

# Investigating Echo-State Networks Dynamics by Means of Recurrence Analysis

Filippo Maria Bianchi, Lorenzo Livi, *Member, IEEE*, and Cesare Alippi, *Fellow, IEEE*

**Abstract**—In this paper, we elaborate over the well-known interpretability issue in echo-state networks (ESNs). The idea is to investigate the dynamics of reservoir neurons with time-series analysis techniques developed in complex systems research. Notably, we analyze time series of neuron activations with recurrence plots (RPs) and recurrence quantification analysis (RQA), which permit to visualize and characterize high-dimensional dynamical systems. We show that this approach is useful in a number of ways. First, the 2-D representation offered by RPs provides a visualization of the high-dimensional reservoir dynamics. Our results suggest that, if the network is stable, reservoir and input generate similar line patterns in the respective RPs. Conversely, as the ESN becomes unstable, the patterns in the RP of the reservoir change. As a second result, we show that an RQA measure, called  $L_{\max}$ , is highly correlated with the well-established maximal local Lyapunov exponent. This suggests that complexity measures based on RP diagonal lines distribution can quantify network stability. Finally, our analysis shows that all RQA measures fluctuate on the proximity of the so-called edge of stability, where an ESN typically achieves maximum computational capability. We leverage on this property to determine the edge of stability and show that our criterion is more accurate than two well-known counterparts, both based on the Jacobian matrix of the reservoir. Therefore, we claim that RPs and RQA-based analyses are valuable tools to design an ESN, given a specific problem.

**Index Terms**—Dynamics, echo-state network (ESN), recurrence plot (RP), recurrence quantification analysis (RQA), nonlinear time series analysis.

## I. INTRODUCTION

SINCE the very first recurrent neural network (RNN) architectures, attempts have been made to describe and understand the internal dynamics of the system (see [1] and references therein). Nowadays, such efforts found renewed interest by those researchers trying to “open the black box” [2]–[5]. This comes natural as recent advances in various fields, such as neurosciences and biophysical systems modeling, demand understanding of the inner mechanism that drives the inductive inference to produce novel scientific results [6].

Manuscript received April 26, 2016; revised October 9, 2016; accepted November 16, 2016.

F. M. Bianchi is with the Machine Learning Group, Department of Physics and Technology, University of Tromsø, 9019 Tromsø, Norway (e-mail: filippo.m.bianchi@uit.no).

L. Livi is with the Department of Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, U.K. (e-mail: l.livi@exeter.ac.uk).

C. Alippi is with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133 Milan, Italy, and also with the Faculty of Informatics, Università della Svizzera Italiana, 6900 Lugano, Switzerland (e-mail: cesare.alippi@polimi.it).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2016.2630802

Reservoir computing is a class of state-space models characterized by a fixed state transition structure, the *reservoir*, and a trainable memoryless *readout* layer [7]–[9]. A reservoir must be sufficiently complex to capture all salient features of the inputs, behaving as a time-dependent, nonlinear kernel function, which maps the inputs into a higher dimensional space. The reservoir is typically generated with a pseudorandom procedure, not rarely driven by a set of rules-of-thumb and a trial-and-error approaches. A popular reservoir-computing architecture is the echo-state network (ESN) [10], an RNN with a nontrainable, sparse recurrent reservoir, and an adaptable (usually) linear readout, mapping the reservoir to the output. ESN reservoir characterization and design attracted significant research efforts in the last decade [11]–[13]. This is mostly due to the puzzling behavior of the reservoir, which, although randomly initialized, has shown to be effective in modeling nonlinear dynamical systems of various nature [14]–[18]. This also motivated researchers to better understand, and hence control, the dynamics of reservoirs when they are driven by input signals. The many existing approaches to reservoir design can be roughly partitioned as either supervised or unsupervised [8]. In the former case, the output of the system is considered to design the reservoir and tune the hyperparameters of the network. In unsupervised approaches, instead, the reservoir is controlled by considering algebraic/topological properties of the weight matrix [19]–[21] or criteria based on statistics of the neuron activations [22], [23]. Recently [24], [25], it was shown that topologies designed by the following deterministic criteria produced the state-of-the-art results in all major benchmarks, while having a more contractive dynamics with respect to randomly generated reservoirs. Another interesting reservoir design was proposed in [26]. The authors showed that, for certain tasks, a large (randomly connected) reservoir could be replaced by a single nonlinear neuron with delayed feedback. It is worth citing that also the nature of the neuron activation function (spiking versus analog) has been studied in this context [27], suggesting that networks operating with spiking neurons achieve inferior performance at the current state of research, possibly due to the adopted complex pre and postprocessing strategies. Input-dependent measures to characterize the reservoir dynamics have been proposed in [28], which considered the temporal profile of the Jacobian of the reservoir, rather than static quantities such as the spectral radius. Another method for describing the dynamic profile of the reservoir is proposed in [29]. The authors demonstrated that both transfer entropy (the predictive information provided by a source about a destination that was not already contained

in the destination history) and active information storage (the amount of past information that is relevant to predict the next state) calculated on the reservoir neurons are maximized right on the transition to an unstable regime.

The use of Poincaré recurrence of states provides fundamental information for the analysis of autonomous dynamical systems [30]. This follows from Poincaré's theorem, which guarantees that the states of a dynamic system must recur during its evolution. In other terms, the system trajectory in phase space must return in an arbitrarily small neighborhood of any of the previously visited states. Recurrences contain all relevant information regarding a system behavior in phase space and can be linked also with dynamical invariants (e.g., metric entropy) and features related to stability. However, especially for high-dimensional complex systems, the recurrence time, which is the time elapsed between recurring states, is difficult to calculate even when assuming full analytical knowledge of the system.

Recurrence plots (RPs) [30]–[33], together with the computation of dynamical invariants and heuristic complexity measures called recurrence quantification analysis (RQA), offer a simple yet effective tool to analyze such recurrences starting from a time series derived from the system under analysis. An RP is a visual representation of recurrence time, which also provides information about the duration of the recurrence by inspecting line patterns [34]. RPs are constructed by considering a suitable distance in phase (equivalently, state) space and a threshold used to determine the recurrence/similarity of states during the evolution of the system. Recently, RPs have been extended to study heterogeneous recurrences [35], i.e., qualitatively different types of recurrences in phase space. It is worth citing also the rapidly developing field of recurrence networks [36], where complex network methods are used to analyze time series.

In this paper, we address the interpretability issue of an ESN by analyzing the dynamics of the reservoir neuron activations with RPs and RQA complexity measures. To the best of our knowledge, recurrence analysis has never been applied in reservoir computing. RPs and RQA-based techniques allow the designer to visualize and characterize (high-dimensional) dynamical systems starting from a matrix encoding the recurrences of the system states over time. We show that RPs and RQA-based analyses allow deducing important and consistent conclusions about the behavior of the network. Therefore, we suggest that they can be used as a valuable analysis tool to design a network for the problem at hand. The novelty content of what proposed can be summarized as follows.

- 1) We offer a method to visualize high-dimensional dynamics of input-driven ESNs through a 2-D representation encoding recurrences of states. Our results show that, if the network is stable, line patterns of RPs related to reservoir dynamics and those of the input signal are similar. This relationship is lost as soon as the network becomes unstable.
- 2) We comment that the degree of instability of the network can be quantified by means of a measure derived from the RP. This measure is named  $L_{\max}$  and is based only on the diagonal lines of the RP (hence, it is

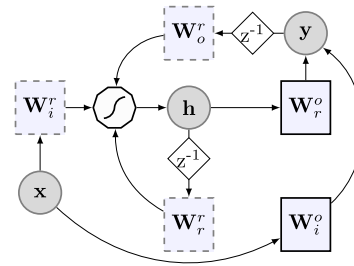


Fig. 1. ESN architecture. Circles represent input  $\mathbf{x}$ , state  $\mathbf{h}$ , and output  $\mathbf{y}$ . Solid squares  $\mathbf{W}_r^o$  and  $\mathbf{W}_i^o$  are the trainable matrices of the readout, while dashed squares  $\mathbf{W}_r^r$ ,  $\mathbf{W}_o^r$  and  $\mathbf{W}_i^r$  are randomly initialized matrices. The polygon represents the nonlinear transformation performed by neurons and  $z^{-1}$  is the unit delay.

computationally cost-effective). To support this claim, we show that  $L_{\max}$  is correlated with the well-established maximal local Lyapunov exponent over the entire range of the ESN control parameters.

- 3) We show that RQA measures can help the application designer to tune important parameters, such as spectral radius and input scaling, in order to push the ESN in a critical state, where computational capability (defined in terms of prediction accuracy and short-term memory capacity) is maximized. Such a critical state, called edge of stability, depends on the dynamics of the network that and, in turn, is influenced by the input signal. Therefore, this cannot be achieved by considering only static measures, e.g., considering only the reservoir topology. The proposed method based on RQA measures, instead, is sensitive also to the nature of the input.

This paper is structured as follows. In Section II, we introduce a typical ESN model and discuss some important figures of merit that have been proposed in the literature to characterize a reservoir. In Section III, we present the contribution of this paper, i.e., we link recurrence analysis with ESNs and discuss how this new tool can be used in practice to design a network, exposing both stability and forecasting accuracy properties. In Section IV, we discuss experimental results that support our claims. Finally, Section V offers concluding remarks and future research directions.

## II. DYNAMICS OF THE ESN RESERVOIR

A schematic of an ESN is shown in Fig. 1. An ESN consists of a reservoir of  $N_r$  nodes characterized by a nonlinear transfer function  $f(\cdot)$ . At time instant  $k$ , the network is driven by the input  $\mathbf{x}[k] \in \mathbb{R}^{N_i}$  and produces the output  $\mathbf{y}[k] \in \mathbb{R}^{N_o}$ , being  $N_i$  and  $N_o$  the dimensionalities of input and output, respectively. The weight matrices  $\mathbf{W}_r^r \in \mathbb{R}^{N_r \times N_r}$  (reservoir internal connections),  $\mathbf{W}_i^r \in \mathbb{R}^{N_i \times N_r}$  (input-to-reservoir connections), and  $\mathbf{W}_o^r \in \mathbb{R}^{N_o \times N_r}$  (output-to-reservoir feedback connections) contain values in the  $[-1, 1]$  interval drawn from a uniform distribution; a Gaussian distribution is another common choice.

An ESN is a discrete-time nonlinear system with feedback, whose model reads

$$\mathbf{h}[k] = f(\mathbf{W}_r^r \mathbf{h}[k-1] + \mathbf{W}_i^r \mathbf{x}[k] + \mathbf{W}_o^r \mathbf{y}[k-1]) \quad (1)$$

$$\mathbf{y}[k] = g(\mathbf{W}_r^o \mathbf{h}[k] + \mathbf{W}_i^o \mathbf{x}[k]). \quad (2)$$

Activation functions  $f(\cdot)$  and  $g(\cdot)$ , both applied componentwise, are typically implemented as a sigmoidal (*tanh*) and identity function, respectively. The output weight matrices  $\mathbf{W}_r^o \in \mathbb{R}^{N_r \times N_o}$  and  $\mathbf{W}_i^o \in \mathbb{R}^{N_i \times N_o}$ , which connect, respectively, reservoir and input to the output, represent the readout of the network. The standard training procedure for such matrices requires solving a straightforward regularized least-square problem [37].

Even though the three weight matrices  $\mathbf{W}_r^r$ ,  $\mathbf{W}_o^r$ , and  $\mathbf{W}_i^r$  are generated randomly, they can be suitably designed and scaled to obtain desired properties. For instance,  $\mathbf{W}_r^o$  can be scaled with a multiplicative constant  $\omega_o$ ; in this paper, we set  $\omega_o = 0$ , which has the effect of removing the output feedback connection.  $\mathbf{W}_i^r$  is controlled by the scalar parameter  $\omega_i$ . Since the gain of the sigmoid nonlinearity in the neurons is the largest around the origin, the scaling coefficient  $\omega_i$  of  $\mathbf{W}_i^r$  determines the amount of nonlinearity introduced by the processing units. In particular, inputs far from zero tend to drive the activation of the neurons toward saturation where they show more nonlinearity. Finally, the spectral radius of  $\mathbf{W}_r^r$ , denoted as  $\rho(\mathbf{W}_r^r)$  or  $\rho$ , controls important properties as discussed in the sequel.

An ESN is typically designed, so that the influence of past inputs on the network gradually fades away and the initial state of the reservoir is eventually washed out. This is granted by the echo-state property (ESP), which ensures that, given any input sequence taken from a compact set, future trajectories of any two different initial states become indistinguishable. ESP was originally investigated in [37] and successively in [38]; we refer the interested reader to [39] for a more recent definition, where also the influence of input is explicitly accounted for. In ESNs with no output feedback, as in our case, the state update of (1) reduces to

$$\mathbf{h}[k] = f(\mathbf{W}_r^r \mathbf{h}[k-1] + \mathbf{W}_i^r \mathbf{x}[k]). \quad (3)$$

In order to study the stability of the network, we analyze the Jacobian of the state update (3) of the reservoir and generate a derived measure, the maximal local Lyapunov exponent ( $\lambda$ ). Such a quantity is used to approximate (for an autonomous system) the separation rate in phase space of trajectories having very similar initial conditions.  $\lambda$  was proposed to characterize a reservoir, and demonstrated its efficacy in designing a suitable network configuration in several applications [7], [28], [40].  $\lambda$  is calculated by considering the Jacobian at time  $k$ , which, if neurons are implemented with a *tanh* activation function, can be conveniently expressed as

$$\mathbf{J}(h[k]) = \begin{bmatrix} 1 - (h_1[k])^2 & 0 & \dots & 0 \\ 0 & 1 - (h_2[k])^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 - (h_{N_r}[k])^2 \end{bmatrix} \mathbf{W}_r^r \quad (4)$$

where  $h_l[k]$  is the activation of the  $l$ th neuron, with  $l = 1, 2, \dots, N_r$ .  $\lambda$  is then computed as

$$\lambda = \max_{n=1, \dots, N_r} \frac{1}{K} \sum_{k=1}^K \log(r_n[k]) \quad (5)$$

where  $r_n[k]$  is the absolute value of the  $n$ th eigenvalue of  $\mathbf{J}(h[k])$  and  $K$  is the total number of time steps in the considered trajectory. Typically, the stable–unstable transition is detected numerically by considering the sign of  $\lambda$  (5). In autonomous systems,  $\lambda > 0$  indicates that the dynamics is chaotic.

Local, first-order approximations provided by (4) are useful also for studying the stability of a (simplified) reservoir operating around the zero state,  $\mathbf{0}$ . In fact, implementing  $f(\cdot)$  as a *tanh* assures  $f(\mathbf{0}) = \mathbf{0}$ , i.e.,  $\mathbf{0}$  is a fixed point of the ESN dynamics. Therefore, by linearizing (3) around  $\mathbf{0}$  and assuming a zero input, we obtain from (4)

$$\mathbf{h}[k] = \mathbf{J}(\mathbf{0})\mathbf{h}[k-1] = \mathbf{W}_r^r \mathbf{h}[k-1]. \quad (6)$$

Linear stability analysis of (6) suggests that, if  $\rho(\mathbf{W}_r^r) < 1$ , then the dynamics around  $\mathbf{0}$  is stable. In the more general case, the nonlinearity of the sigmoid functions in (3) forces the norm of the state vector of the reservoir to remain bounded. Therefore, the condition  $\rho(\mathbf{W}_r^r) < 1$  loses its significance and does not guarantee stability when the system deviates from a small region around  $\mathbf{0}$  [40]. This means that, actually, it is possible to find reservoirs (3) having  $\rho(\mathbf{W}_r^r) > 1$  while still possessing the ESP with stability. In fact, the effective local gain decreases when the operating point of the neurons shifts toward the positive/negative branch of the sigmoid, where stabilizing saturation effects start to influence the excitability of reservoir dynamics [38]. In the more realistic and useful scenario where the input driving the network is a generic (nonzero) signal, a sufficient condition for the ESP is met if  $\mathbf{W}_r^r$  is diagonally Schur-stable, i.e., if there exists a positive-definite diagonal matrix,  $\mathbf{P}$ , such that  $(\mathbf{W}_r^r)^T \mathbf{P} \mathbf{W}_r^r - \mathbf{P}$  is negative-definite [38]. However, this recipe is fairly restrictive in practice. In many cases, such a sufficient condition might generate reservoirs that are not rich enough in terms of provided dynamics, since the use of a conservative scaling factor might compromise the amount of memory in the network, and thus the ability to accurately model a given problem. Therefore, for most practical purposes, the necessary condition  $\rho(\mathbf{W}_r^r) < 1$  is considered “sufficient in practice,” because if the spectral radius is less than 1, the state update map is contractive with high probability, regardless of the input and given a sufficiently large reservoir [21].

The number of reservoir neurons and the bounds on  $\rho$  can be used for a Naïve quantification of the computational capability of a reservoir [38]. However, those are static measures that only consider the algebraic properties of  $\mathbf{W}_r^r$ , without considering other factors, such as the input scaling  $\omega_i$  and the particular properties of the given input signals. Moreover, it is still not clear how, in a mathematical sense, these stability bounds relate to the actual ESN dynamics when processing nontrivial input signals [39]. In this context, the idea of pushing the system toward the so-called “edge of

stability” (also called edge of criticality) has been explored. Langton [41], Bertschinger and Natschläger [42], and Legenstein and Maass [43] show that several dynamical systems, among which randomly connected RNNs, achieved the highest computational capabilities when moving toward the unstable (sometime even chaotic) regime, where the ESP is lost and the system enters into an oscillatory behavior. This justifies the use of spectral radii above the unity in some practical applications. It is important to remark that, although the critical state might maximize the computational capability, there are also tasks that either require very little in terms of computation or reservoirs with fast reaction times and fast locking into attractor states (e.g., multistable switching circuits [44]).

Further descriptors used for characterizing the dynamics of a reservoir are based on information-theoretic quantities, such as the (average) transfer entropy and the active information storage one [29]. The authors have shown that such quantities peak right when  $\lambda > 0$ . In addition, also the minimal singular value of the Jacobian (4), denoted as  $\eta$ , was demonstrated to be an accurate predictor of ESN performance, providing more accurate information regarding the ESN dynamics than both  $\lambda$  and  $\rho$  [28]. Finding hyperparameters that maximize  $\eta$  results in a dynamical system that is far from singularity, it has many degrees of freedom, a good excitability and separates well the input signals in phase space [28].

Complementary to  $\lambda$ ,  $\eta$ , and the aforementioned information-theoretic descriptors, in this paper, we propose to use RPs and the related RQA measures for characterizing the dynamics of an ESN.

### III. ANALYZING ESN DYNAMICS BY INVESTIGATING RECURRENCES OF NEURON ACTIVATIONS

Here, we discuss how the input-driven dynamics of an ESN reservoir (3) can be fruitfully analyzed by means of RP-based techniques.

The sequence of ESN states can be seen as a multivariate time series  $\mathbf{h}$ , consisting of  $N_r$  state variables coming from the reservoir neuron activations. An RP is constructed by calculating a  $K \times K$  binary matrix  $\mathbf{R}$ . The generic element  $R_{ij}$  is defined as

$$R_{ij} = \Theta(\tau_{RP} - d(\mathbf{h}[i], \mathbf{h}[j])), \quad 1 \leq i, j \leq K \quad (7)$$

where  $K$  is the length of the time series,  $d(\cdot, \cdot)$  is a dissimilarity measure operating in phase space, and  $\Theta(\cdot)$  the Heaviside function:  $\Theta(x) = 0$  if  $x < 0$ ;  $\Theta(x) = 1$  otherwise.  $\tau_{RP} > 0$  is a user-defined threshold used to identify recurrences.  $\tau_{RP}$  can be defined in different ways, but typically, it represents a percentage of the average or the maximum phase space distance between the states. The selection of an “optimal” threshold  $\tau_{RP}$  is a problem-dependent issue. However, we comment that, usually,  $\tau_{RP}$  has an impact only on the quantitative information derived from an RP but does not affect the general properties of the system as seen through recurrence analysis.  $\mathbf{R}$  is constructed by considering also a nonnegative dissimilarity measure  $d(\cdot, \cdot)$ , which evaluates the distance between states. Typical examples include the Euclidean, Manhattan, or max-norm distances. Several options have been exploited in the

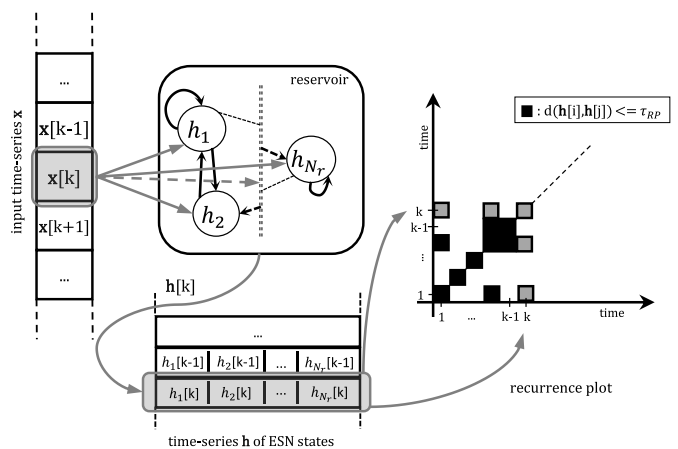


Fig. 2. When  $\mathbf{x}[k]$  is fed as input to the  $N_r$  neurons of the ESN reservoir, the internal state is updated to  $\mathbf{h}[k] = [h_1[k], h_2[k], \dots, h_{N_r}[k]]^T$ , where  $h_n[k]$  is the output of the  $n$ th neuron at time  $t$ . Once the time series  $\mathbf{h}$  is generated, the RP is constructed by using a threshold  $\tau_{RP}$  and a dissimilarity measure  $d(\cdot, \cdot)$ . If  $d(\mathbf{h}[k], \mathbf{h}[i]) \leq \tau_{RP}$ , the cell of the RP in position  $(k, i)$  is colored in black, otherwise it is left white. The elements in gray highlight the operations performed at time step  $k$ .

literature for  $\tau_{RP}$  and  $d(\cdot, \cdot)$ ; we refer the reader to [30] for further detailed discussions. Fig. 2 shows the algorithmic steps required to generate an RP on ESN states.

Depending on the properties of the analyzed time series, different line patterns emerge in an RP [34]. Besides providing an immediate visualization of the system properties, starting from  $\mathbf{R}$ , it is possible to compute also several complexity measures, those associated with RQA. Such measures are defined by the distribution of both vertical/horizontal and diagonal line structures present in the RP and provide a numerical characterization of the properties associated with the underlying dynamics.

In Section III-A, we provide details about the RQA measures considered in this paper. In Section III-B, we discuss how a network designer can address the two important issues of stability and unsupervised maximization of computational capability through RQA. Finally, in Section III-C, we explain how to interpret the RP and related RQA measures relative to ESN state sequences driven by different classes of inputs.

#### A. RQA Complexity Measures

We assume here to have generated  $\mathbf{R}$  according to (7) on a time series of length  $K$ . Many of the following measures are based on the histograms  $P(l)$  and  $P(v)$ , counting, respectively, the number of diagonal and vertical lines having specific lengths  $l$  and  $v$ :

$$P(l) = \sum_{i,j=1}^{K-l} (1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k}$$

$$P(v) = \sum_{i,j=1}^{K-v} (1 - R_{i,j})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k}$$

Such RQA measures capture different aspects of the dynamics (e.g., features related to the stability and complexity of the system evolution). Abbreviations and notation are kept consistent with [30].

**RR:** The recurrence rate is a measure of the density of recurrences in  $\mathbf{R}$

$$RR = \frac{1}{K^2} \sum_{i,j=1}^K R_{ij}. \quad (8)$$

RR corresponds to the correlation sum [45]. RR could also guide the selection of  $\tau_{RP}$ , e.g., if ones wants to preserve the recurrence rate when performing tests in different conditions.

**DET:** A measure of the determinism level of the system based on the percentage of diagonal lines of minimum length  $l_{\min}$  is defined as

$$DET = \frac{\sum_{l=l_{\min}}^K lP(l)}{\sum_{l=1}^K lP(l)} \in [0, 1]. \quad (9)$$

DET is close to unity in a periodic system and close to zero in systems with no time-dependence. The threshold  $l_{\min}$  can be used also to discriminate different forms of determinism, such as periodic and chaotic motions. In fact, it is known that, due to exponential divergence in phase space, diagonal lines are very short for chaotic systems.

**$L_{\max}$ .** Maximum diagonal line length

$$L_{\max} = \max\{l_i\}_{i=1}^{N_l}, \quad N_l = \sum_{l \geq l_{\min}} P(l) \quad (10)$$

where  $l_i$  is the length of the  $i$ th diagonal line,  $N_l$  is the total number of diagonal lines, and  $1 \leq L_{\max} \leq \sqrt{2}K$ . This measure is related to the mean exponential divergence in phase space. A measure of divergence can be obtained as

$$DIV = 1/L_{\max} \in (0, 1]. \quad (11)$$

As formally discussed in [30], it is possible to relate  $L_{\max}$  with the correlation entropy of the system—a dynamical invariant measure. Notably, correlation entropy can be estimated by using the diagonal lines distribution. The correlation entropy is a lower bound for the sum of the positive Lyapunov exponents, thus providing a formal connection with the analysis of sensitivity on initial conditions.

**LAM:** Laminarity is a descriptor of the presence of laminar phases in the system

$$LAM = \frac{\sum_{v=v_{\min}}^K vP(v)}{\sum_{v=1}^K vP(v)} \in [0, 1] \quad (12)$$

where  $v_{\min}$  is a threshold for the minimal vertical line length to be considered. Laminar phases denote the states of the system that do not change or change very slowly for a number of consecutive time steps.

**ENTR:** This quantity measures the complexity of an RP with respect to the diagonal lines distribution

$$ENTR = - \sum_{l=l_{\min}}^K p(l) \ln(p(l)) \quad (13)$$

where  $p(l) = P(l)/N_l$ . As usual,  $0 \leq ENTR \leq \log(K)$ . Signals with no time dependence present  $ENTR \simeq 0$ , i.e., the diagonal lines distribution is fully concentrated on very short lines (e.g., single dots). Conversely, such a measure assumes high values when the diagonal line distribution becomes heterogeneous.

**SWRP:** Entropy of a weighted RP. Such a measure has been recently introduced in [32] and provides an alternative to ENTR (13). It quantifies the complexity of scalar distributions of *strengths* for each time step, providing a solution for the border effects that might appear in the computation of ENTR. The computation of SWRP considers the similarity matrix  $\tilde{\mathbf{S}}$ , whose elements are defined as  $\tilde{S}_{ij} = \exp(-d(\mathbf{h}[i], \mathbf{h}[j]))$ , where  $d(\cdot, \cdot)$  is the same as in (7). The results are not influenced by  $\tau_{RP}$  and do not depend directly on the length of the time series. The strengths are used to measure density heterogeneity of a given point in phase space. The  $i$ th strength is defined as  $s_i = \sum_{j=1}^K \tilde{S}_{ij}$ . The distribution of the strengths is represented by histograms with a number  $B$  of bins defined by the user. SWRP measure reads as

$$SWRP = - \sum_{b=1}^B p(b) \ln p(b) \quad (14)$$

where  $P(b)$  is the number of elements in the  $b$ th bin and  $p(b) = P(b)/K$  is the respective probability; as for the other entropy measures,  $0 \leq SWRP \leq \log(B)$ .

## B. Design of a Stable and Effective Network

In the sequel, we discuss how recurrence analysis on the RP built from  $\mathbf{h}$ , can be used to: 1) assess network stability for a given configuration and 2) tune network hyperparameters, so that high computational capability can be achieved.

**Stability Issue:** As mentioned before, the degree of stability can be measured by  $L_{\max}$  (10): the higher  $L_{\max}$ , the more stable the system. In addition to  $L_{\max}$ , the designer can have an immediate visual interpretation of stability by inspecting RP: short and erratic diagonal lines in RPs denote instability/chaoticity, while long diagonal lines denote regularity (e.g., a periodic motion).  $L_{\max}$  is anticorrelated with  $\lambda$ , and hence, it can be considered as a reliable indicator for the (input-dependent) degree of network stability. This is confirmed by our experiments, where we highlight also that, if the network is stable, the dynamics of input and reservoir produce compatible line patterns in the respective RPs. Instead, if the network is unstable, this similarity is lost. Therefore, RPs can be used by the designer as visual tools to analyze the response of the network to a specific input.

**Computational Capability Issue:** A widespread criterion to determine the (signal-dependent) edge of stability relies on the sign of  $\lambda$ , which becomes positive when the system enters into a (globally) unstable regime. We propose, instead, to use information derived from RR (8), DET (9), LAM (12), ENTR (13), and SWRP (14) to determine the edge of stability. In fact, we observe that, when such indices start to fluctuate, the network achieves high prediction accuracy. Moving an RNN to the edge of stability provides a guideline to the network designer for setting critical hyperparameters, such as  $\rho$  and  $\omega_i$ , in the presence of input. On the other hand, unstable ESNs initialized with different weight matrices, yet configured with the same  $\rho$  and  $\omega_i$ , yield very different RQA values. This means that, in unstable regimes, RQA measures are very sensitive to different network initializations. Accordingly, we identify the edge of stability as the configurations

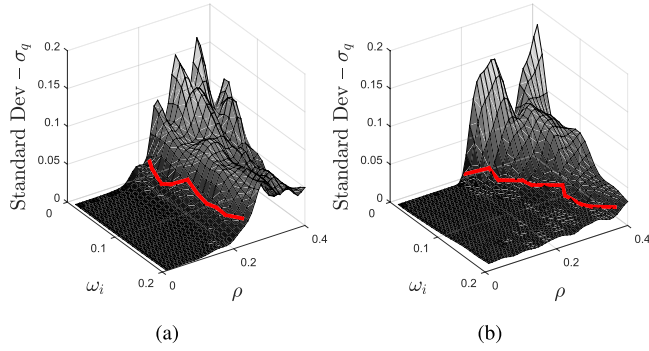


Fig. 3. Standard deviations of an RQA measure  $q$  for different values of  $\rho$  and  $\omega_i$  in two ESNs driven by different input signals. The edge of stability is identified in correspondence of the parameter configurations (shown as a red line) for which the standard deviation  $\sigma_q(\rho, \omega_i)$  increases abruptly, assuming values greater than its mean value  $\bar{\sigma}_q$ .

of  $\rho$  and  $\omega_i$ , which yield a sudden increment in the standard deviation of the RQA measures. Let  $P$  and  $\Omega$  be the domains for  $\rho$  and  $\omega_i$ , respectively. For a given RQA measure  $q$ ,  $\bar{\sigma}_q$  is the average value of its standard deviation computed over  $P \times \Omega$ . We define the edge of stability (relative to  $q$ ) as a set of pairs  $\{(\rho^1, \omega_i^1), (\rho^2, \omega_i^2), \dots, (\rho^{|\Omega|}, \omega_i^{|\Omega|})\} \subset P \times \Omega$ , for which  $q$  has a standard deviation greater than  $\bar{\sigma}_q$ . Specifically, for every  $\omega_i^j \in \Omega$ , we select the largest  $\rho^j \in P$  such that

$$\sigma_q(\rho^j, \omega_i^j) \leq \bar{\sigma}_q \quad (15)$$

is satisfied.  $\sigma_q(\rho^j, \omega_i^j)$  is the standard deviation of  $q$ , obtained from different randomly initialized ESNs configured with  $(\rho^j, \omega_i^j)$ . Fig. 3 shows an illustrative example. We underline that other criteria based on fluctuations of RQA measures could be conceived. However, as we will show in the experiments, such a criterion offers a more accurate description for the edge of stability than that obtainable with  $\lambda$  and  $\eta$ .

### C. Visualization and Classification of Reservoir Dynamics

In the following, we discuss how RPs permit to visualize, and hence classify, the reservoir dynamics when the ESN is fed with inputs possessing specific characteristics. In doing so, we assume given a stable ESN described by (3); RPs are constructed following the procedure depicted in Fig. 2. Although there exists many different types of signals/systems (with related sub-classes) [30], here we focus on the ability to discriminate between five important classes for the input signals: 1) with/without time-dependence; 2) periodic/nonperiodic motions; 3) laminar behaviors; 4) chaotic dynamics; and 5) nonstationary processes.

We refer to the examples in Fig. 4 to discuss the characteristics of an RP, when the ESN is fed by a signal belonging to the aforementioned classes. It is important to underline that RPs offer visual information and the validity of the following comments, although based on phenomenological observations, is general and not application-specific. In addition, we stress that RP-based analyses are applicable also to short time series (with tens/hundreds of records).

*Time Dependence:* A uniformly distributed RP denotes absence of a time-dependence in the time series

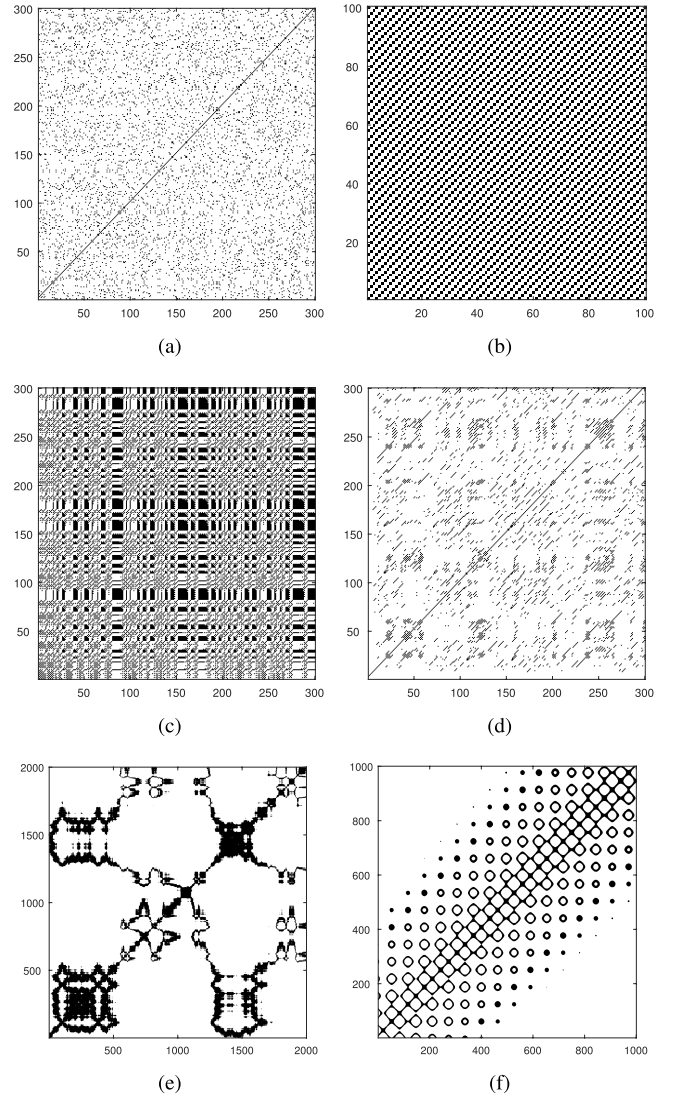


Fig. 4. RPs generated by state sequences  $\mathbf{h}$  of ESN fed with input signals associated with the considered classes. Both axes represent time. (a) Gaussian white noise;  $\tau_{RP} = 0.4$ . (b) Periodic;  $\tau_{RP} = 0.2$ . (c) LM: laminar states;  $\tau_{RP} = 0.5$ . (d) LM: chaos;  $\tau_{RP} = 0.2$ . (e) Brownian motion;  $\tau_{RP} = 0.2$ . (f) Drift;  $\tau_{RP} = 0.2$ .

(e.g., an uncorrelated signal). It is possible to rely on specific RQA measures, such as DET (9), ENTR (13), and SWRP (14), to numerically quantify the amount of time dependence. All three indices would yield very low values (close to zero) when there is no time-dependence in the signal. Instead, in a periodic signal characterized by a strong time dependence, DET would be very high, while ENTR and SWRP would still be low. In fact, ENTR and SWRP are measures of complexity, which is low in signals without a temporal structure. As an example, in Fig. 4(a), we examine the RP generated by feeding the ESN with Gaussian noise, a typical example of signal with no time dependence. We observe a uniform RP for the reservoir states, which is peculiar for all signals composed by realizations of statistically independent variables.

*Periodicity:* Periodic systems induce long diagonal lines and the vertical spacing provides the characteristic period of the oscillation. A periodic system is characterized by high

values for DET (9) and  $L_{\max}$  (10). In addition, as stressed before, it has low complexity as expressed by low values of ENTR and SWRP. In Fig. 4(b), we show an example of a periodic motion generated by a sinusoid having a single dominating frequency. The regularity of the diagonal lines can be immediately recognized from the figure.

*Laminarity:* A system presents laminar phases if its state does not change or varies very slowly over a number of successive time steps. Laminar phases can be visually recognized in an RP by the presence of (fairly) large black rectangles. Every system possessing laminar phases is characterized by high values for LAM (12). To provide an example (but results are general and apply also to other systems), we consider the logistic map (LM)

$$\mathbf{x}[n+1] = \tau_{\text{LM}}\mathbf{x}[n](1 - \mathbf{x}[n]) \quad (16)$$

where usually  $\tau_{\text{LM}} \in (0, 4]$ ; here, LM is configured with initial condition  $\mathbf{x}[0] = 0.5$ . In Fig. 4(c), we show the RP for  $\tau_{\text{LM}} = 3.679$ . The system exhibits chaos–chaos transitions in that configuration. In fact, the related RP is compatible with the one of a (mildly) chaotic system, showing also the presence of laminar phases visualized as large black rectangles.

*Chaoticity:* RPs are effective to detect chaotic dynamics, which are characterized by the presence of erratic and very short diagonal lines. As a consequence, RR (8) would be very low. The two complexity measures ENTR and SWRP are also useful to determine the degree of chaoticity: the higher their values, the more chaotic/complex the system. Chaos is characterized by trajectories diverging exponentially fast. This can be quantified with  $L_{\max}$  (10) and DIV (11), whose values would be, respectively, very low and close to one for systems with a high degree of chaoticity. As an example, we consider a chaotic system obtained by LM configured with  $\tau_{\text{LM}} = 4$ . The reservoir dynamics, as shown in the RP in Fig. 4(d), denotes fully developed chaos, as indicated by the presence of only short and erratic diagonal lines.

*Nonstationarity:* Peculiar line patterns observed for all nonstationary signals include large white areas with irregular patterns denoting abrupt changes in the dynamics. In Fig. 4(e), we provide an example obtained by feeding the ESN with Brownian motion. This latter is a nonstationary stochastic process that corresponds to a random walk, whose increments are drawn from a Gaussian distribution. Drift is another typical form of nonstationarity, which is visually recognized in an RP by the fading of recurrences in the top-left and bottom-right corners. In Fig. 4(f), we show an example of drift by adding a linear trend to a sinusoid. Nonstationarity can be numerically detected by considering an RQA measure called TREND (not considered in our study) and by analyzing the variation of RQA measures when time-delay is applied to the signal (see [30] for technical details).

#### IV. EXPERIMENTS

For the experiments, we consider two time series generated, respectively, by an oscillatory and by the Mackey–Glass (MG) dynamical system. We chose these two signals, since both of them are often considered as benchmarks for prediction in the ESN literature [7], [37], [46] and they exemplify a very

regular and a mildly chaotic system, respectively. We perform two kinds of test. In the first one—Section IV-A—we use RPs to visualize the dynamics of reservoirs when driven by a given input signal. When the reservoir operates in a stable regime, RPs of reservoir and input show similar line patterns.

In the second experiment—Section IV-B—we leverage on RQA to tune two critical hyperparameters, the spectral radius ( $\rho$ ) and input scaling ( $\omega_i$ ), in an ESN driven by input signal. We study RQA measures obtained for different settings of  $\rho$  and  $\omega_i$ , with the aim to: 1) assess the network stability and 2) determine in an unsupervised way a setting for these two hyperparameters, where the predictive power of the network is maximized. Note that we focus on the prediction of the time series generated by MG system [28], rather than on learning its chaotic attractor from data [37]. The latter task requires enabling the output feedback, which leads to internal dynamics and to a Jacobian matrix different from the ones considered in this paper.

In both experiments, we consider an ESN with no output feedback ( $\omega_o = 0$ ), configured with a standard setting: uniformly distributed weights in  $[-1, 1]$  for  $W_i^r$  and  $W_r^r$ , percentage of nonzero connections in  $W_r^r$  of 25%. The readout is trained by ridge regression with regularization parameter set to 0.1. According to the standard drop-out procedure, we discarded the first 100 elements of  $\mathbf{h}$  in order to get rid of the ESN transient states. In the first experiment, we set the number of reservoir neurons to  $N_r = 50$ . In preliminary tests, we noted that the number of reservoir neurons, even if very small (two neurons), does not affect the results obtained for the analysis of recurrences. Therefore, here we do not study the effect of the network size. We used the Manhattan distance to evaluate dissimilarities in phase space

$$d(\mathbf{h}[j], \mathbf{h}[i]) = \sum_{n=1}^{N_r} |h_n[j] - h_n[i]|. \quad (17)$$

The threshold  $\tau_{\text{RP}}$  is set proportional to the average dissimilarity value between the states in  $\mathbf{h}$ . For example,  $\tau_{\text{RP}} = 0.1$  indicates that  $\tau_{\text{RP}}$  is set to 10% of the average distance between all states. In few cases, to better visualize the RPs, we increase  $\tau_{\text{RP}}$  beyond values typically used in the literature. However, as stated before, our main aim here is to show the agreement between RQA measures and the stability–instability transition together with the corresponding prediction accuracy of the network, rather than providing an exact quantification of its characteristics. Our results are easily reproducible with the ESN<sup>1</sup> and RP<sup>2</sup> toolboxes, freely available on the Web.

##### A. Visualization of ESN Dynamics

The first signal is a sinusoid defined as  $\mathbf{x}[k] = \sin(\psi k)$ , where  $k = 1, 2, \dots, 5000$  and  $\psi = 3/50$ .

Fig. 5(a) shows the RPs relative to the input signal, which has been embedded into a 2-D phase space. In Fig. 5(b), we report the RP of the reservoir, configured with a conservative setting of  $\rho$ . The resulting RP is compatible with a periodic system having a single frequency. This suggests

<sup>1</sup><http://www.reservoir-computing.org/node/129>

<sup>2</sup><http://www.recurrence-plot.tk/>

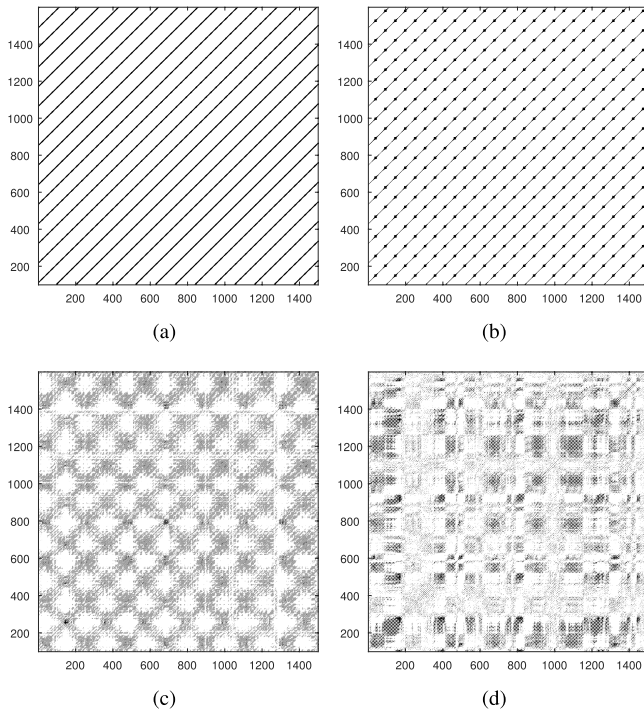


Fig. 5. RP for the input signal and sequence of states of the reservoir. When  $\rho = 0.99$ , the activations are compatible with the input dynamics. When  $\rho$  exceeds one, the activations denote instability patterns. (a) RP of the input signal. (b)  $\rho = 0.99$  and  $\tau_{RP} = 0.1$ . (c)  $\rho = 1.5$  and  $\tau_{RP} = 0.8$ . (d)  $\rho = 2$  and  $\tau_{RP} = 0.8$ .

that, although the (randomly initialized) reservoir performs a nonlinear mapping, its resulting dynamics preserves the one of the input. In fact, the vertical spacing between the diagonal lines is almost the same (98 time intervals). The small difference is due to a slight discrepancy in the period of the two systems (i.e., input and reservoir), whose cause was formally explained in [39]. There, the authors demonstrate that when an ESN is driven by a periodic signal with period  $\vartheta$ , the induced network dynamics asymptotically becomes periodic as well, with period  $r\vartheta$ ,  $r > 0$ . On the other hand, when  $\rho$  is pushed beyond unity, the RPs are comparable with the ones of unstable/chaotic systems with a degree of instability monotonically related to the value of  $\rho$ . Please note that, while in Fig. 5(a) and (b) we used  $\tau_{RP} = 0.2$  and  $\tau_{RP} = 0.1$ , respectively, in Fig. 5(c) and (d), to improve readability, we used a larger threshold,  $\tau_{RP} = 0.8$ .

The second input signal is given by a time series generated from the MG system

$$\frac{dx}{dt} = \frac{\alpha x(t - \tau_{MG})}{1 + x(t - \tau_{MG})^{10}} - \beta x(t). \quad (18)$$

We obtained a time series of 150 000 time steps using  $\tau_{MG} = 17$ ,  $\alpha = 0.2$ ,  $\beta = 0.1$ , initial condition  $x(0) = 1.2$ , and 0.1 as integration step. In Fig. 6, we report the RPs of the input signal and those related to the three different settings of the ESN with an increasing value of  $\rho$ . As in the previous case, we notice a similarity between RPs of input [Fig. 6(a)] and reservoir [Fig. 6(b)]. As the reservoir is pushed toward instability by increasing  $\rho$ , it is possible to observe the usual incremental transition toward a chaotic regime in

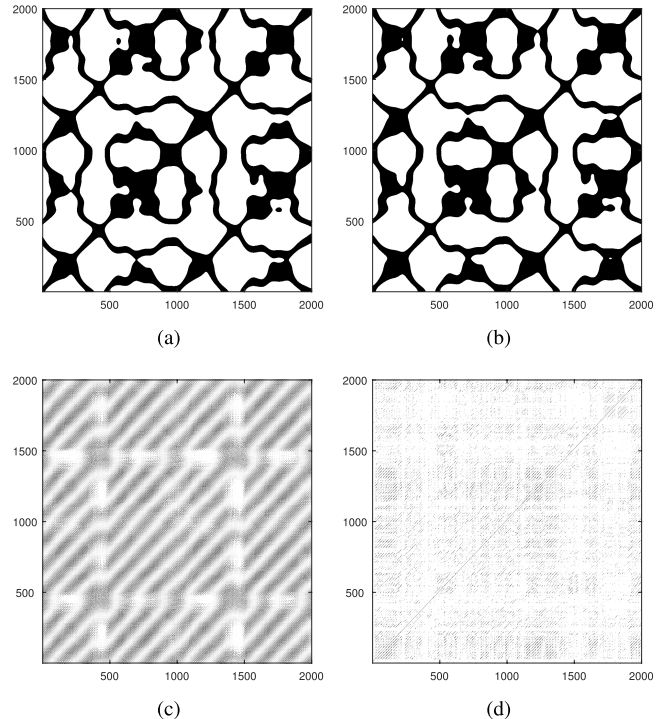


Fig. 6. RPs of the MG time series and of the reservoir states. (a) RP of the input signal. (b)  $\rho = 0.9$  and  $\tau_{RP} = 0.5$ . (c)  $\rho = 1.5$  and  $\tau_{RP} = 0.5$ . (d)  $\rho = 2$  and  $\tau_{RP} = 0.5$ .

the related RPs. In Fig. 6(c) and (d),  $\rho$  is set to 1.5 and 2, respectively. In both the cases, we observe typical line patterns of unstable systems. In particular, for  $\rho = 2$ , the system becomes fully chaotic and the RP contains very short and erratic diagonal lines [see Fig. 4(d) for a visual comparison].

### B. Characterization of Reservoirs With RQA

In this second experiment, we show how to leverage on RQA for identifying the configuration of ESN hyperparameters yielding maximum computational capability. We evaluate the accuracy of our unsupervised methodology by considering the same sinusoidal and MG time series. We compare the results with  $\lambda$  (5) and the minimal singular value of the Jacobian,  $\eta$ , whose average value is computed over time; both are widely used in the literature for identifying an optimal network setting in an unsupervised way. Specifically, we consider the regions in the hyperparameter space where  $\lambda$  crosses zero and where  $\eta$  is maximized as indicators to determine the (input-dependent) edge of stability of the network. Prediction error is measured with the normalized root mean squared error

$$\text{NRMSE} = \sqrt{\frac{\langle \|\mathbf{y} - \mathbf{d}\|^2 \rangle}{\langle \|\mathbf{y} - \langle \mathbf{d} \rangle \|^2 \rangle}} \quad (19)$$

being  $\mathbf{y}$  the ESN output (2) and  $\mathbf{d}$  the desired one. The prediction accuracy is defined as  $\gamma = \max\{0, 1 - \text{NRMSE}\}$ .

In the following, we monitor how  $\gamma$ ,  $\lambda$ ,  $\eta$ , and RQA measures vary as the values of  $\rho$  and  $\omega_i$  are changed. First, we compare  $\lambda$  with  $L_{\max}$ , the value of the longest diagonal line in an RP—see (10) and related discussion.  $L_{\max}$  is a global indicator of stability that we show here to be highly



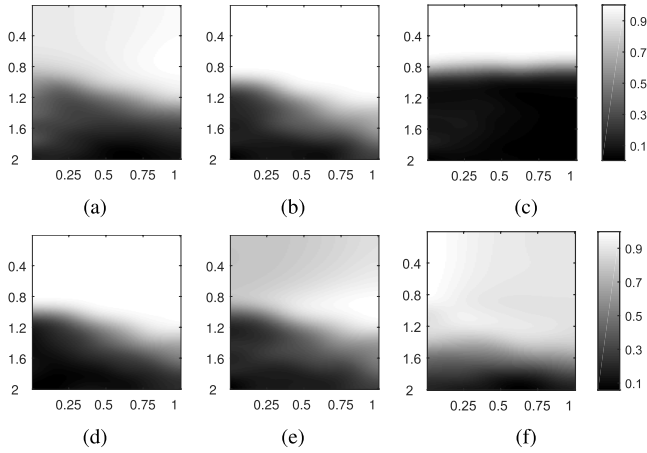


Fig. 7. RQA measures for sinusoid while changing  $\rho$  (vertical axis) and  $\omega_i$  (horizontal axis). (a) RR. (b) DET. (c)  $L_{\max}$ . (d) LAM. (e) ENTR. (f) SWRP.

correlated with  $\lambda$ . Successively, we show that the edge of stability determined according to (15) identifies more accurately the region of the hyperparameter space where  $\gamma$  is maximized. To produce more interpretable results, for both input signals, we map independently the values of each RQA measure, obtained for the different values of  $\rho$  and  $\omega_i$ , in the  $[0, 1]$  interval by using a unity-based normalization. It is important to note that this operation neglects the possibility to perform a more in-depth analysis of RQA measures; which, however, is not the scope of this paper. Nonetheless, this choice allows the network designer to immediately recognize areas of interest and transitions in the related plots. We always use 50 bins for computing the SWRP measure (14). For each configuration of  $\rho$  and  $\omega_i$ , we perform 15 different simulations with independent random initialization of  $\mathbf{W}_i^r$  and  $\mathbf{W}_r^r$ . For this second experiment, since the number of reservoir neurons has an impact on the prediction performance, we adopted two different configurations for  $N_r$ .

*Sinusoidal Signal:* The ESN is trained to perform a 25-step ahead prediction and it is tested on a time series of 1500 time steps; we use  $N_r = 75$ . We evaluate the network performance by varying  $\rho \in P = [0.1, 2]$  and  $\omega_i \in \Omega = [0.1, 1]$ , both discretized with resolution 0.1.

In Fig. 7, we show the RQA measures calculated by varying  $\rho$  and  $\omega_i$ . The ESN, when  $\rho \leq 1$ , shows high levels of regularity and stability, as expressed by RR, DET, LAM, and  $L_{\max}$ . When  $\rho \simeq 1$ , we note that  $L_{\max}$  starts to decrease, suggesting that we are approaching the unstable phase. ENTR and SWRP behave similarly, although they are maximized for different configurations of  $\rho$  and  $\omega_i$ . In fact, ENTR assumes the relative maximum value for larger input scaling when  $\rho \simeq 1$ . SWRP instead, is maximized at smaller values of  $\omega_i$  and  $\rho$ .

Fig. 8 shows a 2-D section of Fig. 7, obtained by selecting a specific input scaling  $\omega_i = 0.8$ . In this way, we can visually assess the agreement between  $L_{\max}$  and  $\lambda$ , and also the increasing RQA variability (i.e., standard deviation) to detect the edge of stability. In Fig. 8(a), we report  $\lambda$ ,  $L_{\max}$ , and DIV, as we vary  $\rho$ . To improve visualization,  $\lambda$  is rescaled by dividing it for its maximum (absolute) value.  $\lambda$  and  $L_{\max}$  are

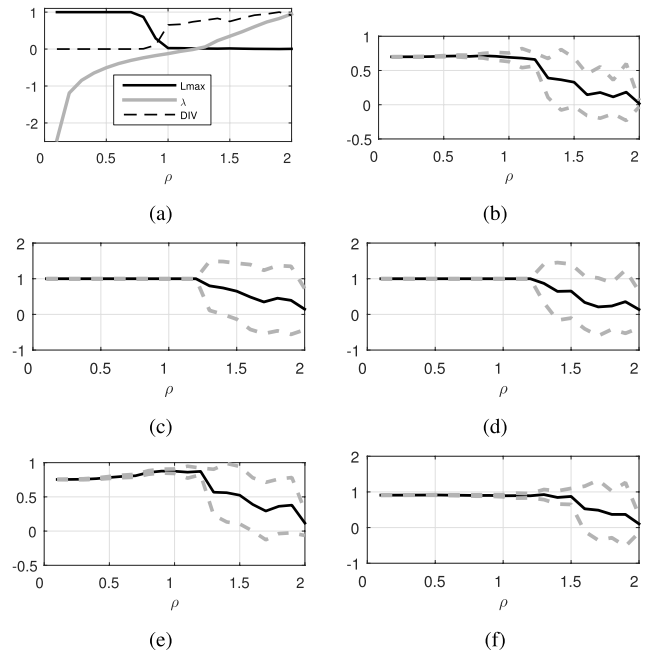


Fig. 8. RQA measures of sinusoidal input for a fixed input scaling,  $\omega_i = 0.8$ . The mean value of each measure is drawn with a solid black line and the standard deviation with gray dashed lines. All RQA measures are stable around their mean values until  $\rho$  is pushed beyond 1. (a) Rescaled, mean value of  $\lambda$  (gray solid line), the mean value of  $L_{\max}$  (solid black line), and the mean value of DIV (dashed black line). (a)  $L_{\max}$ ,  $\lambda$  and DIV. (b) RR. (c) DET. (d) LAM. (e) ENTR. (f) SWRP.

TABLE I  
CORRELATIONS BETWEEN  $\lambda$ , DIV, AND  $L_{\max}$

	$\lambda \setminus L_{\max}$	$\lambda \setminus \text{DIV}$
<b>Sin</b>	-0.74	0.53
<b>MG</b>	-0.65	0.57

anticorrelated with a (Pearson) correlation equal to  $-0.78$ : the value of  $L_{\max}$  decreases as  $\rho$  increases, while  $\lambda$ , as expected, increases with  $\rho$ . In addition, we observe a positive correlation (0.79) between  $\lambda$  and DIV. We stress that the agreement between  $\lambda$  and  $L_{\max}$  is consistent for the entire range of  $\omega_i$ , as shown in Fig. 9(d) and quantified in Table I. This confirms that statistics on RP diagonal lines offer reliable complexity measures to characterize network stability. Please note that the correlations in Table I are computed on two matrices in  $\mathbb{R}^{|P|} \times \mathbb{R}^{|\Omega|}$ . Since the matrix elements are not interrelated, the matrices are represented as vectors with dimension  $|P| \cdot |\Omega|$  and the correlation is computed on such vectors.

Let us comment the results obtained by considering RR, DET, LAM, ENTR, and SWRP to determine of the edge of stability. All these RQA measures denote low fluctuations until  $\lambda$  becomes positive at  $\rho \simeq 1.2$ —see related panels in Fig. 8. Values of  $\gamma$  are shown in Fig. 9(a), where we note that for low values of  $\omega_i$ , we obtain a slightly inferior accuracy. In addition, the bright area in Fig. 9(a) is larger for higher values of  $\omega_i$  and it includes also several configurations with  $\rho = 1.2$ . This is justified by the fact that high-amplitude signals tend to saturate nonlinear activation functions and cause the poles to shrink toward the origin. This results in a system with

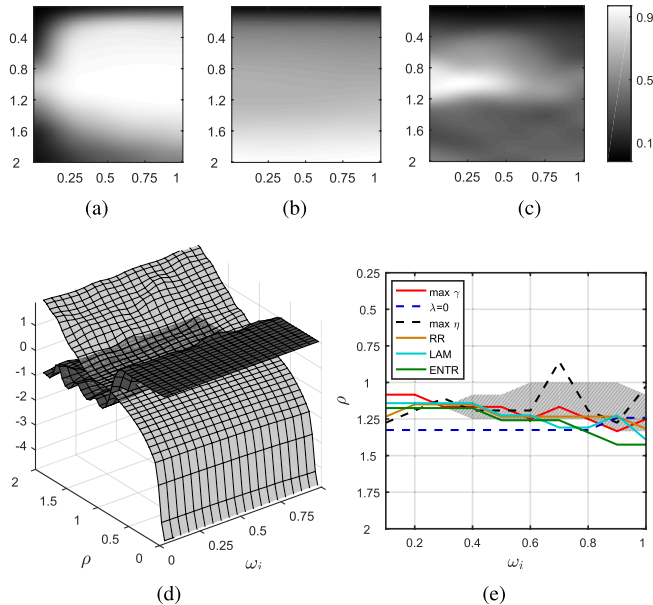


Fig. 9. Sinusoidal input. Prediction accuracy (a)  $\gamma$ , (b)  $\lambda$ , and (c)  $\eta$ , calculated for different values of  $\rho$  and  $\omega_i$ . (d)  $\lambda$  (light gray surface) assumes positive values in correspondence of the gap where  $L_{\max}$  (dark gray surface) goes to 0. (e) Edge of stability is identified as the configurations of  $\rho$  and  $\omega_i$ , where  $\lambda$  crosses 0,  $\eta$  is maximized, and where RQA measures start to fluctuate (15). For the sake of readability, here we show only three RQA measures; see Table II for detailed results. The gray area in (e) shows configurations with high accuracy and it corresponds to the brightest area in (a), where  $\gamma$  assumes larger values; the red line indicates where  $\gamma$  is maximum.

a larger stability margin and a more contractive dynamics, hence allowing the use of larger values of  $\rho$ . In Fig. 9(e), we depict different lines in the  $\rho$ - $\omega_i$  plane, each one denoting the edge of stability determined according to the criteria based on  $\lambda$ ,  $\eta$ , and RQA measures. In the same plot, with a solid red line, we indicate the configurations where the performance  $\gamma$  is maximized. To quantify the accuracy in identifying the edge of stability, in Table II, we report the average distances of the different edges from the configurations maximizing  $\gamma$  (red line). Our results show that, in all cases except for SWRP [not reported in Fig. 9(e)], RQA measures determine a more accurate edge of stability with respect to the one obtainable from the criteria based on  $\lambda$  and  $\eta$ . Notably, RR produces differences from  $\gamma$  that are significantly smaller than those of  $\lambda$  and  $\eta$  ( $p < 0.05$ ); the same holds for DET, LAM, and ENTR ( $p < 0.05$ ), but not for SWRP ( $p > 0.05$ ), which cannot accurately identify the edge. The low performance of the criterion based on  $\lambda$  ( $p > 0.05$ ) might be related to its low sensitivity on  $\omega_i$ , as shown in Fig. 9(b). Instead,  $\eta$  ( $p < 0.05$ ) is more sensitive to  $\omega_i$ , as shown in Fig. 9(c), and produces a better estimation of the edge of stability. The statistical significance has been assessed with the  $t$ -test.

In Table III, we report the dependences between edges identified by the RQA measures and  $\gamma$ . For the sinusoidal input, linear correlations of the edges with  $\gamma$  are not very satisfactory, although all correlations are significantly different from zero, with the only exception of SWRP. However, the estimated (normalized) mutual information indicates, in general, the presence of relevant nonlinear dependences. Even in this case, SWRP shows the highest dependence with  $\gamma$ .

TABLE II  
AVERAGE DISTANCES AND STANDARD DEVIATIONS BETWEEN THE CONFIGURATIONS WHERE  $\gamma$  IS MAXIMIZED AND THE INDICATORS USED TO DETERMINE THE EDGE OF STABILITY:  $\eta$  IS MAXIMUM,  $\lambda$  CROSSES 0, AND  $\sigma_q \geq \bar{\sigma}_q$  FOR RQA MEASURES. RESULTS FOR RQA SHOWN IN BOLD ARE STATISTICALLY SIGNIFICANT ( $p < 0.05$ )

	$\eta$	$\lambda$	RR	DET	LAM	ENTR	SWRP
MG Sin	<b>1.4±1.26</b>	1.7±0.95	<b>0.7±0.67</b>	<b>1.0±0.82</b>	<b>0.9±0.74</b>	<b>0.8±0.63</b>	2.5±0.85
MG	<b>3.7±2.21</b>	7.3±4.78	<b>2.0±0.94</b>	<b>3.1±1.29</b>	<b>2.6±1.35</b>	<b>2.9±1.29</b>	<b>3.0±0.94</b>

TABLE III  
DEPENDENCE BETWEEN  $\gamma$  AND RQA MEASURES. FIRST ROW: CORRELATIONS. SECOND ROW: ESTIMATED MUTUAL INFORMATION

	RR	DET	LAM	ENTR	SWRP
Sin	0.35	0.42	0.40	0.43	0.55
	0.66	0.69	0.67	0.68	0.81
MG	0.57	0.40	0.38	0.62	0.68
	0.77	0.55	0.54	0.78	0.80

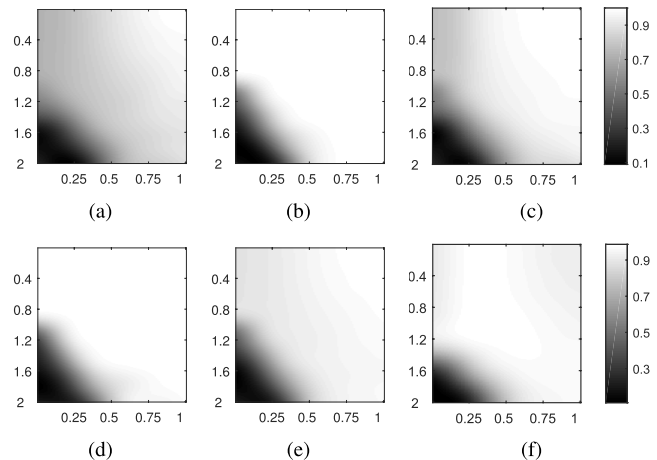


Fig. 10. RQA measures of MG while varying  $\rho$  (vertical axis) and  $\omega_i$  (horizontal axis). (a) RR. (b) DET. (c)  $L_{\max}$ . (d) LAM. (e) ENTR. (f) SWRP.

It is interesting to note that SWRP is a measure of entropy, and in our framework, it quantifies the heterogeneity of the neurons activations. It is well-known that RNNs (and then also ESNs) are effective if their internal dynamical patterns are sufficiently rich. In the literature, this is usually expressed in terms of entropy of neuron activations [22], explaining the relation between SWRP and  $\gamma$ .

*MG Time Series:* The prediction of the MG time series is a common benchmark where ESNs achieve good performance in terms of prediction accuracy [47], [48]. For this test, we train the ESN to perform a 20-step ahead prediction on a test time series of 2000 time steps; we set the size of the reservoir to  $N_r = 100$ ; intervals for  $P$  and  $\Omega$  are defined as before.

In Fig. 10, we show the outcomes for six RQA measures as both  $\rho$  and  $\omega_i$  change. In contrast to the sinusoidal input, we immediately observe a higher sensitivity on  $\omega_i$ . This can be explained by the requirement of a higher degree of nonlinearity for modeling the MG system. The dynamics of the reservoir,

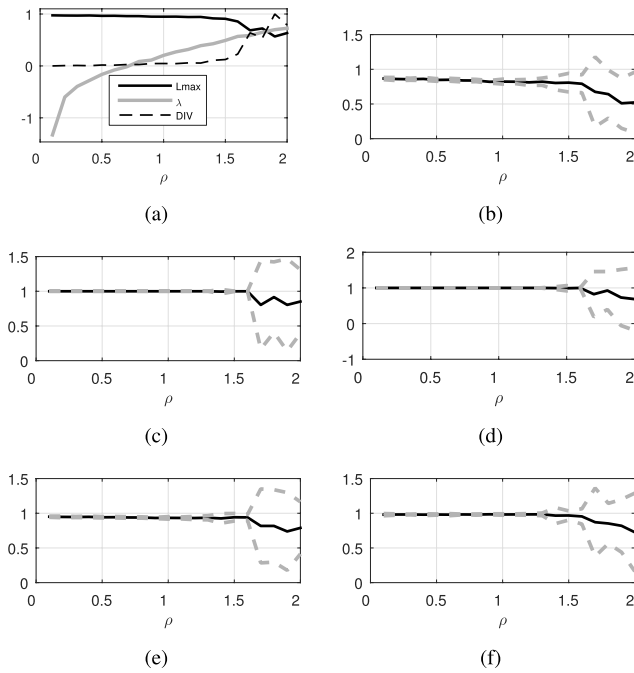


Fig. 11. RQA measures for MG by considering a fixed input scaling,  $\omega_i = 0.5$ . Solid black lines represent the mean value of each measure; gray dashed lines the standard deviation. The standard deviation of every RQA increases significantly as soon as the network becomes unstable. (a) Rescaled, mean value of  $\lambda$  (gray solid line), the mean value of  $L_{\max}$  (solid black line), and the mean value of DIV (dashed black line). (a)  $L_{\max}$ ,  $\lambda$  and DIV. (b) RR. (c) DET. (d) LAM. (e) ENTR. (f) SWRP.

for the entire range of  $\omega_i$ , appears highly deterministic with laminar phases (as expressed by DET and LAM, respectively). We note a consistent maximization for SWRP and ENTR, which assume high values for a large set of configurations that include also values for  $\rho \simeq 2$  when  $\omega_i > 0.5$ . For what concerns RR, instead, we note that it is maximized only for very small values of  $\rho$  and high values of  $\omega_i$ , suggesting that density of recurrences (i.e., RR) is very sensitive to  $\rho$  for the MG time series.

In Fig. 11, we show the mean and standard deviation of the RQA measures when  $\omega_i = 0.5$  is fixed. As before, by monitoring these quantities, we can assess the network stability and locate the edge of stability, where there is an increment in the fluctuations. In Fig. 11(a), we show the mean value of  $\lambda$ , which has been rescaled also in this case,  $L_{\max}$ , and DIV as we vary  $\rho$ . As for the sinusoidal input,  $\lambda$  and  $L_{\max}$  show a good anticorrelation, with a value of  $-0.64$ , while  $\lambda$  and DIV are correlated with a value of  $0.60$ . The agreement between  $L_{\max}$  and  $\lambda$  is confirmed for all values of  $\omega_i$ , as shown in Fig. 12(d) and numerically quantified in the second row of Table I.

Let us now consider the issue of determining the edge of stability. In Fig. 12(a), we note that high prediction accuracy  $\gamma$  is obtained for configurations located in the center of the plot toward the right-hand side, corresponding to  $\rho \simeq 1.5$  and  $\omega_i \geq 0.8$ ; for example, when  $\omega_i = 0.5$ , the best accuracy is obtained for  $1 \leq \rho \leq 1.3$ . By referring to Fig. 11(a), when  $\omega_i = 0.5$   $\lambda$  becomes positive for  $\rho \geq 0.7$ . On the other hand,  $\eta$  [Fig. 12(c)] and the RQA measures start to fluctuate for much higher values of  $\rho$ , close to 1.5.

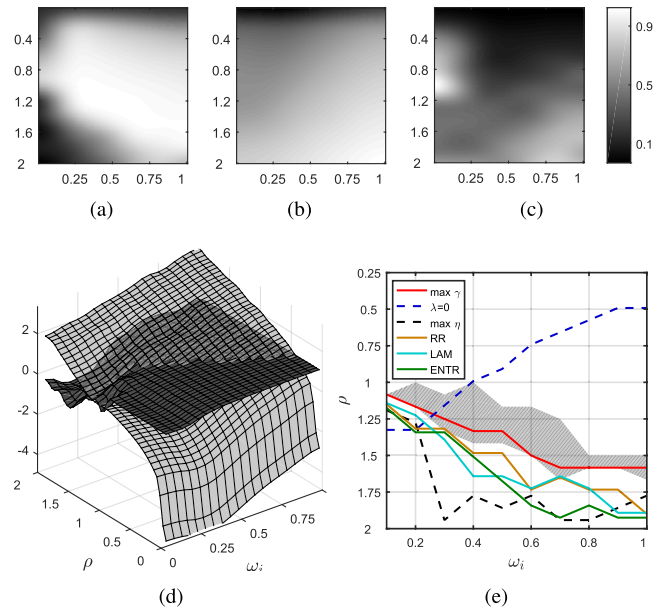


Fig. 12. MG time series. Prediction accuracy (a)  $\gamma$ , (b)  $\lambda$ , and (c)  $\eta$ , calculated for different values of  $\rho$  and  $\omega_i$ . (d)  $\lambda$  (light gray surface) assumes positive values in correspondence of the gap where  $L_{\max}$  (dark gray surface) goes to 0. (e) Edge of stability is identified as the configurations of  $\rho$  and  $\omega_i$ , where  $\lambda$  crosses 0,  $\eta$  is maximized, and where RQA measures start to fluctuate (15). For the sake of readability, here we show only three RQA measures; see Table II for detailed results. The gray area in (e) shows configurations with high accuracy and it corresponds to the brightest area in (a), where  $\gamma$  assumes larger values; the red line indicates where  $\gamma$  is maximum.

As clearly shown in Fig. 12(e),  $\lambda$  underestimates the location of the edge of stability for almost all values of  $\omega_i > 0.2$ . All RQA measures instead show a very good correlation with the red line denoting the best performance,  $\gamma$ , even though they slightly overestimate the edge. This comparison is made formal in Table II, where we show that all RQA measures have smaller distances from  $\gamma$  than both  $\lambda$  and  $\eta$ , with differences that are statistically significant. Notably,  $p$ -values are lower than 0.05 for RR, LAM, ENTR, DET, and SWRP, while for  $\lambda$  and  $\eta$ , we obtain  $p > 0.05$ . Therefore, in the MG time-series case, the information provided by RQA measures regarding the edge of stability is significantly more accurate than the one of both  $\lambda$  and  $\eta$ . As for the sinusoid, we note a low sensitivity of  $\lambda$  with respect to  $\omega_i$  [Fig. 12(b)]. Instead, the results obtained by considering  $\eta$  [Fig. 12(c)] lead to overestimate the edge of stability, especially for values of  $\omega_i \in [0.3, 0.5]$  [Fig. 12(e)]. Finally, it is worth to note that differences between  $\gamma$  and  $\eta$  are statistically significant, showing that  $\eta$  performs better also in the MG case with respect to  $\lambda$ .

Let us now consider the dependences between the edges identified by the RQA-based criterion and the regions of hyperparameter space where  $\gamma$  is maximized. As shown in Table III, even in this case, there is a good linear agreement between SWRP and  $\gamma$ , with a correlation of 0.68. Also ENTR achieves a good correlation of 0.62 as, in fact, ENTR and SWRP behave similarly for the MG time series (see Fig. 10). Nonlinear dependence, as indicated by mutual information, shows high values for RR, ENTR, and SWRP.

## V. CONCLUSION

In this paper, we investigated the possibility to analyze the dynamics of ESNs by means of recurrence plots and related RQA complexity measures. Being a nonautonomous system, ESN dynamics depend not only on static properties, such as the spectral radius of the reservoir matrix, but also on the particular input driving the network. We have shown how RPs can be used to analyze the evolution of the network states, when qualitatively different signals are presented as input. Our results indicate that the recurrences elaborated from the reservoir dynamics clearly capture the amount of instability in the network. This suggests that RPs can be used as a visual tool to design a stable network. Our results highlighted a strong statistical agreement between the maximal local Lyapunov exponent and  $L_{\max}$ , an RQA complexity measure based on the diagonal lines in RPs. Therefore,  $L_{\max}$  can be used to define novel, input-dependent stability criteria for reservoirs based on statistics of state recurrences.

We have also proposed an unsupervised method based on RQA to identify a hyperparameter configuration that brings the network on the edge of stability, where maximal computation capability is expected to be reached. We defined a criterion based on the observation that, when the network approaches the edge of stability, average values of all RQA measures experience a sudden change. This resulted also in a rapid increment of fluctuations of the RQA measures. To evaluate the performance of our approach, we considered prediction problems on two typical benchmarks, a sinusoid and the MG time series. Specifically, we compared the agreement of the stability edges yielded by our methodology, with region in the hyperparameter space where the supervised prediction accuracy is maximized. We also compared the performance of our method with two other unsupervised counterparts, which define the edge of stability according to the maximum local Lyapunov exponent  $\lambda$  and the minimum singular value  $\eta$  of the reservoir Jacobian matrix. Experimental results showed that both criteria were outperformed by the approach proposed in this paper based on fluctuations of RQA measures.

Our results suggest allocating future research effort in a number of ways. First, it is worth to study reservoirs initialized according to different (either statistical or deterministic) rules. In addition, it could be interesting to derive a connection between the network short-term memory capacity and RQA measures. We also plan to consider invariant measures elaborated from RPs: the correlation entropy provides a natural lower bound to the sum of positive Lyapunov exponents. Some initial results in this direction have been reported in [49]. It might be also interesting to study joint recurrences, occurring between input and reservoir. This might lead to the discovery of (hidden) synchronization mechanisms. Finally, we suggest that RP-based analyses could be extended to fully trainable RNNs. The recurrent layer would not be modeled as a multidimensional time series, but instead as a time-varying graph. To this end, we hypothesize a possible interplay between graph matching [50] and recurrence analysis tools.

## REFERENCES

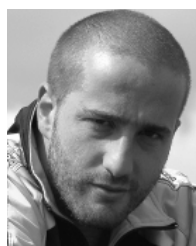
- [1] Y. Zhang and J. Wang, "Global exponential stability of recurrent neural networks for synthesizing linear feedback control systems via pole assignment," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 633–644, May 2002.
- [2] U. D. Schiller and J. J. Steil, "Analyzing the weight dynamics of recurrent learning algorithms," *Neurocomputing*, vol. 63, pp. 5–23, Jan. 2005.
- [3] R. L. Marichal and J. D. Piñeiro, "Analysis of multiple quasi-periodic orbits in recurrent neural networks," *Neurocomputing*, vol. 162, pp. 85–95, Aug. 2015.
- [4] D. Sussillo and O. Barak, "Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks," *Neural Comput.*, vol. 25, no. 3, pp. 626–649, 2013.
- [5] Y. Shen and J. Wang, "An improved algebraic criterion for global exponential stability of recurrent neural networks with time-varying delays," *IEEE Trans. Neural Netw.*, vol. 19, no. 3, pp. 528–531, Mar. 2008.
- [6] D. Sussillo, "Neural circuits as computational dynamical systems," *Current Opinion Neurobiol.*, vol. 25, pp. 156–163, Apr. 2014.
- [7] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Netw.*, vol. 20, no. 3, pp. 391–403, 2007.
- [8] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, 2009.
- [9] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," *Sci. Rep.*, vol. 2, Jul. 2012, Art. no. 514.
- [10] H. Jaeger, "Adaptive nonlinear system identification with echo state networks," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA, USA: MIT Press, 2003, pp. 593–600.
- [11] T. Strauss, W. Wustlich, and R. Labahn, "Design strategies for weight matrices of echo state networks," *Neural Comput.*, vol. 24, no. 12, pp. 3246–3276, Dec. 2012.
- [12] C. Gallicchio and A. Micheli, "Architectural and Markovian factors of echo state networks," *Neural Netw.*, vol. 24, no. 5, pp. 440–456, 2011.
- [13] Q. Ma, W. Chen, J. Wei, and Z. Yu, "Direct model of memory properties and the linear reservoir topologies in echo state networks," *Appl. Soft Comput.*, vol. 22, pp. 622–628, Sep. 2014.
- [14] S. Scardapane, D. Wang, and M. Panella, "A decentralized training algorithm for echo state networks in distributed big data applications," *Neural Netw.*, vol. 78, pp. 65–74, Jun. 2015.
- [15] F. M. Bianchi, S. Scardapane, A. Uncini, A. Rizzi, and A. Sadeghian, "Prediction of telephone calls load using echo state network with exogenous variables," *Neural Netw.*, vol. 71, pp. 204–213, Nov. 2015.
- [16] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, "Short-term electric load forecasting using echo state networks and PCA decomposition," *IEEE Access*, vol. 3, pp. 1931–1943, Oct. 2015.
- [17] S. I. Han and J. M. Lee, "Fuzzy echo state neural networks and funnel dynamic surface control for prescribed performance of a nonlinear dynamic system," *IEEE Trans. Ind. Electron.*, vol. 61, no. 2, pp. 1099–1112, Feb. 2014.
- [18] M. D. Skowronski and J. G. Harris, "Automatic speech recognition using a predictive echo state network classifier," *Neural Netw.*, vol. 20, no. 3, pp. 414–423, Apr. 2007.
- [19] Y. Xue, L. Yang, and S. Haykin, "Decoupled echo state networks with lateral inhibition," *Neural Netw.*, vol. 20, no. 3, pp. 365–376, Apr. 2007.
- [20] L. Boccatto, R. Attux, and F. J. Von Zuben, "Self-organization and lateral interaction in echo state network reservoirs," *Neurocomputing*, vol. 138, pp. 297–309, Aug. 2014.
- [21] B. Zhang, D. J. Miller, and Y. Wang, "Nonlinear system modeling with random matrices: Echo state networks revisited," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 175–182, Jan. 2012.
- [22] M. C. Ozturk, D. Xu, and J. C. Principe, "Analysis and design of echo state networks," *Neural Comput.*, vol. 19, no. 1, pp. 111–138, 2007.
- [23] B. Schrauwen, M. Wardermann, D. Verstraeten, J. J. Steil, and D. Stroobandt, "Improving reservoirs using intrinsic plasticity," *Neurocomputing*, vol. 71, nos. 7–9, pp. 1159–1171, Mar. 2008.
- [24] A. Rodan and P. Tiño, "Simple deterministically constructed cycle reservoirs with regular jumps," *Neural Comput.*, vol. 24, no. 7, pp. 1822–1852, 2012.
- [25] A. Rodan and P. Tino, "Minimum complexity echo state network," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 131–144, Jan. 2011.

- [26] L. Appeltant *et al.*, “Information processing using a single dynamical node as complex system,” *Nature Commun.*, vol. 2, no. 13, p. 468, Sep. 2011.
- [27] L. Büsing, B. Schrauwen, and R. Legenstein, “Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons,” *Neural Comput.*, vol. 22, no. 5, pp. 1272–1311, 2010.
- [28] D. Verstraeten and B. Schrauwen, “On the quantification of dynamics in reservoir computing,” in *Artificial Neural Networks*, vol. 5768, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds. Heidelberg, Germany: Springer, 2009, pp. 985–994.
- [29] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada, “Information processing in echo state networks at the edge of chaos,” *Theory Biosci.*, vol. 131, no. 3, pp. 205–213, 2012.
- [30] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Phys. Rep.*, vol. 438, nos. 5–6, pp. 237–329, 2007.
- [31] N. Marwan, “How to avoid potential pitfalls in recurrence plot based data analysis,” *Int. J. Bifurcation Chaos*, vol. 21, no. 4, pp. 1003–1017, 2011.
- [32] D. Eroglu *et al.*, “Entropy of weighted recurrence plots,” *Phys. Rev. E*, vol. 90, no. 4, p. 042919, 2014.
- [33] N. Marwan, S. Schinkel, and J. Kurths, “Recurrence plots 25 years later—Gaining confidence in dynamical transitions,” *Europhys. Lett.*, vol. 101, no. 2, p. 20007, 2013.
- [34] N. Marwan and J. Kurths, “Line structures in recurrence plots,” *Phys. Lett. A*, vol. 336, nos. 4–5, pp. 349–357, 2005.
- [35] H. Yang and Y. Chen, “Heterogeneous recurrence monitoring and control of nonlinear stochastic processes,” *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 24, no. 1, p. 013138, 2014.
- [36] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, “Recurrence networks—A novel paradigm for nonlinear time series analysis,” *New J. Phys.*, vol. 12, no. 3, p. 033025, 2010.
- [37] H. Jaeger, “The ‘echo state’ approach to analysing and training recurrent neural networks—with an erratum note,” *German Nat. Res. Center Inf. Technol.*, Bonn, Germany, GMD Tech. Rep. 148, 2001, p. 34.
- [38] I. B. Yildiz, H. Jaeger, and S. J. Kiebel, “Re-visiting the echo state property,” *Neural Netw.*, vol. 35, pp. 1–9, Nov. 2012.
- [39] G. Manjunath and H. Jaeger, “Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks,” *Neural Comput.*, vol. 25, no. 3, pp. 671–696, 2013.
- [40] D. Verstraeten, B. Schrauwen, and D. Stroobandt, “Reservoir-based techniques for speech recognition,” in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Vancouver, BC, Canada, Jul. 2006, pp. 1050–1053.
- [41] C. G. Langton, “Computation at the edge of chaos: Phase transitions and emergent computation,” *Phys. D, Nonlinear Phenomena*, vol. 42, nos. 1–3, pp. 12–37, 1990.
- [42] N. Bertschinger and T. Natschläger, “Real-time computation at the edge of chaos in recurrent neural networks,” *Neural Comput.*, vol. 16, no. 7, pp. 1413–1436, 2004.
- [43] R. Legenstein and W. Maass, “Edge of chaos and prediction of computational performance for neural circuit models,” *Neural Netw.*, vol. 20, no. 3, pp. 323–334, 2007.
- [44] H. Jaeger, “Short term memory in echo state networks,” *German Nat. Res. Center Inf. Technol.*, Germany, Tech. Rep. GMD Technical Report 152, 2001.
- [45] E. Bradley and H. Kantz, “Nonlinear time-series analysis revisited,” *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 9, p. 097610, 2015.
- [46] J. J. Steil, “Memory in backpropagation-decorrelation  $O(N)$  efficient online recurrent learning,” in *Artificial Neural Networks: Formal Models and Their Applications*, W. Duch, J. Kacprzyk, E. Oja, and S. Zadrozny, Eds. Berlin, Germany: Springer, 2005, pp. 649–654.
- [47] Z. Shi and M. Han, “Support vector echo-state machine for chaotic time-series prediction,” *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 359–372, Mar. 2007.
- [48] H. Jaeger and H. Haas, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication,” *Science*, vol. 304, no. 5667, pp. 78–80, Apr. 2004.
- [49] L. Løkse, F. M. Bianchi, and R. Jenssen. (2016). “Training echo state networks with regularization through dimensionality reduction.” [Online]. Available: <https://arxiv.org/abs/1608.04622>
- [50] L. Livi and A. Rizzi, “The graph matching problem,” *Pattern Anal. Appl.*, vol. 16, no. 3, pp. 253–283, 2013.



**Filippo Maria Bianchi** received the B.Sc. degree in computer engineering, the M.Sc. degree in artificial intelligence and robotics, and the Ph.D. degree in machine learning from the Sapienza University of Rome, Rome, Italy, in 2009, 2012, and 2016, respectively.

He was a Research Assistant with the Computer Science Department, Ryerson University, Toronto, ON, Canada. He is currently a Post-Doctoral Fellow with the Machine Learning Group, University of Tromsø, Tromsø, Norway. His current research interests in machine learning and pattern recognition, include graph matching, clustering, classification, time series analysis, reservoir computing, and deep learning.



**Lorenzo Livi** (M’14) received the B.Sc. and M.Sc. degrees from the Department of Computer Science and the Ph.D. degree from the Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, Rome, Italy, in 2007, 2010, and 2014, respectively.

He has been with the ICT industry during his studies. From 2014 to 2016, he was a Post-Doctoral Fellow with Ryerson University, Toronto, ON, Canada. In 2016, he was a Post-Doctoral Fellow with the Politecnico di Milano, Milan, Italy, and the Università della Svizzera Italiana, Lugano, Switzerland. He is currently a Lecturer (Assistant Professor) in data analytics with the Department of Computer Science, University of Exeter, Exeter, U.K. His current research interests include computational intelligence methods, time series analysis, and complex dynamical systems, with a focus on applications in computational biochemistry and biophysics.

He is a member of the Editorial Board of *Applied Soft Computing* (Elsevier) and a regular Reviewer of several international journals, including the IEEE TRANSACTIONS ON FUZZY SYSTEMS and *Information Sciences* (Elsevier).



**Cesare Alippi** (F’99) is currently a Full Professor of Information Processing Systems with the Politecnico di Milano, Milan, Italy, and the Università della Svizzera Italiana, Lugano, Switzerland. He has authored a monograph on *Intelligence for Embedded Systems* in 2014 (Springer) and co-authored over 200 papers in international journals and conference proceedings. He holds five patents. His current research interests include adaptation and learning in nonstationary environments and Intelligence for embedded systems.

He is a Distinguished Lecturer of the IEEE Computational Intelligence Society (CIS) and a Vice-President Education of the IEEE CIS. He is also a member of the Board of Governors of the International Neural Networks Society. In 2016, he received the Gabor Award from the International Neural Networks Society and the IEEE Computational Intelligence Society Outstanding Transactions on Neural Networks and Learning Systems Paper Award. He also received the IBM Faculty Award in 2013 and the IEEE Instrumentation and Measurement Society Young Engineer Award in 2004. He was an Associate Editor of the IEEE *Computational Intelligence Magazine*, the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENTS, and the IEEE TRANSACTIONS ON NEURAL NETWORKS.