidden variables that are not measured directly [2], and difficulties in assessing and enforcing known symmetries and stability [3, 4, 5]. These limitations motivate the development of new techniques to advance the current state-of-the-art in model discovery. The current literature is largely based on optimization techniques [6] and artificial intelligence (AI) [7] algorithms with the aim of discovering new models, governing equations, conservation laws and symmetries for complex systems [8, 9, 10, 11, 1, 12, 13] for a variety of applications, ranging from system identification [14, 1, 15, 16, 17, 18], forecasting [19, 20] and reconstruction [21, 22, 23] to control [24, 25].

One of the new challenges of data-driven methods and learning-based approaches concerns the exploitation and development of state-of-the-art techniques to tackle the modeling of complex systems such as the atmosphere, the ocean, and the climate. These fields encountered a sensing revolution in the last years, which sets up the community with databases that enable the use of AI. In this context, significant advances in model reduction techniques [26, 27], interpolation and reconstruction of gap-free observations [21, 28, 29] and short term forecast [30] have been achieved by AI models. These models, when compared to classical physical models, issued for instance from the Navier-Stokes equations, are easier to handle from a computational point of view, they also often offer a more flexible application range with the drawback of being bad learners. AI models often fail to generalize to situations (or initial conditions) that were never seen in the training data. These models also often fail at matching physical models in their ability to simulate the dynamics of the observations. From a pure learning perspective, generalization, and long-term simulation performance of data-driven schemes are linked to physical properties of the underlying dynamics (such as conserved quantities), and most of the time, these properties are not revealed by the complexity of the training data.

In this context, a central question concerns the ability of data-driven models to capture the behavior of a system from partial and imperfect observations. Takens theorem [31] states the conditions under which a delay embedding representation of the observations is diffeomorphic to the underlying hidden states. Such a representation, when determined, captures essential information about the hidden states, which makes it possible to learn a model [2]. Unfortunately, the derivation of a dynamical system from such a representation remains a challenge, since no explicit relationships between the reconstructed phase space variables and the original hidden variables have been clearly identified. To address this limitation, the Neural embedding of Dynamical Systems (NbedDyn), proposed in [32], allows for both reconstructing and forecasting the phase space of the unseen dynamical system. However, this model suffers from generalization and stability issues, especially for initial conditions that are not near the data used to train the model.

It has been shown that embedding prior physical knowledge, such as invariances, symmetries, and conservation laws, into the learning process of dynamical systems may improve the generalizability and stability of the resulting models [3, 33, 34, 35, 36, 4]. Within this new field of research, there is an increasing focus on incorporating stability and boundedness constraints into data-driven models. This interest is driven by a combination of works, on different classes of models ranging from sparse regression techniques [4] to deep learning architectures [37, 38], showing that such constraints can be necessary in order to promote stability and generalization of the models.

Following these developments, we propose a robust framework for the modeling of geophysical dynamical systems from data. This framework relies on the learning of an embedding of the observations governed by a neural ODE (NODE) model [13]. In this work, we constrain our models to be globally bounded by restricting the form of the NODE to be linear-quadratic, which enables the use of a generalization of the direct Lyapunov method given by the Schlegel-Noack boundedness theorem [39]; similar approaches have recently been used to assess the stability of data-driven models [5] and has been incorporated into sparse nonlinear models [4]. In different case studies, we show that once these constraints are satisfied, the trained representation can reproduce realistic trajectories with respect to the training data using a closed-loop prediction setting. The boundedness constraint also improves the generalization capabilities of the trained model for arbitrary initial conditions, even those lying outside the attractor spanned by the training data. This latter property is extremely important since the generalization performance is a critical feature for data-driven schemes. Overall, key aspects of the proposed framework are three-fold:

- we define a physics-constrained learning-based workflow for the modeling and forecasting of partially observed dynamical systems, with an emphasis on geophysical dynamics;

- the proposed framework relies on learning an implicit higher-dimensional embedding. The dynamics of this embedding are constrained to be bounded by construction, which promotes the generalization of models for arbitrary initial conditions, even those far from the training data;

- the relevance of the proposed scheme is tested against state-of-the-art methods for short-term forecasting and long-term simulation on several case studies, namely the Lorenz 63 and Lorenz 96 systems and the shallow water equation dynamics. Specifically, our model is compared to the sparse regression (SR) technique [1], to a recurrent neural network (RNN) and to an extended dynamic mode decomposition (EDMD) technique based on delay embedding observables [40]. Through all experiments, the proposed framework is the only model able to achieve simultaneously both an accurate short-term forecast and long-term simulation of the dynamics.

Regarding the data-driven identification of real-world geophysical dynamics, we believe that this proposed framework provides a promising approach for learning consistent models in terms of short-term and long-term forecasts through the implementation of physical constraints from prior knowledge of the conservation laws governing the dynamics.

The paper is organized as follows. We start in section 2 by introducing state-of-the-art literature on data-driven dynamical systems and physics-based learning. We then briefly discuss, in section 3, stability criteria, both from a dynamical systems theory and an identification perspective. We also introduce a generalization of the direct Lyapunov theorem as proposed in [39]. Section 4 presents the proposed framework, followed by the experiments and results in Section 5. We close this paper with conclusions and perspectives for future works in Section 6.

2 Related work

Data-driven identification of dynamical representations. As stated above, recent advances in learning-based approaches motivate numerous investigations to apply these tools for the identification of dynamical representations from data. A full review of state-of-the-art techniques is out of the scope of this paper, and we concentrate on providing a high-level overview, organized by the nature of the training data. Specifically, when provided with observations that form an embedding of a deterministic time-evolving system, several works including symbolic or dictionary-based approaches [41, 42, 43, 1, 44], neural-network and deep learning approaches [45, 46, 13, 47] and model-free representations [48, 12] demonstrated the ability of data-driven techniques to reverse engineer the governing equations from a sequence of observations. When provided with noisy and sparse observations, studies have investigated either deep learning generative architectures, such as variational autoencoders [49, 50, 51], or classical Bayesian filtering schemes such as the ensemble Kalman filter [52], in deriving reconstructions of the irregular observations that can be useful from an identification perspective [53, 54, 17].

When provided with partial observations, *i.e.* with no one-to-one mapping between the observation space and the underlying dynamical system, most state-of-the-art techniques do not apply since the variability of the observations is influenced by some missing components, making the temporal evolution of the observations stochastic in the observation space. From this point of view, the deep learning literature in time series forecasting mainly relies on some sort of recurrent neural network (RNN) architectures [55, 56, 57] which can be motivated from a dynamical system perspective as a temporal propagator of a delay embedding of the observations. Recent work has connected SINDy with time-delay autoencoders to reconstruct hidden variables, with promising results [2].

Methods to learn linear propagators of non-linear, and potentially high-dimensional dynamical systems have also emerged in the context of the modern

Koopman theory [58]. Finding a suitably chosen transformation/embedding of the observations that can be approximately propagated linearly in time has motivated several works [59, 60, 61, 62, 63, 64, 40, 65, 66, 67] on a variety of applications, including control [68] and prediction [65].

Physics-informed data-driven dynamical representations. In the context of dynamical model identification, several works have shown that including physical knowledge has a positive impact on the data-driven models [69, 3, 46, 70, 35, 71, 34, 36, 72, 73, 74, 75, 76, 77]. For instance, [1] proposed a sparse regression framework for the derivation of interpretable dynamical representations from data and [3] extended this framework to satisfy physical constraints, such as known energy preserving symmetries in the dynamics. In neural-networks and deep learning, significant effort has been focused on understanding residual networks [78, 79] as numerical integration schemes of differential equations [80, 15, 47, 81, 82]. In this line of study, neural ordinary differential equations [13, 83, 84, 85, 86, 87] have shown great success in merging neural networks and ODE identification techniques.

There is also a growing interest in developing hybrid representations that merge both a physical model and a machine learning component. Typical examples can be found in the context of closure modeling [88, 89, 90, 91, 92, 27] in which, broadly speaking, a machine learning model is optimized on data to correct a physical model.

Stability and boundedness of data-driven dynamical systems. Exploiting stability constraints when learning dynamical systems from data usually improves the long-term stability and generalization performance of the trained models [93, 94, 95, 96, 4]. Recent works further investigated Lyapunov functions of dynamical systems for control applications [97] or to force the stability around a global equilibrium point [37]. In this context, [38] proposed a new framework that guarantees exponential stability of a NODE model. Promoting Lyapunov stability in autoencoders was also investigated in [98] with applications to reduced-order models of two-dimensional turbulent fluid flows. Beyond the direct Lyapunov method, the Schlegel-Noack boundedness theorem [39] was also considered in [4] to promote the global stability of linear quadratic data-driven dynamical systems.

3 From Stability to Boundedness, in the context of identification

This section introduces some common stability criteria for the analysis of nonlinear models. We briefly discuss the implementation of these criteria and focus on boundedness constraints as we further explore their relevance in regularizing the learning of chaotic dynamics.

3.1 Stability of limit-sets

Let us assume a continuous s-dimensional dynamical system $\mathbf{z}_t$ governed by an autonomous ODE $\dot{\mathbf{z}}_t = f(\mathbf{z}_t)$ with Φ_t the corresponding flow $\Phi_t(\mathbf{z}_{t_0}) = \mathbf{z}_{t_0} + \int_{t_0}^{t} f(\mathbf{z}_w)dw$. The trajectories of this dynamical system are assumed to be asymptotic to a limit-set $\mathcal{S}$ of dimension p contained in $\mathbb{R}^s$.

Stability theory addresses the characterization of the asymptotic behavior of a set of solutions of a differential equation with respect to a given limit-set $\mathcal{S}$. Formally, and assuming for the sake of simplicity that $\mathcal{S}$ is an equilibrium point $\mathbf{z}_{eq}$, we may distinguish the following stability definitions:

- The equilibrium point $\mathbf{z}_{eq}$ is globally (respectively locally) stable if $\forall \mathbf{z}_0 \in \mathbb{R}^s$ (respectively $\mathbf{z}_0 \in \mathcal{U} \subset \mathbb{R}^s$), $\Phi_t(\mathbf{z}_0) \longrightarrow \mathbf{z}_{eq} + \epsilon$ with $|\epsilon| > 0$ and finite;
- The equilibrium point $\mathbf{z}_{eq}$ is globally (respectively locally) asymptotically stable if $\forall \mathbf{z}_0 \in \mathbb{R}^s$ (respectively $\mathbf{z}_0 \in \mathcal{U} \subset \mathbb{R}^s$), $\Phi_t(\mathbf{z}_0) \longrightarrow \mathbf{z}_{eq}$ as $t \longrightarrow \infty$;
- The equilibrium point $\mathbf{z}_{eq}$ is globally (respectively locally) exponentially stable if $\forall \mathbf{z}_0 \in \mathbb{R}^s$ (respectively $\mathbf{z}_0 \in \mathcal{U} \subset \mathbb{R}^s$), $|\Phi_t(\mathbf{z}_0) - \mathbf{z}_{eq}| \leq C|\mathbf{z}_0 - \mathbf{z}_{eq}|e^{-\alpha t}$ with α a convergence rate, C a positive constant and $|\cdot|$ a given norm in $\mathbb{R}^s$.

Broadly speaking, we may state that a limit-set is i) stable if two nearby trajectories stay close by the action of the vector field, ii) asymptotically stable if a sufficiently close trajectory is attracted to the limit-set, and finally, iii) exponentially stable if a trajectory converges to the limit-set with an exponential decay rate.

Depending on the limit-set of a given ODE, there are several approaches to analyze its stability. A great introduction to classical stability criteria is given in [99], starting from the classical eigenvalues of a linear (or linearized) system around an equilibrium point and finishing with the Lyapunov exponents of chaotic attractors. From an identification perspective, and given some observation data, stability criteria can be used as constraints, either to reproduce an observed asymptotic behavior or to avoid blowups, as long as the attractor revealed by the observations is not strange *i.e.* chaotic. Chaotic solutions of differential equations are only revealed through criteria that exploit long-term simulations of the system such as Lyapunov exponents. Therefore, they cannot be characterized based on the dynamical equation (such as the eigenvalues of an equilibrium point).

Overall, one cannot explicitly constrain a system to be chaotic, simply by manipulating its differential equation. This point appears critical since numerous geophysical systems can be chaotic. Alternatively, we propose to relax the problem by moving from constraining chaos to constraining the boundedness of a system. Such a boundedness constraint applies to every single limit-set mentioned above and would avoid blowups of data-driven models. As highlighted by [99], this constraint is also a natural feature of real-world geophysical systems as, from an experimental perspective, blowups cannot be observed in nature. For this purpose, we start by introducing the direct Lyapunov stability criterion, since it provides global stability properties on non-linear dynamics. A generalization of this approach to the boundedness of linear quadratic ODEs as proposed in [39] is then presented.

3.2 Lyapunov stability of dynamical systems

The direct Lyapunov stability method [100] was introduced to study the stability of any dynamical system that admits an equilibrium point at the origin. It uses a scalar function of the state $V : \mathbb{R}^s \longrightarrow \mathbb{R}$ as follows:

$$\begin{aligned} &V(\mathbf{z}) = 0 \text{ if and only if } \mathbf{z} = 0 \\ &V(\mathbf{z}) > 0 \text{ if and only if } \mathbf{z} \neq 0 \\ &\dot{V}(\mathbf{z}) \ \leq 0 \ \forall \mathbf{z} \neq 0 \end{aligned} \tag{1}$$

If $V(\mathbf{z})$ satisfies the above conditions $\forall \mathbf{z} \in \mathbb{R}^s$ (respectively $\forall \mathbf{z} \in \mathcal{U} \subset \mathbb{R}^s$) the system is globally (respectively locally) stable. Furthermore, if $\dot{V}(\mathbf{z}) < 0 \forall \mathbf{z} \neq 0$ the asymptotic stability of the system is also guaranteed.

We can apply this criterion to any non-linear model without resorting to any linearization, which makes it particularly appealing. Yet, finding an appropriate function is far from being straightforward, and several works proposed candidate Lyapunov functions for various types of problems [101].

3.3 Generalization to the boundedness of Linear Quadratic Models (LQMs)

The direct Lyapunov stability method is restricted to dynamical systems with an equilibrium point at the origin. This property is restrictive since it does not apply to other dynamical regimes such as periodic and chaotic solutions. Furthermore, the choice of the Lyapunov function being non-systematic, the exploitation of a direct Lyapunov constraint for the data-driven identification of dynamical systems is restricted to a small class of parametric models. Fortunately, Schlegel and Noack [39] proposed a generalization of the direct Lyapunov method on a class of parametric differential equations. This work also introduces the associated Lyapunov function, along with a condition for the existence of a monotonically attracting trapping region in the phase space. These trapping regions were later incorporated into nonlinear model identification to improve stability [4].

A trapping region is a domain in the phase space that locks trajectories of a dynamical system, *i.e.*, once a trajectory enters a trapping region, it will stay in this domain as the system evolves [102]. When this region is globally monotonically attractive, all trajectories in the phase space will converge to the trapping region. We may point out that a trapping region can contain a single or multiple limit-sets. The class of models for which the proposed criterion in [39] is valid is LQMs. They can be encountered, for instance, by a spectral discretization of the Navier-Stokes equation, which makes them appealing in the context of reduced order models (ROMs). Formally, as proposed in [39], let us rewrite the dynamical system governing $\mathbf{z}$ as a linear quadratic model

$$\dot{\mathbf{z}}_t = \mathbf{c} + \mathbf{L}\mathbf{z}_t + [\mathbf{z}_t^T\mathbf{Q}^{(1)}\mathbf{z}_t, ..., \mathbf{z}_t^T\mathbf{Q}^{(s)}\mathbf{z}_t]^T \tag{2}$$

where $\mathbf{c} \in \mathbb{R}^s$, $\mathbf{L} \in \mathbb{R}^{s\times s}$ and $\mathbf{Q}^{(i)} = [q_{i,j,k}]^s_{j,k=1}, i = 1, ..., s$. The s symmetric quadratic matrices $\mathbf{Q}^{(i)}$, $i = 1, ..., s$ are supposed to be energy preserving, *i.e.*

$$q^i_{j,k} + q^j_{i,k} + q^k_{i,j} = 0, i, j, k = 1, ..., s \tag{3}$$

Let us also consider a shifted variable $\bar{\mathbf{z}} = \mathbf{z} - \mathbf{m}$ with $\mathbf{m} \in \mathbb{R}^s$ an arbitrary finite state. The dynamical equation of the shifted state can be written as

$$\dot{\bar{\mathbf{z}}}_t = \mathbf{d} + \mathbf{A}\bar{\mathbf{z}}_t + [\bar{\mathbf{z}}_t^T\mathbf{Q}^{(1)}\bar{\mathbf{z}}_t, ..., \bar{\mathbf{z}}_t^T\mathbf{Q}^{(s)}\bar{\mathbf{z}}_t]^T \tag{4}$$

with

$$\mathbf{d} = \mathbf{c} + \mathbf{L}\mathbf{m} + [\mathbf{m}^T\mathbf{Q}^{(1)}\mathbf{m}, ..., \mathbf{m}^T\mathbf{Q}^{(s)}\mathbf{m}]^T \tag{5}$$

and

$$\mathbf{A} = \Big(a_{ij}\Big) = \Big(l_{ij} + \sum_{k=1}^{s}(q_{i,j,k} + q_{i,k,j})m_k\Big) \tag{6}$$

The evolution of the fluctuation energy $K = \frac{1}{2}\sum_{i=1}^{s}\bar{\mathbf{z}}_i^2$ of the shifted system is considered as a Lyapunov function [39]. The time derivative of this quantity can be written as:

$$\dot{K} = [\nabla_{\bar{\mathbf{z}}}\mathbf{K}]^T\dot{\bar{\mathbf{z}}} = \bar{\mathbf{z}}^T\mathbf{A}_s\bar{\mathbf{z}} + \mathbf{d}^T\bar{\mathbf{z}} \tag{7}$$

where $\mathbf{A}_s = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$. The contribution of the quadratic terms to the fluctuation energy K is zero due to the energy-preserving condition (3). A sufficient condition for the existence of a monotonically attracting trapping region is the existence of a finite $\mathbf{m}$ such that $\mathbf{A}_s$ has only negative eigenvalues. In the next section, we enforce this condition in the proposed data-driven scheme for partially-observed systems in order to train long-term bounded dynamical models.

4 Bounded Neural Embedding

This section first briefly outlines the neural embedding of dynamical systems —NbedDyn—[32], and introduces the bounded neural embedding framework.

4.1 Neural embedding of dynamical systems

The NbedDyn model allows one to jointly derive a geometric reconstruction of the unseen phase space from partial observations and a corresponding dynamical system model. When compared to classical state-of-the-art models in deep learning (e.g., recurrent neural networks, latent variable models based on variational autoencoders [17], etc.), NbedDyn formulates the latent-variable-reconstruction and dynamic prediction as a single optimization problem (15).

Let us consider a dynamical system governed by an autonomous ODE:

$$\dot{\mathbf{z}}_t = f(\mathbf{z}_t) \tag{8}$$

For most applications, the true state $\mathbf{z}_t \in \mathbb{R}^s$ of the system is unknown and limited to a series of observations $\{\mathbf{x}_t\}$:

$$\mathbf{x}_t = \mathcal{H}(\mathbf{z}_t) \tag{9}$$

where $\mathcal{H} : \mathbb{R}^s \to \mathbb{R}^n$ is an observation operator that does not satisfy the conditions [103] under which the predictable deterministic dynamics expressed in the space of $\mathbf{z}$ are still deterministic in the observation space.

The NbedDyn technique tackles this problem by searching an augmented space, where the states are governed by diffeomorphic flows and can be mapped to the observations $\mathbf{x}_t$. For any given operator $\mathcal{H}$ of a deterministic dynamical system, Takens theorem [31] guarantees that such an augmented space exists. However, instead of using a delay embedding, NbedDyn defines a d_E-dimensional augmented space with states $\mathbf{u}_t \in \mathbb{R}^{d_E}$ as follows:

$$\mathbf{u}_t^T = [\mathcal{M}(\mathbf{x}_t^T), \mathbf{y}_t^T] \tag{10}$$

where $\mathcal{M}$ is an order reduction operator such as $\mathcal{M}(\mathbf{x}_t) \in \mathbb{R}^r$ with $r \leq n$ and $\mathbf{y}_t \in \mathbb{R}^{d_E - r}$ are stated as the latent states. They account for the information of the unobserved components of the true state $\mathbf{z}_t$.

The corresponding dynamics and observation operator are defined as:

$$\dot{\mathbf{u}}_t = f_\theta(\mathbf{u_t}) \tag{11}$$

$$\mathbf{x}_t = \mathcal{M}^{-1}(\mathbf{G}\mathbf{u_t}) \tag{12}$$

where the dynamical operator f_θ belongs to a given family of neural network operators parameterized by a parameter vector θ. $\mathbf{G}$ is a projection matrix that satisfies $\mathcal{M}(\mathbf{x}_t) = \mathbf{G}\mathbf{u_t}$. Using an integration scheme (in all the experiments, the flow of the proposed framework is built on a classical Runge-Kutta 4 integration scheme), we can associate f_θ with a diffeomorphic mapping, to derive the flow Φ_θ of the NODE model (11) as follows:

$$\Phi_{\theta,t}(\mathbf{u}_{t-1}) = \mathbf{u}_{t-1} + \int_{t-1}^{t} f_\theta(\mathbf{u}_{t-1}) \tag{13}$$

From Eqs. (11), (12) and (13), we define a state space model:

$$\left\{ \begin{array}{lcl} \mathbf{u}_t & = & \Phi_{\theta,t}(\mathbf{u}_{t-1}) \\ \mathbf{x}_t & = & \mathcal{M}^{-1}(\mathbf{G}\mathbf{u_t}) \end{array} \right. \tag{14}$$

Given an observation time series of size $N+1$ $\{\mathbf{x}_0, \ldots, \mathbf{x}_{t_N}\}$, the Neural Embedding scheme aims to minimize the forecasting error of the observations with respect to the model parameters and the latent states as follows:

$$\hat{\theta}, \mathbf{y}_{t_0:t_{N-1}} = \arg \min_{\theta, \{\mathbf{y}_t\}_t} \sum_{i=1}^{N} \|\mathbf{x}_{t_i} - \mathcal{M}^{-1}(\mathbf{G}\Phi_{\theta,t_i}\left(\mathbf{u}_{t_{i-1}}\right))\,\|^2 + \lambda_1 \|\mathbf{u}_{t_i} - \Phi_{\theta,t_i}(\mathbf{u}_{t_{i-1}})\|^2 \tag{15}$$

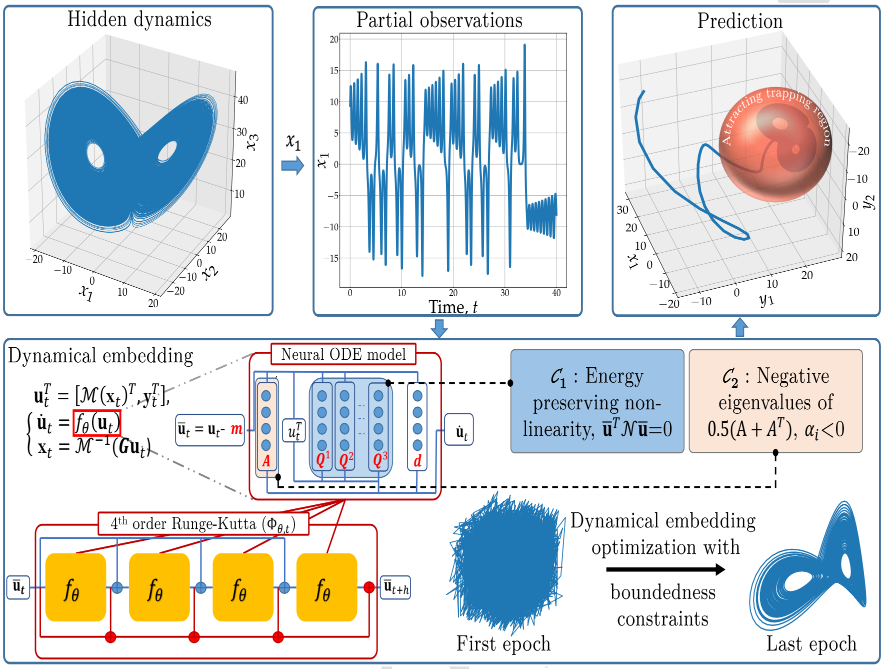

Hidden dynamics
x_1
x_2
x_3
Partial observations
x_1
Time, t
Prediction
Attracting trapping region
y_1
y_2
x_1
Dynamical embedding
$\mathbf{u}_t^T = [\mathcal{M}(\mathbf{x}_t)^T, \mathbf{y}_t^T]$,
$\dot{\mathbf{u}}_t = f_\theta(\mathbf{u}_t)$
$\mathbf{x}_t = \mathcal{M}^{-1}(\mathbf{G}\mathbf{u}_t)$
Neural ODE model
$\bar{\mathbf{u}}_t = \mathbf{u}_t - \mathbf{m}$
u_t^T
A
Q^1
Q^2
Q^3
d
$\dot{\mathbf{u}}_t$
$\mathcal{C}_1$: Energy preserving non-linearity, $\bar{\mathbf{u}}^T \mathcal{N} \bar{\mathbf{u}} = 0$
$\mathcal{C}_2$: Negative eigenvalues of $0.5(A + A^T)$, $\alpha_i < 0$
4th order Runge-Kutta ($\Phi_{\theta,t}$)
$\bar{\mathbf{u}}_t$
f_θ
$\bar{\mathbf{u}}_{t+h}$
First epoch
Dynamical embedding optimization with boundedness constraints
Last epoch

Figure 1: ***Sketch of the proposed architecture.*** Given a sequence of observations issued from an unknown higher dimensional dynamical system, the observations are projected into a higher dimensional space parameterized by a linear quadratic neural ODE model. The parameters of the models, highlighted in red, as well as the latent states are optimized to minimize the forecasting of the observations while satisfying the boundedness constraints as proposed in [39].

with λ_1 a weighting parameter and $\|\cdot\|$ the L^2 norm.

After training, the NbedDyn model can be used to forecast new observations by determining an initial condition of the unobserved component $\mathbf{y}_t$ and performing a numerical integration of the neural ODE model (13). We infer the initial condition using a minimization of an objective function similar to (15) but only with respect to the latent states $\{\mathbf{y}_t\}$. This minimization can be seen as a variational data assimilation problem, with partial observations of the state-space variables and known dynamical and observation operators [104].

4.2 Constrained dynamical embedding

In practice, the optimization problem (15) often leads to solutions that minimize the forecasting cost but have incorrect asymptotic behaviors. Specifically, when using the neural ODE model to simulate the dynamics of the observed variables, the asymptotic behavior of the resulting time series can differ significantly from that of the observations. Furthermore, even when the asymptotic behavior of the NbedDyn models are correct, they typically have an extremely small trapping domain, which may limit the effectiveness of such models, especially for data-assimilation applications. These issues relate to the generalization capabilities of the data-driven models, which may be improved by adding regularization to (15). Physical and/or stability constraints are commonly enforced. However, these regularization techniques are depend on the specific problem.

In this context, we propose to include boundedness constraints in the neural embedding model [32] and investigate their effectiveness. In contrast to specialized physical constraints or stability constraints, which may only apply to specific problems, boundedness is present in every system, as blowups are not present in reality. However, boundedness constraints depend on the form of the neural ODE model f_θ. Therefore, we restrict the form of the model f_θ, to be linear-quadratic, where it is possible to apply the attracting trapping region condition based on the Schlegel-Noack boundedness theorem [39]. The combination of Volterra theory for non-linear differential equations [105] and lifting transformation techniques [106, 107] have shown that linear-quadratic models can faithfully represent (in a higher dimension) the dynamics of non-linear differential equations. In this context, the present work studies the restriction of the neural ODE operator f_θ in (11) to a linear-quadratic architecture with boundedness constraints enforced via the Schlegel-Noack boundedness theorem [39].

A linear quadratic NODE model can be, similarly to (2), written as follows:

$$\dot{\mathbf{u}}_t = \mathbf{c} + \mathbf{L}\mathbf{u}_t + [\mathbf{u}_t^T\mathbf{Q}^{(1)}\mathbf{u}_t, ..., \mathbf{u}_t^T\mathbf{Q}^{(d_E)}\mathbf{u}_t]^T \tag{16}$$

where $\mathbf{c} \in \mathbb{R}^{d_E}$, $\mathbf{L} \in \mathbb{R}^{d_E \times d_E}$ and $\mathbf{Q}^{(i)} = [q_{i,j,k}]_{j,k=1}^{d_E}, i = 1, ..., d_E$. The above approximate model is shifted according to $\bar{\mathbf{u}} = \mathbf{u} - \mathbf{m}$ with $\mathbf{m} \in \mathbb{R}^{d_E}$. The approximate dynamical equation of the shifted state can be written as

$$\dot{\bar{\mathbf{u}}}_t = \mathbf{d} + \mathbf{A}\bar{\mathbf{u}}_t + [\bar{\mathbf{u}}_t^T\mathbf{Q}^{(1)}\mathbf{u}_t, ..., \bar{\mathbf{u}}_t^T\mathbf{Q}^{(d_E)}\bar{\mathbf{u}}_t]^T \tag{17}$$

with $\mathbf{d}$ and $\mathbf{A}$ computed according to (5) and (6) respectively.

The training setting comes to jointly learning the model parameters $\theta = \{\mathbf{c}, \mathbf{L}, \mathbf{Q^1}, \mathbf{Q^2}, \cdots, \mathbf{Q^{d_E}}, \mathbf{m}\}$ and the latent states $\mathbf{y}$ according to the following constrained optimization problem

$$\begin{aligned} \hat{\theta}, \mathbf{y}_{t_0:t_{N-1}} = \arg\min_{\theta, \{\mathbf{y}_t\}_t} & \sum_{i=1}^{N} \|\mathbf{x}_{t_i} - \mathcal{M}^{-1}(\mathbf{G}(\Phi_{\theta,t_i}\left(\mathbf{u}_{t_{i-1}}\right) + \mathbf{m}))\|^2 \\ & + \lambda_1\|\mathbf{u}_{t_i} - \Phi_{\theta,t_i}(\mathbf{u}_{t_{i-1}})\|^2 \\ & + \lambda_2\mathcal{C}_1 \\ & + \lambda_3\mathcal{C}_2 \end{aligned} \tag{18}$$

with $\mathcal{C}_1 = \sum_{i,j,k=1}^{s} \|q_{i,j,k} + q_{i,k,j} + b_{j,i,k} + b_{j,k,i} + b_{k,i,j} + b_{k,j,i}\|^2$ and $\mathcal{C}_2 = \sum_{i=1}^{s} \text{Max}(\alpha_i, 0)/\text{Max}(\alpha_i + 1, 0)$ where $\alpha_i, i = 1, ..., d_E$ the eigenvalues of the matrix $\mathbf{A}_s = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$. The variables $\lambda_{1,2,3}$ are constant weighting parameters. This loss function corresponds to the initial NbedDyn loss given by (15) with two additional constraints $\mathcal{C}_1$ and $\mathcal{C}_2$. The first constraint $\mathcal{C}_1$ steams from the energy-preserving condition given in (3). It forces the contribution of the quadratic terms of f_θ to the fluctuation energy to sum up to zero. The second constraint, $\mathcal{C}_2$, ensures that the eigenvalues of $\mathbf{A}_s$ are negative. Satisfying these constraints guarantees that the model f_θ is bounded through the existence of a monotonically attracting trapping region that includes the limit-set revealed by the minimization of the forecasting loss. Overall, Fig. 1 highlights the main mechanisms of the proposed framework.

5 Numerical experiments

5.1 Experimental setup

Numerical experiments are performed to evaluate the proposed framework for the forecasting and simulation of partially-observed systems. We consider three case-studies: the Lorenz-63, Lorenz-96, and Shallow Water Equation (SWE) dynamics. For these dynamical systems, we use data sampled from the steady-state attractor of the dynamics. Specifically, we run the dynamical models up to 5 Lorenz times in the Lorenz 63 and 96 experiments, and we use the last point of the simulation as an initial condition for the training data to ensure that the system is on the attractor. In the SWE experiment, we achieve a similar result by omitting the initial transient dynamics. Thus, the choice of initial condition in the training data should have a minor impact on the models. In the testing phase, we treat two scenarios, the first one (S1) consists in picking an initial condition close to the attractor revealed by the reconstructed phase space, and a second scenario (S2) where the initial conditions are picked far from the attractor revealed by the training data. The first scenario evaluates the prediction and simulation performance of the models for initial conditions that are inferred from the minimization of a variational data-assimilation problem as explained in section 4.1. The second scenario evaluates the generalization ability of the models to initial conditions that were never seen in the training set. Overall, a good model is the one able at producing a good prediction score in (S1), and a similar long-term simulation with respect to the ones of the true dynamics in both (S1) and (S2). In this context, we compare the Largest Lyapunov exponent [1], and/or the spectral properties of the simulations for all the case studies. We also provide the short-term prediction performance of the models for (S1). In both cases, the provided scores are averaged over 100 realizations. These realizations are generated from the minimization of a data assimilation cost over observations that were randomly sampled in the test set

[1] The largest Lyapunov exponents λ_1 are only reported when the models have an asymptotic behavior that is bounded and unstable which means that the identified systems are chaotic. When the models produce periodic/quasi-periodic simulations, we report $\lambda_1 \leq 0$ and when they diverge to infinity we omit the Lyapunov exponents.

for (S1). For (S2) we randomly sample the initial conditions from a Gaussian distribution with 0 mean and a standard deviation of 20, 10, and 10 for the Lorenz 63, Lorenz 96, and the Shallow Water Equation experiments, respectively. In all the experiments that involve a stochastic gradient descent, we trained an ensemble of 5 models and provided scores that include the mean and a standard deviation of the performance of each model. The code associated to this work can be found at https://github.com/CIA-Oceanix/Bounded-NbedDyn.

5.2 Lorenz 63

The Lorenz-63 dynamical system is a 3-dimensional model governed by the following ODE:

$$\begin{cases} \frac{dz_{t,1}}{dt} &= \sigma\left(z_{t,2} - z_{t,2}\right) \\ \frac{dz_{t,2}}{dt} &= \rho z_{t,1} - z_{t,2} - z_{t,1}z_{t,3} \\ \frac{dz_{t,3}}{dt} &= z_{t,1}z_{t,2} - \beta z_{t,3} \end{cases} \tag{19}$$

Under parametrization $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$, this system exhibits chaotic dynamics with a strange attractor [108].

We simulate a Lorenz-63 state sequence of 5000 time steps, sampled regularly at $dt = 0.01$. This simulation is computed using the LOSDA ODE solver [109]. The training sequence consists of the initial 4000 measurements of the first Lorenz-63 variable, *i.e.*, $\mathcal{H} = [1, 0, 0]$ and $\mathbf{x}_t = \mathcal{H}(z_t) = \mathbf{z}_{t,1}$. The remaining time steps are used as a test set. In this experiment, the projection operator in (10) is the identity matrix, *i.e.* $\mathcal{M}(\cdot) = I_1$. Please refer appendix B and C respectively for similar experiments with varying sizes and time samplings of the training datasets.

Parameterization of the data-driven models: For benchmarking purposes, we perform a quantitative comparison with other leading approaches using delay embedding representations [31]. The parameters of the delay embedding representation, namely the lag τ and the dimension d_E of the augmented space, were chosen using standard techniques. Specifically, the lag parameter was computed using both the mutual information and correlation techniques [110], respectively denoted as τ_{MI} and τ_{Corr}. Regarding the dimension of the embedding representation, we used the Whitney's embedding condition $d_E = 2\text{p} + 1$ with p the dimension of the hidden limit-set. The delay embedding dimension was also computed using the false nearest neighbors (FNN) method [111]. We also tested arbitrary parameters for the delay embedding dimension. Given the delay embedding representation, we tested the Sparse Regression (SR) method [1] on a second-order polynomial representation of the delay embedding states. The time integration of the SR model is carried out using the LOSDA solver [109]. Regarding deep learning models, we compare our method to a stacked Bidirectional LSTM (RNN). The RNN model includes 4 LSTM layers with 10 hidden units. The input sequence of the model is an observation sequence of size 30.We skip the evaluation of the RNN for (S2) (here and in all the considered case studies) since this model is not expected to work for input sequences that are randomly sampled and were never seen in the training data.

The proposed framework is tested with a dimension of the augmented state space $d_E = 3$, *i.e.*, two latent states are concatenated to the observed variable. This dimension is the one given by the FNN algorithm and corresponds to the true dimension of the Lorenz 63 system. Using a d_E that is smaller than 3 would result in an inaccurate model, and a higher d_E adds spurious states that increase the complexity of the model (please refer to appendix A for an analysis of the impact of the dimension of the embedding on the performance of the model). Overall the augmented state $\mathbf{u}$ is constructed as $\mathbf{u}_t = [\mathbf{z}_{1,t}, \mathbf{y}_{1,t}, \mathbf{y}_{2,t}]^T$. The parameters of the data-driven model governing $\mathbf{u}$, as well as the latent states $\mathbf{y}_{1,t}$ and $\mathbf{y}_{2,t}$ are optimized based on both i) the initial NbedDyn formulation presented as proposed in [32] (*i.e.*, using the optimization criterion given by (15)) and ii) the constrained version introduced in this work, illustrated by criterion (18).

Model		$t_0 + dt$	$t_0 + 4dt$	λ_1 (S1)	λ_1 (S2)
SR	$\tau_{MI} = 18$, $d_E(FNN) = 3$	0.040 ± 0.045	0.138 ± 0.146	/	/
	$\tau_{MI} = 18$, d_E(Whitney)$= 6$	0.031 ± 0.030	0.124 ± 0.138	/	/
	$\tau_{Corr} = 35$, d_E(FNN)$= 3$	0.305 ± 0.196	1.171 ± 0.815	≤ 0	≤ 0
	$\tau_{Corr} = 35$, d_E(Whitney)$= 6$	0.237 ± 0.250	1.114 ± 1.541	/	/
	$\tau = 6$, $d_E = 3$	0.004 ± 0.004	0.009 ± 0.008	≤ 0	≤ 0
RNN (averaged)		0.068 ± 0.056	0.175 ± 0.137	≤ 0	/
RNN (best)		0.055 ± 0.055	0.118 ± 0.055	≤ 0	/
RNN (worst)		0.083 ± 0.067	0.236 ± 0.067	≤ 0	/
NbedDyn	$d_E = 3$	0.021 ± 0.027	0.083 ± 0.104	0.967 ± 0.028	/
NbedDyn (best)	$d_E = 3$	0.021 ± 0.026	0.082 ± 0.102	0.922 ± 0.027	/
NbedDyn (worst)	$d_E = 3$	0.021 ± 0.029	0.084 ± 0.108	0.922 ± 0.033	/
Constrained NbedDyn	$d_E = 3$	0.012 ± 0.008	0.056 ± 0.040	0.914 ± 0.003	0.833 ± 0.003
Constrained NbedDyn (best)	$d_E = 3$	0.012 ± 0.008	0.055 ± 0.040	0.914 ± 0.015	0.902 ± 0.018
Constrained NbedDyn (worst)	$d_E = 3$	0.013 ± 0.007	0.060 ± 0.038	0.912 ± 0.015	0.894 ± 0.020

Table 1: ***Forecasting performance on the test set of the benchmarked data-driven models for Lorenz-63 case-study***: first two columns: root mean square error (RMSE) for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under (S2) (The true largest Lyapunov exponent of the Lorenz 63 model is 0.91 [112]). The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

Forecasting performance of the proposed data-driven models: We evaluate in Table 1 the short-term forecast as well as the topological structure of the simulated limit-sets (illustrated for instance through the largest Lyapunov exponent). Regarding the short-term forecast, properly tuning the delay embedding parameters for the SR technique is key in order to achieve the best possible prediction score. This tuning is significantly more challenging in more complex case studies (we do not compare the SR technique in the next case studies due to poor results). The proposed Constrained NbedDyn model outperforms the remaining benchmark, including the classical NbedDyn formulation as proposed in [32]. We suspect that the boundedness constraints regularize the variational data assimilation step carried in the prediction experiment which results in better prediction performance. When considering the long-term simulation performance of the tested data-driven models, when the initial condition is inside the spanned attractor of the augmented states, the dynamical model optimized using the criterion (15) gives trajectories that are bounded and with similar topological characteristics to the true Lorenz 63 model. However, when the initial condition is far from the spanned attractor, the model optimized through (15) diverges to infinity. From a machine learning perspective, this is the direct consequence of a poor generalization performance to states that are far from the attractor spanned by the training data. In this situation, our model involves several attracting regions of chaotic and unstable solutions. When the initial condition is far from the spanned attractor, the state evolution is dominated by a positive energy growth, which makes our model diverge to infinity. The constrained model, on the other hand, satisfies elementary conservation constraints that are present in the actual Lorenz 63 system and leads to a bounded behavior with a large attracting region of the chaotic limit-set. We further illustrate these conclusions with forecasting examples depicted in Figure 2 with initial conditions lying both inside and outside the attractor. These results also hold for different configurations of the training data as shown in appendix B and C. Increasing the size of the training data-set and reducing the time-sampling of the observations does not impact the topological scores of the classical NbedDyn models which stresses the importance of the proposed constrained version.

5.3 Lorenz 96

The Lorenz-96 system [113] with periodic boundary conditions is governed by the following ODE:

$$\frac{d\mathbf{z}_{t,i}}{dt} = (\mathbf{z}_{t,i+1} - \mathbf{z}_{t,i-2})\mathbf{z}_{t,i-1} - \mathbf{z}_{t,i} + F \tag{20}$$

with $\mathbf{z}_{t,-1} = \mathbf{z}_{t,s}$, $\mathbf{z}_{t,s+1} = \mathbf{z}_{t,1}$.

The Lorenz-96 state sequences is also simulated using the LOSDA ODE solver [109] with $F = 8$ and a dimension $s = 40$. Similar to the previous experiment, the numerical simulation consists of 5000 time steps sampled at $dt = 0.01$. The initial 4000 measurements of the first 20 states, *i.e.*, $\mathbf{x}_t = \mathcal{H}(z_t) = [\mathbf{z}_{1,t}, \cdots, \mathbf{z}_{20,t}]^T$ are used as training data. The remaining time steps are used as a test set. In this experiment, the projection operator in (10) is the identity matrix, *i.e.* $\mathcal{M}(\cdot) = I_{20}$. We also report similar experiments with varying sizes and time samplings of the training data in the appendix B and C respectively.

Parametrization of the data-driven models : The proposed framework, is tested with a dimension of the augmented state space $d_E = 40$ *i.e.*, 20 latent states are concatenated to the observed variable. Please refer to appendix A for an analysis of the dimension of the embedding for the Lorenz 96 case study. In this experiment, the quadratic part of the NODE model f_θ is convolutional and corresponds to the true non-linear interactions in (20) with trainable weights. Such representation significantly accelerates the computation of the constraint

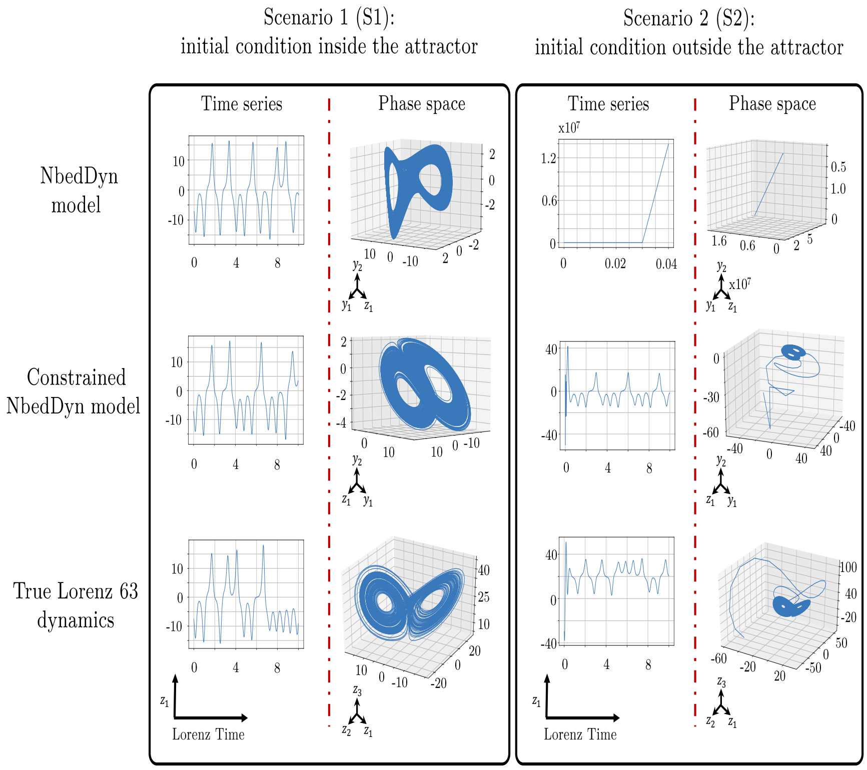


Figure 2: ***Forecasting example of the NbedDyn models for Lorenz 63 case-study.*** First row, NbedDyn model as presented in [32]; second row, proposed constrained NbedDyn model; third row, true Lorenz 63 model. The left (resp. right) panel illustrates the forecasting performance of the NbedDyn for an initial condition inside (resp. outside) the attractor. It is worth noting that the scales of the forecasted latent states $\mathbf{y}_1$ and $\mathbf{y}_2$ are different from the ones of the unseen states $\mathbf{z}_1$ and $\mathbf{z}_2$. This stems from the fact that given partial observations of the system, several models can reflect the variability of the observations while being diffeomorphic to the actual governing dynamics.

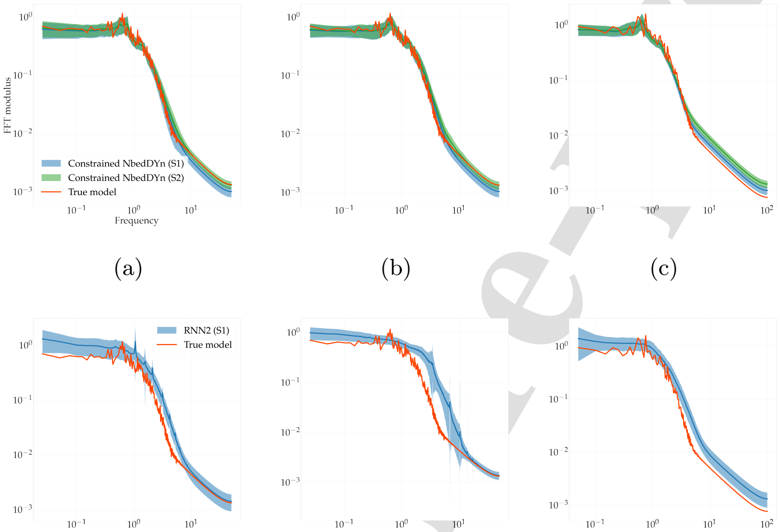
$\mathcal{C}_1$ in (18) [2] We compare both the constrained and unconstrained versions of the NbedDyn to a stacked bidirectional LSTM built under two configurations RNN1 and RNN2. We use the same configuration as the one of the Lorenz 63 experiments with 4 LSTM layers with 10 hidden units in RNN1 and increase the complexity of the model with 10 LSTM layers with 100 hidden units in RNN2.

Model	$t_0 + dt$	$t_0 + 4dt$	λ_1 (S1)	λ_1 (S2)
RNN1	2.129 ± 0.417	2.236 ± 0.408	≤ 0	/
RNN1 (best)	2.102 ± 0.437	2.194 ± 0.437	≤ 0	/
RNN1 (worst)	2.186 ± 0.416	2.329 ± 0.416	≤ 0	/
RNN2	0.392 ± 0.115	0.688 ± 0.184	2.826 ± 1.923	/
RNN2 (best)	0.380 ± 0.095	0.679 ± 0.095	0.913 ± 0.152	/
RNN2 (worst)	0.404 ± 0.131	0.697 ± 0.131	4.740 ± 0.231	/
NbedDyn	0.459 ± 0.932	/	/	/
NbedDyn (best)	0.323 ± 1.801	/	/	/
NbedDyn (worst)	0.722 ± 0.203	/	/	/
Constrained NbedDyn	0.173 ± 0.110	0.640 ± 0.419	1.650 ± 0.311	1.207 ± 0.309
Constrained NbedDyn (best)	0.164 ± 0.080	0.591 ± 0.261	0.784 ± 0.101	0.801 ± 0.113
Constrained NbedDyn (worst)	0.190 ± 0.129	0.754 ± 0.557	1.512 ± 0.204	1.540 ± 0.183

Table 2: ***Forecasting performance on the test set of the data-driven models for Lorenz-96 dynamics where only the first 20 state variables are observed***: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2 (The true largest Lyapunov exponent of the Lorenz 96 model is 1.67 [54]). The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 4000 time-steps.

Forecasting performance of the proposed data-driven models: This experiment highlights the impact of the boundedness constraints on both the short-term prediction and simulation performance of the models. The proposed Constrained NbedDyn is able to outperform all the state-of-the-art models on both prediction and simulation tasks. Interestingly, in the short-term prediction experiment, the classical NbedDyn formulation diverges to infinity, for some realizations, before reaching 4 prediction time steps. This highlights that the proposed constrained version of the NbedDyn is necessary not only for simulation but also for prediction applications as the models without boundedness constraints can diverge to infinity even after short prediction steps. When considering the attractor reconstruction based on long-term simulation of the data-driven models, the constrained version of the NbedDyn model is the only model able to unfold the Lorenz 96 attractor in both (S1) and (S2). The other data-driven models, including the unconstrained NbedDyn either diverge to infinity or generate trajectories that do not match the Lorenz 96 hidden attractor (please refer to appendix B and C for additional experiments).

[2] Other experiments, not included here, show that a full linear quadratic NODE model also converges (but at a slower rate) to solutions with the same properties as the convolutional model evaluated in this experiment.



(a) (b) (c)

(d) (e) (f)

Figure 3: ***Mean Fast Fourier Transform (FFT) modulus of the benchmarked data-driven models for Lorenz-96 case-study***. We report the FFT modulus of simulations of the Constrained NbedDyn generated under both (S1) and (S2). In (a) the Constrained NbedDyn models are trained on 4000 samples with $dt = 0.01$, in (b) the models are trained on 20000 samples with $dt = 0.01$ and in (c) the training is carried on a sequence of 8000 samples with $dt = 0.005$. We compare these spectrums to the ones computed using the RNN2 model under the same training configuration in (d), (e) and (f) respectively. The solid lines represent the FFT averaged over five training runs and 100 realizations for each scheme. The light-color interval is the standard deviation of the different models. The unconstrained NbedDyn trajectories diverge after a short forecasting time, thus, its FFT is omitted.

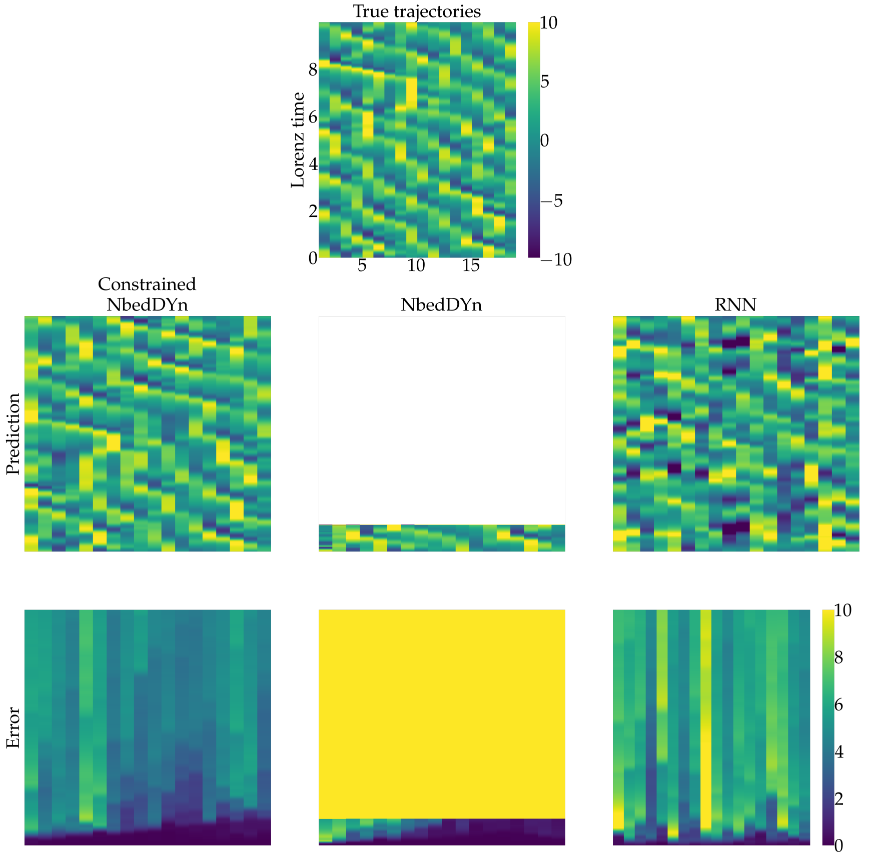


Figure 4: ***Forecasting example of the benchmarked data-driven models for Lorenz-96 case-study***. First row: ground-truth Lorenz 96 time series; Second row : data-driven model simulations over a test period; Third row: associated RMSE map. For each subplot, the vertical axis refers to the time axis and the horizontal axis to the 20-dimensional observation state. As unconstrained NbedDyn trajectories blow up after a short time horizon, no observation state can be depicted.

We further illustrate the quantitative analysis conclusions through the visual comparison of the FFT modulus as well as a forecasting example in figures 3 and 4 respectively. The proposed architecture shows a better match to the true trajectory both in the temporal and spectral domains. Specifically, the constrained NbedDyn trajectory, although diverges from the ground truth (due to the chaotic nature of the attractor), keeps a similar spatio-temporal behavior illustrated by the prediction example in Fig. 4 as well as the spectrum in Fig. 3. Interestingly, this performance is truly due to the boundedness constraints since the unconstrained version, and even using the same ODE parameterization, diverges to infinity after a short simulation time.

5.4 Patch shallow Water equation (PSWE)

As a third case study, we consider SWE dynamics governed by the following set of equations:

$$\begin{cases} \frac{\partial v_x}{\partial t} - F_{v_x} = -g\frac{\partial \eta}{\partial x} \\ \frac{\partial v_y}{\partial t} - F_{v_y} = -g\frac{\partial \eta}{\partial y} \\ \frac{\partial \eta}{\partial t} + \frac{\partial (\eta + H)v_x}{\partial x} + \frac{\partial (\eta + H)v_y}{\partial y} \quad = 0 \end{cases} \tag{21}$$

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 is the acceleration due to gravity, which is taken to be equal to $9.81m/s^2$. H corresponds to the total depth (here $H = 100m$) 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 (21) is carried using a finite difference method. The length of the domain is set to $1000km \times 1000km$ with a corresponding regular discretization of 80×80. The temporal step size dt is set to satisfy the Courant–Friedrichs–Lewy condition ($dt = 40.41$ seconds). As an observed process, we consider a low-resolution version of a patch of size $250km \times 250km$ of the sea surface elevation η. The patch is located in the center of the 2D domain. The low-resolution projection is computed from an 8-dimensional EOF decomposition, which amounts to capture 80% of the total variance. Regarding the training configuration, 100000 simulation time steps are generated. The transient (first 2500) time steps are omitted, and we use the post-transient first 49701 time-steps as training data. The remaining sequence is used as a test set.

This case study combines numerous layers of complexity when compared to the previous experiments. Specifically, the dynamics of the considered $250km \times 250km$ domain depend on the neighboring regions. Its variability is rigorously governed by the Shallow Water Equation in the full domain. As such, one cannot derive an autonomous differential system for the subdomain. Besides, through the EOF-based filtering, the observed process only refers to coarse-scale dynamics. Reduced-order modeling schemes [114] and learning-based closure models [115] would be typical solutions to address this issue, which would rely on the approximation of the observed dynamics by an autonomous system in the observation space or in a lower-dimensional projection of this space. We show in this experiment that the proposed framework can account for unknown boundary conditions and fine-scale dynamics by retrieving the most appropriate missing information to forecast future coarse-scale observations within the considered patch. We may remind that the learning of the latent states only relies on the observed variables.

Parametrization of the data-driven models: All the tested models are built on the 8-dimensional EOF basis used to filter the observations. In the proposed framework, this setting corresponds to a specific parameterisation of the projection operator $\mathcal{M}(\cdot)$ in (10) where $\mathcal{M}(\mathbf{x}_t) \in \mathbb{R}^8$ is an EOF projection. The NbedDyn model (both the constrained and unconstrained settings) is constructed with a dimension of the augmented state $d_E = 18$. This dimension allows one to have both a good short-term forecast and a good reproduction of the dynamics. Please refer to appendix A for an analysis of the dimension of the embedding for the SWE case study. Our framework is compared to a stacked Bidirectional LSTM (RNN) and to a linear regression model. The RNN includes 10 LSTM layers with 100 hidden units. The input sequence of the RNN is an observation sequence of size 40. The linear model, that we call here Delay EDMD, is built on a Singular Value Decomposition (SVD) of delay embedding coordinates of the observations [40]. The delay embedding is computed, for every EOF component, using a lag embedding equal to one-time step and a dimension equal to 200. The dimension of the SVD is set to 150, which accounts for over 99.99% of the total variance of the delay embedding representation. Finally, we only test the second scenario (S2) for the constrained NbedDyn model as the classical NbedDYn diverges to infinity and the RNN is not expected to work on random initial conditions.

Model	$t_0 + dt$	$t_0 + 4dt$
RNN	$7.30E-5 \pm 3.72E-5$	$8.89E-4 \pm 3.97E-4$
NbedDyn	$3.02E-4 \pm 1.88E-4$	$1.84E-3 \pm 1.24E-3$
Delay EDMD	$1.18E-3 \pm 2.54E-4$	$1.18E-3 \pm 2.45E-4$
Constrained NbedDyn	$2.84E-4 \pm 1.61E-4$	$1.67E-3 \pm 9.90E-4$

Table 3: ***Forecasting performance of the benchmarked data-driven models for PSWE case-study***: we report the mean RMSE mertrics for different forecasting time-steps. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1.

Forecasting performance of the proposed data-driven models: regarding the short-term forecasting performance, reported for instance in table 3, the RNN model achieves the best prediction performance. This model is however sensible to the size of the training data as shown in appendix B. We also report a forecasting examples in figure 5 where the short-term forecast, highlighted in the blue panel, are similar to the true state, up to 27 minutes where only the Delay EDMD technique is still qualitatively close to the ground truth. The analysis of the long-term simulated states, bordered by a red panel in Figure 5, gives in the other hand a different conclusion. While the proposed constrained NbedDyn model keeps simulating realistic states, both the unconstrained NbedDyn and RNN models get stuck either at equilibrium points or at very slow manifolds. The linear Delay EDMD model converges to the origin after a finite amount of time, which is expected since finite-dimensional linear models can not reflect a chaotic behavior. We can draw the same analysis from the mean FFT modulus of the forecasted EOF modes, as well as the mean radially averaged PSD of the 2D fields in Fig. 6. When considering the spatial spectrum, only the proposed framework is able to match perfectly the spectrum of the ground-truth. This result illustrates the ability of our model to capture all the spatial scales of the dynamics and to keep simulating these scales after long simulation times. Regarding the temporal spectrum, the constrained NbedDyn and the Delay EDMD models[3] match the spectrum of the ground truth both in the medium and high-frequency range. The low-frequency range in the other hand is not well captured by the model. This suggests that we may need to consider an optimization objective (18) with longer prediction times than the one used in this work to embed realistic low-frequency patterns. The RNN model has a temporal spectrum that corresponds to a periodic signal that does not much any of the ground truth frequencies.

6 Conclusion

Two central challenges in data-driven modeling involve how to handle partial observations and how to guarantee stability in the resulting models. In this paper, we show on several case studies that combining an augmented neural ODE formulation and boundedness constraints, it is possible to considerably improve the data-driven identification and forecasting of partially-observed dynamics. The proposed framework is assessed on several case studies with respect to state-of-the-art forecasting and system identification techniques. Specifically, for classical chaotic ODEs (Lorenz 63 and 96), the proposed methodology matches state-of-the-art short-term forecasting performances while ensuring realistic long-term patterns for partially-observed settings. We also report a more complex case study, in which we have considered low-resolution observations of a subdomain of the SWE with unknown boundary conditions. In all the experiments, we found that, for various configurations of training data, the boundedness constraints are key for the NbedDyn model to achieve good short-term predictions and asymptotic behaviors.

Future works will involve extending the proposed framework to other non-linear models through the generalization of the boundedness constraints. Foremost, polynomial models could be considered by imposing, similar to the

[3] The computation of the delay EDMD time spectrum is carried over a realizations that contain the transient of the dynamics.

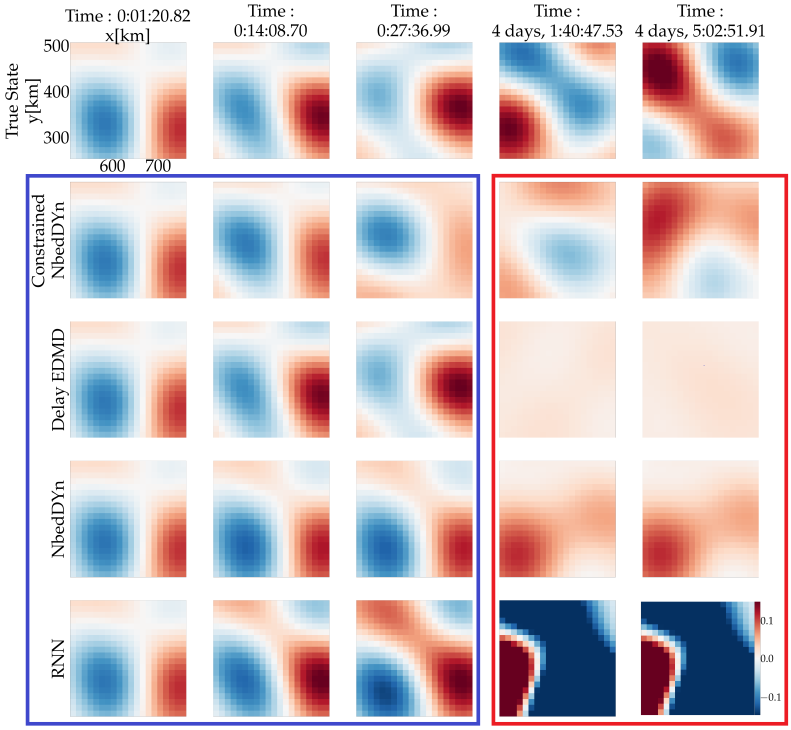
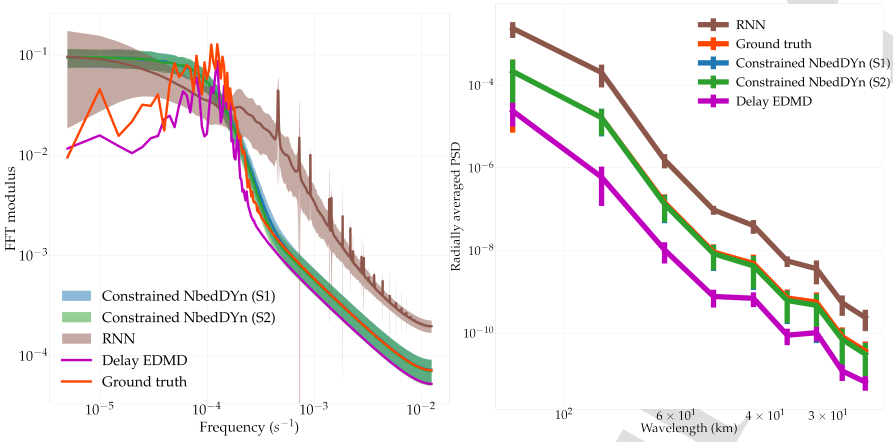


Figure 5: ***Prediction example over the test period of the data-driven models for the PSWE case-study.*** The blue panel illustrates a short-term forecast for the benchmarked models and the red panel highlights the associated long-term simulations. A spectral analysis of the simulated trajectories, both in space and time, are given in figure 6.



(a) (b)

Figure 6: ***Spatio-temporal spectrums of the data-driven models with respect to the ground truth.*** We report the FFT modulus of the time series in (a) and the radially-averaged Power Spectral Density (PSD) in (b). The spectrums are averaged over 100 realizations of 5 runs. Each realization goes up to 5000 time steps.

quadratic terms of the proposed architecture, an energy-preserving non-linearity. Other non-linear architectures may still need further investigation, since the fluctuation energy may not be the appropriate Lyapunov function to consider. The recent works of [37] may be useful for finding Lyapunov functions that guarantee stability and boundedness. We also hope to extend the results of this study to more realistic ocean flows, especially ocean surface dynamics. Such dynamics involve, similar to the PSWE case study, but at a higher level of complexity, unseen components, missing forcing, unknown boundary conditions and unresolved spatio-temporal scales.

To recall, explicitly constraining chaos within a data-driven formulation is rigorously intractable since such a behavior is only revealed, and thus potentially constrained using long-term simulations. From this point of view, the proposed framework is not guaranteed to reach an expected chaotic evolution, since this is not explicitly constrained within the framework. Specifically, when considering the bounded NbedDyn model, the learning criterion allows for any bounded limit-set, including stable ones, as long as the short-term forecast of the observations is minimized. This criterion does not guarantee a replication of the chaotic observations, and long-term simulations of the model can lead to undesirable stable limit-sets. Interestingly, stable limit-sets can be fully characterized by the dynamical equation, without resorting to brute force long-term simulation of the model. Thus, these limit sets can be propagated to infinity. However, sending stable limit-sets to infinity may result in new questions about how observation data may help to regularize the learning stage.

Acknowledgements

This work was supported by LEFE program (LEFE MANU project IA-OAC), CNES (grant OSTST DUACS-HR) and ANR Projects Melody and OceaniX. It benefited from HPC and GPU resources from Azure (Microsoft Azure grant) and from GENCI-IDRIS (Grant 2020-101030). SLB acknowledges valuable discussions with Alan Kaptanoglu.

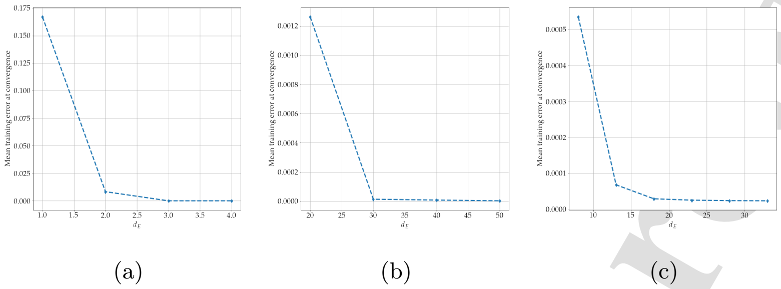
A Analysis of the dimension of the constrained NbedDyn model

The choice of the dimension d_E is linked to the number of independent variables that can be used to model the dynamics using in our context, a bounded autonomous linear quadratic ODE. An initial guess for d_E when given scalar measurements, can be derived by using the false nearest neighbors technique [111]. However, when the measurements are higher dimensional it is harder to derive an estimate of the value of d_E using classical methods.

In this context, we study the direct impact of d_E on the performance of the constrained NbedDyn model. Figure 7 shows the impact of d_E on the training error between the observations and the model simulation. Other criteria such as Lyapunov exponents could be used, but the training error provides a direct measure of the effectiveness of the embedding dimension in the training phase (an example of the study of the Lyapunov exponents for various values of d_E is given in [32] for the NbedDyn model). The first evaluation of the training error is reported in figure 7 for d_E equal to the dimension of $\mathcal{M}(\mathbf{x}_t)$. In this experiments, no latent states $\mathbf{y}_t$ are used and the embedding $\mathbf{u}_t = \mathcal{M}(\mathbf{x}_t)$. In such situations, the NODE model 16 can not perfectly fit the data. Furthermore, at this particular value of d_E, the models are more likely to have a bad asymptotic behavior. As the dimension increases, this training error decreases until it converges for a particular value. We use in the experimental section $d_E = 3$ for Lorenz 63, $d_E = 40$ for Lorenz 96 and $d_E = 18$ for the PSWE dynamics since they provide a good trade-off between short term prediction performance, good asymptotic behaviour and low complexity.

B Additional experiments with varying data-sets sizes

We analyze the prediction and simulation performance of the Constrained NbedDyn model with respect to the size of the training data. Figure 8, highlights the evolution of the RMSE of the short term forecast and the largest Lyapunov exponent in the Lorenz 63 and 96 experiments as we increase the size of the training data (please refer to tables 1, 4 and 5 for a quantitative analysis on the Lorenz 63 case study and to tables 2, 8 and 9 for the Lorenz 96 experiments). The RNN and NbedDyn models scale, as expected, with the amount of training data



(a)

(b)

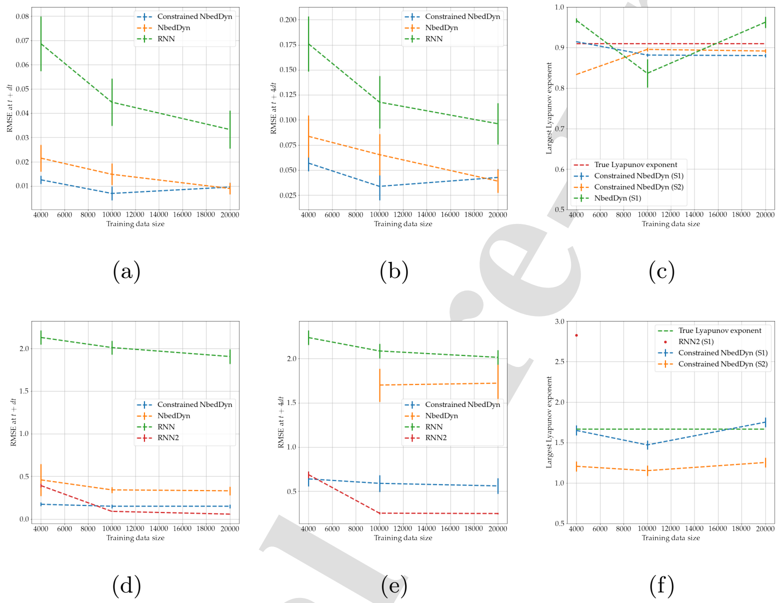
(c)

Figure 7: ***Mean training error at convergence*** . We report the mean training error at convergence of the constrained NbedDyn model for different dimensions d_E of the embedding. This error is averaged over an ensemble of 5 models and we highlight here both the mean and standard deviations for the Lorenz 63, Lorenz 96 and PSWE case studies in (a), (b) and (c) respectively.

in the short term forecast experiment (highlighted, for example in the Lorenz 63 case study, by figures 8a and 8b). Furthermore, figure 8b shows that increasing the data-size allows to improve the generalization properties of the NbedDyn for short-term forecast applications (when trained on 4000 time-steps at $dt = 0.01$ the NbedDyn diverges after short prediction times for some realization, increasing the size of the dataset allows to increase the predictability of the NbedDyn). The constrained nbedDyn on the other hand is able to provide good short term predictions for all the tested experiments. Finally, the amount of data studied here does not impact the ability of these models to simulate chaotic dynamics 8c (the RNN model fails at producing the correct chaotic dynamics and the NbedDyn fails at (S2) for all the considered case studies and data sizes). The Constrained NbedDyn model achieves good simulation performance for all the sizes of the training datasets which suggests that the simple parameterization of the model (linear quadratic), as well as the boundedness constraints, are key in deriving a model that generalizes well for unseen initial conditions given (when compared to other state-of-the-art models) small amounts of training data.

C Additional experiments with varying time sampling

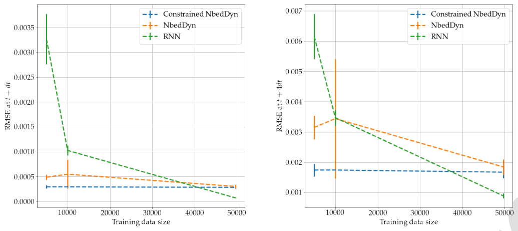
In this experiment, we analyze the prediction and simulation performance of the Constrained NbedDyn model with respect to the sampling dt of the training data. Figure 10, highlights the evolution of the RMSE of the short-term forecast and the largest Lyapunov exponent as we increase the sampling of the observations. Overall, both the constrained NbedDyn and the RNN benchmark have better short term prediction performance (as highlighted in figures 10a and 10b) with a smaller dt. This is expected since as we decrease dt, the variability of the data is



(a) (b) (c)

(d) (e) (f)

Figure 8: ***Performance of the benchmark models with respect to the size of the training data***. We report the RMSE at $t0 + dt$, at $t_0 + 4dt$ and the largest Lyapunov exponent for the data-driven models with respect to the size training data in figures (a) (b) and (c) respectively. The Largest Lyapunov exponent of the RNN is not represented since it remains always negative. Then one of the NbedDyn in (S2) is also omitted since the trajectories diverge. The time sampling of the datasets is set to $dt = 0.01$. The standard deviation of the scores is scaled by a factor of five to fit in the figures. Please refer to tables 1, 4 and 5 for a quantitative analysis on the Lorenz 63 case study and to tables 2, 8 and 9 for the ones of the the Lorenz 96 experiments.



(a) (b)

Figure 9: ***Performance of the benchmark models with respect to the size of the training data in the PSWE experiment.*** We report the RMSE at $t0 + dt$, at $t_0 + 4dt$ with respect to the size training data in figures (a) and (b). The standard deviation of the scores is scaled by a factor of five to fit in the figures. Please refer to tables 3, 12 and 13 for a quantitative analysis.

slower and more predictable. Regarding the simulation performance, increasing the resolution of the data does not impact the largest Lyapunov exponents given in figure 10c in a simple interpretable way. We suspect that these tested time resolutions are good enough for unfolding true dynamics from the observations. Therefore, decreasing the resolution of the data does not impact the simulation performance of the model.

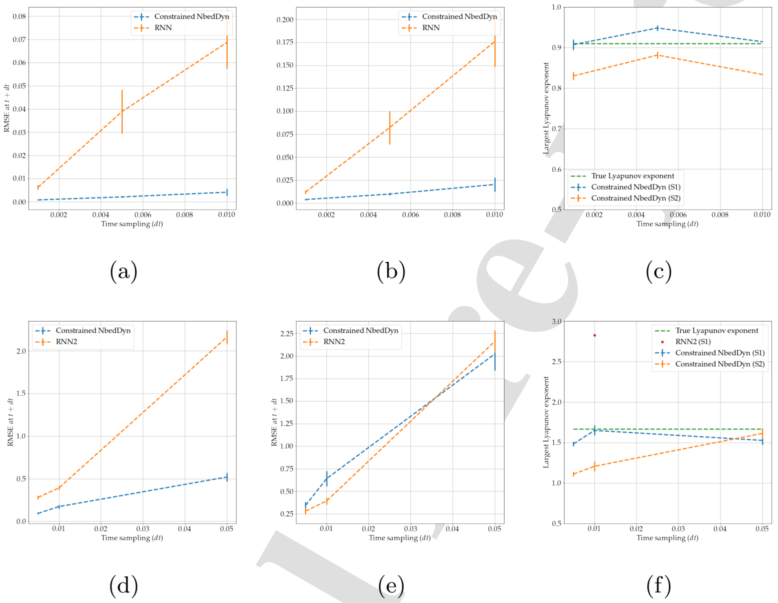
D Training

The trainable parameters of the Constrained NbedDyn models *i.e.* the linear quadratic NODE and initial conditions of the latent states are initially sampled from a uniform distribution. The training of all models is carried using the Adam optimizer. We use a varying learning rate (from 0.1 to 0.001) in all the experiments to speed up the training. Regarding the weighting parameters $\{\lambda_i\}_{i=1}^{i=3}$, we set $\lambda_1 = 1$ during all the training. The weights responsible for the boundedness constraints were set at higher values in the beginning of the training *i.e.* $\lambda_2 = 100$ and $\lambda_3 = 1000$ and then reduced to $\lambda_2 = 1$ when $\lambda_3 \mathcal{C}_2 = 0$. The training is stopped using cross-validation.

E Additional tables

References

[1] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, April 2016.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 10: ***Performance of the benchmark models with respect to the sampling of the training data***. We report the RMSE at $t0 + dt$, at $t_0 + 4dt$ and the largest lyapunov exponent for the data-driven models with respect to the time sampling of the data in figures (a) (b) and (c) for the Lorenz 63 and (d), (e) and (f) for the Lorenz 96 system. The size of the training data goes up to 40 Lorenz time. For the Lorenz 63 experiment, the Largest Lyapunov exponent of the RNN is not represented since it remains always negative. Regarding the Lorenz 96 experiments, we only show the closest positive lyapunov exponent for the RNN model. The standard deviation of the scores is scaled by a factor of five to fit in the figures (please refer to tables 1, 7, 6, 11, 2 and 10 for a quantitative comparison of these scores).

Model		$t_0 + dt$	$t_0 + 4dt$	λ_1	λ_1
SR	$\tau_{MI} = 17$, $d_E(FNN) = 3$	0.0387 ± 0.0436	0.1304 ± 0.1391	≤ 0	≤ 0
	$\tau_{MI} = 17$, d_E(Whitney)$= 6$	0.0235 ± 0.0238	0.0838 ± 0.0895	≤ 0	≤ 0
	$\tau_{Corr} = 31$, d_E(FNN)$= 3$	0.1909 ± 0.1896	0.7587 ± 0.8209	≤ 0	≤ 0
	$\tau_{Corr} = 31$, d_E(Whitney)$= 6$	0.1092 ± 0.1106	0.5458 ± 0.6695	≤ 0	≤ 0
	$\tau = 6$, $d_E = 3$	0.0043 ± 0.0040	0.0090 ± 0.0066	≤ 0	≤ 0
RNN (averaged)		0.0445 ± 0.0489	0.1177 ± 0.1301	≤ 0	
RNN (best)		0.0230 ± 0.0207	0.0621 ± 0.0207	≤ 0	
RNN (worst)		0.0966 ± 0.0732	0.2521 ± 0.0732	≤ 0	
NbedDyn	$d_E = 3$	0.0148 ± 0.0222	0.0653 ± 0.1032	0.8365 ± 0.1740	NaN
NbedDyn (best)	$d_E = 3$	0.0091 ± 0.0298	0.0368 ± 0.1568	0.5372 ± 0.0373	NaN
NbedDyn (worst)	$d_E = 3$	0.0391 ± 0.0382	0.1918 ± 0.1731	NaN	NaN
Constrained NbedDyn	$d_E = 3$	0.0069 ± 0.0141	0.0337 ± 0.0695	0.8816 ± 0.0172	0.8956 ± 0.0221
Constrained NbedDyn (best)	$d_E = 3$	0.0053 ± 0.0123	0.0261 ± 0.0595	0.9153 ± 0.0147	0.8933 ± 0.0244
Constrained NbedDyn (worst)	$d_E = 3$	0.0079 ± 0.0181	0.0384 ± 0.0900	0.9234 ± 0.0164	0.8977 ± 0.0184

Table 4: ***Forecasting performance on the test set of the benchmarked data-driven models for Lorenz-63 case-study, 10000 training samples with*** $dt = 0.01$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

[2] Joseph Bakarji, Kathleen Champion, J Nathan Kutz, and Steven L Brunton. Discovering governing equations from partial measurements with deep delay autoencoders. *arXiv preprint arXiv:2201.05136*, 2022.

[3] Jean-Christophe Loiseau and Steven L Brunton. Constrained sparse galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, 2018.

[4] Alan A Kaptanoglu, Jared L Callaham, Aleksandr Aravkin, Christopher J Hansen, and Steven L Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Physical Review Fluids*, 6(9):094401, 2021.

[5] Boris Kramer. Stability domains for quadratic-bilinear reduced-order models. *SIAM Journal on Applied Dynamical Systems*, 20(2):981–996, 2021.

[6] Ruoyu Sun. Optimization for deep learning: theory and algorithms. *arXiv preprint arXiv:1912.08957*, 2019.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[8] M Grewal and Keith Glover. Identifiability of linear and nonlinear dynamical systems. *IEEE Transactions on automatic control*, 21(6):833–837, 1976.

[9] Keith Glover and Jan Willems. Parametrizations of linear dynamical systems: Canonical forms and identifiability. *IEEE Transactions on Automatic Control*, 19(6):640–646, 1974.

Model		$t_0 + dt$	$t_0 + 4dt$	λ_1	λ_1
SR	$\tau_{MI} = 17$, $d_E(FNN) = 3$	0.0385 ± 0.0437	0.129 ± 0.1389	≤ 0	≤ 0
	$\tau_{MI} = 17$, d_E(Whitney)$= 6$	0.0255 ± 0.0231	0.0894 ± 0.0815	≤ 0	≤ 0
	$\tau_{Corr} = 32$, d_E(FNN)$= 3$	0.2177 ± 0.1962	0.8621 ± 0.8516	≤ 0	≤ 0
	$\tau_{Corr} = 32$, d_E(Whitney)$= 6$	0.1101 ± 0.1106	0.5008 ± 0.6921	≤ 0	≤ 0
	$\tau = 6$, $d_E = 3$	0.0043 ± 0.0040	0.0090 ± 0.0066	≤ 0	≤ 0
RNN (averaged)		0.0332 ± 0.0394	0.0962 ± 0.1033	≤ 0	
RNN (best)		0.0134 ± 0.0120	0.0412 ± 0.0120	≤ 0	
RNN (worst)		0.0863 ± 0.0560	0.2238 ± 0.0560	≤ 0	
NbedDyn	$d_E = 3$	0.0089 ± 0.0122	0.0390 ± 0.0583	0.9627 ± 0.0690	NaN
NbedDyn (best)	$d_E = 3$	0.0062 ± 0.0071	0.02540 ± 0.0285	0.9420± 0.0208	NaN
NbedDyn (worst)	$d_E = 3$	0.0293 ± 0.0241	0.1458 ± 0.1201	0.8803 ± 0.0247	NaN
Constrained NbedDyn	$d_E = 3$	0.0095 ± 0.0086	0.0426 ± 0.0418	0.8803 ± 0.0190	0.8916 ± 0.0229
Constrained NbedDyn (best)	$d_E = 3$	0.0076 ± 0.0054	0.0358 ± 0.0289	0.9220 ± 0.0202	0.8821 ± 0.0242
Constrained NbedDyn (worst)	$d_E = 3$	0.0110 ± 0.0118	0.0495 ± 0.0585	0.9238 ± 0.0180	0.8953 ± 0.0219

Table 5: ***Forecasting performance on the test set of the benchmarked data-driven models for Lorenz-63 case-study, 20000 training samples with*** $dt = 0.01$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

[10] Bo Wahlberg. On the identification of continuous time dynamical systems. *IFAC Proceedings Volumes*, 21(9):435–440, 1988.

[11] S Narendra Kumpati, Parthasarathy Kannan, et al. Identification and control of dynamical systems using neural networks. *IEEE Transactions on neural networks*, 1(1):4–27, 1990.

[12] Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.*, 120:024102, Jan 2018.

[13] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[14] Olalekan Ogunmolu, Xuejun Gu, Steve Jiang, and Nicholas Gans. Nonlinear systems identification using deep dynamic neural networks. *arXiv preprint arXiv:1610.01439*, 2016.

[15] Ronan Fablet, Said Ouala, and Cedric Herzet. Bilinear residual Neural Network for the identification and forecasting of dynamical systems. *SciRate*, dec 2017.

[16] Duong Nguyen, Said Ouala, Lucas Drumetz, and Ronan Fablet. Assimilation-based learning of chaotic dynamical systems from noisy and partial data. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3862–3866, 2020.

[17] Duong Nguyen, Said Ouala, Lucas Drumetz, and Ronan Fablet. Variational deep learning for the identification and reconstruction of chaotic and

Model		$t_0 + dt$	$t_0 + 4dt$	λ_1	λ_1
RNN (averaged)		0.006 ± 0.005	0.011 ± 0.009	≤ 0	
RNN (best)		0.003 ± 0.002	0.006 ± 0.002	≤ 0	
RNN (worst)		0.007 ± 0.005	0.0155 ± 0.005	≤ 0	
Constrained NbedDyn	$d_E = 3$	0.0008 ± 0.0008	0.0038 ± 0.0041	0.9074 ± 0.0631	0.8300 ± 0.0510
Constrained NbedDyn (best)	$d_E = 3$	0.0005 ± 0.0005	0.0025 ± 0.0024	0.7351 ± 0.0229	0.7265 ± 0.03487
Constrained NbedDyn (worst)	$d_E = 3$	0.0010 ± 0.0008	0.0047 ± 0.0040	0.8644 ± 0.02070	0.8365 ± 0.0291

Table 6: ***Forecasting performance on the test set of the benchmarked data-driven models for Lorenz-63 case-study. The training dataset ranges from 0 to 40 Lorenz time with*** $dt = 0.001$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

Model		$t_0 + dt$	$t_0 + 4dt$	λ_1	λ_1
RNN (averaged)		0.038 ± 0.047	0.082 ± 0.090	≤ 0	
RNN (best)		0.0167 ± 0.013	0.043 ± 0.013	≤ 0	
RNN (worst)		0.061 ± 0.066	0.123 ± 0.066	≤ 0	
Constrained NbedDyn	$d_E = 3$	0.002 ± 0.001	0.009 ± 0.008	0.948 ± 0.0349	0.881 ± 0.0414
Constrained NbedDyn (best)	$d_E = 3$	0.001 ± 0.001	0.008 ± 0.008	0.901 ± 0.034	0.868 ± 0.0436
Constrained NbedDyn (worst)	$d_E = 3$	0.002 ± 0.001	0.010 ± 0.008	0.926 ± 0.033	0.877 ± 0.042

Table 7: ***Forecasting performance on the test set of the benchmarked data-driven models for Lorenz-63 case-study. The training dataset ranges from 0 to 40 Lorenz time with*** $dt = 0.005$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

stochastic dynamical systems from noisy and partial observations. *arXiv preprint arXiv:2009.02296*, 2020.

[18] Said Ouala, Laurent Debreu, Ananda Pascual, Bertrand Chapron, Fabrice Collard, Lucile Gaultier, and Ronan Fablet. Learning runge-kutta integration schemes for ode simulation and identification. *arXiv preprint arXiv:2105.04999*, 2021.

[19] Anne Braakmann-Folgmann, Ribana Roscher, Susanne Wenzel, Bernd Uebbing, and Jürgen Kusche. Sea Level Anomaly Prediction using Recurrent Neural Networks. *arXiv:1710.07099 [cs]*, oct 2017. arXiv: 1710.07099.

[20] Bryan Lim and Stefan Zohren. Time series forecasting with deep learning: A survey, 2020.

Model	$t_0 + dt$	$t_0 + 4dt$	λ_1 (S1)	λ_1 (S2)
RNN1	2.011 ± 0.408	2.085 ± 0.401	≤ 0	
RNN1 (best)	1.917 ± 0.424	2.029 ± 0.424	≤ 0	
RNN1 (worst)	2.064 ± 0.431	2.131 ± 0.431	≤ 0	
RNN2	0.090 ± 0.021	0.253 ± 0.060	10.407 ± 10.35	
RNN2 (best)	0.085 ± 0.018	0.250 ± 0.018	2.369 ± 0.091	
RNN2 (worst)	0.097 ± 0.0255	0.263 ± 0.025	27.959 ± 1.253	
NbedDyn	0.342 ± 0.175	1.70 ± 0.935	NaN	NaN
NbedDyn (best)	0.333 ± 0.161	1.532 ± 1.191	NaN	NaN
NbedDyn (worst)	0.365 ± 0.161	1.894 ± 1.191	NaN	NaN
Constrained NbedDyn	0.150 ± 0.109	0.590 ± 0.474	1.471 ± 0.270	1.153 ± 0.311
Constrained NbedDyn (best)	0.142 ± 0.091	0.546 ± 0.362	0.935 ± 0.120	0.926 ± 0.110
Constrained NbedDyn (worst)	0.158 ± 0.104	0.679 ± 0.464	0.898 ± 0.099	0.919 ± 0.100

Table 8: ***Forecasting performance on the test set of the data-driven models for Lorenz-96 dynamics where only the first 20 state variables are observed, 10000 training samples with*** $dt = 0.01$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

Model	$t_0 + dt$	$t_0 + 4dt$	λ_1 (S1)	λ_1 (S2)
RNN1	1.906 ± 0.420	2.013 ± 0.407	≤ 0	
RNN1 (best)	1.862 ± 0.407	1.942 ± 0.407	≤ 0	
RNN1 (worst)	1.958 ± 0.427	2.087 ± 0.427	≤ 0	
RNN2	0.0575 ± 0.011	0.249 ± 0.044	27.736 ± 12.600	
RNN2 (best)	0.0549 ± 0.008	0.248 ± 0.008	40.320 ± 0.467	
RNN2 (worst)	0.0602 ± 0.0132	0.250 ± 0.013	15.151 ± 0.761	
NbedDyn	0.331 ± 0.251	1.721 ± 1.176	NaN	NaN
NbedDyn (best)	0.272 ± 0.143	1.396 ± 0.761	NaN	NaN
NbedDyn (worst)	0.428 ± 0.198	2.299 ± 1.378	NaN	NaN
Constrained NbedDyn	0.150 ± 0.111	0.561 ± 0.449	1.752 ± 0.298	1.254 ± 0.307
Constrained NbedDyn (best)	0.138 ± 0.121	0.452 ± 0.546	0.976 ± 0.121	0.975 ± 0.131
Constrained NbedDyn (worst)	0.162 ± 0.121	0.670 ± 0.546	0.976 ± 0.121	0.975 ± 0.131

Table 9: ***Forecasting performance on the test set of the data-driven models for Lorenz-96 dynamics where only the first 20 state variables are observed, 20000 training samples with*** $dt = 0.01$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

[21] Said Ouala, Ronan Fablet, Cédric Herzet, Bertrand Chapron, Ananda Pascual, Fabrice Collard, and Lucile Gaultier. Neural network based kalman filters for the spatio-temporal interpolation of satellite-derived sea

Model	$t_0 + dt$	$t_0 + 4dt$	λ_1 (S1)	λ_1 (S2)
RNN2	0.281 ± 0.108	0.497 ± 0.180	10.961 ± 23.32	
RNN2 (best)	0.251 ± 0.106	0.455 ± 0.106	16.273 ± 0.0	
RNN2 (worst)	0.294 ± 0.090	0.532 ± 0.090	22.848 ± 0.0	
Constrained NbedDyn	0.093 ± 0.053	0.343 ± 0.195	1.483 ± 0.142	1.110 ± 0.138
Constrained NbedDyn (best)	0.0869 ± 0.048	0.319 ± 0.173	1.092 ± 0.115	1.180 ± 0.099
Constrained NbedDyn (worst)	0.097 ± 0.060	0.360 ± 0.222	1.000 ± 0.123	0.997 ± 0.136

Table 10: ***Forecasting performance on the test set of the data-driven models for Lorenz-96 dynamics where only the first 20 state variables are observed. The training dataset ranges from 0 to 40 Lorenz time with*** $dt = 0.005$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

Model	$t_0 + dt$	$t_0 + 4dt$	λ_1 (S1)	λ_1 (S2)
RNN2	2.161 ± 0.402	3.444 ± 0.615	51.465 ± 8.88	
RNN2 (best)	2.069 ± 0.360	3.318 ± 0.360	53.602 ± 3.631	
RNN2 (worst)	2.252 ± 0.434	3.597 ± 0.434	52.127 ± 4.205	
Constrained NbedDyn	0.520 ± 0.249	2.022 ± 0.925	1.527 ± 0.275	1.613 ± 0.314
Constrained NbedDyn (best)	0.487 ± 0.243	1.841 ± 0.921	1.655 ± 0.141	1.670 ± 0.124
Constrained NbedDyn (worst)	0.546 ± 0.229	2.150 ± 0.995	1.610 ± 0.105	1.777 ± 0.132

Table 11: ***Forecasting performance on the test set of the data-driven models for Lorenz-96 dynamics where only the first 20 state variables are observed. The training dataset ranges from 0 to 40 Lorenz time with*** $dt = 0.05$: first two columns: mean RMSE for different forecasting time-steps, third column: largest Lyapunov exponent under (S1), fourth column: largest Lyapunov exponent under S2. The results are averaged over 100 realizations for an ensemble of 5 models as explained in section 5.1. The Lyapunov exponents are computed from a simulation of 10000 time-steps.

surface temperature. *Remote Sens.*, 10(12):1864, Nov 2018.

[22] William Gilpin. Deep reconstruction of strange attractors from time series, 2020.

[23] Ronan Fablet, Bertrand Chapron, Lucas Drumetz, Etienne Mémin, Olivier Pannekoucke, and François Rousseau. Learning variational data assimilation models and solvers. *arXiv preprint arXiv:2007.12941*, 2020.

[24] Kangbeom Cheon, Jaehoon Kim, Moussa Hamadache, and Dongik Lee. On replacing pid controller with deep learning controller for dc motor system. *Journal of Automation and Control Engineering Vol*, 3(6), 2015.

Model	$t_0 + dt$	$t_0 + 4dt$
RNN	$1.02E-3 \pm 4.78E-4$	$3.47E-3 \pm 1.34E-3$
NbedDyn	$5.49E-4 \pm 1.44E-3$	$3.44E-3 \pm 9.86E-3$
Delay EDMD	$1.18E-3 \pm 2.54E-4$	$1.18E-3 \pm 2.45E-4$
Constrained NbedDyn	$2.90E-4 \pm 1.81E-4$	$1.71E-3 \pm 1.10E-3$

Table 12: ***Forecasting performance of the benchmarked data-driven models for PSWE case-study***: we report the mean RMSE mertrics for different forecasting time-steps for the test set. We use 10000 training samples.

Model	$t_0 + dt$	$t_0 + 4dt$
RNN	$3.26E-3 \pm 2.53E-3$	$6.15E-3 \pm 3.72E-3$
NbedDyn	$4.88E-4 \pm 2.73E-4$	$3.14E-3 \pm 1.95E-3$
Delay EDMD	$1.18E-3 \pm 2.54E-4$	$1.18E-3 \pm 2.45E-4$
Constrained NbedDyn	$2.95E-4 \pm 1.69E-4$	$1.74E-3 \pm 1.04E-3$

Table 13: ***Forecasting performance of the benchmarked data-driven models for PSWE case-study***: we report the mean RMSE mertrics for different forecasting time-steps for the test set. We use 5000 training samples.

[25] Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A*, 474(2219):20180335, 2018.

[26] Omer San and Romit Maulik. Extreme learning machine for reduced order modeling of turbulent geophysical flows. *Physical Review E*, 97(4):042322, 2018.

[27] Abhinav Gupta and Pierre FJ Lermusiaux. Neural closure models for dynamical systems. *Proceedings of the Royal Society A*, 477(2252):20201004, 2021.

[28] Said Ouala, Cedric Herzet, and Ronan Fablet. Sea surface temperature prediction and reconstruction using patch-level neural network representations. *arXiv:1806.00144 [cs, stat]*, may 2018. arXiv: 1806.00144.

[29] Ronan Fablet, Lucas Drumetz, and Francois Rousseau. Joint learning of variational representations and solvers for inverse problems with partially-observed data, 2020.

[30] Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, 2021.

[31] Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young, editors, *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.

[32] S Ouala, D Nguyen, L Drumetz, B Chapron, A Pascual, F Collard, L Gaultier, and R Fablet. Learning latent dynamics for partially observed chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10):103121, 2020.

[33] Alan A Kaptanoglu, Kyle D Morgan, Chris J Hansen, and Steven L Brunton. Physics-constrained, low-dimensional models for mhd: First-principles and data-driven approaches. *arXiv preprint arXiv:2004.10389*, 2020.

[34] Yifei Guan, Steven L Brunton, and Igor Novosselov. Sparse nonlinear models of chaotic electroconvection. *arXiv preprint arXiv:2009.11862*, 2020.

[35] Rui Wang, Robin Walters, and Rose Yu. Incorporating symmetry into deep dynamics models for improved generalization. *arXiv preprint arXiv:2002.03061*, 2020.

[36] Rui Wang, Karthik Kashinath, Mustafa Mustafa, Adrian Albert, and Rose Yu. Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1457–1466, 2020.

[37] Gaurav Manek and J. Zico Kolter. Learning stable deep dynamics models, 2020.

[38] Ivan Dario Jimenez Rodriguez, Aaron D. Ames, and Yisong Yue. Lyanet: A lyapunov framework for training neural odes, 2022.

[39] Michael Schlegel and Bernd R. Noack. On long-term boundedness of galerkin models. *Journal of Fluid Mechanics*, 765:325–352, 2015.

[40] Mason Kamb, Eurika Kaiser, Steven L Brunton, and J Nathan Kutz. Time-delay observables for koopman: Theory and applications. *SIAM Journal on Applied Dynamical Systems*, 19(2):886–917, 2020.

[41] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[42] Johan Paduart, Lieve Lauwers, Jan Swevers, Kris Smolders, Johan Schoukens, and Rik Pintelon. Identification of nonlinear systems using Polynomial Nonlinear State Space models. *Automatica*, 46(4):647–656, April 2010.

[43] Wen-Xu Wang, Rui Yang, Ying-Cheng Lai, Vassilios Kovanis, and Celso Grebogi. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Phys. Rev. Lett.*, 106:154101, Apr 2011.

[44] Ye Yuan, Xiuchuan Tang, Wei Zhou, Wei Pan, Xiuting Li, Hai-Tao Zhang, Han Ding, and Jorge Goncalves. Data driven discovery of cyber physical systems. *Nature communications*, 10(1):1–9, 2019.

[45] Steffen Wiewel, Moritz Becher, and Nils Thuerey. Latent-space physics: Towards learning the temporal evolution of fluid flow, 2018.

[46] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[47] S. Ouala, A. Pascual, and R. Fablet. Residual integration neural network. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3622–3626, May 2019.

[48] Guorui Shen, Jürgen Kurths, and Ye Yuan. Sequence-to-sequence prediction of spatiotemporal systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(2):023102, 2020.

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, december 2015. arXiv: 1512.03385.

[50] Marco Fraccaro, Søren Kaae Sønderby, Ulrich Paquet, and Ole Winther. Sequential Neural Models with Stochastic Layers. *arXiv:1605.07571 [cs, stat]*, may 2016. arXiv: 1605.07571.

[51] Rahul G. Krishnan, Uri Shalit, and David Sontag. Structured Inference Networks for Nonlinear State Space Models. *arXiv:1609.09869 [cs, stat]*, September 2016. arXiv: 1609.09869.

[52] Geir Evensen. *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media, 2009.

[53] Marc Bocquet, Julien Brajard, Alberto Carrassi, and Laurent Bertino. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear Processes in Geophysics*, 26(3):143–162, 2019.

[54] Julien Brajard, Alberto Carassi, Marc Bocquet, and Laurent Bertino. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the lorenz 96 model. *arXiv preprint arXiv:2001.01520*, 2020.

[55] D Nagesh Kumar, K Srinivasa Raju, and T Sathish. River flow forecasting using recurrent neural networks. *Water resources management*, 18(2):143–161, 2004.

[56] Sumit Kumar, Lasani Hussain, Sekhar Banarjee, and Motahar Reza. Energy load forecasting using deep learning approach-lstm and gru in spark cluster. In *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, pages 1–4. IEEE, 2018.

[57] Yi-Ting Tsai, Yu-Ren Zeng, and Yue-Shan Chang. Air pollution forecasting using rnn with lstm. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 1074–1079. IEEE, 2018.

[58] B. O. Koopman. Hamiltonian systems and transformations in hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 17(5):315–318, 1931.

[59] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[60] Steven L Brunton, Bingni W Brunton, Joshua L Proctor, and J Nathan Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PloS one*, 11(2):e0150171, 2016.

[61] Steven L Brunton, Bingni W Brunton, Joshua L Proctor, Eurika Kaiser, and J Nathan Kutz. Chaos as an intermittently forced linear system. *Nature communications*, 8(1):1–9, 2017.

[62] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning koopman invariant subspaces for dynamic mode decomposition. *arXiv preprint arXiv:1710.04340*, 2017.

[63] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):1–10, 2018.

[64] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)*, pages 4832–4839. IEEE, 2019.

[65] Henning Lange, Steven L Brunton, and Nathan Kutz. From fourier to koopman: Spectral methods for long-term time series prediction. *arXiv preprint arXiv:2004.00574*, 2020.

[66] Omri Azencot, N Benjamin Erichson, Vanessa Lin, and Michael Mahoney. Forecasting sequential data using consistent koopman autoencoders. In *International Conference on Machine Learning*, pages 475–485. PMLR, 2020.

[67] Julian Rice, Wenwei Xu, and Andrew August. Analyzing koopman approaches to physics-informed machine learning for long-term sea-surface temperature forecasting. *arXiv preprint arXiv:2010.00399*, 2020.

[68] Alexandre Mauroy, Igor Mezić, and Yoshihiko Susuki. *The Koopman Operator in Systems and Control: Concepts, Methodologies, and Applications*, volume 484. Springer Nature, 2020.

[69] Andrew J Majda and John Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26(1):201, 2012.

[70] Andreas Mardt, Luca Pasquali, Frank Noé, and Hao Wu. Deep learning markov and koopman models with physical constraints. In *Mathematical and Scientific Machine Learning*, pages 451–475. PMLR, 2020.

[71] Hao Ma, Xiangyu Hu, Yuxuan Zhang, Nils Thuerey, and Oskar J. Haidn. A combined data-driven and physics-driven method for steady heat conduction prediction using deep convolutional neural networks, 2020.

[72] Xiaowei Jin, Shengze Cai, Hui Li, and George Em Karniadakis. Nsfnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations. *Journal of Computational Physics*, 426:109951, 2021.

[73] Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Karniadakis. Physics-informed neural networks (pinns) for heat transfer problems. *Journal of Heat Transfer*, 2021.

[74] K Kashinath, M Mustafa, A Albert, JL Wu, C Jiang, S Esmaeilzadeh, K Azizzadenesheli, R Wang, A Chattopadhyay, A Singh, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.

[75] Mohammadamin Mahmoudabadbozchelou, Marco Caggioni, Setareh Shahsavari, William H Hartt, George Em Karniadakis, and Safa Jamali. Data-driven physics-informed constitutive metamodeling of complex fluids: A multifidelity neural network (mfnn) framework. *Journal of Rheology*, 65(2):179–198, 2021.

[76] Ehsan Kharazmi, Zhongqiang Zhang, and George Em Karniadakis. hp-vpinns: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021.

[77] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

[78] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[79] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[80] E Weinan. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 5(1):1–11, 2017.

[81] Lars Ruthotto and Eldad Haber. Deep neural networks motivated by partial differential equations. *Journal of Mathematical Imaging and Vision*, 62(3):352–364, 2020.

[82] François Rousseau, Lucas Drumetz, and Ronan Fablet. Residual networks as flows of diffeomorphisms. *Journal of Mathematical Imaging and Vision*, 62(3):365–375, 2020.

[83] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. *arXiv preprint arXiv:1904.01681*, 2019.

[84] Hanshu Yan, Jiawei Du, Vincent YF Tan, and Jiashi Feng. On robustness of neural ordinary differential equations. *arXiv preprint arXiv:1910.05513*, 2019.

[85] Amir Gholami, Kurt Keutzer, and George Biros. Anode: Unconditionally accurate memory-efficient gradients for neural odes, 2019.

[86] Tianjun Zhang, Zhewei Yao, Amir Gholami, Kurt Keutzer, Joseph Gonzalez, George Biros, and Michael Mahoney. Anodev2: A coupled neural ode evolution framework, 2019.

[87] Aaron Tuor, Jan Drgona, and Draguna Vrabie. Constrained neural ordinary differential equations with stability guarantees. *arXiv preprint arXiv:2004.10883*, 2020.

[88] Julia Ling, Andrew Kurzawski, and Jeremy Templeton. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, 807:155–166, 2016.

[89] Andrea Beck, David Flad, and Claus-Dieter Munz. Deep neural networks for data-driven les closure models. *Journal of Computational Physics*, 398:108910, 2019.

[90] Laure Zanna and Thomas Bolton. Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, 47(17):e2020GL088376, 2020.

[91] Hugo Frezat, Guillaume Balarac, Julien Le Sommer, Ronan Fablet, and Redouane Lguensat. Physical invariance in neural networks for subgrid-scale scalar flux modeling. *arXiv preprint arXiv:2010.04663*, 2020.

[92] Alexis-Tzianni G Charalampopoulos and Themistoklis P Sapsis. Machine-learning energy-preserving nonlocal closures for turbulent fluid flows and inertial tracers. *arXiv preprint arXiv:2102.07639*, 2021.

[93] Klaus Neumann, Andre Lemme, and Jochen J Steil. Neural learning of stable dynamical systems based on data-driven lyapunov candidates. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1216–1222. IEEE, 2013.

[94] Mouhacine Benosman, Boris Kramer, Petros T Boufounos, and Piyush Grover. Learning-based reduced order model stabilization for partial differential equations: Application to the coupled burgers' equation. In *2016 American Control Conference (ACC)*, pages 1673–1678. IEEE, 2016.

[95] Giorgos Mamakoukas, Orest Xherija, and TD Murphey. Memory-efficient learning of stable linear dynamical systems for prediction and control. *arXiv preprint arXiv:2006.03937*, 2020.

[96] Yu Wang, Qitong Gao, and Miroslav Pajic. Deep learning for stable monotone dynamical systems. *arXiv preprint arXiv:2006.06417*, 2020.

[97] Andrew J. Taylor, Victor D. Dorobantu, Hoang M. Le, Yisong Yue, and Aaron D. Ames. Episodic learning with control lyapunov functions for uncertain robotic systems, 2019.

[98] N Benjamin Erichson, Michael Muehlebach, and Michael W Mahoney. Physics-informed autoencoders for lyapunov-stable fluid flow prediction. *arXiv preprint arXiv:1905.10866*, 2019.

[99] Thomas S. Parker and Leon O. Chua. *Stability of Limit Sets*, pages 57–82. Springer New York, New York, NY, 1989.

[100] Aleksandr Mikhailovich Lyapunov. The general problem of the stability of motion. *International journal of control*, 55(3):531–534, 1992.

[101] Michael Malisoff and Frédéric Mazenc. *Constructions of strict Lyapunov functions*. Springer Science & Business Media, 2009.

[102] James D Meiss. *Differential dynamical systems*. SIAM, 2007.

[103] Tim Sauer, James A. Yorke, and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65(3):579–616, Nov 1991.

[104] Peter Lynch and Xiang-Yu Huang. *Initialization*, pages 241–260. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[105] RE Langer. Vito volterra, theory of functionals and of integral and integro-differential equations. *Bulletin of the American Mathematical Society*, 38(9. P1):623–623, 1932.

[106] Chenjie Gu. Qlmor: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(9):1307–1320, 2011.

[107] Elizabeth Qian, Boris Kramer, Benjamin Peherstorfer, and Karen Willcox. Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401, 2020.

[108] Edward N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, March 1963.

[109] A. C. Hindmarsh. ODEPACK, a systematized collection of ODE solvers. *IMACS Transactions on Scientific Computation*, 1:55–64, 1983.

[110] Henry D. I. Abarbanel. *Choosing Time Delays*, pages 25–37. Springer New York, New York, NY, 1996.

[111] Henry D. I. Abarbanel. *Choosing the Dimension of Reconstructed Phase Space*, pages 39–67. Springer New York, New York, NY, 1996.

[112] Julien Clinton Sprott. *Chaos and Time-Series Analysis.* Oxford University Press, Inc., New York, NY, USA, 2003.

[113] Edward N Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.

[114] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced basis methods for partial differential equations: an introduction*, volume 92. Springer, 2015.

[115] Hugo Frezat, Guillaume Balarac, Julien Le Sommer, Ronan Fablet, and Redouane Lguensat. Physical invariance in neural networks for subgrid-scale scalar flux modeling. *Physical Review Fluids*, 6(2):024607, 2021.

Bounded nonlinear forecasts of partially observed geophysical systems with physics-constrained deep learning

Physica D: Nonlinear Phenomena (Major revision)

August 22, 2022

Said Ouala[1], Steven L. Brunton[2], Bertrand Chapron [3],
Ananda Pascual [4], Fabrice Collard [5], Lucile Gaultier [5], Ronan Fablet [1]
Contact: said.ouala@imt-atlantique.fr

[1]Lab-STICC, IMT Atlantique, University of Bretagne Loire, France.
[2] University of Washington, USA.
[3]Ifremer, LOPS, Brest, France.
[4]IMEDEA, UIB-CSIC, Esporles, Spain.
[3]ODL, Brest, France.

Research Highlights:

Overall, our key contributions comprise:

- Neural ODEs are studied for the modeling of partially observed dynamical systems;
- The models are constrained to remain bounded using the Schlegel boundedness theorem;
- This framework allows deriving dynamical models with good asymptotic behaviors;
- These models also have good generalization properties for unprecedented initial conditions;

AUTHORSHIP STATEMENT

Manuscript title: ____

____

____

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *Hong Kong Journal of Occupational Therapy*.

Authorship contributions

Please indicate the specific contributions made by each author (list the authors' initials followed by their surnames, e.g., Y.L. Cheung). The name of each author must appear at least once in each of the three categories below.

Category 1

Conception and design of study: ____________, _____________, ____________, _____________;

acquisition of data: _____________, _____________, _____________, _____________;

analysis and/or interpretation of data: ____________, ___________, ____________, ___________.

Category 2

Drafting the manuscript: ______________, ______________, ______________, ______________;

revising the manuscript critically for important intellectual content: ____________, ____________,

_____________, _____________.

Category 3

Approval of the version of the manuscript to be published (the names of all authors must be listed):

_______________, _______________, _______________, _______________, _______________,

_______________, _______________, _______________, _______________, _______________.

Acknowledgements

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

This statement is signed by all the authors (*a photocopy of this form may be used if there are more than 10 authors***):**

Author's name (typed)	Author's signature	Date
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________
______________________	______________________	____________________

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: