

Consecutive shooting during adjoint optimization enables highly accurate estimation of unobserved state variables of high-dimensional chaotic systems

Kota Shiozawa^{1,*}, Inga Kottlarz^{2,3,4,5,*}, Isao T. Tokuda¹, Ulrich Parlitz^{2,3,4,†} and Sebastian Herzog^{4,6}

¹Graduate School of Science and Engineering, *Ritsumeikan University*, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan

²Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, 37077 Göttingen, Germany

³Institute for the Dynamics of Complex Systems, *University of Göttingen*, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

⁴German Center for Cardiovascular Research (DZHK), Partner Site Lower Saxony, Robert-Koch-Straße 42a, 37075 Göttingen, Germany

⁵Living Systems Institute, *University of Exeter*, Stocker Road, Exeter EX4 4QD, United Kingdom

⁶Department for Computational Neuroscience, Third Institute of Physics—Biophysics, Friedrich-Hund-Platz 1, *University of Göttingen*, 37077 Göttingen, Germany



(Received 27 August 2025; accepted 5 December 2025; published 9 January 2026)

Recent advancements in machine learning have led to the development of sophisticated algorithms for automatic differentiation. These software tools simplify the implementation of the adjoint method by automating the derivation of adjoint equations, enabling researchers to apply the adjoint method to a wide range of problems without requiring extensive expertise in the underlying mathematics. One particularly significant application is the state estimation of partially observable systems. We show that for this application the performance of the adjoint method can be significantly improved by a shooting protocol in which the end of the previous segment is used as the initial condition for estimating the next segment, independent of the success of the estimate for the previous segment. The superior performance of this adjoint optimization-based consecutive shooting method is demonstrated using Lorenz-96 models with dimensions from 9 to 300.

DOI: [10.1103/8c7w-dpdt](https://doi.org/10.1103/8c7w-dpdt)

I. INTRODUCTION

State estimation is a ubiquitous problem in various fields of science and engineering [1–4]. In control theory, for instance, knowing the exact system state is essential for controlling the target, e.g., via feedback [3,5,6]. In the weather forecast, future predictions are made based on current and past states that should contain precise information on the climate system [7]. In practice, however, system's physical state often cannot be entirely determined by direct observations, and some of the state variables have to be estimated from available direct observations of the system. Thus, the goal of the state estimation is to infer hidden variables of the system from a series of measurements observed over time in the presence of statistical noise and other uncertainties. Here, the hidden variables include the state variables and the parameters, which represent the system dynamics. Noise and perturbations are inevitably caused by instruments and environmental factors during the measurements. System uncertainties arise from inaccuracies in the modeling of the system dynamics and variations in the system parameters. In addition to these difficulties, chaotic dynamics impose another fundamental issue,

where a slight estimation error in the initial condition grows exponentially in time [8]. This indicates that, even if the governing equations of the system are fully and explicitly known and the observations contain no noise, the exact estimation of the system state may become extremely difficult. In this sense, state estimation of chaotic systems has been a challenging issue over the last decades, and several approaches have been devised including different methods for data assimilation using (extended) Kalman filters, variational techniques (3DVar and 4DVAR), synchronization, or multiple shooting [4,9–12].

In many previous studies, it turned out that a particularly challenging example for these techniques is the estimation of unobserved state variables and parameters of the D -dimensional Lorenz-96 system [see Eq. (1) in Sec. II], which generates (very) high-dimensional chaotic dynamics. Abarbanel *et al.* [13] found that the estimation of the complete state vector of a D -dimensional Lorenz-96 system by synchronizing the model equations to an L_c -dimensional observed time series (via diffusive coupling) is only successful if $L_c > 0.4D$. For $D = 9$, this means that time series from at least four variables of the system have to be measured. Similar numbers of necessary observations were found in a study on statistical data assimilation by Ye *et al.* [14], where also the model parameter F has been estimated. Stroud and Bengtsson [15] presented successful state reconstruction from observations of every third variable of a 40-dimensional Lorenz-96 system using a Bayesian approach to state and parameter estimation. Using an improved synchronization scheme based on delay embedding, Rey *et al.* [16,17] showed that all state variables of a nine-dimensional Lorenz-96 model can be estimated

*These authors contributed equally to this work.

†Contact author: ulrich.parlitz@ds.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by Max Planck Society.

from a univariate time series of one variable only. This result was confirmed by an observability analysis [18], where for an optimization-based (4D-VAR) estimation algorithm it was demonstrated that estimation success depends crucially on the proper choice of the initial conditions of the estimation algorithm. Due to an intricate basin structure, estimating all variables of a nine-dimensional Lorenz-96 from a measurement of every k th variable turned out to be very challenging for $k > 3$. A rigorous mathematical study by Law *et al.* [19] showed that observing 2/3 of the variables is sufficient to ensure stable estimation results (using 3D-VAR). These authors also presented numerical results for a 60-dimensional Lorenz-96 system indicating that already seven observations may be sufficient in practice if adaptive estimation algorithms are used.

In a more recent study using a combination of conventional data assimilation and reservoir computing, a 6-dimensional Lorenz-96 system with three observed variables and a 40-dimensional model with 15 observed variables were used as examples [20]. A combination of data assimilation and machine learning applied to the Lorenz-96 system was studied by Brajard, *et al.* [21], who found that forecasting performance declined significantly when fewer than every second variable was observed. Guillot *et al.* [22] applied ensemble Kalman and particle filters to the Lorenz-96 model for partial observations, where every second variable was used.

Fablet *et al.* [23] proposed a data assimilation algorithm based on a physics-informed machine learning framework, and present successful state reconstructions using 50% of the variables of a 40-dimensional Lorenz-96 system.

Amemiya *et al.* [24] applied data assimilation to a two-scale Lorenz-96 system, where only partial information about the dynamics of the slow variables is available, and the resulting model bias is corrected by a neural network. This scenario of an imperfect model was also studied earlier by Cocucci *et al.* [25] using two-scale and one-scale Lorenz-96 models with full state observations.

The encountered challenges and limitations to reconstruct the full state vector of the Lorenz-96 system from sparse observation of a few state variables only motivated us to study the performance of adjoint optimization techniques applied to this estimation problem. To do so, we use a recently published implementation of this estimation method called adoptODE [26], which will be presented in Sec. III. By combining this algorithm with a consecutive shooting approach, we devised a very efficient state estimation algorithm that can even for very high-dimensional Lorenz-96 systems successfully cope with sparse observations consisting, for example, of every third variable as a measured signal. The same approach can be applied to any other spatiotemporal system including partial differential equations observed at specific locations, only, as illustrated for the Kuramoto-Sivashinsky equation in Appendix E.

II. THE LORENZ-96 SYSTEM

The Lorenz-96 model (1) is a simplified mathematical model designed to simulate atmospheric dynamics. It was introduced by Edward Lorenz in 1996 [27], and it is used

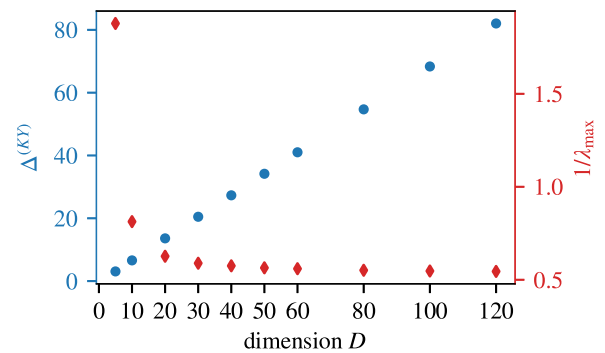


FIG. 1. Kaplan-Yorke dimension $\Delta^{(KY)}$ and Lyapunov time $t_\lambda = 1/\lambda_{\max}$ of the Lorenz-96 system for different system dimensions D . The left axis (blue dots) shows the Kaplan-Yorke dimension $\Delta^{(KY)}$ for a given dimensionality, while the right axis (red diamonds) displays the Lyapunov time $t_\lambda = 1/\lambda_{\max}$.

primarily for studying chaotic behavior in dynamical systems, especially in the context of weather and climate modeling. Its dynamical equations read

$$\frac{dy_i}{dt} = (y_{i+1} - y_{i-2})y_{i-1} - y_i + F, \quad (1)$$

where $y_i \in \mathbb{R}$ ($i = 1, \dots, D$) is a component of the system state $y = [y_1, \dots, y_D]$ and $F \in \mathbb{R}$ is a forcing constant. The components of y are arranged in a ring with periodic boundary conditions $y_{-1} = y_{D-1}$, $y_0 = y_D$, and $y_{D+1} = y_1$.

The Lorenz-96 model can be simulated for arbitrary dimensionalities $D \in [4, \infty)$. The attractor dimension (Kaplan-Yorke dimension) of the system depends linearly on D (see Fig. 1, c.f. [28] for another value of F). The smallest and largest Lyapunov exponents [29–31] initially grow and shrink for increasing D , respectively, but ultimately appear to converge.

Both for comparing results of state reconstruction and prediction to other approaches, and for finding a suitable length of training data, it is important to obtain a characteristic system timescale. The Lyapunov time $t_\lambda = 1/\lambda_{\max}$ is one such timescale, although not immediately accessible in the case of experimental data. It does serve as a good benchmark for comparing our state reconstruction results later on.

Other approaches of determining a characteristic system timescale from measured time series include finding the first zero crossing or minimum of the autocorrelation or the first minimum of the self-mutual information (see Appendix A). In the following, we will use the first zero crossing of the autocorrelation function as a characteristic timescale of the Lorenz-96 system to choose a suitable length of the time series used for state reconstruction.

III. ADJOINT OPTIMIZATION

A. The adoptODE framework

AdoptODE [26] is a framework developed in JAX [32] that utilizes the adjoint method [33–42] for estimating unknowns of differential equations, like parameters or initial conditions. The required Jacobians for the optimization process are computed using automatic differentiation [43]. The core principle

of adoptODE involves solving the equations of motion (EOM) at each optimization step to obtain the system's trajectory, which is influenced by the current parameter estimates and initial conditions. A loss function, denoted by $\mathcal{L}$, is then used to measure the discrepancy between the calculated trajectory and the actual data. The choice of the loss function is flexible and can be tailored to the user's needs, such as the Euclidean distance between the observed and inferred trajectories. Subsequently, adoptODE solves the EOM in reverse time order, simultaneously computing the adjoint state $a(t)$, which represents the sensitivity of the loss function to changes in the system state y at time t and the gradients g based on the loss function $\mathcal{L}$.

Let

$$\frac{dy(t)}{dt} = f(y, t, \theta) \quad (2)$$

be a system of ordinary differential equations given by the user, where $y \in \mathbb{R}^N$ is the true (possibly unobserved) state of the system, t the time, and $\theta \in \mathbb{R}^P$ all associated parameters.

We introduce an observation function h , which maps the true state $y(t)$ (and potentially time t and parameters θ) to the observable quantities $\eta(t) \in \mathbb{R}^M$ (where $M \leq N$ typically):

$$\eta(t) = h(y(t), t, \theta). \quad (3)$$

For simplicity, we can often write $\eta(t) = h(y(t))$ if the observation process itself does not explicitly depend on t or θ beyond their influence on $y(t)$.

The aim is to optimize the parameters θ such that the predicted observable trajectory $\eta(t)$ [derived from $y(t)$ via Eq. (3)] corresponds to a measured trajectory $\eta^*(t)$ also specified (e.g., measured) by the user. For this purpose, a loss functional $\mathcal{L}(Y, Y^*)$ is defined that is to be minimized, where $Y = (\eta(y(\hat{t}_k), \hat{t}_k, \theta))_{k \in \mathcal{K}}$ represents the sequence of predicted observables at discrete time points $\{\hat{t}_k\}_{k \in \mathcal{K}}$, and $Y^* = (\eta^*(\hat{t}_k))_{k \in \mathcal{K}}$ is the sequence of corresponding measured observables. Since $\mathcal{L}$ depends directly on θ (via y and h), a naive optimization would imply that P ODE systems with N variables each have to be solved. This linear scaling with P is not efficiently manageable for cases with a complicated right-hand side of Eq. (2) and is the main motivation for using the adjoint method [33–42]. Instead of solving P ODEs adoptODE uses an augmented system and its derivative:

$$s(t) = \begin{pmatrix} y(t) \\ a(t) \\ \frac{\partial \mathcal{L}(\dots)}{\partial \theta(t)} \end{pmatrix}, \quad \frac{d}{dt}s(t) = \begin{pmatrix} f(y, t, \theta) \\ -a(t)\mathcal{J}_{f,y}(y, t, \theta) \\ -a(t)\mathcal{J}_{f,\theta}(y, t, \theta) \end{pmatrix}. \quad (4)$$

The derivative of the augmented state, $\frac{d}{dt}s(t)$, consists of three components, each with a specific role in the backward integration process used by the adjoint method:

(1) $\frac{d}{dt}y(t) = f(y, t, \theta)$: This is the original system dynamics. When integrating the augmented system backward, this equation is formally run in reverse to reconstruct the forward trajectory $y(t)$ (though in practice, the forward trajectory is often precomputed and stored, or recomputed in segments, to ensure stability and accuracy).

(2) $\frac{d}{dt}a(t) = -a(t)\mathcal{J}_{f,y}(y, t, \theta)$: This describes the backward evolution of the adjoint state $a(t)$. The term $\mathcal{J}_{f,y}(y, t, \theta)$ is the Jacobian of the system dynamics f with respect to the

state y . It quantifies how a small perturbation in the state $y(t)$ affects its own rate of change $\frac{d}{dt}y(t)$. Multiplied by $-a(t)$, this term governs how the sensitivity of the loss with respect to the state propagates backward through time.

(3) $\frac{d}{dt}(\frac{\partial \mathcal{L}(\dots)}{\partial \theta(t)}) = -a(t)\mathcal{J}_{f,\theta}(y, t, \theta)$: This component describes the rate at which the gradient of the loss function $\mathcal{L}$ with respect to the parameters θ accumulates during the backward integration. The term $\mathcal{J}_{f,\theta}(y, t, \theta)$ is the Jacobian of f with respect to the parameters θ . It represents how a change in parameters θ instantaneously influences the rate of change of the system state, $\frac{d}{dt}y(t)$. This instantaneous influence on the state's evolution is then weighted by the adjoint state $a(t)$ [which, as per Eq. (5), measures the sensitivity of the total loss to changes in $y(t)$], thereby determining the contribution of parameter changes at time t to the overall gradient $\frac{\partial \mathcal{L}}{\partial \theta}$.

The adjoint state $a(t)$ itself, as defined in Eq. (5):

$$a(t) = \frac{\mathcal{D}\mathcal{L}((\eta(y(\hat{t}_k), \hat{t}_k, \theta))_{k \in \mathcal{K}}, (\eta^*(\hat{t}_k))_{k \in \mathcal{K}})}{\mathcal{D}y(t)} \quad (5)$$

is the lasting or total derivative of the loss function $\mathcal{L}$ with respect to the system state $y(t)$. This means it quantifies how the overall loss $\mathcal{L}$ (which depends on observations η at various times $\hat{t}_k$) changes if the system's state y is perturbed specifically at time t . This perturbation at $y(t)$ has two consequences: (1) It directly affects the observation $\eta(t) = h(y(t), t, \theta)$ if t is an observation time point. (2) It alters the trajectory of $y(\tau)$ for all subsequent times $\tau > t$ according to the system dynamics $f(y, \tau, \theta)$, thereby affecting all future observations $\eta(\hat{t}_k)$ for $\hat{t}_k > t$. The adjoint state $a(t)$ encapsulates the combined impact of these effects on the total loss.

B. State and parameter estimation from partial observations

Here, we deal with the problem of estimating unobserved state variables from partial observations. The system state is divided into observed and unobserved variables $y(t) = \{y_{\text{obs}}(t), y_{\text{uno}}(t)\}$, which defines the observation function as $\eta^*(t) = h(y(t)) = y_{\text{obs}}(t)$. From a time series of the observed state variables $\{\eta^*(t)\}$, we estimate the parameters $\theta = \{y_{\text{uno}}(0), p\}$ that are composed of the initial condition of the unobserved state variables $y_{\text{uno}}(0)$ and the unknown system parameters p . Once θ is estimated, the system state can be reconstructed from the estimated initial condition using the estimated system parameters. Although adoptODE is a powerful tool for estimating parameters of differential equations, it is a challenging issue to estimate unobserved state variables of chaotic system. Due to their sensitive dependence on initial conditions, the trajectory generated from the estimated initial condition may quickly deviate from the true trajectory. To address this issue, we divide a measured time series into short segments and apply adoptODE to each segment. By estimating the initial condition of each segment, the trajectories for the corresponding segment can be reconstructed from the estimated initial condition. If the segment length is short enough, the deviation from the true trajectories should remain small.

We now describe our method in detail. Suppose we have a multivariate M -dimensional time series $\{\eta^*(t)\}$ that is divided into nonoverlapping segments $\{\eta^*(n\Delta t)\}_{n=0}^{l-1}$ of length $t_S = l\Delta t$, where Δt represents the sampling time interval. As a choice for t_S , the first zero crossing of the system's

autocorrelation function is used. For each segment, we use adoptODE to estimate the initial condition of the unobserved state variables and the unknown system parameters. Since the estimation results depend on the initial guesses provided to adoptODE, we generate N_{guess} different initial guesses for each segment.

For each segment of the time series, the estimation error is calculated as

$$E_{\text{measured}} = \frac{1}{lM} \sum_{n=0}^{l-1} \|h(y^*(n\Delta t)) - h(y(n\Delta t))\|_2^2, \quad (6)$$

where y^* and y represent the true and reconstructed states composed of both observed and unobserved state variables. The estimation error quantifies the difference between the observed variables $\eta^* = h(y^*)$ and the reconstructed variables $\eta = h(y)$. The state and the parameters that provide the minimum estimation error are selected as the final result. The present estimation error is also utilized as the loss function of adoptODE. Furthermore, we define the true error as

$$E_{\text{true}} = \frac{1}{lN} \sum_{n=0}^{l-1} \|y^*(n\Delta t) - y(n\Delta t)\|_2^2, \quad (7)$$

which measures the difference between the true state y^* and the reconstructed state y , including the unobserved variables. Since E_{true} requires a knowledge on the unobserved variables, it cannot be calculated in a practical situation. We used E_{true} only for evaluating the performance of our method. High correlation between E_{true} and E_{measured} implies E_{measured} provides a good measure for evaluating the estimated state and parameters.

IV. SEGMENT-WISE STATE AND PARAMETER ESTIMATION

A. State and parameter estimation of the nine-dimensional Lorenz-96 system

To demonstrate our method, numerical simulations were carried out. For our first example, we generated a true time series by integrating the 9-dimensional Lorenz-96 model (1). The generated time series consisted of 10^4 data points with a sampling time of $\Delta t = 0.01$. We assumed that every third variable, i.e., (y_1, y_4, y_7) was measured, and thus the dimension of the measured time series was $M = 3$. The forcing constant F was considered as an unknown system parameter, with the true value set to $F_{\text{true}} = 8.17$. The observed time series was divided into 250 segments, each consisting of $l = 40$ data points. To each segment adoptODE was applied with $N_{\text{guess}} = 100$ different initial guesses, and the best state and parameter were selected based on E_{measured} . For each initial guess, initial values for the unobserved state were randomly chosen from a uniform distribution in $[-1, 4]$, while the system parameter value was chosen from a uniform distribution in $[4, 20]$. Figures 2(a) and 2(b) show the true state and the observed state, respectively, where the black regions indicate the unobserved states. Figure 2(c) shows the state estimated by adoptODE, while Fig. 2(d) shows the difference between estimated and true states. The reconstructed states are in good agreement with the true ones. Figure 2(e) shows the estimated

system parameter values for each segment which are very close to the true value.

B. Dependence on dimensionality and sparsity of measured data

We next examined how the dimensionality and the sparsity of the measured data influence our method. For simplicity, we assumed that the parameter F was known, and we estimated only the unobserved state variables. For various dimensions ranging from $D = 9$ to $D = 120$, we generated a true time series and extracted the observed state variables. In the Lorenz-96 model, where the state variables are arranged on a ring, we assumed that the observables were uniformly distributed and measured every third to every seventh variable. Taking a 12-dimensional Lorenz-96 model as an example, the measured variables were (y_1, y_4, y_7, y_{10}) , (y_1, y_5, y_9) , and (y_1, y_7) . For each case, we generated 100 different segments, each segment length was set to the first zero crossing of the autocorrelation function. The sampling time step was fixed to $\Delta t = 0.01$. adoptODE was applied once on each segment. Figures 3(a) and 3(b), respectively, show the dependence of E_{true} and E_{measured} on the dimension D . The errors are represented in histograms, where some of them, e.g., $D = 9$, show bimodal distributions, where the lower and upper peaks represent successful and unsuccessful estimations, respectively. The errors concentrating around the lower peak indicate the cases in which our method precisely reconstructed the state with a few iterations, whereas those around the upper peak imply that the method was not so effective. In the case of $D = 9$, i.e., every third variable measured, E_{true} and E_{measured} are distributed bimodally. This indicates that, if we repeat adoptODE multiple times from different initial guesses, successful estimates should be eventually obtained. On the other hand, as the dimensionality or the sparsity of the measured time series increased, the number of successful estimations decreased, and the distributions became unimodal. In this case, multiple applications of adoptODE may not improve the estimation results. Although the distributions of E_{true} and E_{measured} show similar dependencies on the dimensionality and sparsity of the measured time series, there exists a difference between them. Namely, E_{true} tends to take larger values compared to E_{measured} . This tendency is more pronounced for sparser time series, which may cause overfitting of the measured data. As the level of the sparsity of the measured time series was increased, the state estimation became more difficult and E_{true} increased. However, since E_{measured} was calculated only for the measured time series, it was prone to overfitting and tended to underestimate the true error. This implies that E_{measured} misidentified unsuccessful estimations as successful ones, and no longer provided a reliable error measure for the sparse time series.

We next evaluated the success rate of the estimations. An estimation was considered successful if the estimation error was below a threshold and was considered unsuccessful otherwise. To take into account the dependence of the success rate on the threshold, three values of thresholds were considered: 10^{-4} , 10^{-5} , and 10^{-6} , for each of which the success rate was calculated. Figures 3(c) and 3(d) show the success rates derived from E_{true} and E_{measured} , respectively. The markers correspond to success rates calculated using a

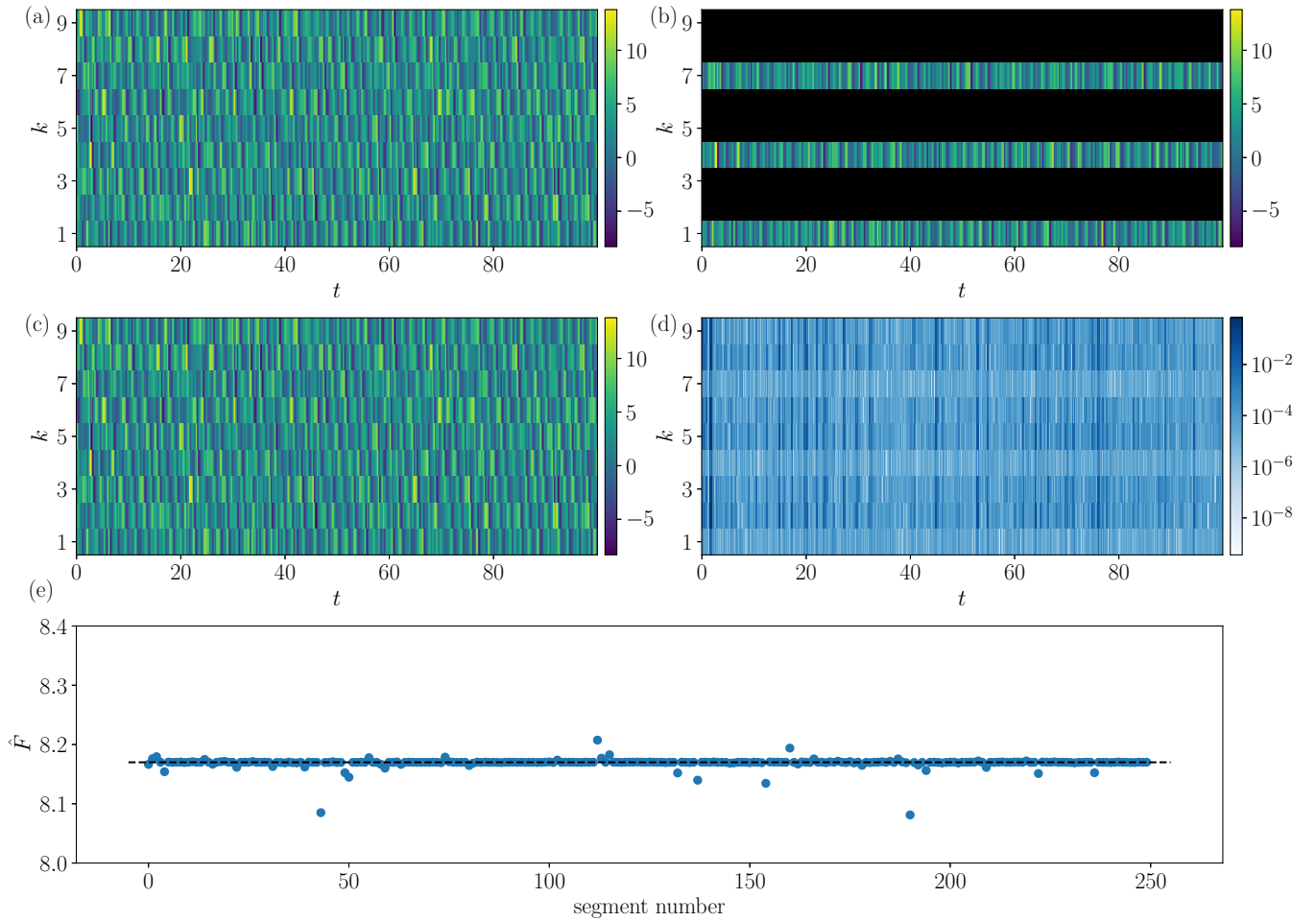


FIG. 2. State and parameter estimation of the nine-dimensional Lorenz-96 model. (a) True states of the 9-dimensional Lorenz-96 model. (b) True states measured as time series (black regions indicate unobserved variables). (c) Estimated state by adoptODE. (d) Absolute error between the true and estimated states. (e) Estimated system parameter for each segment. The dashed line is the true value $F_{\text{true}} = 8.17$. In panels (a)–(d), abscissa and ordinate represent time and the number of state variables, respectively.

threshold value of 10^{-5} . The upper bounds of the error bars represent success rates calculated using a threshold value of 10^{-4} , while the lower bounds correspond to those calculated with a threshold of 10^{-6} . In the case that the dimensionality was $D = 9$ and every third variable was measured, the success rate was approximately 0.54, with very small error bars. This demonstrates that our method can precisely reconstruct the state under these conditions. As the dimensionality increased, however, the success rate decreased rapidly, indicating that the method became less effective. As the sparsity of the measurements increased, not only the success rate decreased, but also the gap between the success rates derived from E_{true} and E_{measured} widened. This suggests that the increased level of sparsity makes the state estimation more challenging and evaluation of the success rate less reliable. Furthermore, the success rates calculated from E_{measured} exhibited larger error bars. This is because the separation between the two peaks in the bimodal distribution, which corresponded to successful and unsuccessful estimations, became less distinct, and determination of the successful state estimation became sensitive to the setting of the threshold.

C. Choosing an initial guess for the optimization procedure

Gradient-based optimization procedures can be highly sensitive to their initial conditions.

There is an abundance of different options for choosing the initial conditions for the optimization. Initially, we found in numerical experiments that drawing the initial condition from a uniform distribution in $[-1, 4]$ yields good results. One could also randomly sample from the amplitude distribution of the observed variables, or sample from a clipped version of it to avoid overestimating the occurrence of amplitudes in the tails of the estimated distribution.

Figure 4 compares the success rates for these different approaches by showing the distributions of the mean squared error after optimizing a single segment of length 0.45 of the 120-dimensional system with every third variable observed. We compare drawing randomly (1) from the full amplitude distribution of observed variables, (2) from a uniform distribution in $[-1, 4]$, (3) from a uniform distribution in the interval of the second and third quartile of the full amplitude distribution, and (4) from the amplitude distribution after clipping the tails before the second and after the third quartile.

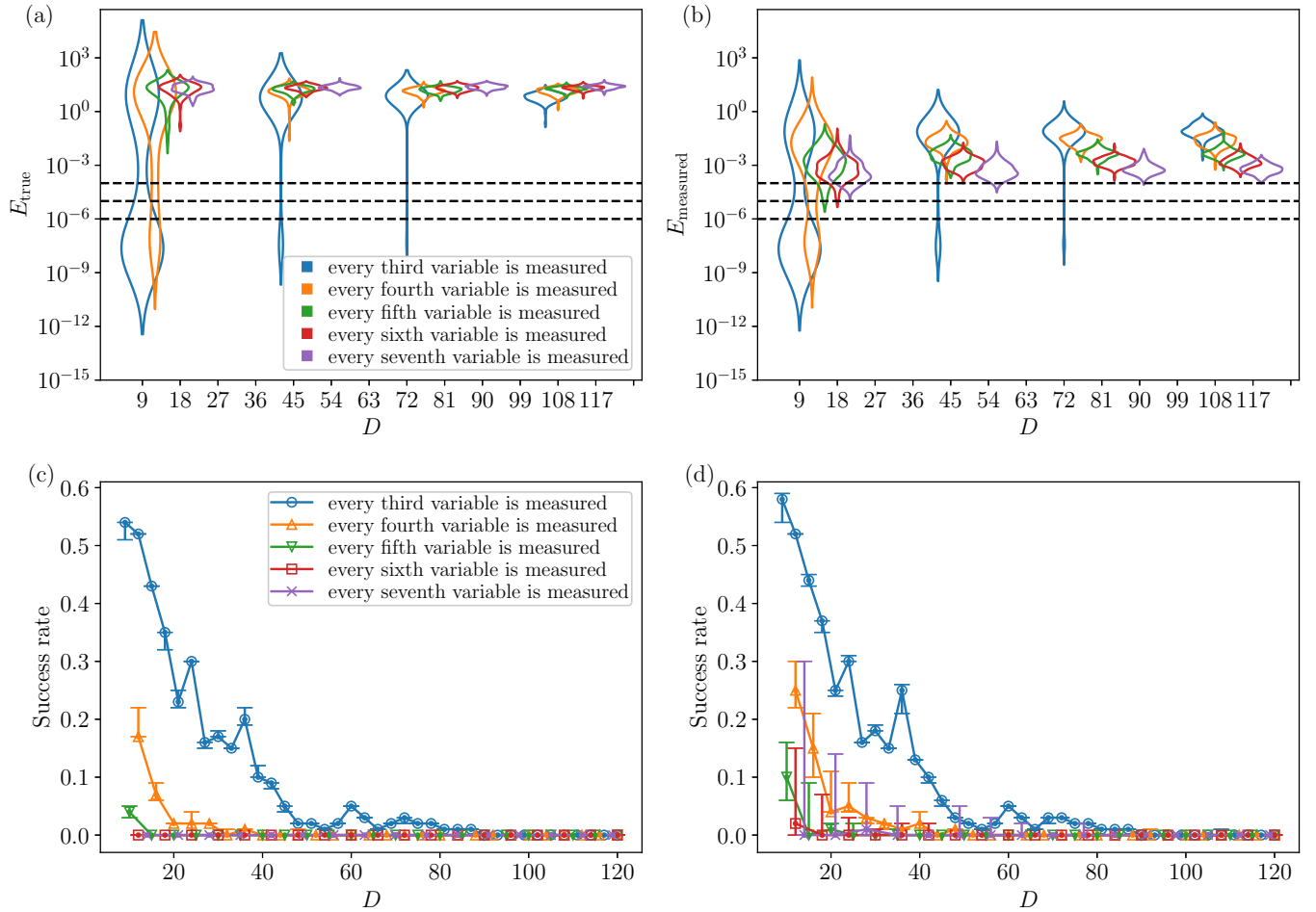


FIG. 3. Performance of the state estimation of the Lorenz-96 model. The dimension was varied from $D = 9$ to 120, in which the Kaplan-Yorke dimension ranged from ~ 6 to ~ 82 . Panels (a) and (b) show distributions of the estimation errors, E_{true} (7) and E_{measured} (6), respectively. Panels (c) and (d) display dependence of the success rate on the dimension D . The success rates in panel (c) were calculated based on E_{true} , whereas those in panel (d) were based on E_{measured} . The state estimation was considered successful if the true error or the estimation error was below a threshold, where three values (10^{-4} , 10^{-5} , 10^{-6}) were set as the threshold. The markers indicate success rates calculated with a threshold of 10^{-5} . The upper and lower bounds of the error bars correspond to success rates calculated with threshold values of 10^{-4} and 10^{-6} , respectively.

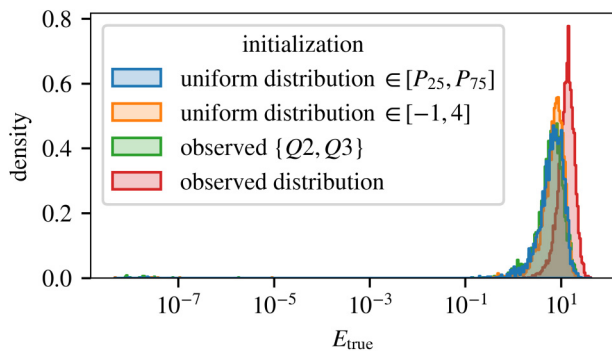


FIG. 4. True mean squared error of the 120-dimensional system with every third variable observed after optimizing a single segment of length 0.45 with different choices for the distribution from which to draw the initial guess for the optimization. For each option, 50 different time series were reconstructed 100 times per time series with a different initial guess. The observed amplitude distribution was estimated from a time series of length 100.0, starting at the beginning of the optimized segment.

We observe that all options that involve clipping the distribution have an approximately equal error distribution, while drawing from the entire observed distribution yields less accurate results. We hypothesize that this is due to not being able to estimate the tails of the amplitude distribution correctly from limited data, leading to overestimation of values in the tail.

D. Consecutive shooting: state estimation using the end of the previous segment

So far, adoptODE has been applied independently for each segment of the time series. We have shown in Fig. 4 that the distribution from which the initial guess for the full system state is drawn has at least some influence on the quality of the state estimation.

The optimization could be improved if more segments are taken into account to capture long-term behavior of the system. For this purpose, we developed an iterative method that utilized adoptODE recursively to consecutive segments.

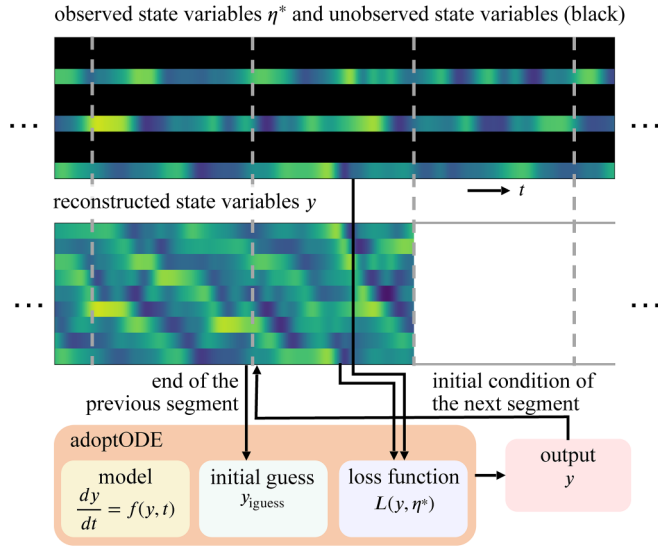


FIG. 5. Main processing steps of consecutive shooting using adoptODE.

In our previous method, the initial guess for unobserved state variables was estimated from a random initial guess. In the iterative method, the final state variables, which are estimated for the previous segment, are used as the initial guess for the initial condition of the next segment as shown in Fig. 5. If the state estimation of the previous segment was successful, the final state of the previous segment should be close to the true initial state of the next segment, thus providing a good initial guess.

We observe that, regardless of whether the estimation of the first segment would be classified as “successful” or not, repeating this process over several consecutive segments of the time series leads to a convergence to the true state after a certain number of segments. This is illustrated in Fig. 6 for a segment length of $t_s = 0.45$ and further demonstrated in Fig. 7 for differently chosen segment lengths.

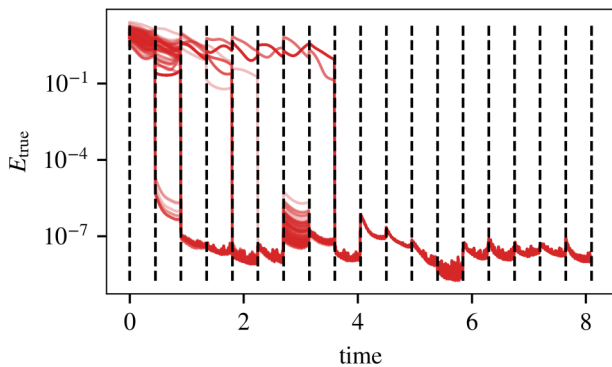


FIG. 6. True mean squared error over time for consecutive optimizations of multiple initial guesses for the same time series. In this example, the system dimension is $D = 120$, the segment length $t_s = 0.45$, and every third variable is observed. Black dashed lines mark the borders between different segments. After four segments, or $t = 3.6$, all optimizations have converged closely to the true state.

We have seen that the method enables convergence to the true state even with sparse observations. This naturally raises the question of how well the system can be predicted beyond the training data. Assuming we now want to start a prediction at the end of the last segment and predict the full system state. Since our method assumes knowledge of the system equations, the long-term prediction power is directly given by the error of the estimated final state of an observed segment (the initial condition of the long-term prediction) and the maximum Lyapunov exponent of the system, as illustrated in Appendix D.

E. Dependence on segment length

Figure 7(a) shows the success rates for reconstructing the full state of the 120-dimensional Lorenz-96 system with every third variable being observed. We repeated our reconstructions for different segment lengths around the first zero crossing of the autocorrelation function, starting at 0.35 system time and ending at 1.0 system time. The first zero crossing of the autocorrelation function is $t_{a0} \approx 0.45$, while the Lyapunov time for the 120-dimensional system lies around $t_\lambda \approx 0.55$.

We clearly observe that with increasing segment length, there is initially an increase in success rate for the shorter segments, followed by a decrease for the longest segment lengths. The best performing segment length is 0.65, that is, ca. $1.2t_\lambda$ or $1.45t_{a0}$. The first zero crossing of the autocorrelation function thus seems to be a good estimate for the order of magnitude of the optimal segment length, but an exploration around this value is recommended.

For longer segment lengths $t_s > 0.45$, we observe increasingly often that for an optimized initial condition of one of the later segments, the true error initially decreases with time within the segment and later on increases. This is discussed further in Appendix C.

V. STATE ESTIMATION FROM NOISY OBSERVATIONS

So far, we have dealt only with noise-free data. Observation noise makes the problem significantly more challenging, as the error in estimating the true state becomes indistinguishable from the noise added to the data. As a result, the estimation error becomes less correlated with the true error, making it difficult to assess whether the state estimation has been successful. Here, we examine how the consecutive shooting approach performs under such noisy conditions. Furthermore, we revisit a method introduced in Sec. III B to address the challenge of evaluating estimation success. By selecting the best estimate from multiple candidates obtained from different initial guesses, it may be possible to identify a successful estimate without access to the true error.

A. Consecutive shooting for noisy data

The difficulty of estimating the full system state from noisy observations comes from the limitation of evaluating the true estimation error that is perturbed by the observation noise. Namely, if the observations deviate on average by σ from the ground truth, the error cannot distinguish their difference, and consequently the estimated state cannot approach to the truth

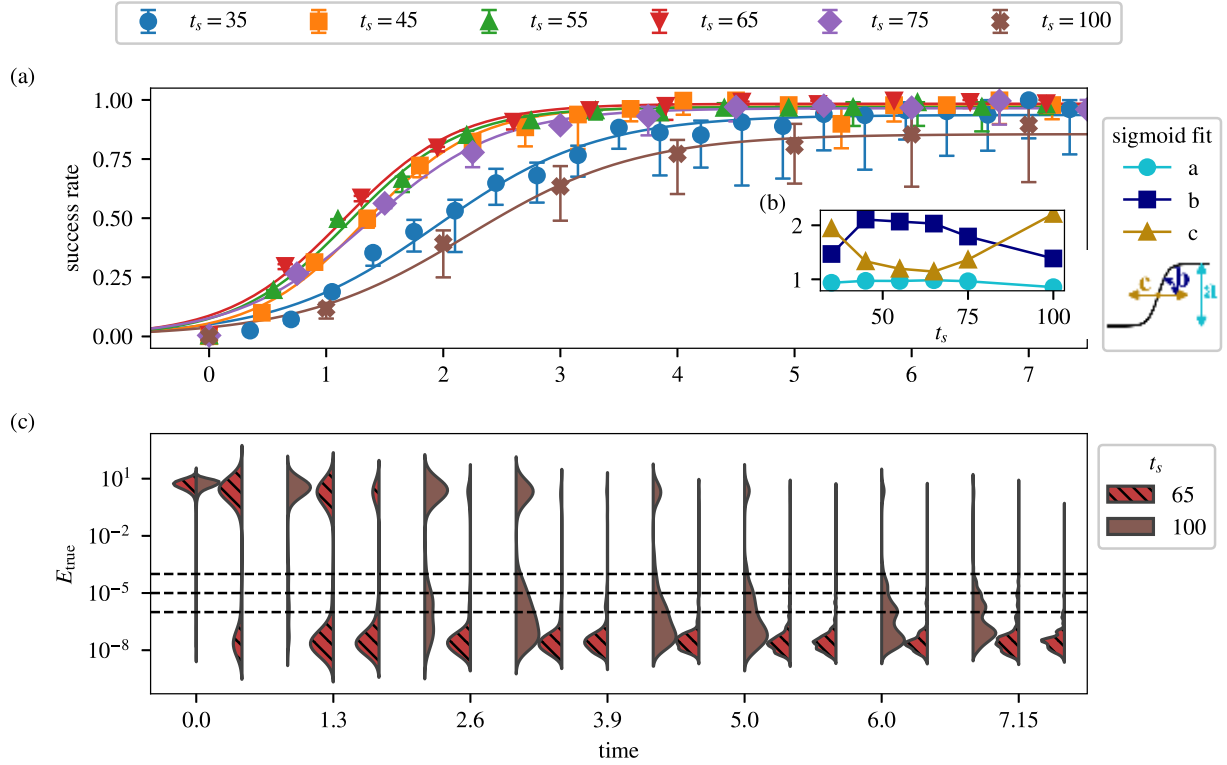


FIG. 7. Success rate of state estimation for the 120-dimensional Lorenz-96 system (every third variable observed) for different segment lengths t_s in case of consecutive shooting (see Sec. IV D). The sampling time step was fixed to $dt = 0.01$. In panel (a), the success rate was calculated for different thresholds: The markers represent success rates calculated using a threshold of 10^{-5} , the lower error corresponds to a threshold of 10^{-6} and the upper one to a threshold of 10^{-4} . The segment lengths $t_s \in \{0.45, 0.55, 0.65\}$ show the highest success rates early on. Panel (b) shows the parameters of fitting a sigmoid curve (fit plotted) to the respective data. Panel (c) shows the distributions of the true mean squared error for consecutive segments and the three error thresholds considered in panel (a), marked by dashed lines. For both displayed segment lengths $t_s = 0.65$ and $t_s = 1.0$, we can see a transition from an (almost) unimodal distribution with high error via a bimodal distribution to an (almost) unimodal distribution with low error.

closer than σ . Since the correlation between E_{measured} and E_{true} is weakened, it is difficult to determine whether the estimation is successful or not.

Here, we examined the effect of observation noise. For simplicity, the parameter value was assumed to be known, and only the state variables were estimated from noisy time series. Note that since the observed state variables are perturbed by noise, we estimate not only the unobserved state variables but also the observed ones. We generated 100 sets of time series (i.e., ground truth) by integrating the 9-dimensional Lorenz-96 model from different initial conditions with a sampling time interval of $\Delta t = 0.01$. By observing every third variable and adding observation noise, we obtained 100 sets of measured time series. The observation noise was uniformly distributed, and the ratio between the noise variance and the signal variance, σ_N^2/σ_S^2 (hereafter, noise-to-signal ratio), was set to 10^{-3} . We applied consecutive shooting to the measured time series with segment length $t_s = 0.4$. Figure 8(a) shows the evolution of E_{true} over 20 shooting steps. Initially, the true errors formed two groups, one corresponding to successful estimations and the other to unsuccessful. As the consecutive shooting step proceeded, the true error converged eventually to the successful group, and by the tenth shooting step, esti-

mations for all 100 time series became successful. This result demonstrates the robustness of the consecutive shooting even in the presence of observation noise. Figure 8(b) shows the corresponding E_{measured} . Unlike E_{true} , successful and unsuccessful estimations could not be separated in the initial phase, since E_{measured} was calculated only from noisy observations. Although the convergence of E_{measured} to lower error level was still discernible, it was less discernible than the case of E_{true} . This result exemplifies the difficulty of judging whether the estimation is successful or not in the presence of observation noise.

We next examined the dependence of consecutive shooting on the noise level. We varied the noise-to-signal ratio from 0.0 to 10^{-1} , and performed the consecutive shooting. Figure 9 shows the distributions of E_{measured} and E_{true} at the 20th step of the consecutive shooting for variable noise-to-signal ratios. Since the lower bound of E_{measured} was determined by the noise variance σ_N^2 , E_{measured} exhibited narrow distributions centered around the noise variance. E_{true} , on the other hand, exhibited much broader distributions, since E_{true} did not have such a lower bound. Despite the difference in the widths of the distributions, E_{true} distributed also unimodally centered around the lower bound of E_{measured} . This implies that the

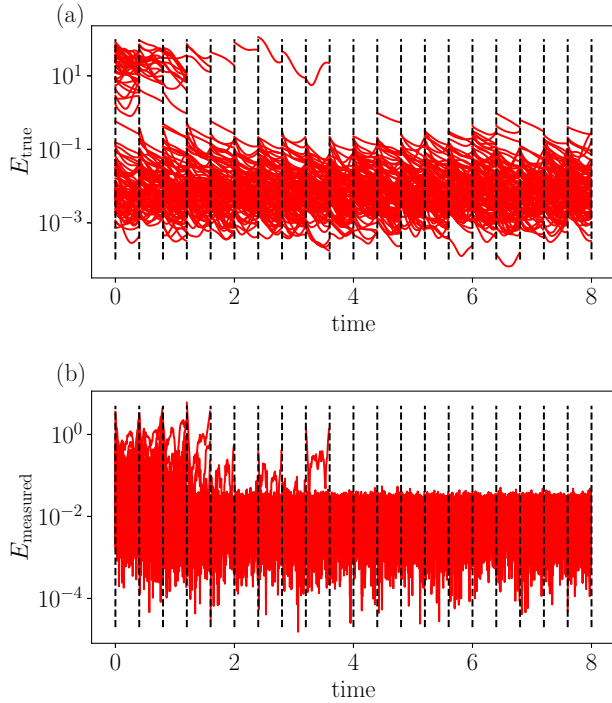


FIG. 8. Application of consecutive shooting to noisy data. True error E_{true} (a) and measured error E_{measured} (b) are plotted over 20 steps of the consecutive shooting. 100 time series were generated by integrating the nine-dimensional Lorenz-96 system from different initial conditions. Every third variable was measured as the data, and observation noise was added with a noise-to-signal ratio of 10^{-3} .

consecutive shooting achieved the minimum possible value of E_{true} in the presence of observation noise.

B. Select the best estimate from multiple trials

Although the consecutive shooting was shown to be robust against observation noise, judging whether the estimation was successful or not remains challenging. To address this issue,

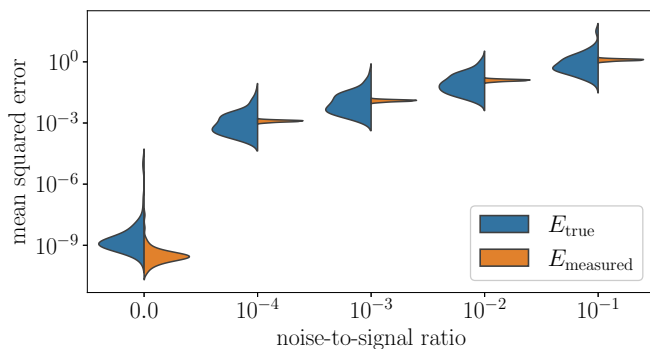


FIG. 9. Dependence of consecutive shooting on noise-to-signal ratio. Distributions of true errors E_{true} (blue) and measured errors E_{measured} (orange) were drawn at the 20th step of the consecutive shooting, which was applied to 100 sets of noisy data. For each data, every third variable was observed from the 9-dimensional Lorenz-96 system. The noise-to-signal ratio was varied from 0.0 to 10^{-1} .

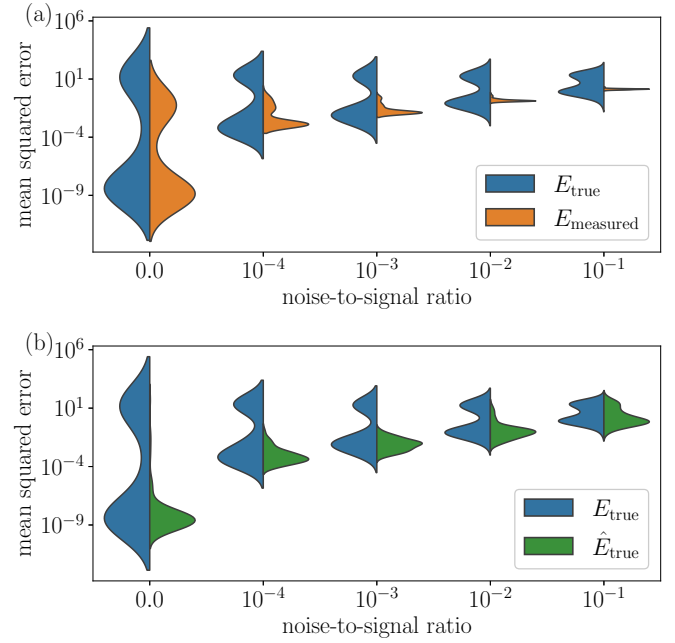


FIG. 10. Results of selecting the best estimate from multiple trials. Distributions of true errors E_{true} (blue), measured errors E_{measured} (orange), and selected best errors $\hat{E}_{\text{true}}$ (green) were drawn. Single shooting was applied to 100 sets of noisy data. For each data, every third variable was observed from the 9-dimensional Lorenz-96 system. The noise-to-signal ratio was varied from 0.0 to 10^{-1} .

we again considered the method presented in Sec. III B. In this method, multiple estimations were performed independently from different initial conditions and the one with the smallest measured error was selected. This selection increased the likelihood of estimating more accurate states, even when direct evaluation of the true error was not possible.

This approach was applied to 100 sets of noisy data, which were generated in a similar manner as the previous simulation. Namely, every third variable was measured from the 9-dimensional Lorenz-96 model (1) as a true time series, which consisted of 4×10^3 data points with a sampling time interval of $\Delta t = 0.01$. To the time series, uniformly distributed random noise was added, where the noise-to-signal ratio was varied from 0.0 to 10^{-1} . The noisy time series was divided into 100 segments, each of which consisted of $l = 40$ data points. To each segment, adoptODE was applied $N_{\text{guess}} = 100$ times with different initial guesses. Figure 10(a) shows the distributions of E_{measured} and E_{true} for variable noise-to-signal ratios. Overall, E_{true} exhibited bimodal distributions with two peaks corresponding to successful and unsuccessful estimates. For small noise-to-signal ratios, E_{measured} also showed bimodal distributions, whereas, as the noise-to-signal ratio increased, it became unimodal with a sharp peak centered around the noise variance. From the unimodal distribution of E_{measured} , it is not straightforward to judge whether the estimation was successful or not.

We then selected, for each segment, the estimation result that gave the minimum E_{measured} among $N_{\text{guess}} = 100$ trials. As a substitute for the true error, the minimum E_{measured} selected by this way was denoted as $\hat{E}_{\text{true}}$. Since only the best

estimates were selected, it is expected that $\hat{E}_{\text{true}}$ should be concentrated around the lower peak of E_{true} , which represented successful estimates. Figure 10(b) compares $\hat{E}_{\text{true}}$ with E_{true} . In the case of noise-free data, the distribution of $\hat{E}_{\text{true}}$ peaked around the lower peak of E_{true} , indicating that the selected estimates were indeed successful. Similar results were obtained in the cases of small noise (i.e., noise-to-signal ratio of 10^{-4} and 10^{-3}). However, as the noise-to-signal ratio was increased to 10^{-2} and 10^{-1} , the tail of $\hat{E}_{\text{true}}$ overlapped with the upper peak of E_{true} . This indicates that a certain proportion of the estimates was unsuccessful, even though they were selected as the best among 100 trials. This unsuccessful proportion grew as the noise level was increased. Although the present approach might be less effective for strong noise, reliable judgment of successful state estimation was still possible as far as the noise intensity was small enough.

VI. CONCLUSION

In this manuscript, we showed two main results: (1) Adjoint optimization can be used for the successful state estimation and prediction of hyperchaotic dynamical systems. (2) Our approach of a state estimation using shooting in consecutive segments leads to convergence of the estimated state to the true state to almost numerical precision in the noiseless case.

We believe that the consecutive shooting approach (2) is promising also for other dynamical systems, and possibly even for other gradient descent-based optimization approaches. This expectation is supported by similar results for the Kuramoto-Sivashinsky equation (see Appendix E) and the complex Ginzburg-Landau equation.

In contrast to a classic shooting-based approach, we do not optimize the same segment or initial condition multiple times, but instead rely on the convergence of the trajectory to the chaotic attractor. From a dynamical systems point of view, this is an interesting phenomenon: Even the “badly” estimated final state of the first segment is a better initial guess for the initial state of the next segment than drawing the guess from an (empirical) amplitude distribution.

From a probabilistic point of view, our knowledge of the model equations allows us to use the observations of the previous segment to refine the estimation of the probability distribution $p(y_i(t = t_S) | \{y_j(t \in [t_0 - t_S, t_S])\})$, where $j \in \{\text{observed variables}\}$ and $i \in \{\text{unobserved variables}\}$. While in a randomly guessed initial condition, the unobserved are treated as independent (of both each other and the observed variables), integrating these independent initial conditions allows to capture dependencies between the variables, so that an initial condition based on this guess is closer to the true attractor than a random, independent guess. This gives us a refined distribution to select an initial condition for the optimization of the next segment, yielding more accurate results, and over several iterations of this process, we can converge to the true system state.

Another angle of view for the convergence of our method is the context of gradient landscapes. While a rigorous convergence proof for high-dimensional chaotic systems is beyond the scope of this work, the effectiveness of the proposed consecutive shooting method can be rationalized by consider-

ing the optimization landscape in the limit of short segment lengths. In general, the loss landscape of chaotic systems becomes highly nonconvex over long integration times due to sensitive dependence on initial conditions, resulting in numerous local minima where gradient-based methods can become trapped. However, over sufficiently short intervals T_S , the system dynamics is well-approximated by their linearization (the tangent linear model). In this regime, the mapping from the initial condition to the observations is approximately linear, rendering the least-squares objective function locally quadratic (convex). In such a convex landscape, gradient descent methods are theoretically guaranteed to converge to the global optimum. Therefore, the success of our method relies on the fact that short segments ensure we are essentially optimizing within a “convex basin.” By selecting T_S near the timescale of the system’s autocorrelation (as shown in Fig. 7)—which is typically on the order of the Lyapunov time—we effectively restrict the optimization to a temporal window where this linear approximation holds validity. The consecutive shooting protocol then functions as an iterative mechanism to bridge these local basins: The solution from the basin of the previous segment is passed as the initial guess into the basin of the next segment. This effectively chains together local linear optimizations to track the global nonlinear trajectory, avoiding the local minima associated with the global folding of the attractor.

In the results presented in this manuscript, we can also see that often, some initial guesses in the consecutive shooting approach already lead to convergence within the first segment. However, getting a reasonable probability of “hitting” one of these initial guesses would warrant 100 or more optimizations with different initial guesses, and then selecting the best one based on the observed error. In contrast, with consecutive shooting, we can almost guarantee a convergence after five or six segments of optimal length and observability. That means one could use a single initial guess, apply the consecutive shooting six times, and almost surely have a highly accurate estimation of the full system state, reducing the number of needed optimizations by an order of magnitude. Of course, the first few segments might not be estimated very well, but once we have a good convergence, nothing forbids us to simply take the initial state of a later segment and simply integrate the equation of motion backward in time to get a better estimation of the initial system state.

This illustrates again how crucial the choice of initial guess is for gradient-descent-based optimization methods. An explanation for why this approach works could be that the optimization algorithm finds initial conditions in the subspace spanned by covariant Lyapunov vectors corresponding to negative Lyapunov exponents. This hypothesis would be supported by the observation that if an estimated initial condition is evolved, the error of the state initially decreases over time.

We want to clearly state that the results presented in this manuscript crucially depend on the knowledge of the system’s equation of motion and its parameters. Without this knowledge, a successful estimation of unobserved variables, and therefore prediction of the time series, becomes even more challenging.

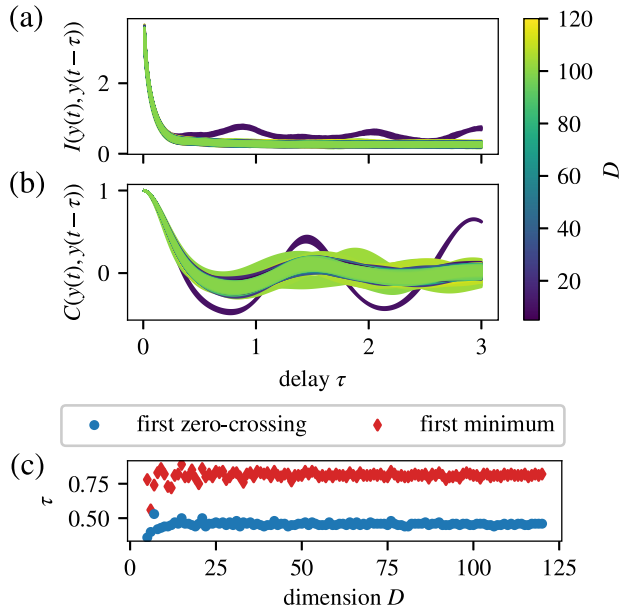


FIG. 11. (a) Self-mutual information and (b) autocorrelation functions of the Lorenz-96 system, and (c) the first zero crossings and minima of the autocorrelation functions for different system dimensions D .

Knowing this, the advances in system estimation, and the results shown here, are primarily of conceptual interest

and not (yet) directly applicable to high-dimensional systems without good knowledge of equations of motion.

ACKNOWLEDGMENTS

I.K. acknowledges being funded by a Ph.D. Fellowship (Promotionsstipendium) by Studienstiftung des Deutschen Volkes. K.S. was supported by a Fellowship from Japan Society for the Promotion of Science (No. 23KJ2098).

K.S. and I.K. performed the formal analysis, data curation, and validation, and wrote the original draft. I.T., S.H., and U.P. conceptualized and supervised the work and wrote the original draft. All authors performed review and editing of the original draft.

DATA AVAILABILITY

The data that support the findings of this article are openly available in the repositories [44,45].

APPENDIX A: CHARACTERISTIC SYSTEM TIMESCALES

Apart from the Lyapunov time t_λ , there exist other ways to estimate a characteristic timescale of a dynamical system, which is more suitable for measured data, where the underlying equation of motion is unknown.

We calculated the autocorrelation and self-mutual information functions for different dimensionalities D using the DynamicalSystems.jl library [46].

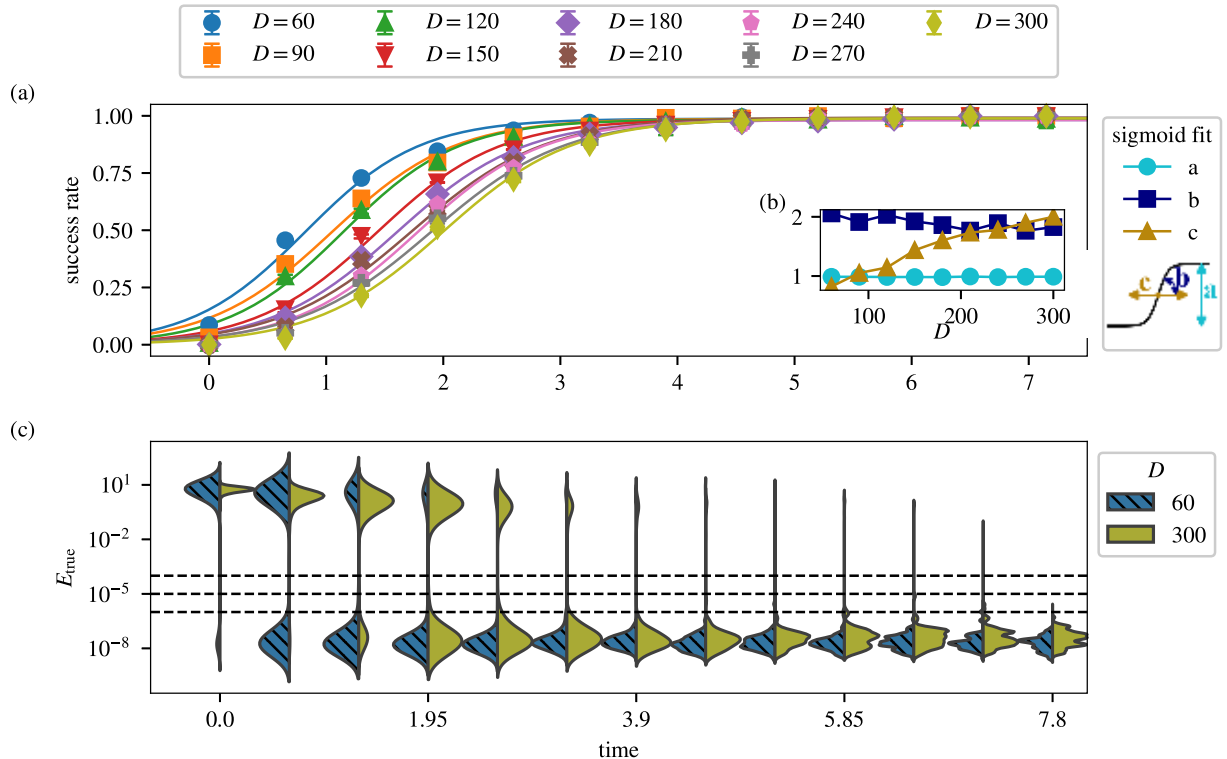


FIG. 12. Success rates for the consecutive shooting approach with varying system dimensionality. The success percentage is again of sigmoidal shape, with increasing inflection point for increasing dimension. The shift of the inflection point to the right is approximately linear with small slope. Even for the highest tested system dimension $D = 300$, we can achieve an almost perfect success rate after $t \approx 4$. The presented simulations were done for $t_s = 0.65$, which was the optimal segment length in Fig. 7. Every third oscillator was observed.

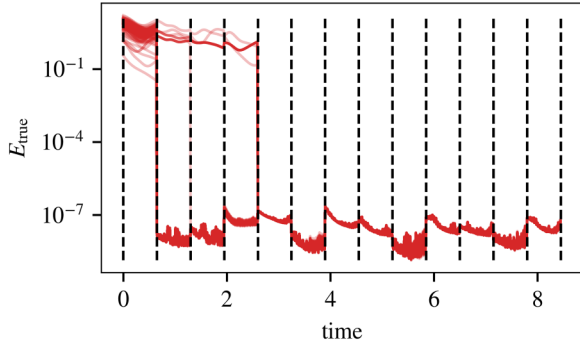


FIG. 13. True mean-squared error over time for consecutive optimizations of multiple initial guesses for the same time series. In this example, the system dimension is $D = 120$, the segment length $t_S = 0.65$, and every third variable is observed. Faint black vertical lines mark the borders between different segments. After four segments, or $t = 2.6$, all optimizations have converged closely to the true state, but for each following segment, the error first decreases over time and then increases toward the end of the segment.

1. Autocorrelation

A quantity that is also easily accessible for experimentally obtained data is the autocorrelation function of the time series. Figures 11(b) and 11(c) show the autocorrelation functions and the respective first zero crossings and first minima for different system dimensions D of the Lorenz-96 system.

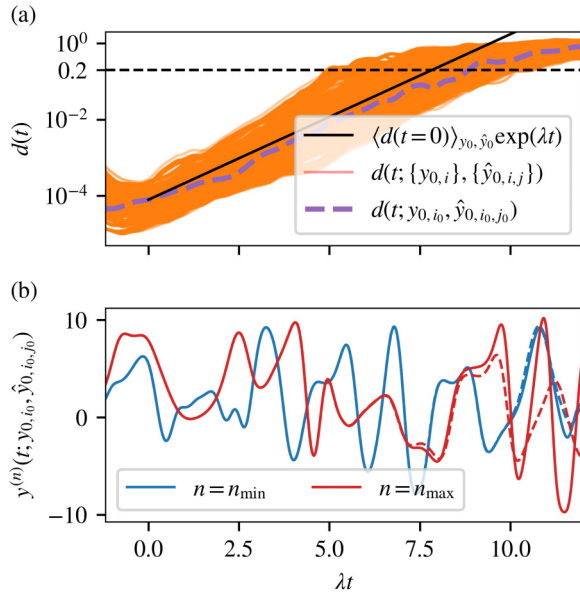


FIG. 14. (a) Time-dependent normalized error $d(t)$ for a free-running prediction starting at $t = 0$. The displayed data is for the 120-dimensional system with segment length $t_S = 0.65$ and every third variable observed. For $t < 0$, a consecutive segment optimization, as described in Sec. IV D, was performed over 13 segments ($\lambda t \approx 15.7$). (b) Illustrative time series of the variables with maximum ($n = n_{\max}$) and minimum ($n = n_{\min}$) average deviation for a representative single prediction. Solid lines mark the ground truth, while dashed lines represent the prediction. The corresponding time series of the normalized error is marked in purple in panel (a).

We can observe that apart from a few outliers for low-system dimensionalities, the first minimum and the zero crossing are almost independent of D .

The default segment length for the performed optimizations was chosen to be approximately equal to the first zero crossing of the autocorrelation function.

2. Self-mutual information

Another common choice to determine a characteristic timescale of a dynamical system from measured data is to calculate the first minimum of the self-mutual information of the time series. Figure 11(a) shows the self-mutual information of time series of the Lorenz-96 system for different dimensionalities D .

Because the self-mutual information does not have a clear minimum in most cases, we refrain from obtaining a characteristic timescale using this quantity.

APPENDIX B: CONSECUTIVE SHOOTING WITH INCREASING SYSTEM DIMENSION

Figure 12 shows the result for the consecutive shooting approach with the ideal segment length of $t_S = 0.65$ according to Fig. 7 and every third variable observed. The system dimension D is varied from 60 to 300.

We observe that even for a very high-system dimension of $D = 300$, the consecutive shooting approach converges after a few segments, and after 13 segments, all optimizations have converged to a successful full state estimation almost surely.

This illustrates the power of the consecutive shooting approach for very high-dimensional data. The time of convergence, using the parameter c of a sigmoid fit as illustrated in the figure as a proxy, increases linearly with D , but only very slowly. It must be checked whether that is the case for other chaotic systems as well, but it is a promising start.

APPENDIX C: DECREASING ERROR OVER TIME IN INDIVIDUAL OPTIMIZED SEGMENTS

Figure 13 shows the true error over time for consecutive optimizations of multiple initial guesses for the same time series of the 120-dimensional system. After all initial guesses have closely converged to the true time series, the error over time displays a distinct pattern for each segment: Even though the initial condition for the segment (the state at the beginning of the segment) deviates from the true state, the error initially decreases with time. This stands in contrast with the fact that the system is, in fact, chaotic, and almost every deviation from the true initial condition should grow exponentially with time. Moreover, after an initial decrease of the error, it increases again toward the end of the segment. The initial decrease happens for all segment lengths considered in Fig. 7, while the increase is more prominent in the longer segment lengths. We hypothesize that the initial decrease is due to the fact that the initial state of the segment lies in the subspace spanned by covariant Lyapunov vectors corresponding to negative Lyapunov exponents.

This would mean that the state initially evolves very closely to the stable manifold of the attractor, and only after

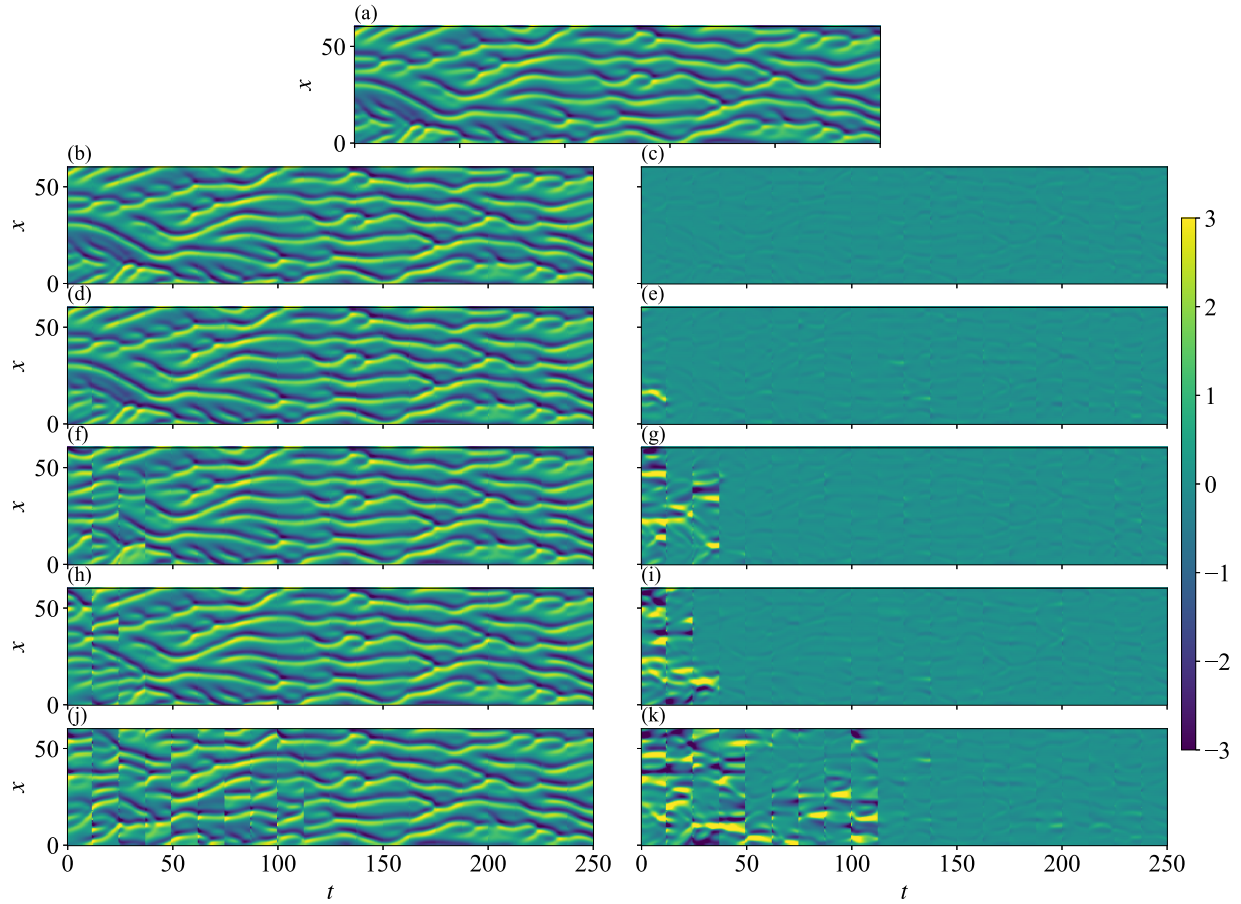


FIG. 15. (a) Spatiotemporal evolution of the KSE (E1), reconstructions (left column), and differences (right column) when observing only every 5th, 6th, 8th, 10th, or 12th grid point (from top to bottom).

some time is “thrown off.” We calculated the slopes of the decrease and increase of the error in the semilogarithmic plots, but could not find any obvious relation to the largest or smallest Lyapunov exponent. Slopes were very spread out for the optimizations for different time series and also across different segments of the same time series.

APPENDIX D: LONG-TERM PREDICTIVE POWER

Figure 14 shows that for a free-running prediction following a consecutive segment optimization, the error grows according to the largest Lyapunov exponent of the system. We

plot the normalized error

$$d(t) = \frac{\|y(t) - \hat{y}(t)\|_2}{\langle \|y(t)\|_2 \rangle_t}. \quad (\text{D1})$$

For this figure, we analyzed data of 50 time series with different, nonconsecutive initial conditions $y_{0,i}$, where for each time series, 100 initial guesses $\hat{y}_{0,i,j}$ were randomly selected from a uniform distribution in the two central quartiles of the amplitude distribution of the system. In Fig. 14(b), selected variables of the system corresponding to the dashed purple-marked error time series are plotted. To illustrate the span of accuracy for individual variables, we selected the variable $y_{n_{\min}}$ as the one with the minimum average deviation, and correspondingly, $y_{n_{\max}}$ as the one with the maximum average deviation. The solid lines represent the ground truth, while the dashed lines are the prediction. While $y_{n_{\max}}$ already shows visible deviation from the ground truth slightly before the normalized error crosses the validity threshold, $y_{n_{\min}}$ barely shows any visible deviation even after the crossing.

We can achieve valid times of at least $5\lambda t$, with an average valid time of approximately $7.5\lambda t$ after less than $16t_\lambda$ of “training data.”

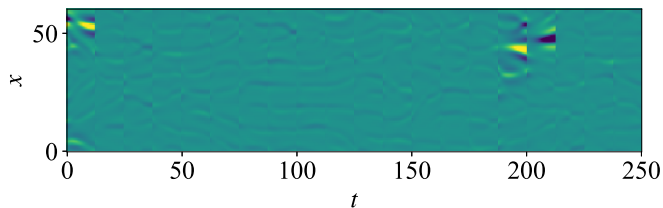


FIG. 16. Difference between the reconstructed and true states of the KSE (E1) when every sixth grid point is observed showing an intermittent increase of the reconstruction error.

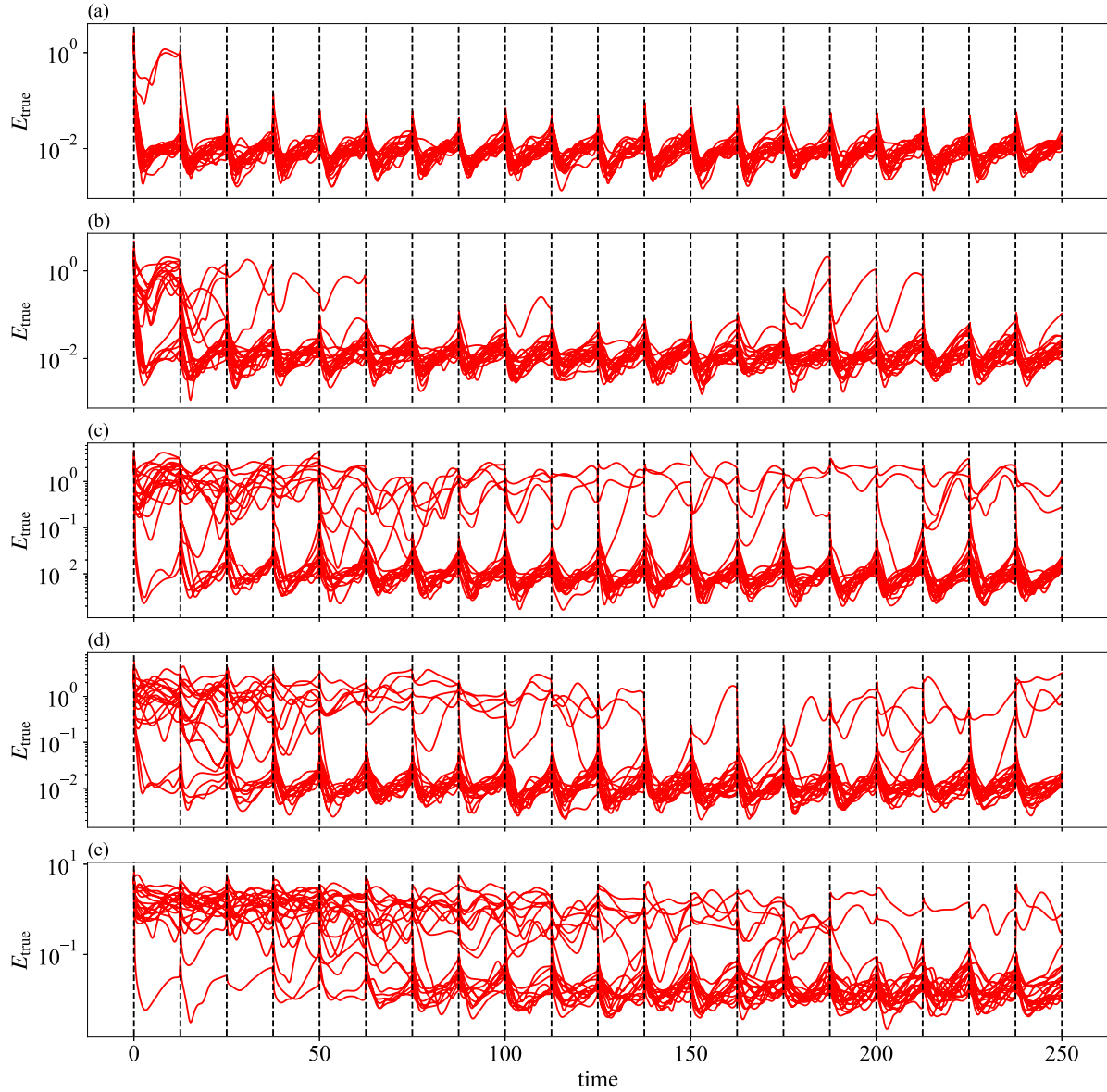


FIG. 17. Errors of consecutive shooting attempts for the KSE (E1) sampled at every 5th, 6th, 8th, 10th, or 12th grid point in the spatial domain (compare Figs. 6 and 13 for the Lorenz-96 model).

APPENDIX E: APPLICATION TO THE KURAMOTO-SIVASHINSKY EQUATION

State estimation in partially observed systems using adoptODE can also be applied to partial differential equations (PDEs). As an example, we considered the one-dimensional Kuramoto-Sivashinsky equation (KSE) [47,48]

$$\partial_t u(x, t) = -\frac{1}{2} \partial_x [u^2(x, t)] - \partial_x^2 u(x, t) - \partial_x^4 u(x, t), \quad (\text{E1})$$

where u is a spatiotemporal variable, which evolves on a one-dimensional domain of length $L = 60$. Solutions were computed numerically using finite differences based on 120 grid points. Training data were generated on a discretized spatial domain using a pseudospectral method and integration by means of an exponential time differencing fourth-order

Runge-Kutta (ETDRK4) method [49]. The estimation algorithm based on adoptODE utilizes the finite-difference method for solving the PDE.

Figure 15 shows the spatiotemporal evolution of the KSE and reconstructions when observing only every 5th, 6th, 8th, 10th, or 12th grid point. While reconstructions using observations from every fifth grid point (i.e., with spatial gaps of size $5 \times 60/120 = 2.5$) provide error-free reconstructions of the dynamics between these grid points, the reconstruction is gradually deteriorating when increasing the gaps between the observation points. In particular, short phases with increased deviations may occur from time to time, as shown in Fig. 16. This scenario is also reflected in the performance of consecutive shooting, which is shown in Fig. 17, where the proportion of successful attempts decreases as the size of the gaps increases.

- [1] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation* (CRC Press, Boca Raton, FL, 2004).
- [2] F. L. Chernousko, *State Estimation for Dynamic Systems* (CRC Press, Boca Raton, FL, 1993).
- [3] R. E. Kalman, A new approach to linear filtering and prediction problems, *J. Basic Eng.* **82**, 35 (1960).
- [4] H. D. I. Abarbanel, *The Statistical Physics of Data Assimilation and Machine Learning* (Cambridge University Press, Cambridge, 2022).
- [5] A. Isidori, *Nonlinear Control Systems: An Introduction* (Springer, Berlin, Heidelberg, 1985).
- [6] H. Nijmeijer and A. Van der Schaft, *Nonlinear Dynamical Control Systems* (Springer, New York, NY, 1990), Vol. 175.
- [7] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. Patil, and J. A. Yorke, A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A: Dyn. Meteorol. Oceanogr.* **56**, 415 (2004).
- [8] G. Datsis and U. Parlitz, *Nonlinear Dynamics: A Concise Introduction Interlaced with Code* (Springer Nature, Cham, Switzerland, 2022).
- [9] G. Evensen, *Data Assimilation* (Springer, Berlin, 2009).
- [10] L. Junge and U. Parlitz, Synchronization and control of coupled Ginzburg-Landau equations using local coupling, *Phys. Rev. E* **61**, 3736 (2000).
- [11] S. Berg, S. Luther, and U. Parlitz, Synchronization based system identification of an extended excitable system, *Chaos* **21**, 033104 (2011).
- [12] E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, Fitting ordinary differential equations to chaotic data, *Phys. Rev. A* **45**, 5524 (1992).
- [13] H. D. Abarbanel, D. R. Creveling, R. Farsian, and M. Kostuk, Dynamical state and parameter estimation, *SIAM J. Appl. Dyn. Syst.* **8**, 1341 (2009).
- [14] J. Ye, N. Kadakia, P. J. Rozdeba, H. D. I. Abarbanel, and J. C. Quinn, Improved variational methods in statistical data assimilation, *Nonlinear Processes Geophys.* **22**, 205 (2015).
- [15] J. R. Stroud and T. Bengtsson, Sequential state and variance estimation within the ensemble Kalman filter, *Mon. Weather Rev.* **135**, 3194 (2007).
- [16] D. Rey, M. Eldridge, M. Kostuk, H. D. Abarbanel, J. Schumann-Bischoff, and U. Parlitz, Accurate state and parameter estimation in nonlinear systems with sparse observations, *Phys. Lett. A* **378**, 869 (2014).
- [17] D. Rey, M. Eldridge, U. Morone, H. D. I. Abarbanel, U. Parlitz, and J. Schumann-Bischoff, Using waveform information in nonlinear data assimilation, *Phys. Rev. E* **90**, 062916 (2014).
- [18] J. Schumann-Bischoff, U. Parlitz, H. D. I. Abarbanel, M. Kostuk, D. Rey, M. Eldridge, and S. Luther, Basin structure of optimization based state and parameter estimation, *Chaos* **25**, 053108 (2015).
- [19] K. J. H. Law, D. Sanz-Alonso, A. Shukla, and A. M. Stuart, Filter accuracy for the Lorenz 96 model: Fixed versus adaptive observation operators, *Physica D* **325**, 1 (2016).
- [20] S. G. Penny, T. A. Smith, T. Chen, J. A. Platt, H. Lin, M. Goodliff, and H. D. I. Abarbanel, Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation, *J. Adv. Model. Earth Syst.* **14**, e2021MS002843 (2022).
- [21] J. Brajard, A. Carrassi, M. Bocquet, and L. 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, *J. Comput. Sci.* **44**, 101171 (2020).
- [22] J. Guillot, P. Ailliot, E. Fr  nod, J. Ruiz, and P. Tandeo, State and stochastic parameters estimation with combined ensemble Kalman and particle filters, *Mon. Weather Rev.* **153**, 1141 (2025).
- [23] R. Fablet, B. Chapron, L. Drumetz, E. M  min, O. Pannekoucke, and F. Rousseau, Learning variational data assimilation models and solvers, *J. Adv. Model. Earth Syst.* **13**, e2021MS002572 (2021).
- [24] A. Amemiya, M. Shlok, and T. Miyoshi, Application of recurrent neural networks to model bias correction: Idealized experiments with the Lorenz-96 Model, *J. Adv. Model. Earth Syst.* **15**, e2022MS003164 (2023).
- [25] T. J. Cocucci, M. Pulido, M. Lucini, and P. Tandeo, Model error covariance estimation in particle and ensemble Kalman filters using an online expectation-maximization algorithm, *Q. J. R. Meteorol. Soc.* **147**, 526 (2021).
- [26] L. Lettermann, A. Jurado, T. Betz, F. W  rg  tter, and S. Herzog, Tutorial: A beginner's guide to building a representative model of dynamical systems using the adjoint method, *Commun. Phys.* **7**, 128 (2024).
- [27] E. Lorenz, Predictability: A problem partly solved, in *Seminar on Predictability, 4-8 September 1995* (ECMWF, Shinfield Park, Reading, 1995), Vol. 1, pp. 1-18.
- [28] I. Kottlarz and U. Parlitz, Ordinal pattern-based complexity analysis of high-dimensional chaotic time series, *Chaos* **33**, 053105 (2023).
- [29] The Lyapunov spectra were calculated using the Dynamical-Systems.jl library [8,46], which implements the method "H₂" in Ref. [30], originally stated in Ref. [31].
- [30] K. Geist, U. Parlitz, and W. Lauterborn, Comparison of different methods for computing Lyapunov exponents, *Prog. Theor. Phys.* **83**, 875 (1990).
- [31] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; A method for computing all of them. Part I: Theory, *Meccanica* **15**, 9 (1980).
- [32] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python+NumPy programs (2018), <http://github.com/google/jax>.
- [33] J. S. Arora and E. J. Haug, Methods of design sensitivity analysis in structural optimization, *AIAA J.* **17**, 970 (1979).
- [34] A. D. Belegundu, Lagrangian approach to design sensitivity analysis, *J. Eng. Mech.* **111**, 680 (1985).
- [35] G. Bianconi, A. Arenas, J. Biamonte, L. D. Carr, B. Kahng, J. Kertesz, J. Kurths, L. L  , C. Masoller, A. E. Motter, M. Perc, F. Radicchi, R. Ramaswamy, F. A. Rodrigues, M. Sales-Pardo, M. S. Miguel, S. Thurner, and T. Yasserli, Complex systems in the spotlight: Next steps after the 2021 Nobel prize in physics, *J. Phys.: Complex.* **4**, 010201 (2023).
- [36] R. M. Errico, What is an adjoint model? *Bull. Am. Meteorol. Soc.* **78**, 2577 (1997).
- [37] P. Ji, J. Ye, Y. Mu, W. Lin, Y. Tian, C. Hens, M. Perc, Y. Tang, J. Sun, and J. Kurths, Signal propagation in complex networks, *Phys. Rep.* **1017**, 1 (2023).

- [38] S. A. Jurovics and J. E. McIntyre, The adjoint method and its application to trajectory optimization, *ARS J.* **32**, 1354 (1962).
- [39] M. Lotkin and H. N. Browne, On the accuracy of the adjoint method of differential corrections, *Am. Math. Monthly* **63**, 97 (1956),.
- [40] R.-E. Plessix, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.* **167**, 495 (2006).
- [41] R. Tomović, *Sensitivity analysis of dynamic systems, Electronic Sciences Series* (McGraw-Hill, New York, 1963).
- [42] R. Tomović, The role of sensitivity analysis in engineering problems, in *Sensitivity Methods in Control Theory*, edited by L. Radanovic (Pergamon, Oxford, 1966), pp. 103–109.
- [43] J. Schumann-Bischoff, S. Luther, and U. Parlitz, Nonlinear system identification employing automatic differentiation, *Commun. Nonlinear Sci. Numer. Simul.* **18**, 2733 (2013).
- [44] <https://gitlab.gwdg.de/sherzog3/adoptODE>.
- [45] <https://github.com/KotaShiozawa/adoptODE-lorenz96>.
- [46] G. Datseris, DynamicalSystems.jl: A Julia software library for chaos and nonlinear dynamics, *J. Open Source Software* **3**, 598 (2018).
- [47] Y. Kuramoto, Diffusion-induced chaos in reaction systems, *Prog. Theor. Phys. Suppl.* **64**, 346 (1978).
- [48] G. I. Sivashinsky, On flame propagation under conditions of stoichiometry, *SIAM J. Appl. Math.* **39**, 67 (1980).
- [49] A.-K. Kassam and L. N. Trefethen, Fourth-order time-stepping for stiff PDEs, *SIAM J. Sci. Comput.* **26**, 1214 (2005).