Reservoir Sizes and Feedback Weights Interact Non-linearly in Echo State Networks

Danil Koryakin and Martin V. Butz

Cognitive Modeling, Department of Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany

Abstract. In this paper we investigate parameter dependencies in the echo state network (ESN). In particular, we investigate the interplay between reservoir sizes and the choice of the average absolute output feedback connection weight values (W_{OFB}). We consider the multiple sine wave oscillator problem and the powered sine problem. The results show that somewhat contrary to basic intuition (1) smaller reservoir sizes often yield better networks with higher probability; (2) large W_{OFB} values paired with comparatively large reservoirs may strongly decrease the likelihood of generating effective networks; (3) the likelihood of generating an effective ESN depends non-linearly on the choice of W_{OFB}: very small and large weight values often yield higher likelihoods of generating effective ESNs than networks resulting from intermediate W_{OFB} choices. While the considered test problems are rather simple, the insights gained need to be considered when designing effective ESNs for the problem at hand. Nonetheless, further studies appear necessary to be able to explain the actual reasons behind the observed phenomena.

Keywords: echo state network, output feedback, size of dynamic reservoir.

1 Introduction

Echo state networks (ESN) are a special kind of the artificial neural networks (ANNs). They are a kind of reservoir computing network, which is a young and intensively investigated area of ANNs. ESNs possess features that make them unique compared to other ANN architectures. The most prominent attribute of ESNs is the simplicity of setting-up and training such a network. The network only adapts the connection weights to the output neurons by linear regression.

Despite their simplicity, ESN have been successfully applied to robotics tasks ([2], [3]) and to speech recognition ([4], [5]). Like standard RNNs, ESNs are often applied to modeling the time-series of typical benchmark problems. One of them is the multiple superimposed oscillator (MSO) considered, for example, in [6]. Despite its visible simplicity, it is widely accepted that it is hard for solving with standard ESNs. Several modifications of the standard ESNs ([7] and [8]) were successfully applied to solve this problem.

A.E.P. Villa et al. (Eds.): ICANN 2012, Part I, LNCS 7552, pp. 499–506, 2012.

In general the basic scheme of the ESNs is relatively tolerant to the choice of their parameters. However, in [1] Herbert Jaeger, the creator of ESNs, underscores the importance of a proper parameter choice: "*However, it should be emphasized that a rough optimization of these few parameters is crucial and task-specific.*" (p. 34) In his work [1], [2], and [10] Jaeger gave several useful recommendations for the correct choice of the ESN parameters. While the *spectral radius* is clearly responsible for ensuring the echo-state property – essentially preventing activity blow-up in the dynamic reservoir – also network size has been mentioned as an important optimization candidate. Nonetheless, also other parameters must be handpicked if the highest ESN performance is strived-for.

In the current work we study the dependency between two macro parameters of ESNs: (a) the reservoir size and (b) the interval for initializing the output feedback weight (W_{OFB}) values. The output feedback weights determine the influence of the current time series dynamics on the unfolding dynamics inside the reservoir. Thus, it is apparent that the output feedback weight values need to be chosen with care to avoid overruling the unfolding dynamics inside the reservoir. The obtained results show that the reservoir size must not be chosen too large to ensure the generation of very accurate ESNs with high probability. Moreover, the results show that large W_{OFB} values paired with large dynamic reservoirs may decrease the likelihood of generating effective ESNs severely. Finally, the likelihood of generating good ESNs depends on the choice of the W_{OFB} interval non-linearly. These results are obtained exemplary on several related, rather small benchmark problems. Thus, future work needs to evaluate the generality of the conclusions - especially for larger and more complex time series prediction problems. However, the insights gained ask more detailed studies and give immediate recommendations on how ESNs should be initialized in the search for effective networks.

2 ESNs

Echo State Networks (ESNs) are reservoir computing networks that maintain an internal reservoir of neural activities over time. This reservoir determines the output of the ESN via output connections. Only the weights of these output connections are trained by means of the least mean square algorithm – or generally any other linear regression algorithm that is guaranteed to minimize the mean squared error between the target and generated output values [10].

Detailed descriptions of ESNs can be found elsewhere, e.g. [1,10]. Here, we focus on the crucial aspects that we manipulated in this paper. According to our experience with ESNs, it is indispensable that a balance is maintained between the reservoir dynamics and the excitation signal coming from the "outside". In ESNs without input neurons, the output feedback connections are the only source of external excitation. Its contribution to the activity of a particular reservoir neuron i can be quantified as

$$C_i^{OFB} := w_{iy}y, \tag{1}$$

where y is the output value in the previous iteration and w_{iy} the output feedback connection weight. The contribution of the reservoir dynamics, on the other hand, can be quantified as

$$C_i^{RD} := \sum_{j=1}^{N} w_{ij} a_j, \tag{2}$$

where a_j is the activity of reservoir neuron j within the dynamic reservoir of size N, and w_{ij} is the connection weight from neuron j to the neuron in question i.

Elsewhere, we have proposed that the balance between both contributions is established when

$$\sigma^2(C_{RD}) \approx \sigma^2(C_{OFB}) \tag{3}$$

where $\sigma^2(X)$ is the variance of the output feedback or the internal reservoir [9]. The condition requires that the external signal does not dominate the reservoir dynamics on average. If $\sigma^2(C_{OFB})$ was systematically larger than $\sigma^2(C_{RD})$, then the external signal would mostly dominate the internal dynamics, preventing longer-term internal dynamics to unfold. On the other hand, if $\sigma^2(C_{OFB})$ was systematically smaller than $\sigma^2(C_{RD})$, then the feedback signal may not be strong enough to excite the internal dynamics, resulting in progressive activity decline.

Since $\sigma^2(C_{OFB})$ directly depends on the target dynamics and $\sigma^2(C_{RD})$ indirectly depends on the target dynamics, it is however generally impossible to assure the proposed balance. Thus, it seems vital to investigate the interaction between crucial parameters in typical benchmark problems, to reveal parameter interdependencies. Here, we focus on the dependency between reservoir size and output feedback weights.

3 Experimental Study

In the following experiments we evaluate the performance of ESNs on three time series functions. In all reported results, 500 ESNs were randomly generated for every considered parameter setting. The considered ESNs did not contain any input neurons and one output neuron. Thus, the dynamic reservoir was driven only by the output feedback via the connections from the single output neuron to every reservoir neuron.

The ESN reservoirs had a random connectivity of 40%. The connections with non-zero weights were randomly distributed in the reservoir. To determine the recurrent connection weights, the non-zero reservoir weights were assigned random values uniformly distributed in the interval $[-1, +1]$. Next, the weights were scaled to yield a spectral radius of 0.8 within the network. The reservoir neurons had TANH as their activation function. Every reservoir neuron was connected to the output neuron, which had a linear activation function. There was no self-recurrent connection for the output neuron. All output feedback weights – those that result into the weighted projection of the activity of the output neuron back into the dynamic reservoir – were assigned random values uniformly distributed in the interval $[-W_{OFB}, +W_{OFB}]$.

To investigate the behavior of the feedback weights, we chose the MSO2, PowerSin7, and PowerSin11 as the target dynamics. The MSO2 is a sum of two sine waves of different frequencies. It is defined as

$$y(t) = \sin(0.2t) + \sin(0.311t), \tag{4}$$

where t is the time step index. Due to the superimposed sine waves, the period of the resulting function is rather large. Thus, the best solution for an ESN would be to have both sinusoidal dynamics present within the reservoir and combine them in an additive way in the output neuron.

The two PowerSin functions, on the other hand, are more simple periodic functions based on a sinusoidal. They exhibit a stronger difficulty due to a higher degree of non-linearity caused by the power function. They are computed as:

$$y(t) = \frac{1}{2} \sin^x(\frac{2\pi t}{T}) \tag{5}$$

where T is the period and x is a degree of non-linearity. We set the period T to 10π. The same value was used in the works of [1] and [11], where the function with $x = 7$ was considered before. In our experiments we varied the degree of non-linearity to check its impact on the ESN behavior. We compare the originally used value of $x = 7$ with $x = 11$, and refer to the two target dynamics with PowerSin7 and PowerSin11, respectively.

A sequence of 700 time steps was generated for every target dynamics. The first 100 steps were used for the washout. The next 300 steps were used for training the output weights. Finally, 300 test steps followed. This sequence was used to train and test every generated ESN. When applying an ESN to the sequence, the states of all its reservoir neurons and the state of its output neuron were set to zero. During the washout and the training phases, "teacher-forcing" was applied, thus feeding the target output activity back into the network. The states of the reservoir neurons were collected to compute the output weights according to the procedure described in [10]. After the determination and assignment of the optimal output weights, the ESN was run 300 further steps without teacher-forcing, thus feeding the predicted output back into the dynamic reservoir. The difference between the predicted output values and the corresponding target output values was evaluated further.

To evaluate the resulting ESNs, we distinguish between *well-performing* and *ill-performing* ESNs. To do so, we define a *small error length* (SEL) criterion, which denotes the number of consecutive time steps in the test phase during which the ESN output does not differ from the target dynamics by more than a predefined threshold. In the following evaluations, we always set this threshold to 1% of the maximum value.

3.1 Reservoir Size Interacts with W_{OFB}

To evaluate the interaction between reservoir sizes and the range W_{OFB} of output feedback weight values, we systematically varied the reservoir sizes and output feedback weights. We now first report the most suitable W_{OFB} values identified. These were deduced by considering 29 reservoir sizes for each target function (reservoir sizes $3, 4, ..., 20, 30, ..., 100, 150, 200, 250$) and considering for each of these reservoir sizes 50 different W_{OFB} values $(0.1, .001, ..., 10^{-50})$. For each of the setting, 500 networks were generated randomly and the number of well-performing networks, as defined above, were recorded.

In this section, we report the most suitable W_{OFB}, which was defined as the W_{OFB} setting that yielded the largest number of networks that satisfied the SEL criterion for each particular reservoir size. The results for all three target dynamics are shown in Fig. 1 (top). For all three target dynamics, the range of suitable W_{OFB} values comprises many orders of magnitude. In smaller reservoirs, relatively large W_{OFB} settings yield the largest number of well-performing ESNs. For larger reservoir sizes, however, the most promising W_{OFB} values drop down to values as low as 10^{-40} for comparatively large reservoirs. Within the range of small reservoir sizes, where larger W_{OFB} values yielded better performance, a slight increase in the W_{OFB} settings with increasing reservoir sizes was also observable. Below, we show that the distribution of the likelihoods of generating well-performing reservoirs progressively shifts from those with larger W_{OFB} values to those with very small values.

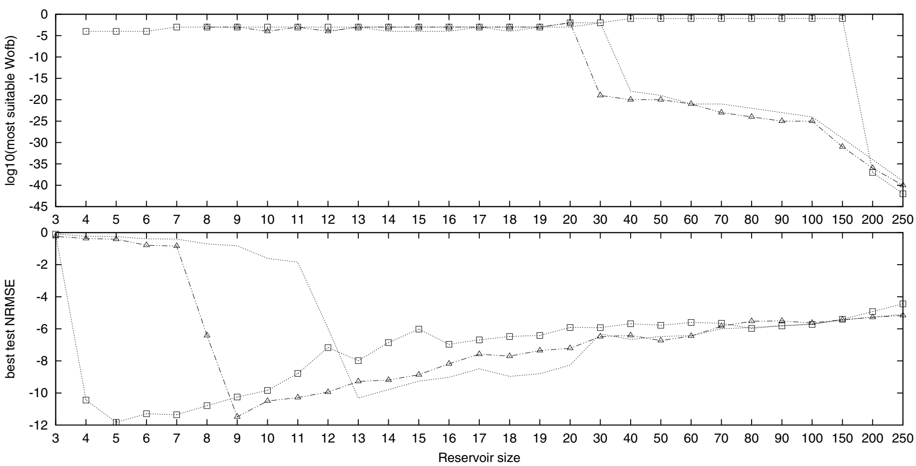

Fig. 1. Most suitable W_{OFB} (top) and the best NRMSE on the test sequence (bottom). The curves of MSO2 are shown lines with squares, PowerSin7 - lines with triangles and PowerSin11 - with solid lines. On all dynamics the smallest error was reached when W_{OFB} were relatively large, about 10^{-4} to 10^{-3}.

Before investigating the distribution of the likelihoods of generating well-performing ESNs dependent on W_{OFB} further, however, we report the normalized root-mean squared error values for the best performing network in the respective settings. Fig. 1, bottom, shows that smaller reservoirs with larger W_{OFB} settings provided the most accurate ESNs. The results show that a critical minimal size of four neurons appears necessary to generate a well-performing

ESN for the MSO2 problem with an NRMSE significantly below 0.01, a reservoir size of eight is necessary for the PowerSin7 function, and a reservoir size of 13 is necessary for the PowerSin11 function. This suggests that a higher non-linearity requires a larger reservoir size. Note that the obtained NRMSE of 1.472×10^{-12} for MSO2 is several orders of magnitude lower than the best results reported in the literature we are aware of, which were obtained in [6] with a performance of 3.92×10^{-8}. Also for the PowerSin7, we obtained a comparatively low NRMSE value of 3.272×10^{-12}.

3.2 W_{OFB}-Dependent ESN Performance

While the results above suggested a sudden drop-off of the most suitable W_{OFB} setting, we now provide further details on the actual number of well-performing networks generated for all W_{OFB} setting for a particular target function and reservoir size. Fig. 2 shows these results for reservoir sizes 20 and 70. The results show that the likelihoods of generating a well-performing reservoir, systematically varies with the W_{OFB} setting. However, this variation is non-linear yielding a "two-hill" distribution. Gaussian-like distributions can only be found for problem-respective very small or very large reservoir sizes. This can be seen for

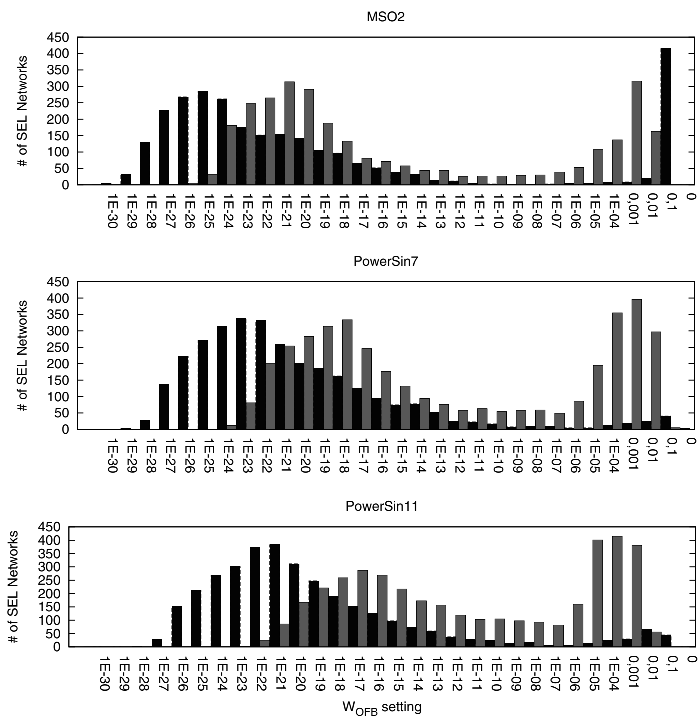

Fig. 2. Distribution of well-performing networks (those that satisfy the SEL criterion with maximally 1% error) for the MSO2, PowerSin7, and PowerSin11 target functions for reservoir sizes 20 (gray) and 70 (black)

the reservoir size 70 for the PowerSinX functions to some degree - increasing the size even further yields only well-performing networks for very small W_{OFB} settings. In the case of very small networks, only comparatively large W_{OFB} settings yield well-performing networks (distribution not shown).

For the case of small W_{OFB} settings, it is apparent that larger networks demand even smaller W_{OFB} values to reach the highest likelihood of generating a well-performing network. For the case of large W_{OFB} settings, on the other hand, larger networks require slightly larger W_{OFB} values – until saturation is reached when W_{OFB} values reach one. The complexity of the function is reflected in these distributions as well. Comparing the PowerSin7 with the PowerSin11 distributions, the slightly higher difficulty of the PowerSin11 function is apparent in that W_{OFB} values need to be slightly higher for generating the PowerSin11 dynamics. The MSO2 with its two dynamics but lower local non-linearities, on the other hand, yields well-performing networks still for much larger reservoir networks with large W_{OFB} settings.

4 Conclusions

In this study we revealed a non-linear dependency between the strength of the driving signal, which is controlled by the output feedback weight ranges W_{OFB}, and the reservoir size. When considering the case of system identification, in which ESN is trained to reproduce particular target dynamics, the ESN dynamics are solely driven the output feedback. The larger the output feedback weights, the stronger is the driving signal. According to the discovered dependency, smaller reservoirs require relatively large feedback weights whereas larger reservoirs need much smaller weights to have a rather high likelihood of generating well-performing ESNs.

The most intriguing observation made, however, is the fact that the likelihood of generating good networks interacts non-linearly with the reservoir size and with the output feedback weights. Given a generally suitable reservoir size, either very low W_{OFB} values (in the order of 10^{-25} to 10^{-15}) or values rather large (in the order of 10^{-5} to 0.1) yield high chances of generating well-performing ESNs. This suggests that there are somewhat two categories of reservoir dynamics that are able to solve the considered dynamics. With the very low W_{OFB} values, the internal dynamics will be on a much lower level mainly fluctuating around values close to zero. With the larger W_{OFB} values, the internal dynamics may also reach saturation values at times. When the reservoir size becomes too large, large W_{OFB} values are not suitable any longer – most likely because neural saturation occurs too frequently, thus disallowing the reproduction of the continuous sinusoidal dynamics. With very small networks, on the other hand, very low W_{OFB} values will have not enough effect to induce sufficient dynamics. Unfortunately, though, it remains rather unclear to us why mid-range W_{OFB} values yield a very low probability of generating well-performing networks.

From an application-oriented perspective, it seems that the very low W_{OFB} values will not yield effective performance given more noisy signals than the ones investigated. Future investigations will explore the impact of noise on the observed ESN performances. It also remains unclear if similar behaviors in the likelihoods of generating well-performing networks can also be observed for more complex target dynamics, such as the Mackey-Glass function or also the superimposed sine wave problem with more than two sinusoidal functions. Nonetheless, the observation that too large reservoirs yield only effective ESNs when the feedback weights are very low suggests that small ESNs are generally preferable. Future research is necessary to shed further light on the non-linear interactions revealed between reservoir sizes and output feedback weights.

Acknowledgments. Funding from the Emmy Noether program (German Research Foundation, DFG, BU1335/3-1) is acknowledged.

References

1. Jaeger, H.: The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, German National Research Center for Information Technology (2001)
2. Jaeger, H.: Short term memory in echo state networks, German National Research Center for Information Technology. Technical Report GMD report 152 (2002)
3. Ishii, K., van der Zant, T., Becanovic, V., Plöger, P.: Identification of Motion with Echo State Network. In: OCEANS 2004, MTTS/IEEE TECHNO-OCEAN 2004, vol. 3, pp. 1205–1210 (2004)
4. Verstraeten, D., Schrauwen, B., D'Haene, M., Stroobandt, D.: An experimental unification of reservoir computing methods. Neural Networks 20(3), 391–403 (2007)
5. Rodan, A., Tino, P.: Minimum complexity echo state network. IEEE Transactions on Neural Networks 22(1), 131–144 (2011), doi:10.1109/TNN.2010.2089641
6. Roeschies, B., Igel, C.: Structure optimization of reservoir networks. Logic Journal of IGPL Advance Access (published September 3, 2009)
7. Xue, Y., Yang, L., Haykin, S.: Decoupled echo state networks with lateral inhibition. Neural Networks, 365–376 (2007)
8. Holzmann, G., Hauser, H.: Echo State Networks with Filter Neurons and a Delay & Sum Readout. Neural Networks 23(2), 244–256 (2010)
9. Koryakin, D., Lohmann, J., Butz, M.V.: Balanced Echo State Networks. Submitted for Publication in Neural Networks (2012)
10. Jaeger, H.: A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach, GMD Report 159, German National Research Center for Information Technology, 48 p. (2002)
11. Ozturk, M.C., Xu, D., Principe, J.C.: Analysis and design of echo state networks. Neural Computation 19(1), 111–138 (2007)