Quantifying robustness and capacity of reservoir computers with consistency profiles

Thomas Lymburn¹[0000–0002–3208–9383], Thomas Jüngling¹[0000–0003–3935–276X], and Michael Small^{1,2}[0000–0001–5378–1582]

¹ Complex Systems Group, Department of Mathematics and Statistics,
The University of Western Australia, 35 Stirling Highway, Crawley, Western
Australia 6009, Australia

² Mineral Resources, CSIRO, Kensington, Western Australia 6151, Australia

Abstract. We study the consistency property in reservoir computers with noise. Consistency quantifies the functional dependence of a driven dynamical system on its input via replica tests. We characterise the high-dimensional profile of consistency in typical reservoirs subject to intrinsic and measurement noise. An integral of the consistency is introduced to measure capacity and act as an effective size of the reservoir. We observe a scaling law in the dependency of the consistency capacity on the noise amplitude and reservoir size, and demonstrate how this measure of capacity explains performance.

Keywords: Reservoir Computing · Consistency Correlation · Signal-to-Noise Ratio · Principal Component Analysis

1 Introduction

The interplay of signal and noise is a well-known concept in engineering, but is much less understood in nonlinear dynamical systems. Every physical system is naturally subject to some form of noise which may be relevant depending on the context. Recently, there has been growing interest in using physical versions of machine learning techniques, due to their potential for increased speed and efficiency. Thus the role of noise in the non-digital physical substrates is an important subject of investigation [18]. In particular, reservoir computing (RC) is a technique that employs a wide range of physical systems for computation [20] and therefore motivates the study of noise effects. In this paper we will apply the concept of consistency, a measure of functional dependence in drive-response systems, to RC. The basic idea of consistency is to drive a nonlinear dynamical system repeatedly with the same signal, and then compare the responses by means of cross-correlation [21]. Hence, the problem of modelling a sophisticated functional relationship is elegantly bypassed [17, 19]. The relevance of consistency for RC has recently been highlighted [2, 13]. We will employ here a high-dimensional extension of the consistency measure to characterise the robustness and the signal processing capacity in reservoirs [12].

Throughout this work we will consider echo state networks (ESNs), a reservoir computing form of recurrent neural networks (RNN) [6, 11]. An ESN is a dynamical system represented as a network. It is typically characterized by untrained and randomly generated weights in both the input and the reservoir. For a network of N nodes with a single input and output, we denote the weight matrices by $\mathbf{V} \in \mathbb{R}^N$, and $\mathbf{W} \in \mathbb{R}^{N \times N}$, respectively. The state vector $\mathbf{x}(t) \in \mathbb{R}^N$ of the network is updated according to

$$\mathbf{x}(t+1) = \tanh(\mathbf{W} \cdot \mathbf{x}(t) + \mathbf{V} \cdot u(t) + \boldsymbol{\beta}), \quad (1)$$

where $u(t)$ is a scalar input sequence with $t \in \mathbb{Z}$ being discrete time, $\boldsymbol{\beta} \in \mathbb{R}^N$ is a vector of biases, and $\tanh(\cdot)$ is a commonly used activation function that is applied elementwise. Only the weights $\mathbf{R} \in \mathbb{R}^N$ of the linear readout $y(t) = \sum_i R_i x_i(t)$ are trained. Typically ridge regression is used to reproduce a target signal $z(t)$, leading to $\mathbf{R} = \operatorname{argmin}_{\mathbf{R}'} \sum_t (\sum_i R'_i x_i(t) - z(t))^2 + \lambda \|\mathbf{R}'\|_2^2$ with λ the ridge parameter.

We apply the consistency measure to quantify the response of ESNs with various network structures. These structures are encoded in the connectivity matrix $\mathbf{W}$ which is usually set to be sparse and random in the typical ESN practise. Here, we will consider a ring (or lattice) structure, sparse random networks, and fully connected networks as reservoirs. In Section 2 we describe how to calculate a consistency profile of the high-dimensional reservoir trajectory. In Section 3 we add intrinsic and measurement noise to the ESNs and investigate the corresponding consistency profiles. In Section 4 we use the consistency profiles to obtain a measure of the network's computational capacity and effective size. We will investigate how this measure depends on the reservoir size and noise amplitude, and relate this to the performance of the reservoir.

2 Consistency profiles

Consistency is a measure of functional dependence in the response of a system to a drive signal [21]. It is determined by driving a system repeatedly with a signal, or driving multiple replicas with an identical driving signal, each starting from different initial conditions. The agreement of the replicas is then measured with a pairwise cross-correlation, the *consistency correlation* [14]. Assuming ergodicity and long enough time series, and denoting the two scalar recordings of the replicas as $x(t)$ and $x'(t)$, the consistency correlation reads $\gamma^2 = \langle \bar{x}(t) \bar{x}'(t) \rangle_t$, where the bar denotes normalisation to zero mean and unit variance, and $\langle \cdot \rangle_t$ is a time average. A key result of consistency theory is that γ poses a fundamental limit on the correlation between the output of the system and any function of the input [9]. The implication of this result in the context of RC is that the performance measured in terms of cross-correlation between readout and target is limited by the consistency level [12]. However, while the readout of a reservoir computer may be scalar, the performance with respect to a general task is determined by its multivariate response time series. The functional dependence of which is characterised by a high-dimensional generalisation of the consistency

correlation, which we call the *consistency profile* or the *consistency spectrum*. It is obtained by a method similar to principal component analysis (PCA) to determine an orthogonal set of characteristic directions with corresponding consistency correlations γ_k^2 , where $k = 1 \dots N$ and N is the dimension of the reservoir. We will briefly recall the theoretical framework of the consistency profile; for a more detailed discussion see Ref. [12].

The initial point of analysis is the stationary N -dimensional response trajectory of the reservoir, $\{x_i(t)\}_{i=1 \dots N}$. We calculate the covariance matrix $[C_{xx}]_{ij} = \langle x_i(t)x_j(t) \rangle_t$ after normalizing to zero mean. Since C_{xx} is positive semi-definite, the eigendecomposition is given by $C_{xx} = Q\Sigma^2Q^\top$. The columns of the orthogonal matrix Q are the directions of the principal components, and the diagonal entries of Σ^2 are their magnitudes, which are the variances in these directions. The transformation $T_o = Q\Sigma^{-1}Q^\top$ yields coordinates in which the reservoir response is isotropic, see Fig. 1. To ensure numerical stability of this transformation in the presence of rank deficiencies, we add a small regularization term $10^{-9} \times \mathbf{1}$ to C_{xx} , with $\mathbf{1}$ denoting the $N \times N$ identity matrix.

Fig. 1. An illustration of the PCA description of the consistent components (red) and full response (black). The directions of maximum and minimum consistency are shown in blue. Panel (a) shows the analysis in the original coordinates and panel (b) shows it in the normalized coordinates after the transform T_o .

The concept of consistency is similar to the signal-to-noise ratio in signal processing, except that consistency is tailored to the response of nonlinear dynamical systems. The core idea is to decompose a signal from a driven dynamical system into a consistent component and an inconsistent component which are linearly independent, $x(t) = s(t) + n(t)$. The term $s(t)$ represents a function of the input signal, and $n(t)$ is noise that does not depend on the input. Given all quantities have zero mean, the consistency correlation is the ratio of variances $\gamma^2 = \langle s^2(t) \rangle_t / \langle x^2(t) \rangle_t$. Extending this result to the high-dimensional reservoir trajectory in the normalized coordinates $\mathbf{x}_o(t) = T_o \mathbf{x}(t)$, the covariance matrix of the consistent component is – under ergodicity and in the limit of long time series – equal to the cross-covariance matrix of two replica [12],

$$[C_c]_{ij} = \langle s_{o,i}(t)s_{o,j}(t) \rangle = \langle x_{o,i}(t)x'_{o,j}(t) \rangle. \quad (2)$$

The consistent (signal) components $\mathbf{s}_o(t)$ form an ellipsoid within the isotropic full response $\mathbf{x}_o(t)$. By performing the decomposition $C_c = Q_c G^2 Q_c^\top$ we obtain the principal components as the orthogonal set of consistency directions. The consistency profile is given by the diagonal elements γ_k^2 of the matrix G^2 , where the index k enumerates the consistency components.

3 Noise in ESN

We will use the consistency profiles to measure how different network structures deal with noise. This will be done for three different network structures (Fig. 2), subject to either intrinsic i.i.d. noise according to Eq. (3), or observational i.i.d. noise as in Eq. (4) (with $\mathbf{x}(t)$ from Eq. (1)). Each component of the noise is denoted as $\nu_i(t) \sim \mathcal{N}(0, 1)$, and $\Omega > 0$ is the noise strength (‘amplitude’). The driving signal $u(t)$ is set to be scalar Gaussian white noise. Other choices of stationary input signals are possible, like for example in Sec. 4.2, and will result in different consistency profiles. The noisy trajectories for each case are

$$\tilde{\mathbf{x}}_{int}(t+1) = \tanh(\mathbf{W} \cdot \tilde{\mathbf{x}}_{int}(t) + \mathbf{V} \cdot u(t) + \boldsymbol{\beta} + \Omega \boldsymbol{\nu}(t)) , \quad (3)$$

$$\tilde{\mathbf{x}}_{obs}(t) = \mathbf{x}(t) + \Omega \boldsymbol{\nu}(t) . \quad (4)$$

Fig. 2. Network types studied here. (a) ring topology, (b) sparse random connectivity and (c) fully connected.

The network types that we use as reservoirs are the unidirectional ring (Fig. 2(a)), sparsely connected random networks (Fig. 2(b)), and fully connected networks (Fig. 2(c)). For the sparse random networks, each potential directed link between pairs of nodes is drawn randomly with a probability of connection given by $p = 10/N$, which ensures that the average degree is 10 regardless of network size. Aside from the adjacency structure, all networks are generated with random internal weights drawn from a Gaussian distribution then scaled to achieve a spectral radius of one. The input connections are chosen uniformly from the interval $[-1, 1]$. All of these structures, including the simpler ring topology, are capable of performing nontrivial tasks [1, 4, 16].

Consistency profiles for these networks are shown in Fig. 3 for a selected network size of $N = 200$. There are no qualitative differences for other reservoir

sizes over several orders of magnitude ($N < 2000$). For both forms of noise there is a wide range of consistency levels in the corresponding principal components. Some components are extremely sensitive to noise, while others maintain almost complete consistency for the displayed noise levels. These consistent components (directions in state space) explain how a reservoir computer can be robust to noise. We also note that there is a significant difference between the profiles of the ring topology and the more highly connected networks, with the ring topology being more robust to noise. We speculate that the remarkable agreement between the profiles of the sparsely connected and the fully connected networks might be related to the average shortest path length.

Fig. 3. The consistency profiles of ESNs with different structures when subject to observational and intrinsic noise. Panels (a-b) show the profiles of a ring network, (c-d) show a sparsely connected network and (e-f) show a fully connected network. The profiles for different noise levels are shown, with $\Omega = 0.01$ shown in red, $\Omega = 0.05$ in green and $\Omega = 0.1$ in blue. All networks have $N = 200$.

4 Capacity

One way to assess the quality of an information processing system is to determine its capacity to compute a range of nonlinear functions. The more capacity a system has, the broader the range of functions it has access to in order to synthesise some target function, and thus the more accurate it may be. There is

no unified measure of the capacity of an information processing system, though several have been proposed [3, 5, 7]. Here we will derive a capacity measure from the consistency profile, which is particularly useful in the presence of noise or chaos in the reservoir.

Given the typically linear readout, a naive measure of capacity of a fully consistent reservoir is given by the number of linearly independent functions in its response trajectory. This number is the rank of the covariance matrix of the multivariate time series and is limited by the number of nodes N . However, this simple capacity measure does not take into account the quality of the set of functions, nor does it apply in the presence of noise.

In the context of consistency, we will not investigate the suitability of the reservoir functions for a particular task, but instead measure how each function is attenuated by noise, and how this affects the performance in general. In the following, all quantities will be normalised to zero mean, and — unless stated otherwise — also to unit variance, to allow for compatibility with standard measures of correlation and covariance. A bias neuron will consequently not be introduced in addition to the reservoir activations. Moreover, all time series are considered to be stationary and sufficiently long to approximate limit values, and to leave aside discussion on statistical aspects including training, validation, and testing methodology. Let $z(t)$ be an arbitrary scalar target signal with $\langle z \rangle = 0$ and $\langle z^2 \rangle = 1$, and $u(t)$ the corresponding scalar input signal, with $\langle u \rangle = 0$ and $\langle u^2 \rangle = 1$, and $\langle \cdot \rangle$ being a time average. We further assume that the target depends on the input via a functional dependence $z(t) = \mathcal{F}[u](t)$, meaning that the present value of $z(t)$ can be completely determined by using the values of $u(t')$ for $t' \leq t$, eventually up to some maximum lag $t - t_m < t'$. Let the reservoir response to the input $u(t)$, including any form of noise, be $\{\tilde{x}_i(t)\}_{i=1\dots N}$ with $\langle \tilde{x}_i \rangle = 0$. Applying the normalisation $T_o = Q\Sigma^{-1}Q^\top$ from the PCA of $\tilde{\mathbf{x}}(t)$ yields the isotropic response $\mathbf{x}_o(t) = T_o\tilde{\mathbf{x}}(t)$. In these coordinates we calculate the cross-covariance matrix C_c to determine the consistent components. From the eigendecomposition $C_c = Q_c G^2 Q_c^\top$, we use the orthogonal transformation $\boldsymbol{\xi}(t) = Q_c^\top \mathbf{x}_o(t)$ to align the main axes of the consistent ellipsoid with our coordinate system. Let the consistent components be $\gamma_k s_k(t)$ with $\langle s_k \rangle = 0$ and $\langle s_k^2 \rangle = 1$, where the consistency correlations γ_k^2 are the diagonal elements of G^2 because of the previous normalisation procedure. Using the least-mean-squares expression for the readout vector $\mathbf{R}$ in the $\boldsymbol{\xi}$ -coordinates, we find that $R_k = \gamma_k \langle s_k z \rangle$, because any correlation from the target can only stem from the consistent components. Denoting the correlations in R_k as $a_k = \langle s_k z \rangle$, we can write the output of the reservoir as

$$y(t) = \sum_k \gamma_k a_k \xi_k(t) . \quad (5)$$

The output has zero mean by construction, and the variance can be determined as $\langle y^2 \rangle = \sum_k \gamma_k^2 a_k^2$, using $\langle \xi_k \xi_l \rangle = \delta_{kl}$. From this, we can calculate the mean-square error (MSE) (which here is equal to the normalised MSE), and the cross-correlation C_{yz} with the target, as $\text{MSE} = 1 - \sum_k \gamma_k^2 a_k^2$ and $C_{yz}^2 = \sum_k \gamma_k^2 a_k^2$, respectively. The recurring expression here is the sum of consistency correlations

Fig. 4. Capacity for networks of different sizes noise amplitudes (b). Panels (a-b) are for observational noise and (c-d) are for intrinsic noise. For panels (a) and (c) the noise amplitude is fixed at $\Omega = 0.5$ and for panel (b) and (d) the network size is $N = 500$. The results are averaged across 10 different realisations of the network structures. Dependencies in (c) and (d) are truncated to display only the reliable range with respect to systematic errors from finite-size estimates.

γ_k^2 and squared target correlations a_k^2 . Considering these in general to be independent, because the former is a measure of the reservoir response only whereas the latter depends on the task, we define the *consistency capacity* as

$$\Theta = \sum_{k=1}^N \gamma_k^2 = \text{Tr}(C_c). \quad (6)$$

The latter identity comes from the fact that the orthogonal transformation Q_c does not change the trace. By construction, the capacity is $0 \leq \Theta \leq N$. As the size of a network with full rank gives the number of functions it produces, this measure of capacity can be thought of as the effective network size for reservoirs with noise, or chaos acting as noise.

4.1 Capacity and noise

We calculate numerically the capacity of ESNs with the different structures described above for a range of N and Ω , and for both intrinsic and observational

noise. Figures 4(a)-(c) show the dependence of the capacity Θ on the reservoir size N for the three network types and fixed noise strength. The capacity increases monotonically with reservoir size, and a power law is indicated for all networks. The power-law exponent of the ring network appears to be slightly larger than that of the other two types, which behave remarkably similarly. All exponents are less than one, in agreement with the upper bound $\Theta \leq N$, while the majority are close to $\Theta \propto \sqrt{N}$. In Figures 4(b)-(d), we show for a fixed size N the dependence of Θ on the noise strength Ω . While for vanishing noise the capacity is $\Theta = N$, it decays monotonically with increasing noise. A power-law scaling is indicated for moderate noise levels, most pronounced for the sparse random and the fully connected reservoirs. For observational noise, we show how this can be traced back to scaling behavior in the principal components of a noise free reservoir.

In the case of uniformly additive measurement noise, the consistency spectrum and capacity can be determined analytically. Let the fully consistent time series of a reservoir be $x_i(t)$, with $\langle x_i \rangle = 0$ for each i , and the noisy observations $\tilde{x}_i(t)$ according to Eq. (4). If the eigenvalues (or singular values) of the covariance matrix C_{xx} of the noise-free time series are known, here denoted as σ_k^2 , the noise intensity is added to the variance of each of these signal components separately. Thus the covariance matrix of the noisy time series reads $C_{\tilde{x}\tilde{x}} = C_{xx} + \Omega^2 \mathbf{1}$. The consistency correlations follow then as the noise-free variance divided by the total variance for each of the original principal components independently

$$\gamma_k^2 = \frac{\sigma_k^2}{\sigma_k^2 + \Omega^2}. \quad (7)$$

It is worth mentioning that the noise amplitude can be related to the ridge parameter λ if regularisation is applied. For time series of length L , the identity $\lambda = \Omega^2 L$ holds. Thus if the coefficients a_i for $\lambda = 0$ are known, the ridge trace can be determined using the readout in Eq. (5) and the consistencies in Eq. (7).

Fig. 5. Scaling of noise-free principal components of a sparse random network reservoir with $N = 300$ nodes, and interaction with additive measurement noise of strength Ω .

Figure 5 shows the principal components σ_k^2 for a noise-free reservoir of the sparse random type. A large part of the spectrum appears to scale with the power law $\sigma_k^2 \propto k^{-2}$. To our knowledge, this behavior has not previously been documented. We do not aim at a theory of this scaling here, but consider it as a phenomenological premise for the scope of our investigation. This implies a level of caution in the analysis given the uncertainty about the validity and range of the observed scaling behavior. In particular, the different reservoir types indicate deviations from the scaling behavior, which thus may be a specific property of a certain class of reservoirs. We also note that the detailed scaling depends on the type of input signal. Assuming the indicated power law holds strictly, we can apply Eq. (7). We set $\sigma_k = bk^{-1}$, with $b > 0$ depending on N , and calculate an infinite sum to account for the quickly decaying tail of the spectrum. We obtain $\Theta = (b\pi/2\Omega) \coth(b\pi/\Omega) - 1/2$. This expression contains two regimes. For $\Omega \ll b$, $\Theta \propto b/\Omega$, and for $\Omega \gg b$, $\Theta \propto b^2/\Omega^2$. The former reveals the power law indicated in Fig. 4(b). An intuitive explanation of how it emerges from the noise-free scaling $\sigma_k \propto k^{-1}$ is given as follows. By increasing the noise level Ω the signal components σ_k are successively covered by noise, as illustrated in Fig. 5. As long as the σ -spectrum is only partially covered, the spacing between the components determines the rate by which a noise increment translates into a change of capacity, such that the σ_k scaling determines the $\Theta(\Omega)$ scaling. The latter power law is not shown in the figures. It emerges when noise exceeds the largest signal component, σ_1 , so that the above mechanism of successive component coverage ceases. Hence, in Eq. (7) all $\gamma_k^2 \approx \sigma_k^2/\Omega^2$, from which follows that $\Theta \propto \Omega^{-2}$. Note that the analytical model does not account for the case of vanishing noise $\Omega \rightarrow 0$, where the predicted capacity diverges instead of saturating to the reservoir size, $\Theta \rightarrow N$. This is due to the infinite sum approximation which requires $\gamma_k \rightarrow 0$ for large k and therefore is not valid as the reservoir becomes fully consistent. The model can be refined by truncating the sum, however, this does not lead to further insight.

Another observation from the noise-free principal components is a scaling with the reservoir size by $\sigma_k^2 \propto N$ (not shown). This is explained by the increased overall signal power due to the input weights vector $\mathbf{V} \in \mathbb{R}^N$, under the assumption that all nodes in the reservoir are highly correlated. In contrast, if the input power was kept constant, e.g. by setting the number of non-zero input weights to N_0 independently of N , the principal components would remain approximately the same throughout the scaling regime. Introducing the N -scaling in the analytical model by setting $b^2 \propto N$, we obtain $\Theta \propto \sqrt{N}$ in the moderate noise regime. This agrees with the observations in Fig. 4(a).

In summary, the scaling law $\Theta \propto \sqrt{N}/\Omega$ was derived using the observation $\sigma_k^2 \propto N/k^2$ from sparse random networks and measurement noise. It holds for this network type very well, especially for large reservoirs, but also for the fully connected network. The behavior of the ring reservoir deviates slightly, although the findings still agree well. Remarkably, the same behavior is indicated for intrinsic noise, even though this noise type is qualitatively different from the observational noise, for which the exact relationship Eq. (7) holds. We conclude

that the generating mechanisms are robust and explain a generic behavior of the consistency capacity of reservoirs with noise.

4.2 Capacity and performance

The size of a reservoir computer is known to be an important factor of its performance, with common practise being to create a reservoir as large as computationally feasible [10]. As the consistency capacity Θ gives a measure of the effective size of a reservoir subject to noise, it is expected to have some predictive power of the performance.

To confirm the relationship between performance and capacity, we apply our reservoirs to a chaotic time series prediction task on which reservoir computers are known to perform well [8, 15]. We will predict the Lorenz system and measure the relationship between capacity and the (normalised) MSE. The input is chosen to be the X -variable such that $u(t) = X(t\Delta)$ and the target is the X -variable τ time steps in the future, $z(t) = X((t + \tau)\Delta)$, where $\Delta = 0.1$ and $\tau = 11$, corresponding to approximately one Lyapunov time. We will perform this task for reservoirs with sizes $N \in [20, 300]$ and with both observational and intrinsic noise with amplitudes $\Omega \in [0, 0.2]$.

Fig. 6. Performance on the Lorenz prediction task given by NMSE for reservoirs of various sizes with various levels of noise added, leading to a range of capacities. The red data points correspond to the reservoirs without added noise. Panel (a) shows the results for a ring network, panel (b) shows the results for a sparsely connected network. The results for the fully connected network are similar to the sparsely connected network and not shown.

The results for ESNs with the ring topology and sparse random connectivity are shown in Figs. 6(a) and 6(b), respectively. Firstly, we notice that the performance is monotonic with the capacity, with some variation due to the different random realisations. Within that variation, all outcomes collapse to a

single curve. This implies that the consistency capacity is a valid predictor of performance. Second, there is a sharp drop in the NMSE as Θ increases from 0 to 50, after which the performance improves gradually. The implication is that a small fraction of the degrees of freedom are responsible for a significant portion of the reservoir's computation. Finally, the performance of the reservoirs with capacity reduced by noise is similar to the performance of a reservoir without noise but with similar Θ (assuming full rank, $\Theta = N$ for $\Omega = 0$, though large ring networks may have a rank-deficient response). This means that the effect of noise in a reservoir is equivalent to removing some of the degrees of freedom and reducing the reservoir size.

5 Conclusion

We have demonstrated how consistency profiles describe reservoir computers with different forms of noise. By doing so we have shown how reservoirs with noise may still perform useful computation, and that the simple degree-one ring network is more robust to noise than networks with higher degrees.

Beyond understanding robustness to noise, the consistency profiles can be used for a measure of the capacity of the reservoir. This capacity measure applies to reservoirs with and without noise, and can be thought of as the effective size of the reservoir. We find that the capacity measure scales as $\Theta \propto \sqrt{N}/\Omega$ and, regardless of the level or type of noise, serves as a performance predictor.

In this work we have only considered a small subset of network structures and Gaussian white noise. The consistency profile may provide interesting insights into the effect of scale-free topologies, local structures, different ESN hyperparameters and different forms of noise. We have also focused on the restricted training algorithm of reservoir computing. The consistency profile may give more insights into the functions produced by RNN with trained internal weights.

Acknowledgements

T.L. is supported by the Australian Government Research Training Program at The University of Western Australia. T.J. acknowledges fruitful discussions about the concept of consistency, as well as inspiring insights into the field of reservoir computing, with Ingo Fischer and members of the Nonlinear Photonics Group during his time at the IFISC (UIB-CSIC) in Palma de Mallorca. M.S. is supported by the Australian Research Council Discovery Project (Grant No. DP180100718).

References

1. Appeltant, L., Soriano, M., Van der Sande, G., Danckaert, J., Massar, S., Dambre, J., Schrauwen, B., Mirasso, C., Fischer, I.: Information processing using a single dynamical node as complex system. *Nature Comm.* **2**, 468 (2011)

2. Bueno, J., Brunner, D., Soriano, M.C., Fischer, I.: Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback. *Opt. Expr.* **25**(3), 2401 (2017)
3. Carroll, T.L., Pecora, L.M.: Network structure effects in reservoir computers. *Chaos* **29**(8), 083130 (2019)
4. Dale, M., Dewhurst, J., O’Keefe, S., Sebald, A., Stepney, S., Trefzer, M.: The role of structure and complexity on reservoir computing quality. In: McQuillan, I., Seki, S. (eds.) *UCNC 2019: Unconventional Computation and Natural Computation. Lecture Notes in Computer Science*, vol. 11493. Springer, Cham (2019)
5. Dambre, J., Verstraeten, D., Schrauwen, B., Massar, S.: Information processing capacity of dynamical systems. *Sci. Rep.* **2**, 514 (2012)
6. Jaeger, H.: The ”echo state” approach to analysing and training recurrent neural networks-with an erratum note. German National Research Center for Information Technology GMD Technical Report **148** (2001)
7. Jaeger, H.: Short term memory in echo state networks. German National Research Center for Information Technology GMD Technical Report **152** (2002)
8. Jaeger, H., Haas, H.: Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**(5667), 78 (2004)
9. Jüngling, T., Soriano, M.C., Oliver, N., Porte, X., Fischer, I.: Consistency properties of chaotic systems driven by time-delayed feedback. *Phys. Rev. E* **97**, 042202 (2018)
10. Lukoševičius, M.: A practical guide to applying echo state networks. In: Montavon, G., Orr, G.B., Müller, K.R. (eds.) *Neural Networks: Tricks of the Trade*. vol. 7700. Springer, Berlin, Heidelberg (2012)
11. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**(3), 127 (2009)
12. Lymburn, T., Khor, A., Stemler, T., Corrêa, D.C., Small, M., Jüngling, T.: Consistency in echo-state networks. *Chaos* **29**(2), 023118 (2019)
13. Nakayama, J., Kanno, K., Uchida, A.: Laser dynamical reservoir computing with consistency: an approach of a chaos mask signal. *Opt. Expr.* **24**(8), 8679 (2016)
14. Oliver, N., Jüngling, T., Fischer, I.: Consistency properties of a chaotic semiconductor laser driven by optical feedback. *Phys. Rev. Lett.* **114**, 123902 (2015)
15. Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E.: Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102 (2018)
16. Rodan, A., Tino, P.: Minimum complexity echo state network. *IEEE Trans. Neural Netw.* **22**(1), 131 (2011)
17. Schumacher, J., Haslinger, R., Pipa, G.: Statistical modeling approach for detecting generalized synchronization. *Phys. Rev. E* **85**, 056215 (2012)
18. Semenova, N., Porte, X., Andreoli, L., Jacquot, M., Larger, L., Brunner, D.: Fundamental aspects of noise in analog-hardware neural networks. *Chaos* **29**(10), 103128 (2019)
19. Soriano, M.C., Van der Sande, G., Fischer, I., Mirasso, C.R.: Synchronization in simple network motifs with negligible correlation and mutual information measures. *Phys. Rev. Lett.* **108**, 134101 (2012)
20. Tanaka, G., Yamane, T., Héroux, J.B., Nakane, R., Kanazawa, N., Takeda, S., Numata, H., Nakano, D., Hirose, A.: Recent advances in physical reservoir computing: A review. *Neural Netw.* **115**, 100 (2019)
21. Uchida, A., McAllister, R., Roy, R.: Consistency of nonlinear system response to complex drive signals. *Phys. Rev. Lett.* **93**, 244102 (2004)