

# Calibrated reservoir computers

Cite as: Chaos **30**, 113134 (2020); <https://doi.org/10.1063/5.0030651>

Submitted: 23 September 2020 . Accepted: 28 October 2020 . Published Online: 18 November 2020

 Y. A. Mabrouk, and C. R ath



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Using mathematics to diagnose, cure, and predict cardiac arrhythmia](#)

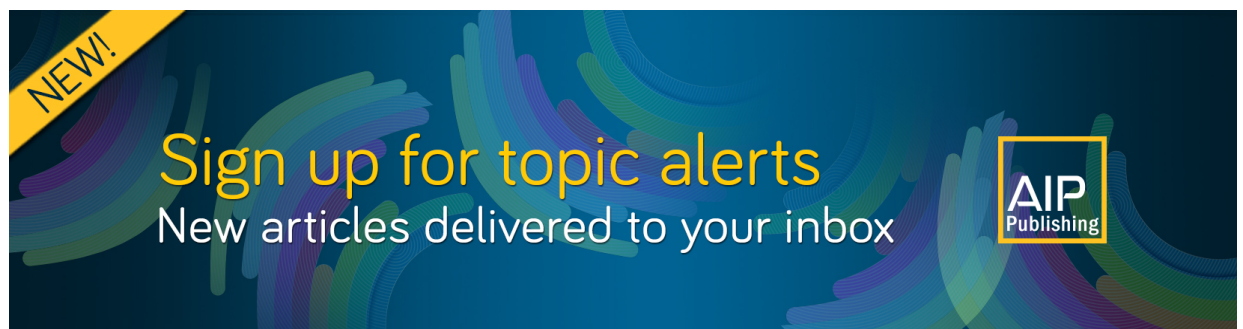
Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 113132 (2020); <https://doi.org/10.1063/5.0021844>

[Network temporality can promote and suppress information spreading](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 113136 (2020); <https://doi.org/10.1063/5.0027758>

[Accuracy of neural networks for the simulation of chaotic dynamics: Precision of training data vs precision of the algorithm](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 113118 (2020); <https://doi.org/10.1063/5.0021264>



**NEW!**

Sign up for topic alerts  
New articles delivered to your inbox

AIP  
Publishing



# Calibrated reservoir computers

Cite as: Chaos 30, 113134 (2020); doi: 10.1063/5.0030651

Submitted: 23 September 2020 · Accepted: 28 October 2020 ·

Published Online: 18 November 2020



View Online



Export Citation



CrossMark

Y. A. Mabrouk<sup>1,2,a)</sup>  and C. R ath<sup>1,2,a)</sup>

## AFFILIATIONS

<sup>1</sup>Fakult t f r Physik, Ludwig-Maximilians-Universit t M nchen, Schellingstra e 4, 80779 M nchen, Germany

<sup>2</sup>Institut f r Materialphysik im Weltraum, Deutsches Zentrum f r Luft-und Raumfahrt, M nchner Stra e 20, 82234 We bling, Germany

<sup>a)</sup>Authors to whom correspondence should be addressed: [youssef.mabrouk@physik.uni-muenchen.de](mailto:youssef.mabrouk@physik.uni-muenchen.de) and [christoph.raeth@dlr.de](mailto:christoph.raeth@dlr.de)

## ABSTRACT

We observe the presence of infinitely fine-scaled alternations within the performance landscape of reservoir computers aimed for chaotic data forecasting. We investigate the emergence of the observed structures by means of variations of the transversal stability of the synchronization manifold relating the observational and internal dynamical states. Finally, we deduce a simple calibration method in order to attenuate the thus evidenced performance uncertainty.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0030651>

Modern scientific measurement devices can resolve a remarkably wide range of temporal scales and thus cover a wide diversity of multi-scale physics. This motivates the study of methods allowing to extract physical laws and reconstruct hidden features from temporal measurement data. With the emergence of statistical learning methods, it is of interest to investigate how these can be used in order to alleviate the often tedious efforts involved with established attractor reconstruction techniques such as the delay embedding method. Among the learning methods widely employed for the analysis of temporal partially observed data, a particularly appealing class is that of reservoir computers. This class is distinguished by the introduction of an internal state variable that exhibits a dynamical behavior driven by the observational data. An advantage of such a method is that the internal state variable can be updated incrementally in step with the incremental measurements of the dynamical system under observation with the absence of any need of preprocessing the data before it is fed into the internal states. Relying on statistical learning methods has, however, the severe disadvantage of leading to hardly interpretable results. Here, we investigate the dependence of the performance of reservoir computers on its empirical fitting parameters. We show the sensitive dependence of performance on the range of scaling of the internal reservoir weights. We characterize the observed sensitivity patterns by investigating the changes in the stability properties of the synchronized state that relates the dynamical internal states to the observational data.

The real-time forecast of nonlinear dynamical systems is thought to be of great potential for a number of interdisciplinary fields.<sup>1</sup> For example, combustion driven high-amplitude oscillations within low emission gas turbines are known to cause serious engine damage that calls for the need of an efficient control scheme.<sup>2</sup> Also, the highly fluctuating electrical power feed of wind turbines is a major difficulty undermining the integration of renewable energy sources to a stable electrical power grid.<sup>3</sup> In such cases, forecasting the system states through modeling the underlying mechanisms has been mostly hindered either by the difficulty of specifying suitable model equations or by a full measurement of the system states in cases where the model equations are known.<sup>4</sup> The lack of analytical solutions and often excessive computational cost of mechanistic models can also often prohibit the employment of knowledge based approaches.

Finding a dynamical rule from a sequence of observations without physical knowledge is a common problem for which a number of machine-learning methods have been developed. The basic distinction of reservoir computers is the introduction of dynamically changing internal states driven by the observational data. In particular, reservoir computers are distinguished by the fact that the internal state dynamics are not adjusted with respect to a specific computation, and that learning solely consists by adjusting a linear filter to extract the needed information from the internal states.<sup>5</sup> This fact results in their computational easiness and convenience for large-scale real-time computations.<sup>6,7</sup> The conceptual

simplicity of this method also makes it well-suited for developing an understanding of its underlying learning mechanisms<sup>8,9</sup> and thus providing a solid basis for applications in the scientific domain. Yet, there is a notable lack of basic understanding in the computational performance and properties of reservoir computers.

The indicated lack of basic understanding is primarily related to the excessive number of empirical parameters that reservoir computers admit.<sup>10</sup> Even in a setting where these parameters are optimized, the network construction necessitates a relevant number of random variables that imply an unwanted variability in the learning performance. For example, it has been reported that a slight change in the randomly generated reservoir connectivity results in large changes in the short-term as well as in the long-term properties of the predicted time series.<sup>11</sup> A similar variability has also been reported for the case of finite-time fluctuations, i.e., when the learned sequence of data is slightly shifted either backward or forward in time.<sup>12</sup> This kind of unreproducible behavior is hardly admissible for applications and thus motivates investigating the question of whether one can establish a reproducible generic result that quantifies the degree of system uncertainty and whether one can find general principles that allow to avoid operation in suboptimal parameter regimes.

In this work, we propose a two-parameter performance study evaluated with respect to forecasting several different synthetic chaotic systems. We demonstrate the presence of fine-scaled patterns of optimal and suboptimal parameters and investigate their relation to the variations of transversal stability of the synchronization manifold that is embedded in the space spanned by the observational and internal dynamical states. We quantify the degree of system uncertainty by calculating the fraction of uncertain parameter points as a function of the parameter precision. The fact that the distinct datasets result in similar uncertainty exponents is an indication that reservoir computers are generically highly sensitive to parameter changes. In what follows, we briefly document the main lines of our implementation. We then present the results of the performance study and discuss its implications on the calibration of reservoir computers.

We consider the situation where we are given a series of observational data  $\{\mathbf{x}_t\}$  for  $t \in [0, T]$  of a variable  $\mathbf{x}_t \in \mathbb{R}^N$  that we assume to obey a deterministic dynamical equation of the form

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t), \quad (1)$$

for some unknown  $f: \mathbb{R}^N \rightarrow \mathbb{R}^N$ . From the recorded past values for  $t \in [0, T]$ , we wish to infer future  $t \geq T$  to high accuracy until a certain lapse of time that characterizes the dynamical system in question. After this time, it is only desired that the forecast remains qualitatively similar to the correct behavior of the system. We will, namely, address cases where  $\mathbf{x}_t$  evolves on a chaotic attractor, implying an exponential temporal growth of errors in the initial conditions, making the inference task challenging. While the direct way of inference would be to approximate  $f$  by a linear combination of a suitably chosen functional basis, we proceed by introducing the dynamical variable  $\mathbf{y}_t \in \mathbb{R}^D$  that is coupled to the observational data through  $g: \mathbb{R}^D \times \mathbb{R}^N \rightarrow \mathbb{R}^D$  according to the driven dynamical system given by

$$\mathbf{y}_{t+1} = g(\mathbf{x}_t, \mathbf{y}_t). \quad (2)$$

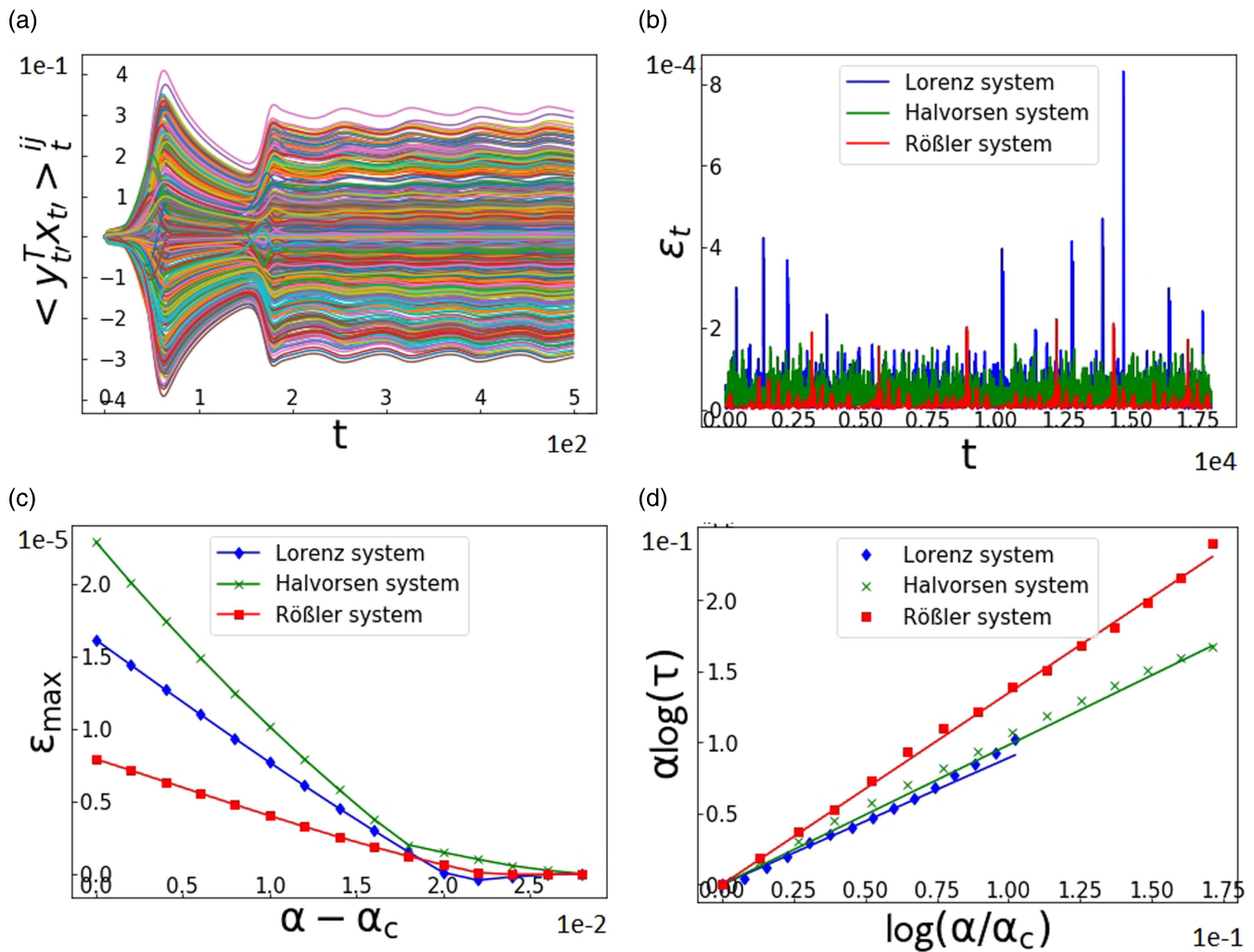
In particular, for a suitably chosen coupling  $g$ , the dynamical system defined by the set of equations (1) and (2) may achieve an asymptotic state of synchronization,<sup>13-15</sup> where the driving and the driven variables  $\mathbf{x}_t$  and  $\mathbf{y}_t$  are related by a static functional relation  $\varphi: \mathbb{R}^D \rightarrow \mathbb{R}^N$  such that  $\|\varphi(\mathbf{y}_t) - \mathbf{x}_t\| \rightarrow 0$  for  $t \rightarrow \infty$ . Once synchronization is established, the inference is reduced to finding a suitable approximation for  $\varphi$ , allowing to evolve the internal state dynamics  $\mathbf{y}_t$  according to the autonomous equation

$$\mathbf{y}_{t+1} = g(\varphi(\mathbf{y}_t), \mathbf{y}_t) \quad (3)$$

and to infer the variable of interest from it systematically through  $\mathbf{x}_t = \varphi(\mathbf{y}_t)$ . Whether synchronization is established can be deduced by monitoring the correlation matrix  $\langle \mathbf{y}_t^\top \mathbf{x}_t \rangle_t$  as time  $t$  increases in  $[0, T]$ . A correlation that settles on a constant value and does not change as time changes is a hint that the relation between  $\mathbf{x}_t$  and  $\mathbf{y}_t$  is independent of time [see Fig. 1(a)]. Once the correlation has settled, the variable change  $\varphi$  can be extracted from it through the leading order estimation  $\varphi(\mathbf{y}) = \langle \mathbf{y}_t^\top \mathbf{y}_t - r \mathbb{I} \rangle_t^{-1} \langle \mathbf{y}_t^\top \mathbf{x}_t \rangle_T \mathbf{y}$ , which results from the Tikhonov regularized ridge regression of  $\varepsilon = \langle \|\varphi(\mathbf{y}_t) - \mathbf{x}_t\| \rangle_T$  with respect to  $\varphi$ . In principal, a uniformly contracting coupling  $\|g(\mathbf{x}, \mathbf{y}_1) - g(\mathbf{x}, \mathbf{y}_2)\| \leq \|\mathbf{y}_1 - \mathbf{y}_2\|$  ensures an initial difference between two internal states driven by the same series  $\{\mathbf{x}_t\}$  to be made arbitrarily small for a sufficiently long series.<sup>14</sup> This ensures the internal states  $\mathbf{y}_t$  to be uniquely related to the observational series  $\{\mathbf{x}_t\}$ . For an internal state dimension at least twice as large as the dimension of observational data, this relation is likely to be invertible,<sup>14</sup> thus implying the establishment of the indicated state of synchronization. Uniform contraction can be, for example, imposed for a sigmoidal coupling

$$g(\mathbf{x}, \mathbf{y}) = \tanh(W_{in} \cdot \mathbf{x} + W \cdot \mathbf{y}), \quad (4)$$

by choosing the largest singular value of  $W$  smaller than 1, where  $W_{in} \in \mathbb{R}^{D \times N}$  and  $W \in \mathbb{R}^{D \times D}$  are two random matrices whose entries are uniformly sampled from the intervals  $[-\alpha, \alpha]$  and  $[-\beta, \beta]$ , respectively. In practice, this condition was shown to be restrictive in the sense that it excludes a good fraction of  $W$  possibilities that imply well defined internal states.<sup>16</sup> A formally less plausible yet practically more favored condition consists in tuning the entries range  $\beta$  of  $W$  close to unity.<sup>17</sup> It has been recently demonstrated that by changing the entries range  $\alpha$  of  $W_{in}$ , the stability properties of the invariant set  $M = \{(\mathbf{x}, \mathbf{y}) : \varphi(\mathbf{y}) = \mathbf{x}\}$  may change.<sup>14</sup> Stability here relates to transversal perturbations  $\varepsilon_t = \|\varphi(\mathbf{y}_t) - \mathbf{x}_t\|$  away from the invariant set  $M$  during the autonomous dynamical regime, rather than the stability of internal states  $\mathbf{y}_t$  when driven by a certain  $\{\mathbf{x}_t\}$  as discussed above. The two regimes are fundamentally distinguished by the approximation errors in the drive signal made by the leading order approximation of  $\varphi$ . For a stable  $M$ , these errors will drive the forecast away from the true signal after a short term while the forecast will maintain statistically consistent properties in the long term. In turn, an unstable  $M$  implies also a correct short-term forecast but the long-term behavior would be inconsistent in this case. Thus, the change of stability properties of the invariant set  $M$  clarifies the variability of the long-term forecasting quality. It stays *a priori* unclear why the short-term forecasting quality shows strong sensitive dependence on the parameters.<sup>11</sup> In particular, this sensitivity was observed for parameter regions where the long-term forecasts remain qualitatively unchanged, i.e., where



**FIG. 1.** Dynamics in the driven regime. Each dataset  $\{x_t\}$  is normalized to zero mean and variance one. For all datasets, the internal state dimension is set to  $D = 500$ , the internal connectivity matrix  $W$  is obtained by the random Erdős–Rényi algorithm with edge creation probability  $p = 2 \times 10^{-1}$ , the regularization parameter is  $r = 10^{-6}$ . (a) Initial transient time evolution of the correlation matrix elements exemplified by the Lorenz system. (b) Long-term time evolution of the transversal error, characterized by short-lived high-amplitude bursts. The time interval  $T$  corresponds to 100 Lyapunov units, where one Lyapunov unit corresponds to the inverse maximal Lyapunov exponent  $1/\lambda_{\max}$ . (c) Maximal bursts amplitude  $\varepsilon_{\max}$  as a function of  $\alpha$ . (d) Interburst duration  $\tau$  as a function of  $\alpha$ . The sampling resolution of  $\alpha$  is  $2 \times 10^{-3}$ .

the set  $M$  is stable. This fact indicates the presence of a parameter window separating stability from instability.

Chaotic dynamical systems for which the attractor  $A$  is constrained to a subset  $M$  of the full degrees of freedom are known to exhibit bifurcations where the attractor first loses transversal stability only in the neighborhood of its unstable periodic orbits of low period<sup>18</sup> and then becomes fully transversally unstable.<sup>19,20</sup> Within such a regime, a typical chaotic trajectory subject to transversal perturbations away from  $M$  would explore the attractor  $A$  for long times until it falls into the neighborhood of a transversally unstable orbit where the transversal errors get magnified. For a globally attracting  $M$ , this mechanism results in the appearance of intermittent bursts<sup>21</sup>

where chaotic trajectories of  $A$  are on the one hand repelled from  $M$  by the transversally unstable orbits and on the other hand attracted back to  $M$  due to the global dynamics of the system. We propose this as a basis to understand the intermittent bursting behavior appearing in the transversal error  $\varepsilon_t = \|\varphi(y_t) - x_t\|$  during the driven dynamical regime. Evidence for our proposal is the scaling law<sup>19</sup> describing the mean interburst time  $\tau$  as well as the maximal bursts amplitude  $\varepsilon_{\max}$  when the parameter  $\alpha$  is varied.

The bursting behavior is illustrated in Fig. 1(b) for the three chaotic flows  $f$  used in our study.<sup>22</sup> The plots show the time evolution of the transversal error  $\varepsilon_t$  during the driven regime. As the parameter  $\alpha$  is varied, the bursts maximal amplitude decreases as a linear

function of  $\alpha$ , until it takes a low constant value corresponding to the standard deviation of the error [Fig. 1(c)], where high-quality synchronization is achieved. This qualitative change in the bursting behavior can be monitored by determining the mean duration between two consecutive bursts<sup>23</sup>  $\tau$ , which is displayed in Fig. 1(d). The scaling shown is obtained by fixing the random samples of  $W$  and  $W_{in}$  and evaluating  $\varepsilon_{\max}$  and  $\tau$  at equidistant values of  $\alpha$  starting from the critical point  $\alpha_c$ , which we define as the point where pronounced scaling begins starting from  $\alpha \ll 1$  and thus identify as the cutoff value where linear scaling is observed in Fig. 1(c). The critical value  $\alpha_c$  for each dataset can be found in Table I. We note that different scaling exponents for the interburst time are expected from the fact that the three datasets have different Lyapunov spectra, thus corresponding to different repulsion rates in the direction parallel to  $M$ .<sup>19</sup>

The evidenced locally repellent sets within the basin of attraction of the invariant set  $M$  have direct consequences regarding tuning the reservoir. By tuning is meant the problem of specifying a pair of values  $(\alpha, \beta)$  for given random matrices  $W_{in}$  and  $W$  such that the short-term forecasting performance is optimal. Optimality is achieved when the duration of accurate short-term forecasts is evidently maximized, and importantly when this duration is not

TABLE I. Summary of the computed scaling laws and the corresponding scaling exponents.

Scaling law	System		
	Lorenz	Halvorsen	Rößler
$\mu(\Delta) \sim \Delta^\nu$	0.038	0.039	0.038
$\alpha \log \tau \sim \log \alpha$	0.89	0.98	1.34
$\varepsilon_{\max} \sim \alpha$	$8.1 \times 10^{-4}$	$1.1 \times 10^{-3}$	$3.7 \times 10^{-4}$
$\alpha_c$	0.26	0.15	0.15
$\lambda_{\max}$	0.9056	0.7899	0.0714

significantly affected by the finite time segment  $\{x_t\}$  learned by the reservoir. Indeed, since the bursts of  $\varepsilon_t$  during the driven regime do not alternate the regression error  $\varepsilon$  due to their short-lived and infrequent nature, the inspection of  $\varepsilon$  as a function of  $(\alpha, \beta)$  leads to a structureless landscape. In contrast, starting the autonomous regime at a time  $t'$  in the order of ten Lyapunov units earlier than  $T$  and inspecting the average  $\varepsilon'$  of  $\varepsilon_t$  within this autonomous regime in  $[T - t', T]$  yields a highly structured pattern. Figure 2(a)

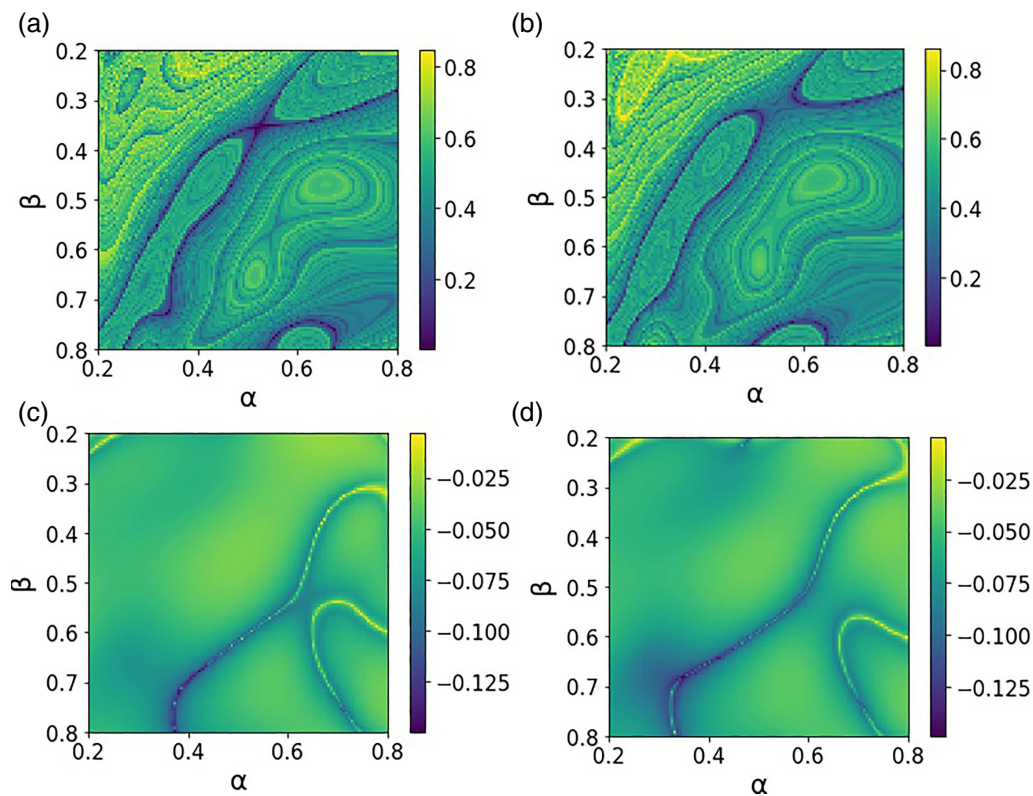


FIG. 2. Performance landscape  $\varepsilon'$  and transversal Lyapunov exponent  $\lambda_{\max}^t$  within  $t' = 10$  Lyapunov units of autonomous evolution after a duration  $T - t'$  corresponding to 100 Lyapunov units of driven evolution. A parameter grid of  $10^2 \times 10^2$  was used to generate each of the shown results. (a)  $\varepsilon'$  (b)  $\varepsilon'$  with the initial condition shifted with respect to (a). (c)  $\lambda_{\max}^t$  (d)  $\lambda_{\max}^t$  with initial condition shifted with respect to (c). The dynamic change can be visualized in the supplementary material via videos showing the patterns for several consecutive values of  $x_0$ . Multimedia view: <https://doi.org/10.1063/5.0030651.1>

(Multimedia view) shows an example of the Lorenz chaotic system. The shown patterns are highly dynamic in the sense that a slight shift of the initial point  $\mathbf{x}_0$  of the learned time series results in clear change of the pattern, as illustrated by Fig. 2(b).

Thus, the difficulty we encounter when trying to extract the optimal pair  $(\alpha, \beta)$  from the resulting  $\varepsilon'$  landscape is that this landscape relates only to the narrow time interval where it was obtained from, i.e., the resultant optimal parameter sets indicate transversal stability only in a narrow portion of  $M$ .

The variations of the transversal stability of the synchronization manifold as well as its existence can be evidenced by calculating the transversal Lyapunov exponents. Since the defining equation of the autonomous regime for the internal states is known, we proceed by calculating the Lyapunov spectrum of the internal states during the autonomous dynamical regime according to the variational equation,

$$\delta_{t+1} = J_g(\varphi(\mathbf{y}_t), \mathbf{y}_t)\delta_t, \quad (5)$$

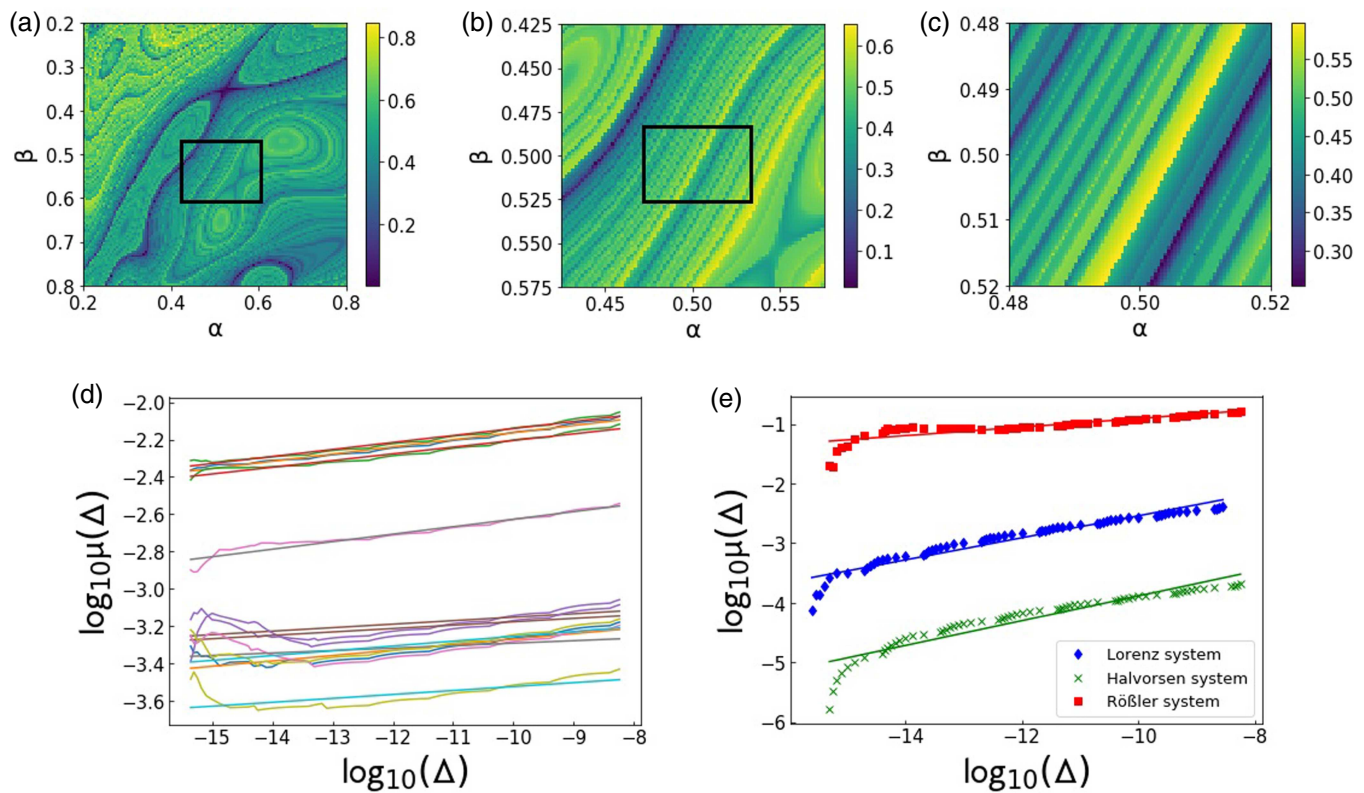
where  $\delta_t \in \mathbb{R}^D$  refers to a small perturbation of the internal state and  $J_g: \mathbb{R}^D \rightarrow \mathbb{R}^D$  indicates the Jacobian matrix of the autonomous internal dynamics. The transversal Lyapunov exponent  $\lambda_{\max}^t$  is identified as the maximal Lyapunov exponent from the above calculated spectrum after elimination of the Lyapunov exponents of the driving signal (we refer to Ref. 14 for a detailed justification). This identification may be ambiguous since it is *a priori* unclear which of the calculated exponents relate to transversal directions, and from the inspection of the full spectrum of  $D$  eigenvalues, only statements about the existence or absence of positive transversal exponents can be made. An ambiguity may thus arise if the order of the spectrum of exponents would change when changing  $\alpha$  and  $\beta$ . The variations of  $\lambda_{\max}^t$  as a function of  $\alpha$  and  $\beta$  are shown in Figs. 2(c) and 2(d). While Figs. 2(a) and 2(b) and Figs. 2(c) and 2(d) show some kind of similarity, a substantial difference is the presence of fine-scaled alternations in the former case. The apparent similarity between the two patterns indicates that the order of exponents is more or less preserved. The absence of small-scale features in the transversal stability patterns may be attributed to the change of order of exponents as  $\alpha$  and  $\beta$  are varied. Notably, the fine-scaled alternations were observed in the performance landscape related to several driving signals. While the mechanism underlying these fine-scaled alternations is intricate, their presence immediately indicates the difficulty of tuning  $(\alpha, \beta)$  with respect to  $\varepsilon'$ .

We insist that this difficulty becomes critical if these fine-scaled alternations have a fractal nature.<sup>24</sup> We mention that fractal parameter sets are already found in the simplest possible nonlinear equation, namely, the quadratic map, where arbitrarily close to a given parameter value for which chaotic solutions are typical, one can find a parameter value that yields periodic solutions. In a similar way, we classify our parameter pairs  $(\alpha, \beta)$  into two classes of optimal or suboptimal choices depending on whether  $\varepsilon'$  is below or above a certain quality threshold  $\delta$ . We suggest that arbitrarily close to any given pair  $(\alpha, \beta)$  belonging to one class, one can find a pair that belongs to the other. Evidence of such a pathological topology can be obtained by calculating the uncertainty exponent  $\gamma$  of the defined parameter sets.<sup>24</sup> The uncertainty exponent characterizes the scaling behavior of the fraction  $\mu(\Delta)$  of uncertain

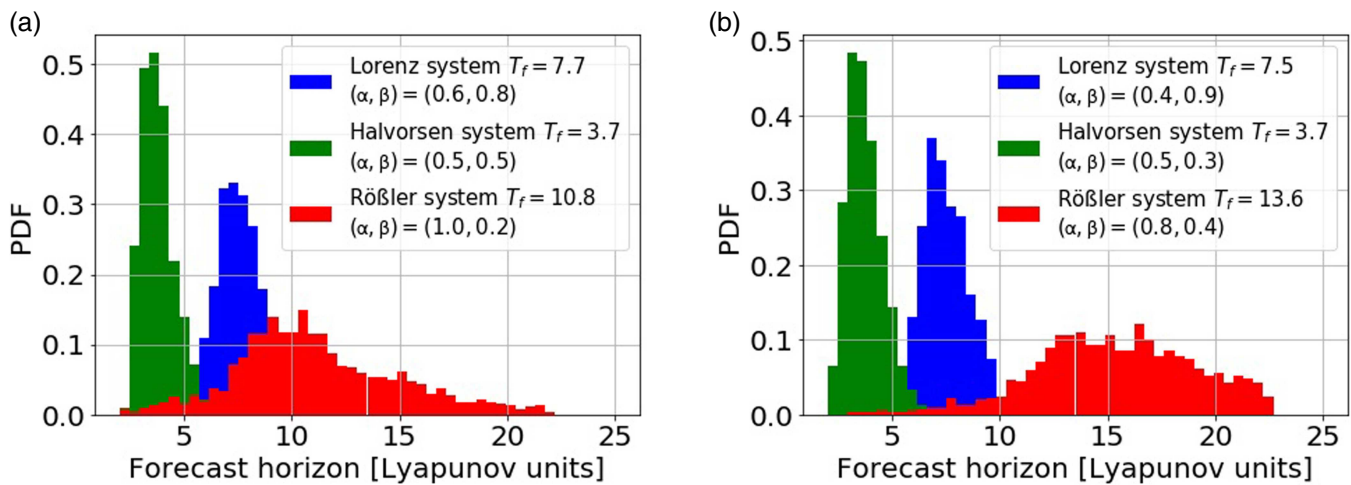
parameter points as the parameter precision  $\Delta$  is varied. A parameter point  $\alpha$  is called uncertain with respect to given precision  $\Delta$  if the points obtained by shifting that point  $\alpha \pm \Delta$  by the given precision belong to a different class. Figure 3 illustrates the alternations with magnified pictures, as well as the scaling behavior of the uncertainty function [Fig. 3(d)]. The indicated scaling shows examples obtained from performing the described procedure for different initial data points  $\mathbf{x}_0$  [Fig. 3(e)], showing that the scaling law  $\mu(\Delta) \sim \Delta^\gamma$  is indeed robust. Notably, for internal states driven by several distinct chaotic systems, the resultant value of the uncertainty exponent  $\gamma = 0.038 \pm 0.002$  occurs to be the same within the indicated error tolerance. The resulting exponents are summarized in Table I. An implication of a scaling exponent  $\gamma < 1$  is that increasing the precision does not significantly decrease the uncertainty.

Thus, we suggest the well known extreme sensitivity of reservoir computers to originate from a seemingly very generic structure of the parameter space. One consequence of such highly sensitive parameters is the inefficiency of gradient methods for finding the optimal pair  $(\alpha, \beta)$ , which minimizes the short-term error  $\varepsilon'$ . One way to reduce this sensitivity is to tune the reservoir by minimizing the error  $\varepsilon'$  averaged with respect several initial points  $\mathbf{x}_0$ . Another way is to maximize the interburst mean period  $\tau$ , which is by definition a long-term average and thus stable against finite-time fluctuations. To show the equivalence between the two methods, we determine the optimal parameter pair  $(\alpha, \beta)$  resulting from the error  $\langle \varepsilon' \rangle_{\mathbf{x}_0}$  averaged over a number of 30 initial data points  $\mathbf{x}_0$ , as well as the parameter pair for which the interburst time  $\tau$  is maximal. We then compare the two optima by evaluating for each optimum the mean and standard deviation of the resulting forecast horizon<sup>25</sup> with respect to a number of  $10^3$  consecutive initial data points  $\mathbf{x}_0$ . The comparison is shown in Fig. 4, suggesting the equivalence of the two described calibration methods. The slight shift in the average of the distribution of the forecast horizon for the Rössler system may be caused by the insufficient number of initial conditions  $\mathbf{x}_0$  used to calculate  $\langle \varepsilon' \rangle_{\mathbf{x}_0}$ . In order to assess the equivalence of the interburst time  $\tau$ , the maximal burst amplitude  $\varepsilon_{\max}$  and the averaged short-term forecast error  $\langle \varepsilon' \rangle_{\mathbf{x}_0}$ , we further consider the variations of these quantities within the full relevant range of scaling of the parameters  $(\alpha, \beta)$ . While Figs. 5(b)–5(c) indicate similar qualitative variations, the landscape of  $\langle \varepsilon' \rangle_{\mathbf{x}_0}$  is different although the location of the minimum is not significantly different in Figs. 5(b)–5(d). Again, the disagreement may be caused by the inefficiency of the estimate of  $\langle \varepsilon' \rangle_{\mathbf{x}_0}$ .

To conclude, we have characterized the learning process as the establishment of a regime of strong temporal correlation between the observational and internal dynamical states. We have characterized the high-amplitude bursts interrupting this regime and investigated the corresponding consequences related to tuning the parameters  $(\alpha, \beta)$ . We have noted that while  $(\alpha, \beta)$  do not seem to affect the performance if the performance is evaluated by means of the regression error  $\varepsilon$ , the inspection of  $\tau$  and  $\varepsilon_{\max}$  as a function of  $(\alpha, \beta)$  shows clear variations. The quantities  $\tau$  and  $\varepsilon_{\max}$  may, therefore, serve as metrics for optimizing  $(\alpha, \beta)$ . Furthermore, while these considerations were related to the driven dynamical regime, we have noted the difficulties that may arise when defining performance via the error  $\varepsilon'$  obtained from the autonomous regime.



**FIG. 3.** Uncertainty analysis. (a) Performance landscape of the Lorenz system. (b) and (c) Blown up versions of (a). (d) Scaling of the uncertainty function. A parameter grid of size  $10^2 \times 10^2$  in the logarithmic scale was used in order to calculate each of the shown scaling behaviors. (e) Variations with respect to initial conditions exemplified for the Lorenz system.



**FIG. 4.** Fluctuations of the forecast horizon due to finite-time fluctuations. (a) Optimum of  $\varepsilon'$  averaged in respect to 30 initial points  $\mathbf{x}_0$ . (b) Optimum of  $\tau$ .

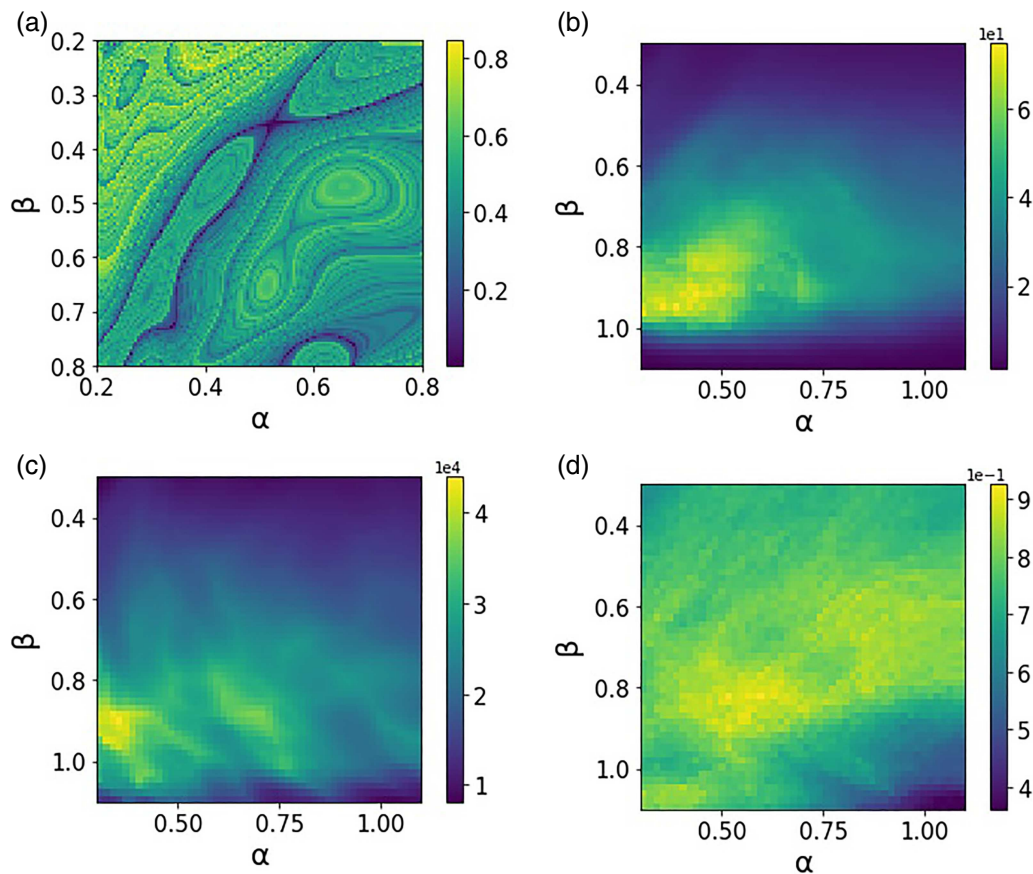


FIG. 5. Comparison between the different performance metrics  $\tau$ ,  $\epsilon_{\max}$ , and  $\langle \epsilon' \rangle$ . (a)  $\epsilon'$ , (b)  $\tau$ , (c)  $\epsilon_{\max}$ , and (d)  $\langle \epsilon' \rangle_{x_0}^{-1}$ .

We have quantified the sensitivity of  $\epsilon'$  by calculating the corresponding uncertainty exponent. The precise mechanism underlying the fine-scaled structures, as well as the reason for the seemingly generic value of the uncertainty exponent, has remained unclear. Future work shall be dedicated to these questions.

#### ACKNOWLEDGMENTS

We wish to thank D. Mohr, P. Huber, J. Aumeier, S. Baur, and J. Herteux for valuable discussions. We wish to acknowledge funding from the Deutsches Zentrum für Luft- und Raumfahrt (DLR).

#### DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### REFERENCES

<sup>1</sup>S. Birkholz, C. Brée, A. Demircan, and G. Steinmeyer, "Predictability of rogue events," *Phys. Rev. Lett.* **114**, 213901 (2015).

<sup>2</sup>V. Nair and R. I. Sujith, "Multifractality in combustion noise: Predicting an impending combustion instability," *J. Fluid Mech.* **747**, 635–655 (2014).

<sup>3</sup>P. Milan, M. Wächter, and J. Peinke, "Turbulent character of wind energy," *Phys. Rev. Lett.* **110**, 138701 (2013).

<sup>4</sup>J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845–848 (1987).

<sup>5</sup>J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," *Phys. Rev. Lett.* **120**, 024102 (2018).

<sup>6</sup>P. Antonik, M. Gulina, J. Pauwels, and S. Massar, "Using a reservoir computer to learn chaotic attractors, with applications to chaos synchronization and cryptography," *Phys. Rev. E* **98**, 012215 (2018).

<sup>7</sup>K. Nakai and Y. Saiki, "Machine-learning inference of fluid variables from data using reservoir computing," *Phys. Rev. E* **98**, 023111 (2018).

<sup>8</sup>H. Fan, J. Jiang, C. Zhang, X. Wang, and Y.-C. Lai, "Long-term prediction of chaotic systems with machine learning," *Phys. Rev. Res.* **2**, 012080 (2020).

<sup>9</sup>J. Jiang and Y.-C. Lai, "Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius," *Phys. Rev. Res.* **1**, 033056 (2019).

<sup>10</sup>O. L. White, D. D. Lee, and H. Sompolinsky, "Short-term memory in orthogonal neural networks," *Phys. Rev. Lett.* **92**, 148102 (2004).

<sup>11</sup>A. Haluszczynski and C. R ath, "Good and bad predictions: Assessing and improving the replication of chaotic attractors by means of reservoir computing," *Chaos* **29**(10), 103143 (2019).



- <sup>12</sup>A. Griffith, A. Pomerance, and D. J. Gauthier, "Forecasting chaotic systems with very low connectivity reservoir computers," *Chaos* **29**(12), 123108 (2019).
- <sup>13</sup>U. Parlitz and A. Hornstein, *Chaos Complex. Lett.* **1**, 135–144 (2005).
- <sup>14</sup>Z. Lu, B. R. Hunt, and E. Ott, "Attractor reconstruction by machine learning," *Chaos* **28**(6), 061104 (2018).
- <sup>15</sup>Z. Lu and D. S. Bassett, "Invertible generalized synchronization: A putative mechanism for implicit learning in biological and artificial neural systems," [arXiv1807.05214](https://arxiv.org/abs/1807.05214) (2019).
- <sup>16</sup>K. Caluwaerts, F. wyffels, S. Dieleman, and B. Schrauwen, "The spectral radius remains a valid indicator of the echo state property for large reservoirs," in *IEEE International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2013), p. 6.
- <sup>17</sup>I. B. Yildiz, H. Jaeger, and S. J. Kiebel, "Re-visiting the echo state property," *Neural Netw.* **35**, 1–9 (2012).
- <sup>18</sup>B. R. Hunt and E. Ott, "Optimal periodic orbits of chaotic systems," *Phys. Rev. Lett.* **76**, 2254–2257 (1996).
- <sup>19</sup>S. C. Venkataramani, B. R. Hunt, E. Ott, D. J. Gauthier, and J. C. Bienfang, "Transitions to bubbling of chaotic systems," *Phys. Rev. Lett.* **77**, 5361–5364 (1996).
- <sup>20</sup>Y.-C. Lai, C. Grebogi, J. A. Yorke, and S. C. Venkataramani, "Riddling bifurcation in chaotic dynamical systems," *Phys. Rev. Lett.* **77**, 55–58 (1996).
- <sup>21</sup>D. J. Gauthier and J. C. Bienfang, "Intermittent loss of synchronization in coupled chaotic oscillators: Toward a new criterion for high-quality synchronization," *Phys. Rev. Lett.* **77**, 1751–1754 (1996).
- <sup>22</sup>J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press Inc., 2003).
- <sup>23</sup>A burst at time  $t$  occurs when the instantaneous value of  $\varepsilon_t$  exceeds the variance of  $\varepsilon_t$ .
- <sup>24</sup>Y.-C. Lai and R. L. Winslow, "Riddled parameter space in spatiotemporal chaotic dynamical systems," *Phys. Rev. Lett.* **72**, 1640–1643 (1994).
- <sup>25</sup>The forecast horizon is defined as the time at which the instantaneous value of  $\varepsilon'_t$  exceeds unity.