

## Research



**Cite this article:** Cenci S, Medeiros LP, Sugihara G, Saavedra S. 2020 Assessing the predictability of nonlinear dynamics under smooth parameter changes. *J. R. Soc. Interface* **17**: 20190627.  
<http://dx.doi.org/10.1098/rsif.2019.0627>

Received: 8 September 2019  
Accepted: 13 December 2019

**Subject Category:**  
Life Sciences—Physics interface

**Subject Areas:**  
systems biology, bioinformatics

**Keywords:**  
time-series analysis, forecasting, nonlinear dynamics, population dynamics

**Author for correspondence:**  
Serguei Saavedra  
e-mail: sersaa@mit.edu

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4799619>.

# Assessing the predictability of nonlinear dynamics under smooth parameter changes

Simone Cenci<sup>1,2</sup>, Lucas P. Medeiros<sup>1</sup>, George Sugihara<sup>3</sup> and Serguei Saavedra<sup>1</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA

<sup>2</sup>DCI, LLC, 201 Spear Street, Suite 250, San Francisco, CA, USA

<sup>3</sup>Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

SC, 0000-0001-6843-468X; LPM, 0000-0002-0320-5058; SS, 0000-0003-1768-363X

Short-term forecasts of nonlinear dynamics are important for risk-assessment studies and to inform sustainable decision-making for physical, biological and financial problems, among others. Generally, the accuracy of short-term forecasts depends upon two main factors: the capacity of learning algorithms to generalize well on unseen data and the intrinsic predictability of the dynamics. While generalization skills of learning algorithms can be assessed with well-established methods, estimating the predictability of the underlying nonlinear generating process from empirical time series remains a big challenge. Here, we show that, in changing environments, the predictability of nonlinear dynamics can be associated with the time-varying stability of the system with respect to smooth changes in model parameters, i.e. its local structural stability. Using synthetic data, we demonstrate that forecasts from locally structurally unstable states in smoothly changing environments can produce significantly large prediction errors, and we provide a systematic methodology to identify these states from data. Finally, we illustrate the practical applicability of our results using an empirical dataset. Overall, this study provides a framework to associate an uncertainty level with short-term forecasts made in smoothly changing environments.

## 1. Introduction

Predicting the behaviour of natural and physical systems has been one of the great challenges of scientific inquiry [1–3]. In the past, the main limitation for the short-term predictability of most natural systems was the lack of an exact knowledge of their governing equations and the uncertainty associated with their parameters [1,4]. Then, in the late 1970s and early 1980s, the pioneering work of Takens [5] showed that important properties of nonlinear systems could be reconstructed solely based on empirical observations, allowing subsequent development of novel non-parametric methods for time-series forecasting [2,6,7]. Indeed, our capacity to build predictive non-parametric models has advanced substantially thanks to a combination of multivariate generalizations of Takens's embedding theorem [8,9] and deep-learning algorithms [4,10,11]. However, the possibility to predict nonlinear dynamics in empirical settings has remained fundamentally limited by the fact that, in changing environments, model parameters are in constant change (e.g. species competition within biological communities increases in periods of food scarcity [12], interest rates are affected by bank policies [13,14], among others) and these changes can have significant effects on the system's dynamics [15].

Unfortunately, the effects of environmental variability on a system's trajectory cannot be easily predicted solely based on past observations. This is simply because predictive models built on properties learned from already observed data cannot generalize well on new data generated by unseen processes. Put differently, minimizing a cost function over training data generated by a given model does not ensure that the cost is also minimized over data generated by a different model

[16]. This is a fundamental limitation arising in the applicability of modern statistical learning theories for the predictability of nonlinear dynamics in changing environments—where model parameters are changing frequently.

A possible way to estimate the effects of environmental variability on a system's dynamics is to use non-parametric scenario exploration [17]: an approach developed to assess the effect of a small change in a physical driver on empirically reconstructed dynamical systems. However, in this approach, the environmental drivers need to be explicitly included in the reconstructed dynamics and, therefore, its applicability may be limited when the drivers of the dynamics are only partially or not at all observed.

To overcome these limitations and to better address the real-world challenge of characterizing predictability in changing environments here we propose a non-parametric approach that focuses on estimating the local structural stability of the underlying data-generating process. This estimate is then used as an indicator of its predictability at any given point in time—the predictability associated with a particular location or 'state' in phase space. The local structural stability of a given dynamical system can be defined as the time-varying stability of the phase portrait under smooth changes of the vector field [18]. That is, the local structural stability of a dynamical system can be interpreted as the resistance to change its trajectory after a smooth change in model parameters. To test these ideas, we combine non-parametric methods and machine learning algorithms to detect time points where a system is more predictable. The remainder of this work is organized as follows. We first present the theoretical background; next we test our theoretical framework on synthetic time-series data; and then we validate our theory on an empirical biological time series.

## 2. Background

To illustrate our approach, we consider nonlinear time series generated by some unknown dynamical system

$$\dot{x} = \mathcal{F}(x, \omega), \quad (2.1)$$

where  $\mathcal{F} \in C^1$  is a nonlinear vector field (i.e. nonlinear dynamical model),  $x \in \mathbb{R}^d$  is the state vector and  $\omega$  is a vector of environment-dependent parameters (e.g. interaction strength between species in population dynamics models, interest rates in financial models). We assumed that environmental changes affect the model dynamics through smooth perturbations of the components of  $\omega$ , but we did not choose a particular parametric form for  $\mathcal{F}$ . In this work, our goal is to estimate the predictability of equation (2.1) from empirical observations and without information on environmental variability, given that in many empirical cases we cannot control or measure all environmental parameters.

In nonlinear systems under changing environments, the accuracy of a forecast of the state variable  $x$  depends both on the power of the learning algorithm (e.g. artificial neural network, local regression) used for prediction and on how predictable the system is. If a perfect learning algorithm were available, the predictability of equation (2.1) would only depend on how changes in  $\mathcal{F}$  or its parameters  $\omega$  (due to environmental changes) affect the value of the future state variables  $x$ . In this case, the predictability of equation (2.1) would depend on the structural stability of  $\mathcal{F}$ , that is,

the stability of this dynamical system against perturbations of its vector field or its parameters. Note that we are extending the classic definition of structural stability [19] (which looks at the capacity to remain in a given global qualitative state) to a local quantitative property [18,20]. Moreover, because we assumed  $\mathcal{F}$  to be nonlinear, the extent to which perturbations to  $\mathcal{F}$  result in changes of  $x$  also depends upon the state of the system (figure 1a). That is, the resistance of a nonlinear dynamical system to changes in its trajectory is state dependent. In the remainder of this work, we use the term local structural stability [18] to refer to structural stability that is state dependent—that can vary with the specific location in phase space.

Following the argument above, we investigate the association between local structural stability and the predictability of nonlinear time series in smoothly changing environments.<sup>1</sup> Estimating the local structural stability of nonlinear systems is in general extremely challenging. Importantly, previous work has found that the inverse of the volume contraction rate (VCR) of the vector field at a given point in the state space is a proxy for its local structural stability [18]. Here, we go a step further and propose that the VCR is also an inverse measure of local predictability in changing environments. The VCR ( $\mathcal{V}$ ) of a continuous dynamical system is the divergence of the vector field  $\mathcal{F}$  [19],

$$\mathcal{V} \equiv \nabla \cdot \mathcal{F} = \text{Tr}(\mathcal{J}), \quad (2.2)$$

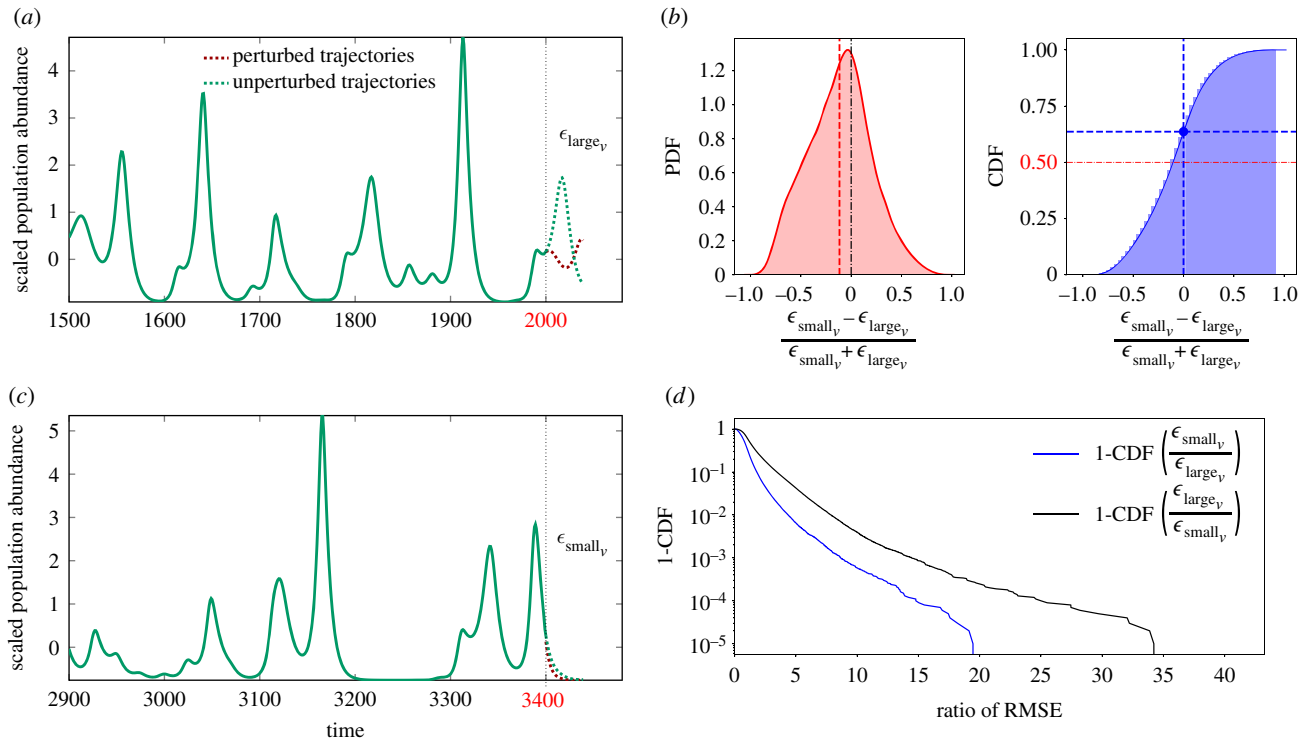
where  $\mathcal{J}$  is the Jacobian matrix of  $\mathcal{F}$  and  $\text{Tr}(\mathcal{J})$  is the trace of this matrix. For example, in a system with three state variables  $x = [x, y, z]$  and a vector field with three components (i.e. three differential equations) we have

$$\mathcal{V} = \frac{\partial \mathcal{F}_x}{\partial x} + \frac{\partial \mathcal{F}_y}{\partial y} + \frac{\partial \mathcal{F}_z}{\partial z}. \quad (2.3)$$

Note that, generally, in nonlinear systems with interacting terms  $\mathcal{V}:\mathcal{V}(x)$  (i.e. the VCR is state dependent). Therefore, unless the equilibrium dynamics is a stable fixed point, the structural stability of the system changes in time.<sup>2</sup> Below, we provide a heuristic argument for using the VCR as an inverse measure of local predictability of nonlinear time series under changing environments.

Let  $\mathcal{S}$  be the state space of equation (2.1), whereas we call  $\omega$  and  $\hat{\omega}$  the unperturbed and perturbed parameters of equation (2.1), respectively. For the purposes of working with empirical data, we can assume that, for any  $x \in \mathcal{S}$  and small perturbations to the parameters (i.e. for  $d(\omega, \hat{\omega}) \sim 0$  under some metric  $d$  such as Euclidean distance),  $\text{Tr}(\mathcal{J}(x, \omega)) \sim \text{Tr}(\hat{\mathcal{J}}(x, \hat{\omega}))$ . That is, for small parameter perturbations, the perturbed and unperturbed dynamics have similar VCRs. Note that this is a very reasonable assumption for continuous and differentiable functions, and it is applicable to many scenarios of environmental change in empirical data [19]. Yet, it is important to recall that this argument is only valid under the assumption of smooth parameter perturbations (i.e. smooth changes in the environmental conditions) and does not apply to strong qualitative changes in the dynamics (e.g. invasion or extinction of a species in an ecosystem).

In a nonlinear dynamical system, the VCR provides a measure of how the volume of the state space changes under the flow of the vector field. Regions in the state space with small VCR correspond to regions under which volumes in the state space do not increase and, therefore, trajectories do not diverge. Within these regions of the state space, two



**Figure 1.** Volume contraction rate (VCR) and local structural stability in changing environments. The time series in (a) and (c) are generated from a numerical integration of a five-dimensional chaotic dynamical system (figure S1). The panels show that the same parameter perturbation occurring at two different points along the model's trajectories (i.e.  $t = 2000$  and  $t = 3400$ ) can have completely different effects on the dynamics. This effect can be measured as the deviation  $\epsilon$  between perturbed (red dotted line) and unperturbed (green dotted line) trajectories. This result is analysed further in (b). The two figures show the PDF and CDF for the percentage difference in the deviation of trajectories (after perturbing the dynamics) at points with small ( $\epsilon_{\text{small}}$ ) and large ( $\epsilon_{\text{large}}$ ) VCRs (the perturbation is the same at the two points). The red line in the left panel indicates the median of the distribution. The blue line in the right panel corresponds to the probability below which  $\epsilon_{\text{small}} < \epsilon_{\text{large}}$ . In (d), we analyse the tail of the distribution. The y-axis is the complement of the cumulative distribution function of the ratio between the two deviations. The panel shows that the probability that a small perturbation leads to a large effect (i.e. large deviation between perturbed and unperturbed trajectories) is much larger at points with large VCRs (top black line) than at points with small VCRs (bottom blue line). This result is validated on other chaotic dynamical systems (see electronic supplementary material, figure S1). Overall, the figure shows that a system's response to a perturbation to the parameters depends on the value of the VCR at the time at which the perturbation occurs. (Online version in colour.)

nearby models (i.e. the model with original parameters and the model with perturbed parameters) have, by the assumption above, similarly low VCRs. Therefore, going from one model (unperturbed) to the other (perturbed), their respective trajectories do not diverge. By contrast, the effects of perturbations occurring at states with a large VCR are, on average, magnified by the flow. This observation allows us to link the VCR to the predictability of the dynamics. That is, because, without prior information on perturbations, a learning algorithm can only predict the unperturbed trajectory, prediction errors will be small when perturbed and unperturbed trajectories do not diverge (i.e. at states with small VCR). On the other hand, when the divergence of the trajectories is magnified by the flow (i.e. when the VCR is large) prediction errors will be larger. As such, this suggests that the VCR can be used to estimate the local predictability of nonlinear time series under changing environments.

Note that the VCR at a given state of a nonlinear dynamical system is the sum of the local Lyapunov exponents (i.e. finite-time Lyapunov exponents computed with the given state as initial condition) [21,22]. Thus, differently from the largest Lyapunov exponent, the VCR contains information about the stretching of nearby trajectories along all the stable and unstable directions. That is to say, the VCR is a linear function of the whole Lyapunov spectrum that captures how volumes in a  $d$ -dimensional space expand or contract. This implies that, for

each point in time, the VCR provides an average rate of divergence or convergence of nearby trajectories, providing a robust estimation of local structural stability and, therefore, of predictability in smoothly changing environments. It is important to note that, because there is always uncertainty associated with the magnitude and direction of perturbations, any predictability metric such as the VCR can only measure the probability of committing small or large prediction errors. Because the ultimate goal of our approach is to analyse empirical data, we would like to highlight that the estimation of the VCR only requires knowing the diagonal elements of the Jacobian matrix of the dynamics—a quantity that can be efficiently estimated from multivariate time series with available non-parametric methods (see discussion below and electronic supplementary material, figure S5, for examples of the inference of the VCR from noisy nonlinear time series [12,23]).

### 3. Prediction of synthetic data

To validate our approach, we provide a numerical illustration using synthetic data. Specifically, we show that the VCR can be used to estimate the local structural stability of the dynamics. Then, we show that structural stability implies predictability. First, to show that the VCR is a valid estimator of local structural stability (i.e. the stability of a system's trajectory in response to smooth parameter perturbations), we

compare the effect of the same parameter perturbation at points on the attractor with a small VCR versus points with a large VCR. For this purpose, we integrate the dynamics of a five-dimensional chaotic dynamical system (see electronic supplementary material, figure S1) and compute the VCR along the trajectory using the analytical Jacobian matrix  $\mathcal{J}$ . Then, we sample one random point with a large VCR and one random point with a small VCR from the system's trajectory. We took the small VCR within the lower 15th percentile of the VCR distribution, whereas the large VCR is within the upper 15th percentile of the same distribution. Then, we perturb a random number of parameters and integrate the dynamics again with the new perturbed parameters using the two sampled points on the attractor as initial conditions.

Parameter perturbations are of the form

$$\hat{\omega}_i = \omega_i + \mathcal{N}\left(0, \frac{\omega_i}{\mathcal{U}(1, \eta)}\right), \quad (3.1)$$

where  $\omega_i \in \boldsymbol{\omega}$ ,  $\mathcal{N}$  and  $\mathcal{U}$  are the normal and uniform distributions, respectively. That is, the perturbation to the parameter is a random number sampled from a Gaussian distribution with zero mean and standard deviation proportional to the value of the parameter itself. By changing the standard deviation of the perturbation (i.e. by changing the value of  $\eta$ ), we can change the overall level of perturbation to the system. In order to preserve the spirit of a non-parametric framework (where in principle there is no knowledge about the governing equations and their parameters), at every realization we also randomly sample which and how many parameters to perturb. Yet, understanding the effects of individual parameters can be a fruitful research area that future work can explore.

After the integration of the new dynamics, we measured the root mean square distance (RMSD) between perturbed and unperturbed trajectories at the small VCR state ( $\epsilon_{\text{small}_V}$ ) and at the large VCR state ( $\epsilon_{\text{large}_V}$ ; see figure 1a,c for an illustrative example). We repeat this same numerical experiment  $10^5$  times. In order to estimate the effect of the same parameter perturbation on two different system states we analyse the distribution of two statistics. First, we compute the percentage difference between RMSD after perturbations at points with small and large VCR, i.e.  $X = (\epsilon_{\text{small}_V} - \epsilon_{\text{large}_V})/(\epsilon_{\text{small}_V} + \epsilon_{\text{large}_V})$ . We show in figure 1b the kernel density estimation of the probability density function (PDF) and cumulative distribution function (CDF) of this statistic. The PDF shows that the mass of the distribution is skewed towards negative values ( $\langle X \rangle \sim -0.12$ ), indicating that, on average,  $\epsilon_{\text{small}_V} < \epsilon_{\text{large}_V}$ . Similarly, the CDF shows that the probability that this inequality holds is clearly greater than 50% (i.e.  $\mathcal{P}(\epsilon_{\text{small}_V} < \epsilon_{\text{large}_V}) \sim 60\%$ ).

Because mean values can be misleading, we also analyse the severity of the extreme events (i.e. the tail risk) and compute the statistics  $Y = (\epsilon_{\text{small}_V})/(\epsilon_{\text{large}_V})$  and  $Z = (\epsilon_{\text{large}_V})/(\epsilon_{\text{small}_V})$ . We then plot the complement of the empirical cumulative distribution function for the random variables  $Y$  and  $Z$ . For a given ratio  $r$  of RMSDs, the condition  $\Pr(Z > r) > \Pr(Y > r)$  implies that points on the attractor with small VCR are more structurally stable than points with large VCR under the exact same parameter perturbation. Figure 1c shows that the complements of the two cumulative distribution functions, i.e.  $\Pr(Z > r)$  and  $\Pr(Y > r)$ , are substantially different, with the former being much larger than the latter. This reveals that the impact of

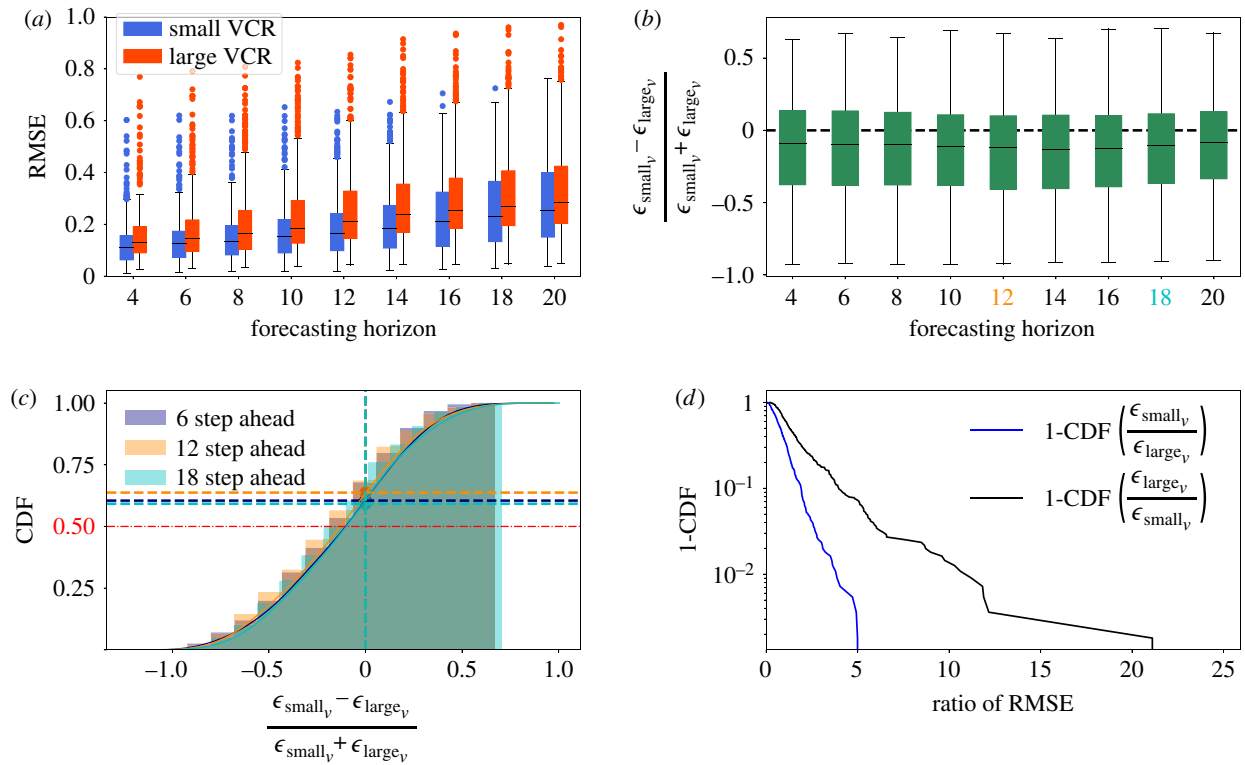
smooth perturbations occurring at points with large VCR is much larger than the impact of perturbations taking place at states with low VCR. To further validate this result, we repeat the same numerical experiment with three additional chaotic dynamical systems (see electronic supplementary material, figure S1). The results are shown in electronic supplementary material, figure S1, and are consistent with our theoretical expectations. These examples provide support to the hypothesis that the VCR is a good estimator of local structural stability in nonlinear dynamical systems [18].

It is worth noting that similar results can be obtained using the largest local Lyapunov exponent as a measure of local structural stability (electronic supplementary material, figure S5 and figure S2). However, as shown in electronic supplementary material, figure S2, which was generated mirroring the simulation done to generate figure 1b, the shift of the mass of the PDF and CDF towards negative values is more significant when predictability is estimated using the VCR. While Lyapunov exponents are not considered a measure of predictability [24], this result is not surprising as a local perturbation in the parameters should induce changes in the trajectories, which will then diverge at a rate measured by the Lyapunov exponents. Yet, as discussed above, the VCR provides a more grounded measure of this divergence because it is a metric constructed by weighting stable and unstable manifolds equally. Moreover, from a practical and computational point of view, the VCR is a more reliable metric to estimate from empirical data than the largest Lyapunov exponent. That is, the computation of the VCR only requires the estimation of the diagonal elements of the Jacobian matrix ( $d$  parameters), while the Lyapunov exponents require the whole matrix ( $d^2$  parameters), making the latter more prone to errors [7,25].

Having corroborated the validity of the VCR as a measure of local structural stability, we now show that, in changing environments, predictions from structurally stable states are less error prone. To illustrate the relationship between local structural stability and predictability, we repeat the same numerical experiment outlined above. However, differently from above, rather than integrating the dynamics, here we forecast the trajectories after each perturbation using a long short-term memory (LSTM) artificial neural network [26,27]. Then, we compute the root mean square error (RMSE), which is widely adopted as a measure of forecasting skill [10,28], between the perturbed and predicted trajectories. Note that the RMSD and the RMSE are mathematically the same, but conceptually different. We used this terminology to distinguish the two types of experiments: divergence (RMSD) and forecasting (RMSE). We focus on short-term predictions (i.e. small forecast horizon) because long-term predictions of nonlinear systems are known to be very unreliable in empirical data even in the absence of parameter perturbations. This analysis is repeated for different levels of perturbations in model parameters (see appendix A for details).

Figure 2a shows the distribution of RMSE of predictions at small and large VCRs as a function of the forecast horizon with a given level of perturbation (i.e.  $\eta = 6$  in equation (3.1); results for a larger level of perturbation are reported in electronic supplementary material, figure S3). This figure illustrates that predictions at states with a small VCR are more accurate than predictions at states with a large VCR, confirming that the VCR is an effective estimator of the uncertainty in short-term predictions of a dynamical system. These results further confirm the link between local structural stability and local





**Figure 2.** Theoretical validation of the volume contraction rate (VCR) as an estimator of predictability in changing environments. (a) The distribution (boxplots) of RMSE in the test set at different forecasting horizons. Predictions are made with a long short-term memory (LSTM) artificial neural network after perturbations at points with small and large VCRs (left and right boxplots, respectively). Perturbations are performed as in the analysis presented in figure 1b, but here we measure the deviations between perturbed and predicted trajectories (i.e. RMSE) instead of deviations between perturbed and unperturbed trajectories (i.e. RMSD). Points at small (large) VCR are sampled from the lower (upper) 15th percentile of the trace of the Jacobian matrix, which is computed numerically at each point on the attractor. The means of the two distributions (i.e. small and large VCRs) are different and the robustness of this result to various sample sizes is discussed in electronic supplementary material, figures S2 and S4. (b) The same distribution shown in figure 1b but as boxplots, which are a function of the forecasting horizon. This panel, together with the CDF in (c), shows that the mass of the distribution is moved towards negative values, i.e. prediction errors after perturbations at small VCRs are on average smaller than prediction errors after perturbation at large VCRs. The blue, orange and cyan lines show the probability mass corresponding to negative values in the percentage difference as a function of forecasting horizon. (d) As expected from figure 1d, the same parameter perturbation has a much greater impact on the predictability of points with large VCRs (top black line) than points with small VCRs (bottom blue line). Overall, the figure shows that the VCR is a consistent measure of predictability in smoothly changing environments. (Online version in colour.)

predictability. The difference in mean RMSE between small and large VCR is statistically significant for all forecast horizons, and this result is robust to the choice of the sample size (see electronic supplementary material, figures S2 and S4). Figure 2b,c is equivalent to figure 1b. The two panels show that, in line with our theoretical expectation, the mass of the distribution is moved towards negative values, i.e. prediction errors after perturbations at small VCR are on average smaller than prediction errors after perturbations at large VCR.

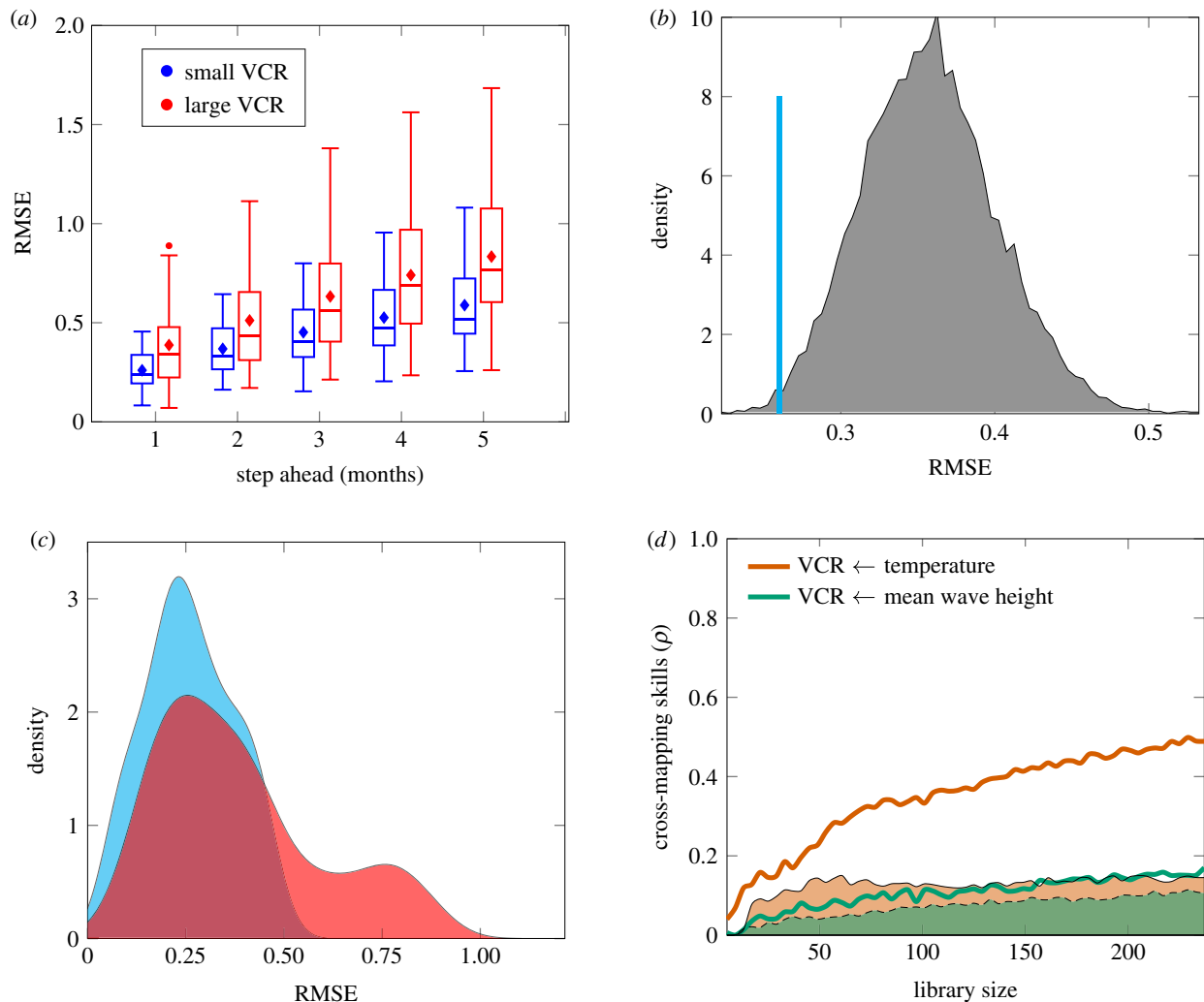
Similarly to the analysis shown in figure 1c, we also estimate the effect on predictability of the same parameter perturbation at different VCR states. Figure 2d shows that the complement of the cumulative distribution function of the ratio of the RMSEs follows our theoretical expectations (figure 1c). That is, in the presence of smooth perturbations to the model parameters (i.e. in smoothly changing environments), the probability of large prediction errors is much larger when predictions are made from states with a large VCR.

## 4. Prediction of empirical data

The theoretical framework and results above support the hypothesis that the VCR can be used to estimate the predictability of nonlinear time series in smoothly changing

environments. We now turn to demonstrate the practical applicability of our approach on empirical data. Specifically, we analyse a biological time series of a rocky intertidal community that exhibits dynamics at the edge of chaos [29]. The community consists of three species (mussels, algae and barnacles) that compete for space on bare rock. Data were collected daily for over 20 years in New Zealand [29].

Following the analysis carried out on synthetic data, we compare the RMSEs after predictions at points with small and large VCRs (see appendix A for further details). Differently from the numerical simulations, the Jacobian matrices at each time step are not available when dealing with empirical data. Therefore, to compute the empirical VCRs non-parametrically, we reconstruct the Jacobian coefficients from the observed time series using the regularized S-map [23], which we also use for forecasting. The regularized S-map is a locally weighted regularized state-space regression model that has been shown to be a robust estimator of Jacobian coefficients from noisy time-series data [7,12,23,25]. See electronic supplementary material, figure S3 for further details on the S-map. Electronic supplementary material, figure S5 shows an illustrative example of inference of VCR from synthetic data and electronic supplementary material, figure S6 shows a comparison of the forecasting skill of the regularized S-map and LSTM. To avoid any cross-contamination between training and test data,



**Figure 3.** Empirical validation of the volume contraction rate as an estimator of predictability in changing environments: NZ rocky intertidal example [30]. (a) The distribution (boxplots) of the root mean square error (RMSE) of the predictions as a function of the forecast horizon (months) for the empirical time series. In line with theoretical expectations (figure 2), the VCR is a good measure of predictability. (b) The results of the significance test described in appendix A; the figure illustrates that the results in (a) are statistically significant. The vertical blue line corresponds to the mean RSME of all predictions made using points with small VCRs (lower 15th percentile of the trace of the inferred Jacobian matrix and predictions are one step ahead), whereas the grey distribution corresponds to the ensemble of mean RMSEs calculated from random samples of the (one step ahead) predictions. The blue curve and red curve in (c) show the kernel density estimation of the distribution of RMSEs in the test set at points with small and large VCRs, respectively. The panel illustrates that very large errors are significantly more likely after predictions at states with large VCRs. Electronic supplementary material, figure S9 shows the robustness of the results to the choice of percentile and prediction horizons. (d) The results of the causality test ( $y$ -axis is the correlation between predicted and estimated) of VCR with temperature (top brown line) and mean wave height (bottom green line). The shaded regions correspond to the expected cross-mapping skills from randomized time series (see appendix A). The panel illustrates that sea temperature has a much greater causal effect on the VCR than mean wave height. (Online version in colour.)

rather than computing the VCR over the whole time series and then sampling points beyond which to forecast (as in the synthetic data analysis section), we compute the VCR sequentially using only past data and then we perform out-of-sample forecasts (see appendix A for further details).

Figure 3a confirms the theoretical results shown in figure 2. In particular, the figure shows that predictions at states (i.e. time points) with a small VCR have, on average, a smaller RMSE than predictions at states with a large VCR. This difference between small and large VCR is statistically significant for all forecast horizons (see electronic supplementary material, table S7). Hence, in line with the results found using synthetic data, the VCR computed empirically from the time series can be used to estimate the local predictability of the data. Note that predictions were made up to five months ahead with a threshold in the VCR set at the 15th percentile. The robustness of the results at different percentiles is shown in electronic supplementary material, figure S9.

To further test the validity of our analysis, we also test for the statistical significance of our results by comparing the observed RMSEs in the lower percentile of the VCR with the distribution of possible mean RMSEs that one would have obtained by using random criteria of predictability. This distribution is computed by randomly sampling multiple subsets of RMSEs, regardless of the VCR at the time of prediction, and then computing the mean of each subset (see appendix A for a more detailed explanation of the test). We found that the VCR is a statistically significant measure of predictability (figure 3b) as long as the separation criterion between small and large VCR is within the 35% of the inferred trace of the Jacobian matrix (see electronic supplementary material, figure S8). Figure 3c shows the kernel density estimation of the distribution of the RMSE in the test set in the lower and upper percentiles of the VCR. The figure clearly illustrates that the two distributions have not only different means but also different shapes. Specifically, even though the distribution

of errors at large VCRs (red region) has a peak at low RMSE, it is clearly more skewed towards larger errors than the distribution of errors at small VCRs (cyan region). This result implies that the probability of large prediction errors is substantially higher when predictions are made at states (i.e. time points) with large VCR. The overlap of the two distributions also implies that other factors, such as noise in the empirical data, most likely affect the local predictability of a time series. Our results are consistent with theoretical expectations (figures 1, 2 and electronic supplementary material, figure S1) and are robust to the choice of the threshold in the VCR (see the distributions in electronic supplementary material, figure S9).

It is important to notice that, unlike the synthetic example, the biological example is not a closed system but is part of a broader environmental context. Thus, an interesting question to ask is how the predictability of this community might be related to relevant environmental variables that are exogenous to the system. To answer this question, we investigate the causal relationship between the VCR with temperature and mean wave height using a non-parametric causality test known as convergent cross mapping [31] (CCM; see appendix A for further details on CCM). It is known that temperature and mean wave height are two key environmental drivers of rocky intertidal communities [18,29,32,33]. Figure 3*d* reveals that sea temperature has a causal influence on the VCR and, therefore, on predictability. However, there is no significant causal relationship between mean wave height and VCR. This result is interesting because it illustrates that environmental variables (e.g. mean wave height) can play a crucial role in the dynamics (e.g. regulate the mortality rate [32]) but have no effect on the predictability of the community dynamics. The statistical significance of the causal links is assessed using surrogate time series (shaded area in figure 3*d*; see appendix A).

## 5. Discussion

Forecasting with nonlinear time series is a central challenge in science and engineering [1–3]. The issue is particularly hard to address when true dynamics are unknown and parameters are likely to change in response to environmental variations. In this context, it is of practical relevance to know when predictions can be trusted the most. To address this issue, here we have provided theoretical arguments, as well as synthetic and empirical evidence in support of the hypothesis that the VCR, a local property of the deterministic skeleton of the dynamics, provides reliable information about the local predictability of nonlinear dynamical systems in smoothly changing environments. Interestingly, analysing synthetic and empirical data, we have found that the distributions of prediction errors at states (i.e. time points) with small and large VCR differ considerably. In particular, the probability of very large prediction errors with respect to the mean is significantly higher when predictions are made at states with a large VCR. We believe that this is an important finding that can have practical implications and that requires further theoretical investigation. For example, with sufficient empirical data, it is in principle possible to compare the current VCR with historical values in order to check if the current VCR is large compared with past values. If this is true, there is a greater risk of making inaccurate forecasts at the current system state.

Despite the potential applicability of the VCR, it is also important to discuss the limitations of this approach. One evident pitfall is that, theoretically, there could be cases in which the VCR stays constant throughout the time. A clear example is the Lorenz system: a chaotic dynamical system whose predictability is notoriously state dependent [34] but, at the same time, has a constant VCR. Yet, while theoretically possible, empirical systems with nonlinear dynamics but constant VCR are probably the exception rather than the rule. For example, a simple process noise (i.e. a noise typically proportional to the square root of the state variable) would generate state-dependent VCRs. A second limitation is the (in)accuracy of the reconstruction of the VCR from empirical data. In the example provided in this article, we have detailed information about the system under investigation. Specifically, all the relevant variables were observed, allowing us to use a fully multivariate embedding from which we know we can reconstruct the VCR accurately using the regularized S-map [12,23]. In cases where not all variables are collected from experiments or observational studies, the dynamics can still be reconstructed from the available observations using Takens's embedding theorem [5]. However, the results will be sensitive to the details of the attractor reconstruction (e.g. the selection of the optimal time lag and embedding dimension). Because the trace of the Jacobian matrix cannot be associated with the VCR if the reconstructed attractor is not embedded in the correct dimensional space, particular care needs to be placed on the application of the methods from reconstructed dynamics. Overall, we have provided a framework to estimate, non-parametrically, the local predictability of empirical nonlinear time series under smooth parameter changes. We believe that this approach and its potential extensions can have important practical applications in management and risk-assessment studies.

**Code availability.** The code to reproduce the figures is available on GitHub at <https://github.com/MITEcology/JRSI-Cenci-Medeiros-Sugihara-Saavedra-2019.git>. The folder also includes a code to estimate the VCR from a multivariate time series.

**Data accessibility.** The data used in the manuscript can be downloaded from [www.pnas.org/content/112/20/6389/tab-figures-data](http://www.pnas.org/content/112/20/6389/tab-figures-data).

**Authors' contributions.** S.C., G.S. and S.S. designed the study, S.C. and L.P.M. performed the study, S.S. supervised the study. S.C., G.S., L.P.M. and S.S. wrote the manuscript.

**Competing interests.** We declare we have no competing interests.

**Funding.** Funding was provided by DoD-Strategic Environmental Research and Development Program 15 RC-2509, Lenfest Ocean Program 200028335, NSF 3DEB-1655203, NSF ABI-1667584, the McQuown Fund and the McQuown Chair in Natural Sciences, University of California, San Diego (G.S.), MIT Research Committee funds and the Mitsui Chair (S.S.).

## Endnotes

<sup>1</sup>Note that our question is different from the problem of stability addressed with sensitivity analysis because, as discussed above, we assume limited information on environmental factors. Also, the problem we are addressing is different from the problem of estimating dynamical stability in chaotic dynamical systems. This other problem focuses on the issue of how perturbations of state variables affect the model trajectories, and it is typically addressed by the computation of Lyapunov exponents. Our work, on the other hand, focuses on the issue of how perturbations of model parameters affect the model's future trajectories.

<sup>2</sup>Note that the VCR also changes during the transient state, regardless of the nature of the fixed point. However, here we only focus on long-term dynamics.

<sup>3</sup>We present this threshold as an illustrative example as in [18] but it has already been shown with synthetic data that the results are robust to the specific choice of the threshold. Nonetheless, for the empirical data we performed an extensive analysis of the robustness to the choice of the threshold.

## Appendix A

### A.1. Analysis of synthetic data

To show that the VCR is a valid measure of predictability in smoothly changing environments, we ran the following analysis. We considered a five-species chaotic population dynamics model as in figure 1a (see electronic supplementary material, figure S1 for details on the model). We integrated the dynamics over  $10^5$  time points and sampled equally spaced points along the trajectories (one every 150 points). Then, we numerically computed the VCR along the attractor using equation (2.2), and we sampled a pair of points in the lower and upper 15th percentiles of this measure.<sup>3</sup> Using these points, we perturbed a random number of parameters with Gaussian random variables at zero mean and with standard deviation proportional to the value of the parameters themselves (see electronic supplementary material, figure S1). Then, we integrated again the dynamics with the new parameters, taking as the initial condition the two perturbed points (i.e. lower and upper percentiles). Note that we applied the very same perturbation to each pair of points. This provided two unperturbed trajectories (i.e. before the two perturbation points) over which we trained an LSTM artificial neural network [26,27] and we then made forecasts over the two perturbed trajectories. Training was performed over the 500 data points preceding the perturbation time, i.e.  $t_{\text{training}} \in [t_{p-500}, \dots, t_p]$ , where  $t_p$  is the time of perturbation. Forecasts were performed up to 20 time steps in the future (i.e. we were interested in short-term forecasts given that long-term predictability of chaotic systems in changing environments is practically unfeasible). Finally, we computed the RMSE at different forecast horizons in the two test sets, i.e. one after the perturbation at the structurally stable points and the second one after the perturbation at the structurally unstable points. We repeated these numerical experiments 500 times at different levels of perturbations. Results of this analysis are shown in figure 2.

### A.2. Analysis of empirical data

We used a multivariate time series of species abundances collected from a 22-year-long observational study of a marine intertidal community from New Zealand. The data can be downloaded from [30]. The community is composed of three species (mussels, algae and barnacles) competing for space on bare rock. Because the three species compete for space, the overall system has four dimensions. It has already been shown that the time series exhibits dynamics at the edge of chaos [30], making it a perfect candidate for our analysis. The time series, as downloaded from [30], is already pre-processed (see [29] for further details on the system).

To test the validity of the VCR as an inverse measure of predictability we run the following analysis: first, we trained the S-map on 150 data points. Then, we computed the VCR from the estimated Jacobians (coefficients of the S-map; see electronic supplementary material and [12]). At this point,

we made predictions up to five months ahead and we saved (i) the VCR before the prediction and (ii) the RMSE of the forecast. Then, we took a new point from the data, fitted the new Jacobian coefficients, computed the new VCR, predicted again and computed the new RMSE. We repeated this numerical experiment up to the final point of the time series. Finally, we plotted the RMSEs as a function of the forecasting horizon for all those points below and above a certain threshold of the VCR. All the forecasts are completely out-of-sample and the VCR is always computed using training data only. The threshold was fixed at the first 15th percentile in the figure shown in the main text. The results at different thresholds are shown in electronic supplementary material, figure S9.

### A.3. Causality test

To test for the existence of a causal link between variables in nonlinear dynamical systems, we used CCM [31], a non-parametric test for causality developed within an empirical dynamic modelling (EDM) framework. Given two variables  $X$  and  $Y$ , CCM uses the results from Takens's theorem to reconstruct two versions of the same manifold, one with embedding of  $X$  and one with embedding of  $Y$ . Then, variable  $X$  is said to be a causal driver of variable  $Y$  if information of  $X$  is contained in the manifold of  $Y$ , i.e.  $Y$  cross maps  $X$ . See [31] for a more detailed presentation of the causality test. To ensure the statistical significance of the causal links with CCM, two conditions need to be satisfied: (i) convergence towards higher cross-mapping skills as the number of data points increases and (ii) CCM skills need to be higher than the one we would obtain with a surrogate version of the time series. Because temperature has a strong seasonal pattern, we used a null model that takes into account seasonality in the data. Specifically, we used the function *make\_surrogate\_season* in the *rEDM* package in R. Note that before performing the causality test, for an accurate reconstruction of the VCR, we have used an ensemble method to compute the Jacobian coefficients [25]; see electronic supplementary material. In electronic supplementary material, figure S10, we show that CCM provides results that meet biological expectations, i.e. the VCR has no causal effect on environmental variables in the biological time series. That is, the VCR does not influence either sea temperature or mean wave height, confirming what would be naturally expected.

### A.4. Statistical tests

As a first test for the validity of the VCR as a measure of predictability, we performed a  $t$ -test on the two distributions (i.e. errors at small and large VCR). Results are shown in electronic supplementary material, table S7. The  $p$ -value of the test is significantly smaller than 5% for all cases. Because the sample size of the predictions in the empirical dataset is relatively small, we ran a second analysis to test if results are statistically significant. First, we set a threshold on the VCR (e.g. 15th percentile as in the main text). Then, we computed the mean RMSE in the test set for all the predictions made at points below this threshold (cyan line in figure 3b). Then, we randomly sampled a subset of the predictions in the test set regardless of the value of the VCR at the time of prediction, and we also computed the mean RMSE in this subset. We chose as sample size the same size as the one in the lower 15th percentile of the VCR. This is effectively a null model that assumes that VCR is not an estimator of



predictability. In other words, the null hypothesis establishes that the mean RMSE observed in low VCR states is expected to happen from any randomly chosen VCR state. Note that our test statistic is the mean RMSE. Hence, the null distribution is not the RMSE distribution, but the mean RMSEs sampled from different VCR states. We repeated this

sampling process  $10^4$  times and we calculated their distribution (grey curve in figure 3b) and a  $p$ -value as area under the curve below the true RMSE (cyan line). The results of the significance test at different thresholds of the VCR and at different forecasting horizons are shown in electronic supplementary material, figure S8.

## References

- Lorenz EN. 1963 Section of planetary sciences: the predictability of hydrodynamic flow. *Trans. N. Y. Acad. Sci.* **25**, 409–432. (doi:10.1111/j.2164-0947.1963.tb01464.x)
- Kantz H, Schreiber T. 2004 *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press.
- Bradley E, Kantz H. 2015 Nonlinear time-series analysis revisited. *Chaos* **25**, 097610. (doi:10.1063/1.4917289)
- Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937. (doi:10.1073/pnas.1517384113)
- Takens F. 1981 *Detecting strange attractors in turbulence*, pp. 366–381. Berlin, Germany: Springer.
- Sugihara G, May RM. 1990 Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**, 734–741. (doi:10.1038/344734a0)
- Sugihara G. 1994 Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A* **348**, 477–495. (doi:10.1098/rsta.1994.0106)
- Deyle ER, Sugihara G. 2011 Generalized theorems for nonlinear state space reconstruction. *PLoS ONE* **6**, e18295. (doi:10.1371/journal.pone.0018295)
- Ye H, Sugihara G. 2016 Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. *Science* **353**, 922–925. (doi:10.1126/science.aag0863)
- Jaeger H, Haas H. 2004 Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80. (doi:10.1126/science.1091277)
- Pathak J, Hunt B, Girvan M, Lu Z, Ott E. 2018 Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102. (doi:10.1103/PhysRevLett.120.024102)
- Deyle ER, May RM, Munch SB, Sugihara G. 2016 Tracking and forecasting ecosystem interactions in real time. *Proc. R. Soc. B* **283**, 20152258. (doi:10.1098/rspb.2015.2258)
- Kontonikas A, MacDonald R, Saggiu A. 2013 Stock market reaction to fed funds rate surprises: state dependence and the financial crisis. *J. Bank. Finance* **37**, 4025–4037. (doi:10.1016/j.jbankfin.2013.06.010)
- Davig T, Gerlach JR. 2006 State-dependent stock market reactions to monetary policy: ‘bubble’ vs. fundamental states. *Int. J. Central Bank.* **2**. (doi:10.2139/ssrn.882151)
- Serquina R, Lai Y-C, Chen Q. 2008 Characterization of nonstationary chaotic systems. *Phys. Rev. E* **77**, 026208. (doi:10.1103/PhysRevE.77.026208)
- Bickel S, Brückner M, Scheffer T. 2007 Discriminative learning for differing training and test distributions. In *Proc. 24th Int. Conf. on Machine Learning*, pp. 81–88. New York, NY: ACM.
- Deyle ER, Maher MC, Hernandez RD, Basu S, Sugihara G. 2016 Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* **113**, 13081–13086. (doi:10.1073/pnas.1607747113)
- Cenci S, Saavedra S. 2019 Non-parametric estimation of the structural stability of non-equilibrium community dynamics. *Nat. Ecol. Evol.* **3**, 912. (doi:10.1038/s41559-019-0879-1)
- Strogatz S. 2014 *Nonlinear dynamics and chaos: with applications in physics biology, chemistry and engineering*. Boca Raton, FL: CRC Press.
- Saavedra S, Rohr RP, Gilarranz LJ, Bascompte J. 2014 How structurally stable are global socioeconomic systems? *J. R. Soc. Interface* **11**, 20140693. (doi:10.1098/rsif.2014.0693)
- Jan A, Claude-henri L. 2003 *Bifurcation and chaos in nonsmooth mechanical systems*. World Scientific Series On Nonlinear Science Series A. Singapore: World Scientific.
- Wei M. 2000 Quantifying local instability and predictability of chaotic dynamical systems by means of local metric entropy. *Int. J. Bifurcation Chaos* **10**, 135–154. (doi:10.1142/S021812740000086)
- Cenci S, Sugihara G, Saavedra S. 2019 Regularized S-map for inference and forecasting with noisy ecological time series. *Methods Ecol. Evol.* **10**, 650–660. (doi:10.1111/2041-210X.13150)
- Smith LA. 1994 Local optimal prediction: exploiting strangeness and the variation of sensitivity to initial condition. *Phil. Trans. R. Soc. Lond. A* **348**, 371–381. (doi:10.1098/rsta.1994.0097)
- Cenci S, Saavedra S. 2018 Uncertainty quantification of the effects of biotic interactions on community dynamics from nonlinear time-series data. *J. R. Soc. Interface* **15**, 20180695. (doi:10.1098/rsif.2018.0695)
- Hochreiter S, Schmidhuber J. 1997 Long short-term memory. *Neural Comput.* **9**, 1735–1780. (doi:10.1162/neco.1997.9.8.1735)
- Gers FA, Schmidhuber J, Cummins F. 1999 Learning to forget: continual prediction with LSTM. *Neural Comput.* **12**, 2451–2471. (doi:10.1162/089976600300015015)
- Perretti CT, Munch SB, Sugihara G. 2013 Model free forecasting outperforms the correct mechanistic model for simulated and experimental data. *Proc. Natl Acad. Sci. USA* **110**, 5253–5257. (doi:10.1073/pnas.1216076110)
- Benincà E, Ballantine B, Ellner SP, Huisman J. 2015 Species fluctuations sustained by a cyclic succession at the edge of chaos. *Proc. Natl Acad. Sci. USA* **112**, 6389–6394. (doi:10.1073/pnas.1421968112)
- Benincà E, Huisman J, Heerkloss R, Jöhnk KD, Branco P, Van Nes EH, Scheffer M, Ellner SP. 2008 Chaos in a long-term experiment with a plankton community. *Nature* **451**, 822–825. (doi:10.1038/nature06512)
- Sugihara G, May R, Ye H, Hsieh C-h, Deyle E, Fogarty M, Munch S. 2012 Detecting causality in complex ecosystems. *Science* **338**, 496–500. (doi:10.1126/science.1227079)
- Helmuth BST, Hofmann GE. 2001 Microhabitats, thermal heterogeneity, and patterns of physiological stress in the rocky intertidal zone. *Biol. Bull.* **201**, 374–384. (doi:10.2307/1543615)
- Harley CDG, Helmuth BST. 2003 Local and regional scale effects of wave exposure, thermal stress, and absolute versus effective shore level on patterns of intertidal zonation. *Limnol. Oceanogr.* **48**, 1498–1508. (doi:10.4319/lo.2003.48.4.1498)
- Nese JM. 1989 Quantifying local predictability in phase space. *Physica D* **35**, 237–250. (doi:10.1016/0167-2789(89)90105-X)

# Supplementary Information: Assessing the predictability of nonlinear dynamics under smooth parameter changes

Simone Cenci<sup>1,2</sup>, Lucas P. Medeiros<sup>1</sup>, George Sugihara<sup>3</sup> and Serguei Saavedra<sup>1¶</sup>

<sup>1</sup> Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA

<sup>2</sup> DCI, LLC, 201 Spear Street, Suite 250, San Francisco, CA, USA

<sup>3</sup> Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

## S1 Chaotic models used in the study

For reproducibility purposes, here we list the chaotic dynamical systems used in this study. The model used in the main text is a 5-species chaotic consumer-resources model [12]:

$$\begin{aligned}\frac{dP_i}{dt} &= \nu_i \lambda_i \frac{P_i C_i}{C_i + C_i^*} - \nu_i P_i \\ \frac{dC_i}{dt} &= \mu_i \kappa_i \frac{C_i R}{R + R^*} - \nu_i \lambda_i \frac{P_i C_i}{C_i + C_i^*} - \mu_i C_i \\ \frac{dR}{dt} &= R \left(1 - \frac{R}{k}\right) - \sum_{i=\{1,2\}} \mu_i \kappa_i \frac{C_i R}{R + R^*},\end{aligned}\tag{S1}$$

with  $i = \{1, 2\}$  and  $\nu_1 = 0.1, \nu_2 = 0.07, \lambda_1 = 3.2, \lambda_2 = 2.9, C_1^* = C_2^* = 0.5, \mu_1 = \mu_2 = 0.15, \kappa_1 = 2.5, \kappa_2 = 2, R^* = 0.3, k = 1.2$ . To generate Figure 1A we have perturbed  $\nu_1 \rightarrow \tilde{\nu}_1 = 0.15$ . For the main analysis instead (i.e., Figure 1B) we perturbed a random number of parameters (sample from a uniform distribution  $\mathcal{U}(3, \text{number of parameters})$ ) with Gaussian noise  $\mathcal{N}(0, \frac{\text{parameter}}{\text{Noise}})$  where  $\text{Noise} = \mathcal{U}(1, 6)$  and  $\mathcal{U}(1, 3)$  for small and large perturbation respectively.

In Figure 1C we have repeated the same analysis we have shown in Figure 1B for different chaotic models. Specifically, from left to right:

---

<sup>¶</sup>To whom correspondence should be addressed. E-mail: sersaa@mit.edu

467

- Model 1 [36]:

$$\begin{aligned}\frac{dx}{dt} &= -y \\ \frac{dy}{dt} &= x + z \\ \frac{dz}{dt} &= xz + ay^2\end{aligned}\tag{S2}$$

468

with  $a = 0.3$ .

469

- Model 2 [37]:

$$\begin{aligned}\frac{dx}{dt} &= ax - yz \\ \frac{dy}{dt} &= -by + xz \\ \frac{dz}{dt} &= -cz + xyz + k\end{aligned}\tag{S3}$$

470

with  $a = 4, b = 9, c = 3.6, k = 4$ .

471

- Model 3 [38] (chaotic Lotka-Volterra):

$$\frac{dx_i}{dt} = r_i x_i \left( 1 - \sum_{j=1}^4 a_{ij} x_j \right),\tag{S4}$$

472

with  $\vec{r} = [1, 0.72, 1.53, 1.27]$  and :

$$\mathcal{A} = \begin{bmatrix} 1 & 1.09 & 1.52 & 0 \\ 0 & 1 & 0.44 & 1.36 \\ 2.33 & 0 & 1 & 0.47 \\ 1.21 & 0.51 & 0.35 & 1 \end{bmatrix}.\tag{S5}$$

---

## S2 Statistical significance of the difference between means in the numerical simulations

To test whether the differences between the mean errors at small and large VCR in Figures 2A-B are statistically significant we ran a t-test. We found that all the results are significant with  $p\text{-value} \ll 0.05$ .

However, in numerical simulations, small  $p$ -values may be caused by large sample sizes. Therefore, to control for this effect, we ran the following analysis: first we fixed a forecasting horizon (e.g., 4 step ahead). Then we considered the results from the whole simulation to be our population of errors at small and large VCR and we sample two random samples of size  $n$  from the two populations (small and large VCR). Then, we computed the mean in the two samples and the significance of their difference with a t-test. If both the mean error at small VCR is smaller than the mean error at large VCR in this sample and the  $p$ -value of the t-test is less than 5%, then we assign a value of 1 (success) to the simulation. Otherwise, we assign a value of 0 (failure). Notice that the conditions for success are both  $\langle RMSE \rangle_{smallV} < \langle RMSE \rangle_{largeV}$  and  $p\text{-value} < 5\%$ . We repeated this simulation 200 times and we computed the accuracy of the test. An accuracy of 1 means that every realization has passed the test, i.e., the means are different. An accuracy of 0.5 means that half of the time the VCR would not have been a good estimator of predictability. Then, we repeated the experiment for every forecasting horizon and for different sample sizes. Results are shown in Figure S4. The figure shows that at a small sample size, at both a very small forecasting horizon and a large forecasting horizon, the accuracy is low (still significantly greater than 0.5). Instead, at a medium forecasting horizon, the accuracy is greater than 90% for any sample size. Overall, this test shows that both the VCR is a reliable measure of predictability in changing environments and the results in Figure 2 are robust to the choice of sample size.

## S3 Estimation of the volume contraction rate from empirical data

As discussed in the main text to compute the VCR we need to estimate the Jacobian coefficients non-parametrically from empirical data. To this end, we used the regularized



502 S-map[7, 12]. The regularized S-map is a locally weighted linear model with an elastic  
 503 net regularization function—a convex mixture of  $L_1$  and  $L_2$  penalty terms [40, 41] intro-  
 504 duced to stabilize the regression in the presence of noise. Specifically, for each point on  
 505 the reconstructed manifold attractor (i.e., for each  $t^* \in \{1, \dots, n\}$  with  $n$  number of obser-  
 506 vations) we solved the following minimization problem:

$$\min_{\mathcal{J} \in \mathbb{R}^{d \times d}} (Y - X\mathcal{J})^T K(X, X^*, \theta) (Y - X\mathcal{J}) + \lambda(\alpha \|\mathcal{J}\|_2^2 + (1 - \alpha) \|\mathcal{J}\|_1), \quad (\text{S6})$$

507 In Eq. (S6)  $\mathcal{J}$  is the Jacobian matrix, and  $X$  is an  $(n - 1) \times d$  data matrix where the point  
 508 removed from the data matrix is the target point  $\vec{x}(t^*)$ . In addition,  $Y$  is the variable to  
 509 be predicted (i.e.,  $Y = X_{t+1} \forall X \neq X^*$ ),  $K(X, X^*, \theta)$  is a kernel function, and  $\lambda$  and  $\alpha$  are  
 510 two regularization parameters. From a preliminary analysis of the empirical time series,  
 511 we have found that the best kernel function to use is the exponential kernel with diagonal  
 512 elements:

$$K(\vec{x}, \vec{x}^*, \theta) = e^{-\theta \frac{\|\vec{x} - \vec{x}^*\|}{d}} \quad (\text{S7})$$

513 Notice that because empirical data are seldom practically identifiable (i.e., the error land-  
 514 scape around the solution of Eq. (S6) is typically flat) we actually compute the Jacobian  
 515 coefficients using a model average technique. That is, we create an ensemble  $\mathcal{M}$  of models  
 516  $\mathcal{L}$  with similar training and test errors and then we compute the weighted average of the  
 517 parameters in the ensemble, i.e., the actual Jacobian matrix at time  $t$  is computed as:

$$\mathcal{J}_{\text{ensemble}}(t) = \mathbb{E}[\mathcal{J}(t)|X] = \sum_{m \in \mathcal{M}} \mathbb{E}[\mathcal{J}(t)|\mathcal{L}_m, X] \mathbb{P}[\mathcal{L}_m|X] \quad \forall t \in [t_o, t_f]. \quad (\text{S8})$$

518 In Reference [25] it has been shown that this ensemble method increases the quality of the  
 519 inference of the jacobian coefficients.

## 520 S4 Locally linear forecasts

521 Because the empirical time series that we use are significantly shorter than the synthetic  
 522 time series, here we use the S-map for both estimating the average VCR and build a predic-  
 523 tive model. Specifically, to make forecast we run the following analysis. The regularized

---

524 S-map provides  $m - 1$  Jacobian coefficients given that the last data point of the training  
 525 set is not trained (i.e., we do not have information about the  $m + 1$  point which is the first  
 526 data in the test set). Therefore, we use the  $(m - 1)$ th coefficient and the  $m$ th data point to  
 527 make predictions about the first data point in the test set (i.e., the  $(m + 1)$ th point in the  
 528 time series). Predictions are made with a locally-linear model [43]:

$$x_i(t + 1) = c_0 + \sum_j \mathcal{J}_{ij}(t - 1)x_j(t). \quad (\text{S9})$$

529 where  $x_i$  is the  $i$  variable in the time series. The intercept is fitted adding a column of  
 530 ones to the data matrix [44]. After each forecast, we use the predicted point to fit the  
 531  $m$ th interaction coefficient, and we repeat the operation for as many times as numbers of  
 532 data points we have in the test set. Note that the forecast is purely out-of-sample because  
 533 this procedure never uses the original test set to select the model parameters or to make  
 534 predictions.

## 535 S5 Local Lyapunov exponents

536 In nonlinear time series analysis literature, the Lyapunov exponents, which measure the  
 537 growth rate of the distance between two infinitesimally close initial conditions, have been  
 538 linked to the predictability of the dynamics in the presence of noise or uncertainty in the  
 539 initial conditions [45, 46, 47]. However, it is unclear to what extent local measures of  
 540 Lyapunov exponents (i.e., finite-time Lyapunov exponents computed for each time point  
 541 along an attractor) are related to predictability under changing environments (i.e., param-  
 542 eter perturbations). In this section, we want to test whether our results could have been  
 543 obtained by simply using the largest local Lyapunov exponents as a measure of structural  
 544 stability.

545 To verify this, we computed the largest local Lyapunov exponent for models 1-3 described  
 546 in section S1. Following standard approaches [48, 49] to compute the local Lyapunov ex-  
 547 ponents, we integrated the tangent dynamics of the system  $\frac{d\vec{y}}{dt} = \mathcal{J}\vec{y}$ , where  $\vec{y}$  is a small  
 548 perturbation of the state vector  $\vec{x}$  (i.e., a tangent vector) and  $\mathcal{J}$  is the Jacobian matrix at  
 549 state  $\vec{x}$ . Then, we computed the largest Lyapunov exponent,  $\lambda_1$ , as the rate at which  $\vec{y}$   
 550 shrunk or expanded after a small amount of time. For a  $d$ -dimensional system, esti-  
 551 mating only the largest Lyapunov exponent,  $\lambda_1$ , is much more accurate than evolving  $d$

---

552 tangent vectors together in order to estimate the whole spectrum of Lyapunov exponents  
553 ( $\lambda_1 > \lambda_2 > \dots > \lambda_d$ ) [48, 49].

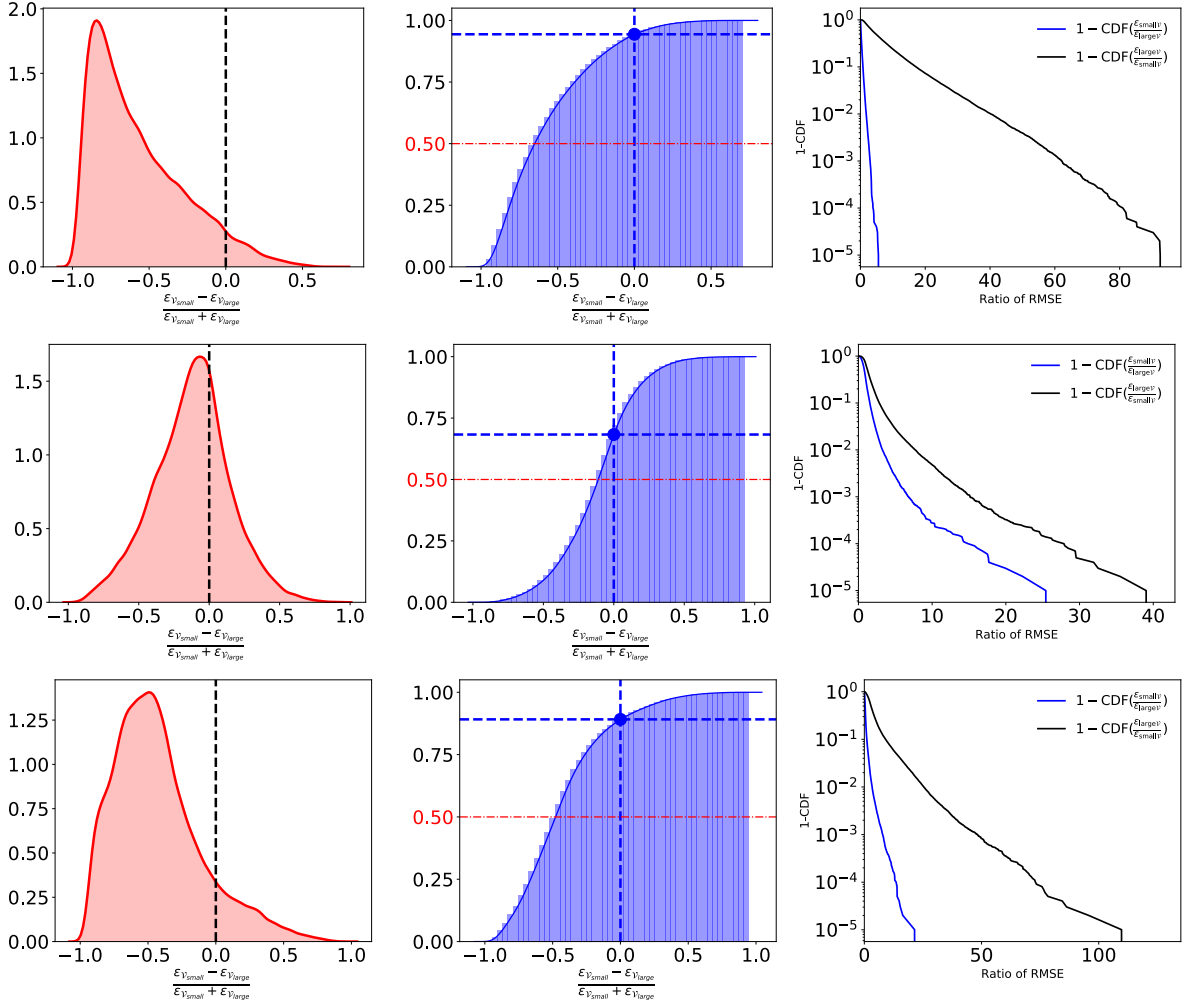
554 After computing the largest local Lyapunov exponent,  $\lambda_1$ , we reproduced the analysis  
555 used to generate Figure 1B but using  $\lambda_1$  instead of the VCR. That is, for a given model,  
556 we integrated its dynamics and computed all Lyapunov exponents for each point in time.  
557 Then, we sampled one random point with a large  $\lambda_1$  and one random point with a small  
558  $\lambda_1$  from the system's trajectory based on the lower and upper 15<sup>th</sup> percentile of the  $\lambda_1$   
559 distribution. Then, we perturbed a random number of parameters and integrated the  
560 dynamics again with the new perturbed parameters using the two sampled points on the  
561 attractor as initial conditions.

562 Figure S2 is an exact replica of Figure 1B and Figure S1. By comparison of Figure S2 with  
563 these two Figures, we can clearly see that the shift of the mass of the PDF and CDF towards  
564 negative values is more substantial when predictability is estimated using the VCR. We  
565 attribute this better performance to the fact that, as discussed in the main text, the VCR is  
566 a linear combination of the Lyapunov spectrum and, therefore, by weighting equally stable  
567 and unstable manifolds it provides a more accurate probabilistic measure of predictability.

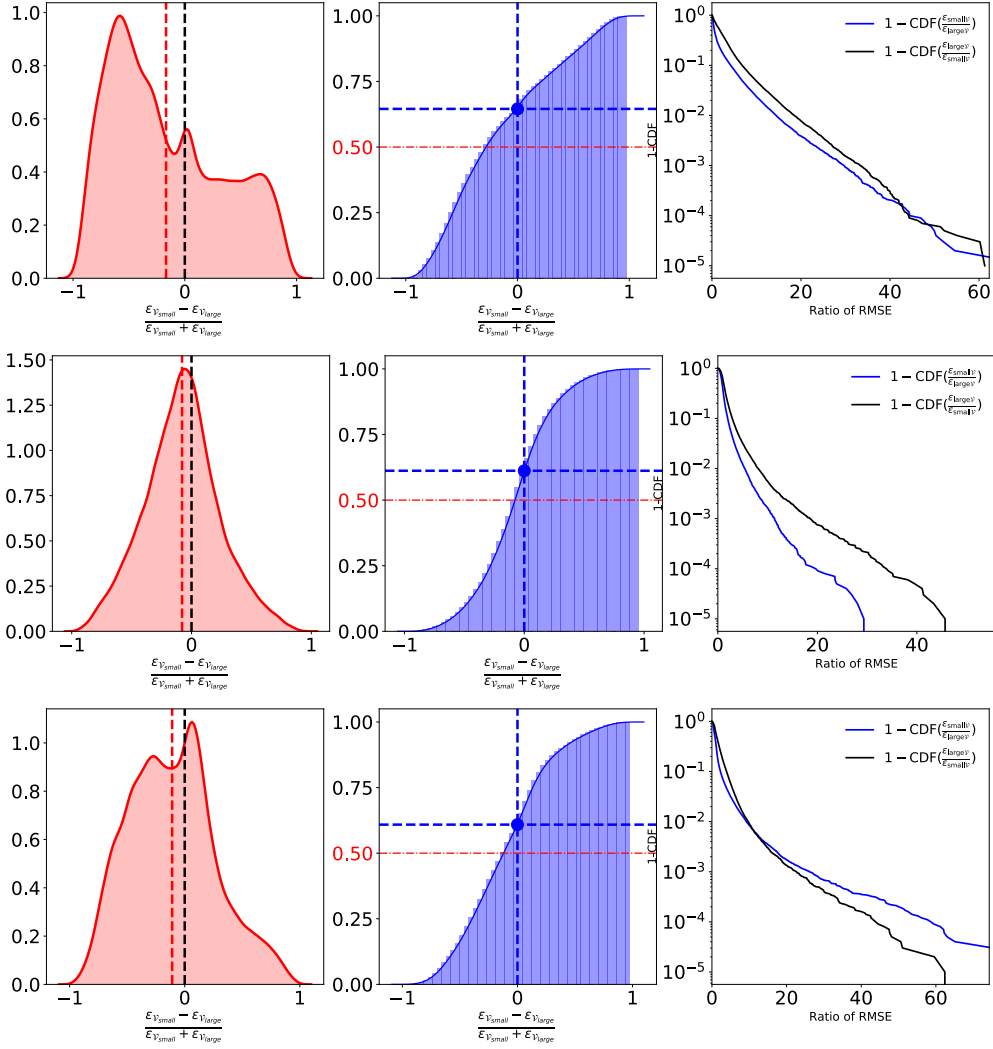
- 
- [35] Deyle, E. R., May, R. M., Munch, S. B. & Sugihara, G., 2016 Tracking and forecasting ecosystem interactions in real time. *Proceedings of the Royal Society of London B: Biological Sciences* **283**. ISSN 0962-8452.
  - [36] Sprott, J. C., 1994 Some simple chaotic flows. *Phys. Rev. E* **50**, R647–R650. (doi:10.1103/PhysRevE.50.R647).
  - [37] Lai, Q., Akgul, A., Li, C., Xu, G. & Ünal, C., 2018 A new chaotic system with multiple attractors: Dynamic analysis, circuit realization and s-box design. *Entropy* **20**.
  - [38] Vano, J. A., Wildenberg, J. C., Anderson, M. B., Noel, J. K. & Sprott, J. C., 2006 Chaos in low-dimensional lotka–volterra models of competition. *Nonlinearity* **19**, 2391.
  - [39] Sugihara, G., 1994 Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **348**, 477–495. ISSN 0962-8428.
  - [40] Zou, H. & Hastie, T., 2005 Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301–320.
  - [41] Li, Q. & Liny, N., 2010 The bayesian elastic net. *Bayesian Analysis* **5**, 151–170.
  - [42] Cenci, S. & Saavedra, S., 2018 Uncertainty quantification of the effects of biotic interactions on community dynamics from nonlinear time-series data. *Journal of The Royal Society Interface* **15**, 20180695.
  - [43] Ye, H., Clark, A., Deyle, E., Munch, S., Keyes, O., Cai, J., White, E., Cowles, J., Stagge, J., Daon, Y. *et al.*, 2017 *rEDM: Applications of Empirical Dynamic Modeling from Time Series*. R package version 0.6.5.
  - [44] Hastie, T., Tibshirani, R. & Friedman, J., 2001 *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York.
  - [45] Ziehmman, C., Smith, L. A. & Kurths, J., 2000 Localized lyapunov exponents and the prediction of predictability. *Physics Letters A* **271**, 237–251.
  - [46] Ding, R. & Li, J., 2007 Nonlinear finite-time lyapunov exponent and predictability. *Physics Letters A* **364**, 396–400.

- 
- [47] Bailey, B. A., 1996 Local lyapunov exponents: predictability depends on where you are. In *Nonlinear Dynamics and Economics: Proceedings of the Tenth International Symposium in Economic Theory and Econometrics*, volume 10, p. 345. Cambridge University Press Cambridge.
- [48] Benettin, G., Galgani, L., Giorgilli, A. & Strelcyn, J.-M., 1980 Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica* **15**, 9–20.
- [49] Ansmann, G., 2018 Efficiently and easily integrating differential equations with jitcode, jitcdde, and jitcsde. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**, 043116.

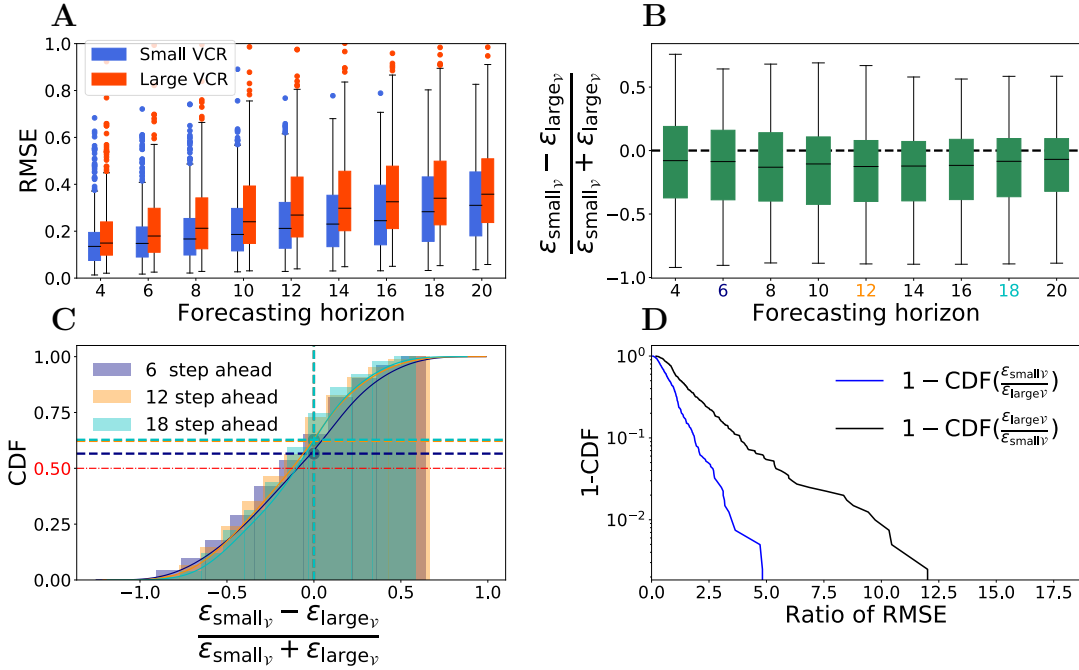




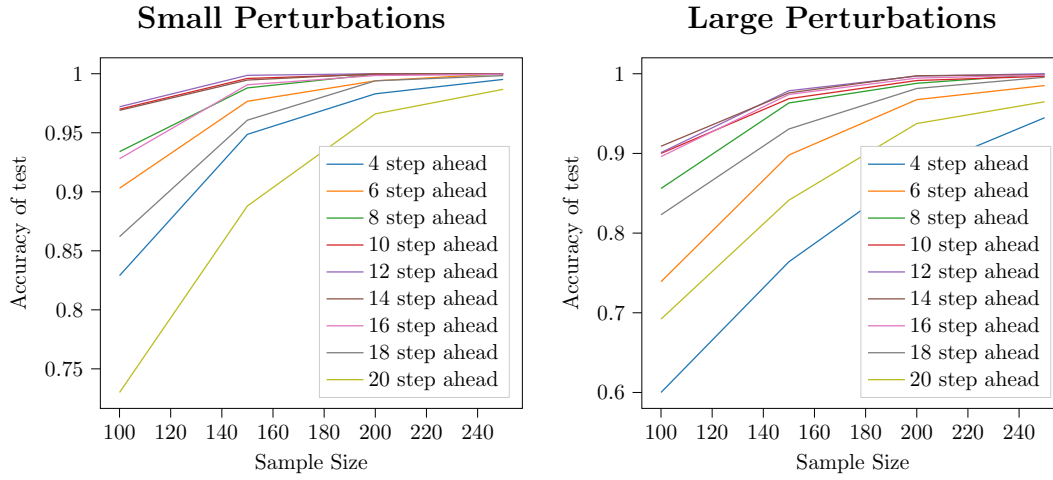
**Fig. S1 | The VCR is a good estimator of local structural stability of non-equilibrium dynamical systems.** Here we repeat the same theoretical analysis discussed in the main text (Figure 1B) using the models described in section S1, i.e., models 1-3 in the different rows of the figure. Overall, the figure provides further support to the hypothesis that the VCR is a good proxy for the local structural stability of non-equilibrium dynamical systems.



**Fig. S2 | The largest local Lyapunov exponent as estimator of the local structural stability of non-equilibrium dynamical systems.** Here we repeat the same theoretical analysis discussed in the main text (Figure 1B) using the models described in section S1 (i.e., models 1-3 in the different rows of the figure), but using the largest Lyapunov exponent ( $\lambda_1$ ) instead of the VCR. Overall, the figure shows that using the largest local Lyapunov exponent rather than the VCR as a measure of local structural stability, we would have obtained qualitatively similar results. However, the shift of the mass of the PDF and CDF towards negative values is sensibly more evident when predictability is estimated using the VCR, Figure S1. Moreover, note that when working with empirical data, the VCR only requires the estimation of the diagonal elements of the Jacobian matrix ( $d$  parameters), while the Lyapunov exponents require the whole matrix ( $d^2$  parameters), making the latter more prone to errors [7, 25].

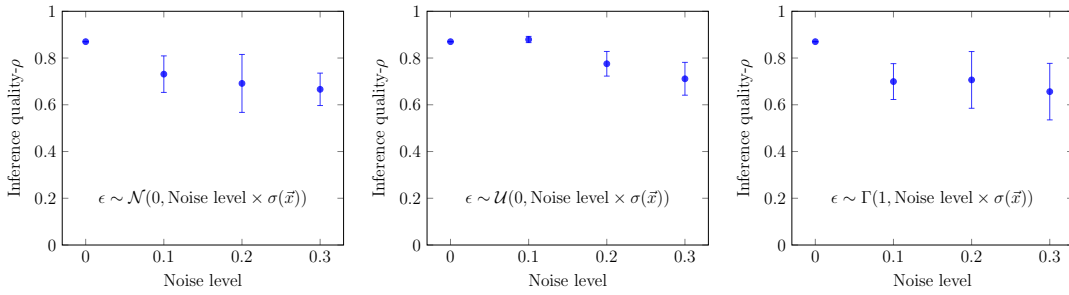


**Fig. S3 | Robustness of the results in Figure 2.** This Figure is an exact replica of Figure 2 but we have used a larger level of perturbation in the simulations ( $\eta = 3$ ). In Eq. (4), Noise =  $\mathcal{U}(1, 6)$  and  $\mathcal{U}(1, 3)$  correspond to small ( $\eta = 6$ ) and large ( $\eta = 3$ ) perturbations, respectively.

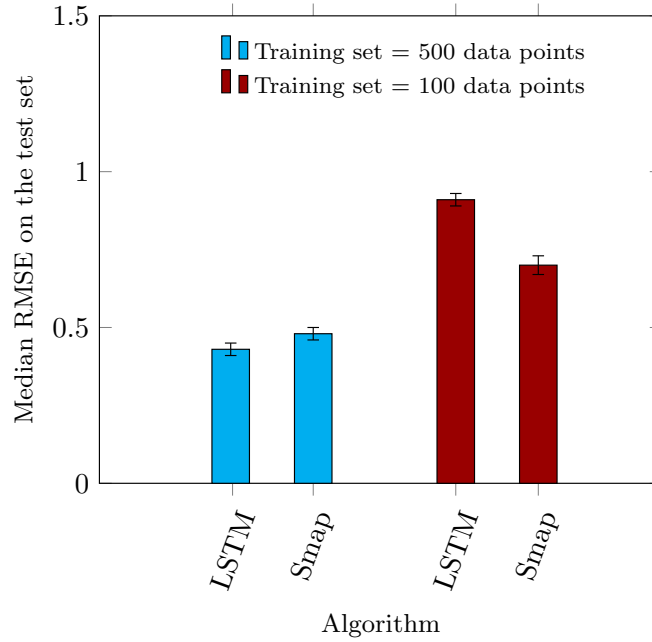


**Fig. S4 | Robustness of the results in Figure 2.** The figure shows the statistical test discussed in Section S2. The x-axis is the sample size used to compute the mean of the sub-population. The y-axis is the accuracy of the test, i.e., the percentage of times the mean error at small and large VCR were statistically significantly different. Overall the figure shows that results are robust to the choice of the sample size in the simulations





**Fig. S5 | Numerical estimation of the VCR from noisy time series.** Here we used the regularized S-map to infer the Jacobian coefficients of the dynamical system used in the main text. Then we compute the trace of the inferred Jacobian and we compare it with the trace of the analytical ones. For the inference we use a training set of 400 data points. The observational noise on the time series,  $\epsilon$ , is normally distributed (left panel), uniformly distributed (central panel) and sample from a gamma distribution (right panel). In the Figure  $\sigma(\vec{x})$  is the standard deviation of the time series. For each distribution and each level of noise the quality of the inference is computed over 100 realizations and the error bar show the standard error of the mean. The figure illustrates that the inference of the trace of the Jacobian matrix is consistently estimated even in the presence of observational noise (notice that performance is measured in terms of Pearson's correlation coefficient as, in order to sample points at low and high VCR we only care about the trend of the trace of the Jacobian).

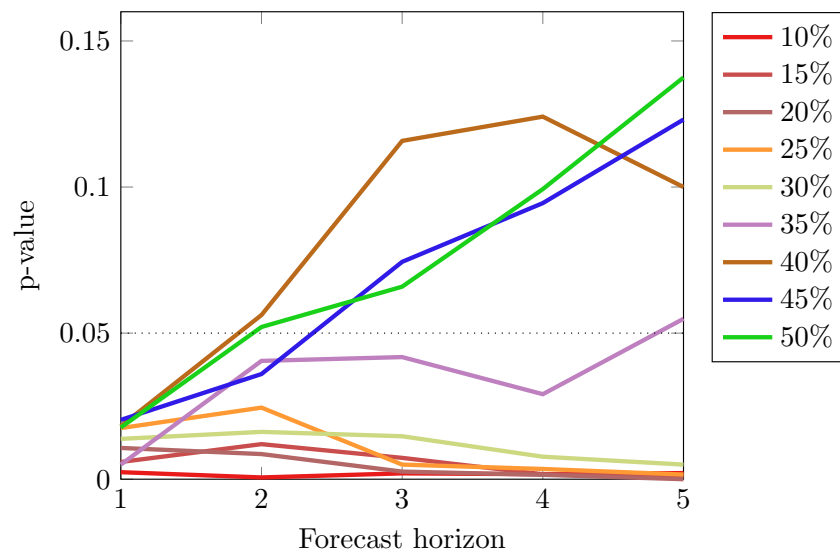


**Fig. S6 | Comparison of the performance of the S-map and the LSTM artificial neural network as a function of the length of the training set.** Performance is measured as the median of the RMSE in the test set over 200 predictions of the population dynamics model used in the main text. The error bars are computed with nonparametric bootstrapping. The figure illustrates that for short training sets, the S-map outperforms the LSTM artificial neural network. Thus, we have used the S-map to forecast the empirical time series. Notice that the median naive error for this time series is  $\sim 1.2$ ; hence, both algorithms perform significantly better than naive forecasts even when trained over as little as 100 data points.

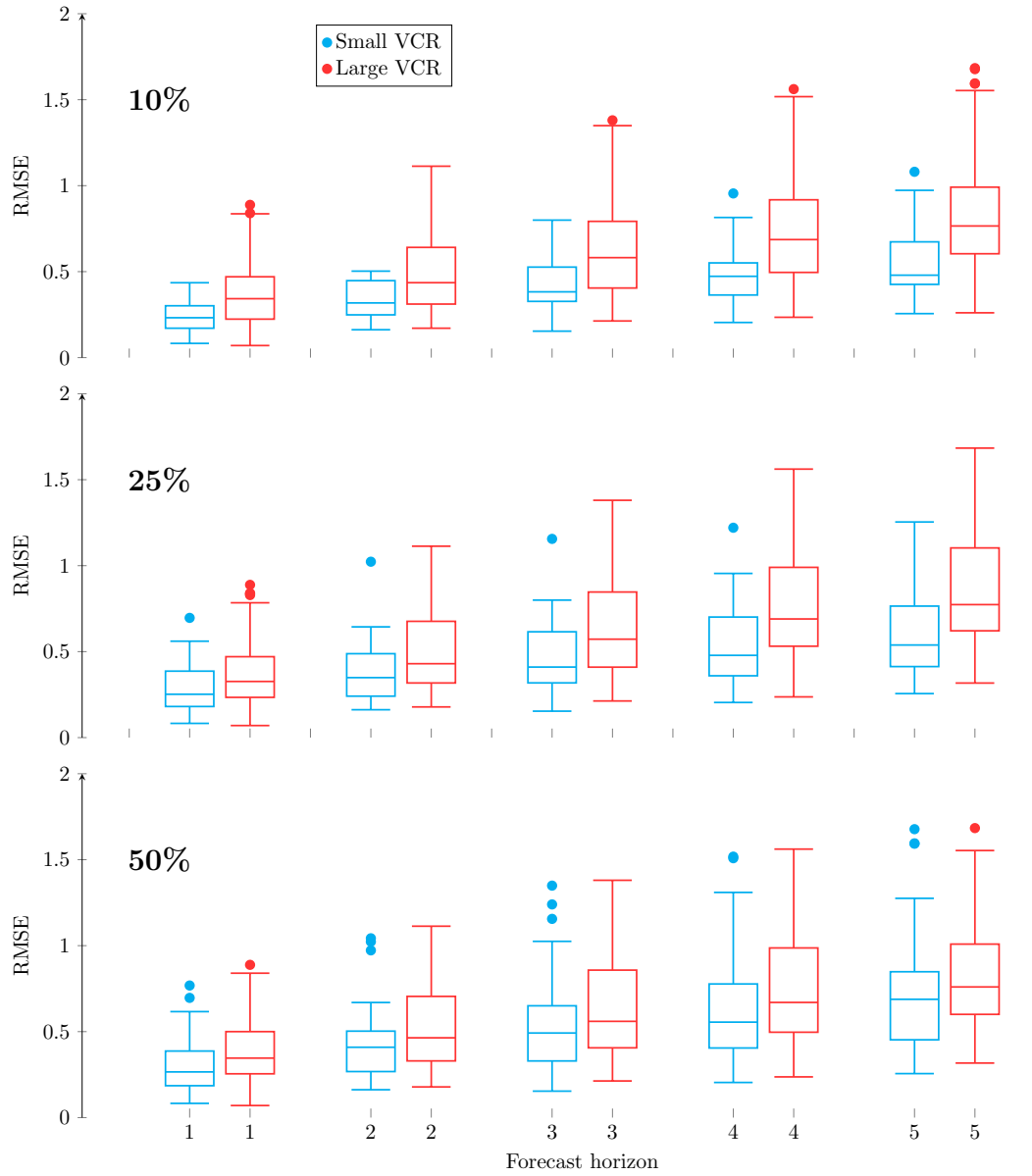
---

Months Ahead	RMSE at Small VCR	RMSE at Large VCR	p-value of t-test
1	0.260059	0.387251	0.006383
2	0.368196	0.511513	0.013863
3	0.451762	0.632516	0.008152
4	0.525750	0.740037	0.004753
5	0.588544	0.833653	0.002567

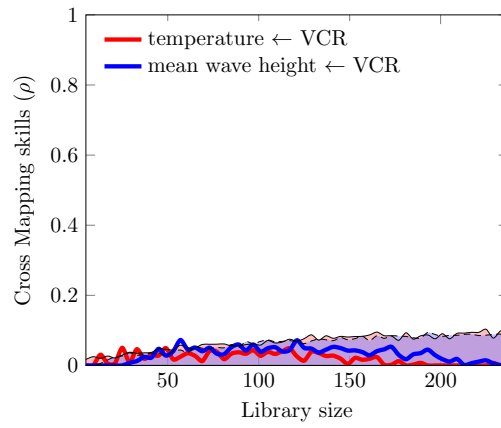
**Fig. S7 T-test on the mean RMSE.** The table shows the mean RMSE at small and large VCR for the empirical time series. The last column shows the p-value of the t-test. The table shows that the two means are indeed significantly different.



**Fig. S8 | Significance test.** The figure shows the p-value of the statistical significance of the VCR as a measure of predictability (Figure 3B) as a function of the forecast horizon at different percentiles dividing small and large VCRs. Using (up to) the 30% percentile of the distribution of the VCR as separation criterion between points with small and large VCR, the VCR is a statistically significant measure of predictability.



**Fig. S9 | Distribution of the prediction error.** The figure shows the distribution of the prediction error (RMSE) in the test set at different percentiles separating what we consider to be a small and large VCRs. Similarly to Figure 3C, the two upper panels show the difference in both mean and shape of the distributions of the error in the lower and upper percentiles of the VCR. Notice that this difference would be lost if we were using a naive 50% as separation criterion between points at small and large VCRs (lower panel).



**Fig. S10 | Validity of the causality test.** Here we show that the causality test provides meaningful results (results that match intuitive expectations). The figure shows that, as naturally expected, the VCR has no causal effect on either temperature nor mean wave height, while the opposite is true (see Figure 3D in main text).