The reliability of recurrence network analysis is influenced by the observability properties of the recorded time series

Leonardo L. Portes, Arthur N. Montanari, Debora C. Correa, Michael Small, and Luis A. Aguirre

ARTICLES YOU MAY BE INTERESTED IN

Effects of stochastic parametrization on extreme value statistics
Chaos: An Interdisciplinary Journal of Nonlinear Science 29, 083102 (2019); https://doi.org/10.1063/1.5095756

Hierarchical network of pulse coupled chemical oscillators with adaptive behavior: Chemical neurocomputer
Chaos: An Interdisciplinary Journal of Nonlinear Science 29, 083104 (2019); https://doi.org/10.1063/1.5099979

Statistical complexity of the time dependent damped L84 model
Chaos: An Interdisciplinary Journal of Nonlinear Science 29, 083105 (2019); https://doi.org/10.1063/1.5107510
The reliability of recurrence network analysis is influenced by the observability properties of the recorded time series

ABSTRACT
Recurrence network analysis (RNA) is a remarkable technique for the detection of dynamical transitions in experimental applications. However, in practical experiments, often only a scalar time series is recorded. This requires the state-space reconstruction from this single time series which, as established by embedding and observability theory, is shown to be hampered if the recorded variable conveys poor observability. In this work, we investigate how RNA metrics are impacted by the observability properties of the recorded time series. Following the framework of Zou et al. [Chaos 20, 043130 (2010)], we use the Rössler and Duffing-Ueda systems as benchmark models for our study. It is shown that usually RNA metrics perform badly with variables of poor observability as for recurrence quantification analysis. An exception is the clustering coefficient, which is rather robust to observability issues. Along with its efficacy to detect dynamical transitions, it is shown to be an efficient tool for RNA—especially when no prior information of the variable observability is available.

I. INTRODUCTION
Our ability to control diverse dynamical systems is intimately related to our ability to observe their state and to characterize them. Not surprisingly, the concepts of controllability and observability emerged together in control theory. Given a parameter space, it is often desired to securely “navigate” through some states while avoiding others that could lead to instabilities, chaotic dynamics, or a poor system performance. Nevertheless, before applying control techniques to navigate along a parameter space, an accurate description of it should be provided, often experimentally by means of time series analysis techniques. If the system equations of motion are known (e.g., from theory or from a model), the construction of a “navigation map” through its different dynamical regimes may be straightforward. However, in many cases, what one may have at disposal are noisy, and usually short, time series from their experimental data.

Recurrence quantification analysis (RQA) and recurrence network analysis (RNA), both based on recurrence plots (RPs), are techniques able to characterize dynamical regimes from rather short time series (as compared to other methods, such as Lyapunov exponents) and are known to be robust against experimental noise. The reliability of RNA for recorded time series with different “levels” of observability is investigated in this paper.
the RQA—which thoroughly reviewed in Ref. 7. Indeed, one of the most interesting RQA traits is its sensitivity to detect bifurcation behavior, especially in chaotic-periodic dynamical transitions.4,6

Following the introduction of the network-based time series analysis paradigm,6 Marwan et al.3 proposed what became a popular method to quantify RPs based on complex networks theory.2 In this approach, the recurrence matrix, from which the RP is a graphical representation, is converted into an adjacency matrix, representing the recurrence network (RN), which describes the structural relationship between edges and nodes of a graph. This allows a whole new set of metrics from graph and complex networks theories to be applied in this context, yielding the so-called RNA, which also retains most of RQA interesting features (including the ability to detect dynamical transitions). While RQA measures are focused on temporal information regarding trajectories in the state-space, RNA measures cover mainly their geometrical properties and spatial dependencies.4,8 Donner et al.9 provide a review for different approaches in the generation of complex networks from time series, uncovering new properties of a dynamical system revealed by RNA when compared to RQA.

A pertinent investigation of the use of RQA and RNA to build navigation maps was conducted by Zou et al.10 in the context of “shrimp” structures. In parameter spaces of nonlinear dynamical systems, a bifurcation scenario provoked by variations of more than one parameter gives rise to a rich variety of dynamical transitions. Shrimps are a common example, being regions of periodic behavior embedded in larger chaotic regimes, which can be found over wide parameter domains. Although the shrimp structure is dependent of its respective nonlinear system, it is usually characterized by a main body with four thin legs that form a “swallow-tail” figure with regular self-similarity and a particular orientation.11,12

The knowledge of the shrimp structure offers an interesting alternative for chaos control in physical systems. By the adjustment of two or more control parameters, it is possible to navigate over the system dynamical domains through these regions of different periodicities without “ever” transiting over chaotic regimes.13,14 For instance, this control strategy was exploited in CO2 laser15 and dissipative kicked rotator16 applications. The interested reader can see Ref. 1 for further details of the occurrence and relevance of shrimps in chaotic systems and in experimental scenarios.

Taking the Rössler system as a benchmark example, Zou et al.14 demonstrated the applicability of RQA and RNA for the construction of a 2-dimensional navigation map able to detect shrimp structures from short and noisy time series. They have shown a clear advantage of RNA metrics over RQA ones when the detection of dynamical transitions are concerned. It is worth noticing that both RQA and RNA were conducted in the “original” Rössler system state-space with its full set of (three) state-space variables recorded, and so no state-space reconstruction was needed. However, in practical scenarios, often the time series of only “one” variable is available for recording. Hence, real-world applications of both RQA and RNA rely “heavily” on time series embedding techniques to reconstruct the system state-space.17 The recorded time series is effectively “the working data” used to generate the recurrence matrix and, hence, a critical step for the application of RQA/RNA (for example, see the RNA applications to time series obtained from marine dust record5 and paleoclimate record18,19).

Observability theory states that not all measurement functions (e.g., the state variables themselves) are equivalent in conveying dynamical information of a system behavior, and some may be very “poor” in this sense.23 A known direct consequence of this, on embedding theory, is that the state-space reconstruction effectiveness can be strongly influenced by the chosen variable observability properties.24-27 If the reconstruction is hampered by an embedded time series obtained from a variable that conveys poor observability (from now on a “poor variable”), further quantification tools might be heavily impaired. This was shown in the context of RQA28 and multivariate singular spectrum analysis (M-SSA) for phase synchronization detection.29 A relevant consequence of the insights from observability theory on embedding-based nonlinear time series analysis techniques is the enhancement of the technique itself. A recent example has been the extension of M-SSA to the new scenario of phase synchronization characterization in networks of spiking and bursting neuron models.30

Hence, the aim of this paper is to investigate the impact of the “quality” of the recorded variable in terms of conveyed observability on the RNA ability to characterize periodic-chaotic transitions. For this, the detection of shrimp structures in a 2-dimensional parameter space of the Rössler system is investigated but in the context of a “single”—measurement state-space reconstruction, which allows one to perform a direct comparison with the benchmark example of Zou and co-workers’ paper.1 A second study using the Duffing-Ueda oscillator is also discussed. The results show that, analogously to the divergent metric in RQA,42 the “clustering coefficient” metric from RNA is quite robust to observability issues of the recorded variable, even when the signal is contaminated by measurement noise. Other tested metrics do not perform as well when the recorded time series conveys poor observability.

The paper is organized as follows. Section II provides a background on observability metrics and RNA, as well as a description of the numerical setup of our benchmark models, the Rössler and Duffing-Ueda systems. Section III presents the numerical results, which are further discussed in Sec. IV. Section V concludes the work.

II. METHODS
A. Observability

This section provides background on observability metrics for nonlinear systems. The reader is referred to Refs. 2, 3, and 26 for further details and references. The challenges of evaluating observability from data, i.e., without knowing the system equations, have been discussed in Ref. 27. Let us consider the continuous-time autonomous nonlinear dynamical system

\[ x = f(x), \]

\[ s = h(x), \] (1)

where \( x \in \mathbb{R}^n \) is the state vector, \( s \in \mathbb{R} \) is the output signal, \( f : \mathbb{R}^n \mapsto \mathbb{R}^n \) and \( h : \mathbb{R}^n \mapsto \mathbb{R} \).

The observability matrix of (1) can be written as

\[ O_L(x) = \begin{bmatrix} \frac{\partial L_0^0 h(x)}{x} \\ \frac{\partial L_0^1 h(x)}{x} \\ \vdots \\ \frac{\partial L_0^{n-1} h(x)}{x} \end{bmatrix}^\top, \] (2)
where the index \( s \) indicates that \( O_i(x) \) refers to the system observed from \( s(t) = h(x) \), and \( \mathcal{C}' \) is the \( j \)th Lie derivative of \( h \) along the vector field \( f \) (Ref. 32, p. 8). System (1) is observable from \( s(t) \) if \( O_i(x) \) has full rank for all \( x \). Otherwise, it is unobservable.

The classical definition of observability yields "yes-no" answers, which does not distinguish the quality of the system variables to convey dynamical information when comparing observable variables. A coefficient \( \delta \) has been proposed to quantify the numerical ill-conditioning of matrix \( O_i(x) \), i.e., how far it is from being rank-deficient.\(^{33,34}\) This index was extended to a nonlinear context\(^{35,36}\) as

\[
\delta_i(x) = \frac{\lambda_{\text{min}}[O_i(x)O_i(x)^T]}{\lambda_{\text{max}}[O_i(x)O_i(x)^T]},
\]

where \( \lambda_{\text{max}}(-) \) and \( \lambda_{\text{min}}(-) \) indicate, respectively, the maximum and minimum eigenvalue of \( O_i(x)O_i(x)^T \) at point \( s(t) \). Then, \( 0 \leq \delta_i(x) \leq 1 \), and the lower bound is reached when the system is not observable at point \( x \). Averaging \( \delta_i(x) \) along a trajectory over the interval \( t \in [t_0, t_1] \) yields a "global" observability coefficient

\[
\delta = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \delta_i(x(t)) \, dt.
\]

The link between observability and embedding theories was achieved by verifying that the observability matrix (2) is equivalent to the Jacobian matrix of the map between the original space and the reconstructed space produced with \( s(t) \). \(^{35,36}\) Hence, the "quality" of the reconstructed space is fundamentally related to the amount of dynamical information that \( s \) conveys from the dynamics.

### B. Recurrence network analysis

A recurrence plot (RP) is a 2-dimensional representation of the recurrent states of an \( n \)-dimensional state-space trajectory. Given a (recorded) discrete-time state-space trajectory \( \tilde{x}(i) \in \mathbb{R}^n \), for \( i = 1, \ldots, N \), of the dynamical system (1), where \( N \) is the trajectory length, the recurrent states are defined by a pairwise evaluation of all states proximities along their state-space trajectories,

\[
R_{ij}(\epsilon) = \Theta(\epsilon - \|\tilde{x}(i) - \tilde{x}(j)\|),
\]

where \( \Theta(\cdot) \) is the Heaviside function, \( \epsilon \in [0, 1] \) is the binary recurrence matrix, \( \|\cdot\| \) is defined as the maximum norm in this paper, and \( \epsilon \) is a threshold that determines the neighborhood size in which a state is considered recurrent. The reader is referred to Ref. 7 for further details on RP and RNA.

If not all state variables are available for data collection, then \( R_{ij}(\epsilon) \) can be inferred from a recorded time series \( s(i) \), for \( i = 1, \ldots, N \), using, for example, a time delay embedding to appropriately reconstruct the system trajectory in the state-space as

\[
\tilde{x}_{\tau}(i) = [s(i), s(i + \tau), s(i + 2\tau), \ldots, s(i + (m - 1)\tau)],
\]

where \( \tilde{x} \) is the reconstructed state-space vector from \( s \), \( m \) is the embedding dimension, and \( \tau \) is the time delay.

A recurrence network (RN) is the graph representation of an adjacency matrix \( A \) built through the recurrence matrix \( R \) of a given time series\(^{37}\)

\[
A_{ij} = R_{ij} - \delta_{ij},
\]

where \( A \in [0, 1]^{N \times N} \) and \( \delta_{ij} \) is the Kronecker delta. This definition implies an undirected and unweighted network with no self-edges, where the state-space vectors are represented as nodes in a network and the recurrences between state-space vectors as edges.

Recurrence network analysis (RNA) is a set of metrics from graph and complex networks theory that provide further topological characterization of the local and global state-space properties of dynamic systems. These metrics, even though they usually share the mathematical definition established by graph theory, receive new interpretations in the context of RNA (see Ref. 13). In the context of shrimp structures, it was shown that periodic-chaotic transitions of the Rössler system could be characterized through the application of the global clustering coefficient \( C \), and the average path length \( L \).\(^{34}\) Here, the assortativity coefficient \( \mathcal{R} \) is considered as well, since we found its behavior relevant to highlight RNA susceptibility to observability (to be shown in Sec. III). Finally, by considering these three RNA metrics, the results here will enable a broader and deeper discussion of different and "nonredundant" properties of dynamic systems captured by the RNA under the influence of observability.

The clustering coefficient \( C_i \) measures the density degree of neighborhood connections of a given node \( v_i \), for \( i = 1, \ldots, N \), and can be defined as\(^{33,34}\)

\[
C_i = \frac{2T_i}{k_i(k_i-1)},
\]

where \( k_i \) is the node degree and \( T_i \) is the number of closed triangles including node \( v_i \). State-space regions of high density of low-order unstable periodic orbits (UPOs) can be detected by this local metric since the accumulation of states along the UPO region increases the clustering formation over the network.\(^{13}\) Likewise, a more homogeneous filling of the state-space region implies less clustering formations.

A global clustering coefficient \( C \) can be taken as the average over all coefficients \( C_i \)

\[
C = \frac{1}{N} \sum_{i=1}^{N} C_i,
\]

As chaotic orbits are associated with a more homogeneous filling of the state-space than periodic cycles, \( C \) presents lower values on chaotic regimes than periodic ones. For a more thorough discussion on the relation between clustering coefficient, network transitivity and the fractal dimension of the attractor, the interested reader should consult Refs. 36 and 37.

The assortativity coefficient \( \mathcal{R} \) measures the node degree correlation between the interconnected pairs of nodes of a graph, i.e., the preference of a network node to attach others of similar node degree. Computed via the Pearson correlation coefficient,\(^{38,39}\) the assortativity coefficient is

\[
\mathcal{R} = \frac{M^{-1} \sum_{i<j} k_i k_j A_{ij} - (M^{-1} \sum_{i<j} \frac{1}{2} (k_i + k_j) A_{ij})^2}{M^{-1} \sum_{i<j} \frac{1}{2} (k_i^2 + k_j^2) A_{ij} - (M^{-1} \sum_{i<j} \frac{1}{2} (k_i + k_j) A_{ij})^2},
\]
where \( M = \sum_{i=1}^{N} A_{ii} \). In the context of RNA, \( R \) verifies the continuity of state-space density, where a higher \( R \) points to slowly changing states and, therefore, periodicity.

The average path length \( \mathcal{L} \) is the average value of the shortest distances \( l_{ij} \) over all pairs of nodes \((v_i, v_j)\)

\[
\mathcal{L} = \frac{2}{N(N-1)} \sum_{i<j} l_{ij}
\]

where the shortest path \( l_{ij} \) in high-order networks is usually computed via search algorithms—hence being a computationally expensive parameter for real applications. In RN, temporal order is discarded. Therefore, network distance measures are not related to the temporal evolution of a dynamic system, but rather to its state-space distances. In this case, \( \mathcal{L} \) indicates the mean state-space separation between two state vectors. Like \( \mathcal{C} \) and \( R \), a chaotic regime implies a more homogeneous filling of the state-space that offers “shortcuts” between two state vectors. Thus, \( \mathcal{L} \) is smaller for chaotic regimes than periodic ones.

### C. Benchmark model systems

We investigate two benchmark systems: the autonomous Rössler system, which is the benchmark we mainly focus our attention to, and the nonautonomous Duffing-Ueda oscillator, to further generalize our results and discussions. In this section, for each benchmark system, we describe the numerical setup for our simulations, illustrate the investigated parameter space, and compute the coefficient of observability of each variable. We also provide our reasons for the choice of RNA and embedding parameters.

#### 1. Autonomous: Rössler system

The Rössler system is defined by

\[
\begin{align*}
\dot{x} &= -y - z, \\
\dot{y} &= x + ay, \\
\dot{z} &= b + z(x - c),
\end{align*}
\]

where \( \mathbf{x} = [x \ y \ z]^T \in \mathbb{R}^3 \).

The parameter space region of interest is \((c, a = b) \in [20, 45] \times [0.2, 0.3]\), corresponding to the one investigated by Zou et al. The parameter space was divided in a \(500 \times 500\) grid, with a total of \(250,000\) \((c, a = b)\) pairs. For each point of the grid, system (12) was numerically integrated via a fourth-order Runge-Kutta integrator, with a time step of \(\Delta t = 0.01\), and initial conditions \((x_0, y_0, z_0)\) randomly sampled from the Gaussian distribution \(N(0, 1) \times N(0, 1) \times N(0, 1)\). Other simulation parameters were set depending on the desired metric to be estimated, aiming for consistency with Ref. 14 (assumed here as the benchmark result). They are discussed below.

**Navigation map.** The largest Lyapunov exponent \( \lambda_{\text{max}} \), computed numerically by the method of Benettin et al., was used to probe the dynamics in the parameter space [see Fig. 1(a)]. A large time series was required for its reliable numerical estimate, with total simulation time \(T_{\text{sim}} = 5000\), sampling time \(T_s = \Delta t = 0.01\), and a discarded initial transient \(T_{\text{trans}} = 300\), yielding an effective time length of \(T_s = 4000\). These parameters imply a time series of \(N = T_{\text{sim}}/T_s = 400,000\) data points. As expected for the aforementioned parameter space, the main shrimp structure is clearly visible (i.e., \( \lambda_{\text{max}} = 0 \)), being characterized by its singular main blue body attached to four thin legs in a swallowlike form pointed toward a particular orientation. This plot is commonly interpreted as periodic isles (shrimps) surrounded by a vast chaotic sea.

**Coefficient of observability.** The observability coefficient (4) was estimated for each state variable \(x, y,\) and \(z\) at each point of the parameter space [see Figs. 1(b)–1(d)]. This estimate requires a much shorter time series with time length \(T_s = 200\), and coarser \(T_s = 0.07\) (see Ref. 27 for details). For any value of \((c, a = b)\), one observes \(\delta_x \geq 0.1375\), \(0.002 \leq \delta_y \leq 0.012\), and \(\delta_z \leq 0.00032\). This consistency of \(\delta_x > \delta_y \gg \delta_z\) implies the observability rank \(\delta_y > x > z\) for the entire navigation map. Hence, the best variable to convey dynamical information from an embedding is \(y\), while the worst is \(z\).

**RNA and embedding parameters.** Following Ref. 14, the sampling time \(T_s = 0.2\) was used for all time series. The estimates through the original state-space \(\mathbf{x}\) (i.e., requiring no embedding) were assumed as the “benchmark” ones to be compared with the estimates through the individual time delay embedding of \(x, y,\) and \(z\) variables. For each point of the parameter space, simulations were performed with \((T_{\text{trans}}, T_{\text{trans}}) = (900, 500)\), yielding a much shorter time series with length \(T_s = 400\) (i.e., \(N = 2000\) data points). The embedding dimension \(m\) can be safely set to 3: (i) the Rössler system lives in a 3-dimensional space and (ii) we are interested in the recurrence structure in state-space. The time delay \(\Delta t\) is related to the self-correlation of the time series and should be chosen in a way that the underlying attractor structure is properly unfolded. Given an embedding dimension \(m = 3\), we set \(\Delta t = 1.6, 1.6, 0.2\) (respectively, for \(x, y,\) and \(z\)) directly through visual inspection of the reconstructed trajectories. The reconstructed state-spaces are shown in Fig. 2.

**The choice of threshold.** The RP, and consequently the RNA metrics, are very susceptible to the threshold \(\epsilon\). If \(\epsilon\) is too small,
a very sparse (or disconnected) RN might be obtained, and if $\epsilon$ is too large, the result will be a highly (or fully) connected RN. This directly affects the RNA measures, especially $R$ and $L$. Several methodologies to select $\epsilon$ have been addressed in the literature. In this paper, we follow Ref. 14, Sec III D by choosing $\epsilon$ adaptively so that the recurrence rate guarantees that the RN is neither too dense nor too sparse—which is particularly useful in applications involving dynamical transitions or sliding-window-based analyses. For a more detailed discussion on the choice of $\epsilon$, we refer the reader to Ref. 37, Sec. 3.6.1.

2. Nonautonomous: Duffing-Ueda oscillator

The Duffing-Ueda oscillator is defined by

$$
\begin{align*}
\dot{x} &= y, \\
\dot{y} &= B \cos(\omega t) - ky - x^3,
\end{align*}
$$

where $\mathbf{x} = [x \ y]^T \in \mathbb{R}^2$. However, in order to avoid self-intersections on this 2-dimensional state-space, it is useful to consider the phase $\phi = \omega t$ as the third state variable on a circular axis [i.e., assuming $\mathbf{x} = [x \ y \ \cos(\phi)]^T \in \mathbb{R}^3$]. Then, the respective third equation to be included in system (13) is $\dot{\phi} = \omega$.

The parameter space region of interest is $\omega = 1$ and $(k, B) \in [0.27, 0.45] \times [126, 135]$, corresponding to the one investigated by Bonatto et al. In the following numerical simulations, to ensure shorter transient time and reduce computational burden, we set initial conditions to $\mathbf{x}(0) = [0.5 \ 0.1 \ \cos(0)]^T$. Numerical integration was carried out using a fourth-order Runge-Kutta integrator, with a time step of $dt = 0.01$.

**Navigation map.** The largest Lyapunov exponent $\lambda_{\text{max}}$ for the aforementioned parameter space is shown in Fig. 3(a). We set parameters to $(T_{\text{sim}}, T_{\text{trans}}, T_p, T_r) = (600, 300, 300, 0.12)$. Likewise to the Rössler system, one can clearly see the main shrimp structure surrounded by the chaotic sea.

**Coefficient of observability** The observability coefficient (4) was estimated for each state variable $x$ and $y$ at each point of the parameter space [see Figs. 3(b)–3(d)], using $(T_p, T_r) = (200, 0.05)$. Since $\delta_k \gg \delta_y$ is consistent for any value of $(k, B)$, then the observability rank is $x \succ y$. Note that since $\delta_k = 1$ for all the parameter space, then the system variable is “fully observable” from $x$. In other words, there is a global diffeomorphism from the space reconstructed from $x$ and the original state space.

**RNA and embedding parameters.** We set numerical simulations parameters to $(T_{\text{sim}}, T_{\text{trans}}, T_p, T_r) = (600, 300, 300, 0.12)$. The reconstructed state-spaces are shown in Fig. 4, with embedding dimension $m = 2$ and time delay $\Delta t = (0.24, 0.24)$. The choice of threshold $\epsilon$ follows the discussion in Sec. II C.

III. RESULTS

Section III A provides a study of the effects of the recorded time series observability on RNA using the autonomous Rössler system as our main benchmark system, as seen in Fig. 5. The effects of observability are also investigated under the influence of measurement noise in Sec. III B. Our main remarks are further confirmed by statistical measures in Sec. III C. To further generalize our results and discussion to different dynamical systems, Sec. III D provides an additional analysis of a nonautonomous system, the Duffing-Ueda oscillator.

For brevity sake, in the following text, the subscript of a metric will indicate the variable used to reconstruct the state space from which the metric was estimated. For instance, $L_y$ refers to the average path length estimated from a space reconstructed from the $y$ variable [see Fig. 5(k)].

A. Single-variable RNA: Rössler system

Figures 5(a), 5(e), and 5(i) show the RNA results for the original state-space $\mathbf{x}$ of Rössler system, that is, no embedding is required. Its reliability as a benchmark scenario, with the current numerical
As predicted from embedding and observability theories, the state-space reconstruction through good observables allows reliable quantification of the system dynamical features in a “practical” context (i.e., regarding the application of time series analysis techniques).

On the other hand, the reliability of the RNA through the $z$-embedding depends on the specific metric. Clearly, the worst performance occurs in the characterization of periodic regimes with $L_z$, which shows an awfully homogeneous chaotic sea wherein the shrimps borders are loosened (less sharp transitions between the shrimps borders and the chaotic sea). Although chaotic-periodic transitions are still detectable by metrics $C_z$ and $R_z$, the inspection of periodic-periodic and chaotic-chaotic transitions over the parameter space is visually impaired when compared to the respective RNA measures performed with the embedding of “good variables” (i.e., $C_{xy}$ and $R_{xy}$). Overall, RNA with $z$ time series has a poorer performance than with $x$ of $y$ time series, as predicted by observability/embedding theories.

**B. The influence of noise**

We investigate the robustness of the RNA metrics under the influence of an additive noise signal $v_t$ randomly sampled from the Gaussian distribution $\mathcal{N}(0, 0.1 \cdot \text{var}(s))$ (i.e., 10% noise level), where $\text{var}(s)$ denotes the variance of the measured variable time series. Figure 6 presents the RNA metrics estimated under additive noise. For brevity sake, we denote a reconstructed state-space from a noisy time series by a tilde, i.e., $\tilde{x}(t) = x(t) + v(t)$. For instance, $L_{\tilde{y}}$ refers to Fig. 6(h).

Clearly, $R$ is severely hampered under the presence of noise in the time series—indeed in the recorded variable. On the other hand, $C$ and $L$ are rather consistent when variables of good observability are considered (i.e., $C_{xy}$ and $L_{xy}$). Despite being robust to the observability issues when no noise is present, $C_z$ is heavily hampered by the addition of noise. Indeed, the presence of high levels of noise masks the windows of periodic behavior, loosening the otherwise clear transition between chaotic and periodic regimes.

**C. Statistical analysis**

We use two statistical analysis methods in order to compare the performance of RNA metrics with respect to the observability properties: the receiver operating characteristics (ROC) and the mutual information (MI).

The ROC curve is a common tool in machine learning used to compare the performance of binary classifiers. It is a graphical representation where the true positive rate (also called sensitivity or recall) is plotted against the false positive rate (also referred as $1 - \text{specificity}$) for different thresholds used in the decision function.

Here, we perform ROC analysis to compare the degree of agreement between the image provided by the “maximum Lyapunov exponent” [Fig. 1(a)] and the RNA measures for each embedding (Fig. 5). To do so, we interpret our images as output of binary classifiers and compare the performance of each embedding with the ground truth as provided by the maximum Lyapunov exponent. Following Ref. 14, we use a fixed value of $\lambda^* = 0.01$ and vary the thresholds of all embedding images. In that sense, we investigate the performance of RNA measures in order to detect the transitions between
FIG. 6. RNA metrics $C$, $R$, and $L$ estimated through the delay embedding from $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ time series of Rössler system under the influence of high noise.

Chaotic ($\lambda_{\text{max}} > \lambda^*$) and periodic ($\lambda_{\text{max}} < \lambda^*$) behaviors according to the observability properties.

Our results for the ROC analysis are summarized in Fig. 7(a). A direct way to use ROC in order to compare classifiers is to measure the area under the curve (AUC). A random classifier will have ROC AUC equal to 0.5, whereas a perfect classifier will have ROC AUC equal to 1. As far as chaotic-periodic transitions are concerned, this statistic confirms that all metrics performed better in the noise-free case when the embedding was performed with a variable of good observability (i.e., with $x$ and $y$). It also highlights that: (i) $L$ is the metric that suffers most to detect dynamical transitions when performed over a reconstructed state-space from a (noise-free) variable of poor observability and (ii) $R$ is very susceptible to the presence of noise, even when variables of good observability are concerned. Note that an AUC $\approx 0.5$, as in $C$, $R$, and $L$, means that the binary classification performs as good as a random classification (compared to Fig. 1). If AUC $< 0.5$, then the classification is inversely proportional to the signal intensities in each image, and $p(a, b)$ is the joint probability distribution from the joint histogram of the signal intensities.

The MI metric will thus measure the amount of information about image $B$ which image $A$ provides. The metric will be high if $I(A; B)$ is a non-negative symmetric metric, that is, $I(A; B) = I(B; A)$ and has boundedness defined as $I(A; B) \leq (H(A) + H(B))/2$, where $H(A)$ and $H(B)$ are the entropies of images $A$ and $B$, respectively.

The MI will thus measure the distance between the joint distribution $p(a, b)$ and the distribution associated to total independence given by $p(a)p(b)$ as per property of the Kullback-Leibler distance. In general, $I(A; B)$ is a non-negative symmetric metric, that is, $I(A; B) = I(B; A)$ and has boundedness defined as $I(A; B) \leq (H(A) + H(B))/2$, where $H(A)$ and $H(B)$ are the entropies of images $A$ and $B$, respectively.

The MI metric will thus measure the amount of information about image $B$ which image $A$ provides. The metric will be high for each pair of images $(A, B)$, consider $A$ as our target image and $B$ as our testing image. The MI is defined as

$$I(A; B) = \sum_{a \in A} \sum_{b \in B} p(a, b) \log \left( \frac{p(a, b)}{p(a)p(b)} \right),$$

where $p(a)$ and $p(b)$ are the marginal probability distributions of the signal intensities in each image, and $p(a, b)$ is the joint probability distribution from the joint histogram of the signal intensities.
when intensities of the joint histogram are highly concentrated in a small number of bins, whereas it will be low when intensities are spread across many bins. We show the results of the MI analysis in Fig. 7(b). While ROC is exclusively concerned with detection of chaotic-periodic transitions (through its binary classification), the MI analysis captures the whole dynamical transitions over the parameter space, including periodic-periodic and chaotic-chaotic ones. From this point-of-view, it shows that \( C \) is the one that suffers the least from the lack of poor observability (for the noise-free case). Some conclusions from the ROC statistic are also perceptible: (i) \( R \) is heavily susceptible to presence of noise and (ii) an overall decay of performance is expected when the variable conveys poor observability and are affected by noise.

These statistics, along with the visual inspection, yield an interesting conclusion: the clustering coefficient \( C \) is the only (studied) metric, that is, both robust to observability issues and presence of noise. While the assortativity \( R \) is also robust to observability issues [as seen in Fig. 5(b)], it is heavily hampered under the presence of noise. On the other hand, although the average shortest path length \( L \) is the most robust to the presence of noise, it is the most susceptible to observability issues [as seen in Fig. 5(b)].

**D. Single-variable RNA: Duffing-Ueda oscillator**

Figures 8 and 9 present, respectively, the RNA results and the statistical analyses for the Duffing-Ueda oscillator. Setting aside the \( R \) metric, the RNA results behave quite similarly to the results shown for the Rössler system: (i) clear dynamical transitions are present when analyzing the original state-space (see results for \( C_x \) and \( L_x \)) or the state-space embedded from a good variable (\( C_x \) and \( L_x \)); (ii) \( L_y \) has a poor performance, hiding the presence of shrimps in an extremely homogeneous chaotic sea; and (iii) \( C \) is shown to be quite robust to the observability issues of the \( y \) variable. As in the Rössler system, Fig. 9(a) shows that \( C \) has the best performance for the original state-space, followed by the good variable \( x \) and then the poor variable \( y \). This is also perceptible through visual inspection of Figs. 8(a)–8(c).

The assortativity metric \( R \) has the worst performance indicator for all three state-spaces, which can be noticed in the blurry results in Figs. 8(d)–8(f). This is not a matter of observability, but rather an intrinsic consequence of the dynamical behavior of the Duffing-Ueda oscillator. The nonautonomous oscillator is driven by an external sinusoidal signal, represented by the \( \phi \) variable, as discussed in Sec. II C 2. Thus, when the system 13 state-space is augmented with a third dimension related to \( \phi \), the oscillator states evolve along a toroidal surface. The linear phase evolution of \( \phi \) increases the continuity of the state-space density, increasing the overall results of \( R \). This increase masks the periodic-chaotic transitions present on the \( xy \)-plane, making the shrimps less visible in the chaotic sea.

**IV. DISCUSSION**

We discuss qualitatively the conclusions raised in Sec. III. For brevity sake, our discussion is focused around the Rössler system, but it can also be extended to our findings for the Duffing-Ueda oscillator.

Poor observability has to do with the location and dimension of singular sets in the Jacobian matrix of the mapping between the reconstructed and original spaces. For instance, the Rössler system has a singularity point on \( z = 0 \), where the attractor is not reconstructable (diffeomorphic) from the \( z \) variable. Only in the absence of singularities there is a global diffeomorphism between the reconstructed and original spaces, which is the case when the Rössler system is reconstructed using the \( y \) variable. However, even if there

![FIG. 9. Performance indicators for the RNA of the Duffing-Ueda oscillator, as described in the caption of Fig. 7.](image)

![FIG. 8. Duffing-Ueda system: RNA metrics clustering coefficient \( C \), assortativity \( R \), and average path length \( L \) estimated through: (a), (d), and (g) the original state-space \( \vec{x} \) and the time delay embedding of each state variable \( x \) [(b), (e), and (h)] and \( y \) [(c), (f), and (i)].](image)
is a singular set, depending on how close and for how long the system state visits its neighborhood, the dynamics of the original system may still be observed from the reconstructed space. That is the case of the $x$ variable of the Rössler system, because although there is a singular plane, that variable conveys good observability of the original dynamics. From observability and embedding theories, the amount of preserved information is reflected in the quality of the reconstructed state-space, whereas a more entangled reconstruction usually points to the choice of a variable that conveys poor observability. It is clear that reconstructions from $x$ and $y$ time series are sufficiently close to the original state-space [see Figs. 2(b) and 2(c)] and, thus, yield similar results for the RNA metrics.

The entangled state-space reconstruction via $z$ [see Fig. 2(d)] is responsible for the poor performance in terms of $L_z$. A more entangled state-space increases the number of shortcuts in the RN, providing an overall smaller $L$ even in chaotic regimes. This approximates the results of $L_z$ in chaotic regimes to the results attained in periodic regimes. Since the detection of chaotic- periodic transition by $L_z$ is not sharp anymore, the shrimps borders in Fig. 2(b) become loosened among the chaotic sea.

On the other hand, the presence of high additive noise decreases the states correlation, especially for chaotic regimes, leading to deceitful results in $R$. Analogously to the $L_z$ metric, the presence of noise approximates the results of $R$ in periodic regimes to those attained in chaotic regimes. Dynamical regimes of periodic and chaotic behavior are not detectable anymore since these zones are blended together, with no clear distinction in Figs. 6(d)–6(f).

In chaotic regimes, the entanglement of the reconstructed state-space via $z$ is especially high in the neighborhood of the singularity point $z = 0$. This increases the cluster formation in these regions and, consequently, the global metric $C_z$. Differently from $R_z$ and $L_z$, the increase of $C_z$ due to embedding of a poor variable is only noticeable over chaotic regimes, being periodic ones rather unaffected when compared to the case with nonembedding. A clear distinction between periodic-chaotic behaviors still remains, showing that $C$ is robust to observability effects with this goal in mind (in the noise-free scenario). When concerning the identification of periodic-chaotic transitions, the robustness of $C_z$ in RNA seems to be related to the robustness of the divergent metric in RQA, found in Ref. 28.

V. CONCLUSION

An established link between embedding and observability theory shows that the state-space reconstruction can be severely hampered by a variable of poor observability. Likewise, Portes et al. showed that, in RQA, most metrics are negatively impacted by reconstructing the state-space via a recorded time series that conveys poor observability. In this work, we follow the framework of Zou et al. to investigate how some highlighted RNA metrics perform under the practical assumption that only a single variable is available for reconstruction. We find that the average shortest path length $L$ and assortativity $R$ do not perform well when the recorded variable available for reconstruction conveys poor observability or the data are corrupted by noise.

On the other hand, the clustering coefficient $C$ is shown to be quite robust to observability issues as long as there is no high level of noise in the time series. Nevertheless, Donner et al. show that the measure $C$ is related to the mean local dimensionality of the corresponding attractor. This relation not only makes it a useful measure to identify the existence of UPOs, but also a proficient tool of RNA to detect dynamic transitions. Assembling these facts, along with its robustness to noise when the variable used conveys good observability, the clustering coefficient $C$ can be depicted as an invaluable tool for experimental analysis of nonlinear systems, especially if a scientist has no knowledge of the observability properties in its studied system.

ACKNOWLEDGMENTS

This work was partly supported by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. L.L.P acknowledges financial support from the Australian Research Council Discovery (Grant No. DP 180100718). A.N.M. and L.L.P acknowledge financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Brazil) (Finance Code 001 and Grant No. 32001010015P8). M.S. acknowledges financial support from the Australian Research Council Discovery (Grant No. DP 180100718). L.A.A. acknowledges financial support from the Conselho de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) (Grant No. 610 302079/2011-4).

REFERENCES