ISSN 1062-8738, Bulletin of the Russian Academy of Sciences: Physics, 2025, Vol. 89, No. 3, pp. 471–475. © Pleiades Publishing, Ltd., 2025.

Analysis of the Effect of a Recurrent Neural Network Size on Modeling and Prediction Accuracy of a Stochastic FitzHugh–Nagumo Neuron

N. D. Kulagin[a, *], **A. V. Andreev**[a], **and A. E. Hramov**[a]

[a] *Immanuel Kant Baltic Federal University, Kaliningrad, 236041 Russia*

**e-mail: kulagin.nikita03@gmail.ru*

Received September 6, 2024; revised October 15, 2024; accepted November 29, 2024

Abstract—We study the performance of a reservoir computing-based network in dynamics forecast of a FitzHugh–Nagumo model driven by white noise versus reservoir size. We show that the accuracy of signal prediction of a model neuron strongly depends on the number of neurons in the reservoir. The most accurate prediction of both the dynamics itself and the simulation of the coherence resonance phenomenon is achieved when using 500 neurons.

Keywords: FitzHugh–Nagumo model, reservoir computing, stochastic system, coherence resonance, nonlinear dynamics

DOI: 10.1134/S106287382471033X

INTRODUCTION

Nowadays, recurrent neural networks have become a widely used tool for predicting various temporal dependencies [1]. They are used to predict the spatiotemporal dynamics of complex nonlinear systems, such as the propagation of pulses in optical fiber [2], the systems of Lorenz and Kuramoto–Sivashinsky differential equations [3], as well as the logistic map and the Hénon map [4]. The chaotic nature of such systems makes them very difficult to predict.

The nonlinear dynamics of the electrical activity of biological neurons are particularly important for prediction. Artificial neural networks have already been successfully used to classify brain neuron activity data [5]. Further development of prediction methods based on neural activity data can be facilitated by studying the possibility of using neural networks to model the dynamics of mathematical models of neurons. Such models include the Hodgkin–Huxley, FitzHugh–Nagumo, and Hindmarsh–Rose systems [6–8]. In such systems in the excitable state, under external noise driving, the effect of coherence resonance can occur, characterized by the presence of a certain noise amplitude at which the coherence of the induced oscillations reaches a maximum [9–11].

To model the behavior of the FitzHugh–Nagumo stochastic neuron, including the phenomenon of coherence resonance, we used a reservoir computing - based recurrent network, which previously demonstrated a high ability to reproduce the dynamics of the original system at different noise levels, having trained on only one of them [12]. Reservoir computing has also been used to predict the dynamics of coupled networks of oscillators [13].

In this paper we investigate the effect of the number of reservoir neurons on the ability of a recurrent network to predict the dynamics of a FitzHugh–Nagumo neuron excited by external white noise, and to model the phenomenon of coherence resonance.

FITZHUGH–NAGUMO MODEL

The FitzHugh–Nagumo model is used as a model neuron for training and testing the reservoir computing-based network. It is described by the following system of differential equations:

$$\dot{x} = x - \frac{x^3}{3} - y + I, \quad \dot{y} = \frac{x + a - by}{\tau} + D\xi(t), \qquad (1)$$

where x denotes the fast variable of the system, which is interpreted as the membrane potential, y is the slow recovery variable, $I = 0.3$ is the value of the external injected current, $a = 0.7$ and $b = 0.8$ are the system parameters. The parameter $\tau = 12.5$ separates the time scales of the fast and slow variables, $\xi(t)$ is a white noise with zero mean and unit variance, and the value D regulates its amplitude. For these parameter values, the FitzHugh–Nagumo system is in the subthreshold steady state, in which the effect of noise $\xi(t)$ is capable of inducing spikes in the system.

The Euler method with an integration step $dt = 0.1$ was used to integrate the system.

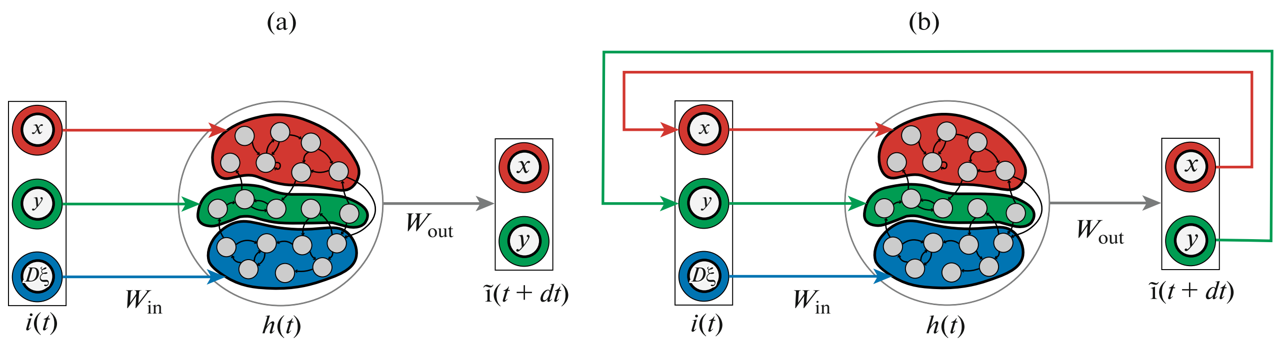

Fig. 1. Schematic representation of the reservoir computing network during the process of training (a) and testing (prediction) (b).

NEURAL NETWORK MODEL

The used neural network model consists of three layers: input neurons layer, recurrent neurons hidden layer (reservoir), and output neurons layer (Fig. 1).

The input neuron layer receives a vector $\mathbf{i}(t)$ as an input, which includes the values of the variables $\mathbf{x}(t)$ and $\mathbf{y}(t)$, as well as the scaled noise value $D\xi(t)$. Each of the input signals is connected to $\frac{N_{\text{RC}}}{3}$ reservoir neurons, where N_{RC} is the total number of recurrent neurons. The connection strengths between the input signals and the reservoir neurons are determined by the input weights matrix $\mathbf{W}_{\text{in}}$, whose values are randomly taken from the uniform interval [–1, 1].

The recurrent connections between the neurons of the reservoir are defined by the matrix $\mathbf{W}_{\text{r}}$, whose initial values are distributed uniformly in the interval [0, 1]. The density of the matrix $\mathbf{W}_{\text{r}}$ is $\frac{d}{N_{\text{RC}}}$, where d is a tunable hyperparameter. $\mathbf{W}_{\text{r}}$ is then rescaled according to the following equation:

$$\mathbf{W}_{\text{r}}^{*} = \frac{\mathbf{W}_{\text{r}}}{\rho_0}\rho, \tag{2}$$

where $\mathbf{W}_{\text{r}}^{*}$ is the rescaled hidden layer weights matrix, ρ_0 is the initial value of the spectral radius $\mathbf{W}_{\text{r}}$, ρ is a hyperparameter that sets the new value of the spectral radius of this matrix. The internal state of the hidden layer $\mathbf{h}$ at time t is defined as follows:

$$\mathbf{h}(t) = \tanh\left(\mathbf{W}_{\text{in}}\mathbf{i}(t) + \mathbf{W}_{\text{r}}\mathbf{h}(t - dt)\right). \tag{3}$$

The output neurons layer takes a vector of the augmented reservoir state $\tilde{\mathbf{h}}(t)$ as an input, which is obtained from $\mathbf{h}(t)$ by squaring half of the values with even indices, and linearly transforms it into the vector of predicted values $\tilde{\mathbf{\imath}}(t + dt)$ according to the equation:

$$\tilde{\mathbf{i}}(t + dt) = \mathbf{W}_{\text{out}}\hat{\mathbf{h}}(t), \tag{4}$$

where $\mathbf{W}_{\text{out}}$ is the output weights matrix, the values of which are determined as a result of training using Tikhonov regularization. In this case, the vector $\tilde{\mathbf{\imath}}(t + dt)$ includes only the values of the variables $\tilde{x}(t + dt)$ and $\tilde{y}(t + dt)$, and no prediction of the scaled values of white noise $D\xi(t)$ occurs.

The scheme of the reservoir computing network used for training is shown in Fig. 1a. The data of the discretized FitzHugh–Nagumo signal for training the neural network is divided into two parts: the transient period (t = 0–1000 s), the values of which are used only to update the state of the reservoir, and the training period (t = 1000–3000 s). During the network training process, at each time step, a vector of input signals $\mathbf{i}(t)$ is fed to the input layer, after which a vector of predicted values of the next time step $\tilde{\mathbf{\imath}}(t + dt)$ is obtained as the output. The data obtained are used to select the output weights by minimizing the following function using the Tikhonov regularization method:

$$\sum_{t=1000}^{3000} \left\|\tilde{\mathbf{\imath}}(t) - \tilde{\mathbf{\imath}}(t)\right\|^2 + \alpha\left\|\mathbf{W}_{\text{out}}\right\|. \tag{5}$$

$\tilde{\mathbf{\imath}}(t)$ denotes the vector of target values, α is a regularization parameter with a set value of 10^{-4}, which serves to prevent overfitting.

The process of testing the trained reservoir computing network is shown in Fig. 2b. Initially, similar to the training process, the signal in the range t = 0–1000 s is continuously fed as the input data, while the output data is discarded, which is necessary to update the state of the reservoir. After that, the prediction process occurs, in which the network begins to use its own predictions $\tilde{\mathbf{\imath}}(t)$ as input signals to predict subsequent values of $\tilde{\mathbf{\imath}}(t + dt)$. White noise data $D\xi(t)$ continues to be fed into the system from the outside.

To optimize the reservoir computing network, a grid search was used, during which the ranges $d \in [10, 20]$ and $\rho \in [0.5, 1.9]$ were explored with step sizes of 1 and 0.1, respectively. For each set of hyperparameters, the network was trained and the testing signal in the range t = 3000–5000 s was predicted. The

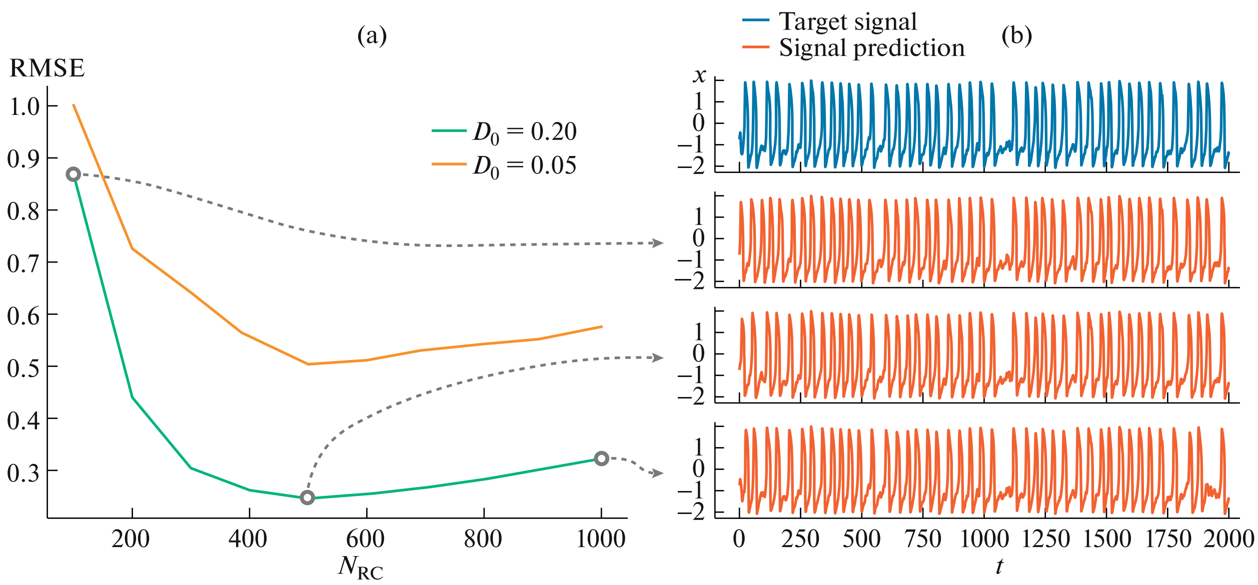

Fig. 2. Dependence of RMSE on the number of reservoir neurons N_{RC} during training of the reservoir computing network on the signal with noise levels $D_0^1 = 0.05$ and $D_0^2 = 0.2$ while predicting signals with similar noise level (a) and temporal dynamics of the FitzHugh–Nagumo stochastic neuron target signal with noise level $D = 0.2$ and predictions of the network with reservoir size N_{RC} equal to 100, 500 and 1000 (b).

root mean square error (RMSE) was used to evaluate the quality of the prediction:

$$\mathrm{RMSE} = \frac{1}{n}\sum_{j=0}^{n}\left(\tilde{x}_j - \overline{x}_j\right)^2, \tag{6}$$

where $\overline{x}$ is the target value of the variable x, n is the total number of predicted points.

RESULTS

To study the effect of the hidden layer size of the reservoir computing network the range of N_{RC} from 100 to 1000 neurons with a step size of 100 was studied. Two options were explored: training the network on the signal of a stochastic neuron with noise amplitudes $D_0 = 0.05$ and $D_0 = 0.2$. The RMSE values when the optimized networks predict signals with $D = D_0$ in the range of $t = 3000–53000$ s are shown in Fig. 2a. The accuracy at the minimum reservoir size considered $N_{RC} = 100$ turns out to be relatively low for both trained networks, but the prediction quality increases drastically with an increase in N_{RC} and reaches a maximum at 500. A further increase in N_{RC} leads to a gradual deterioration in the accuracy of reproducing the dynamics. It is also noticeable that the network trained on a signal with $D_0 = 0.2$ in all cases reproduces the temporal dynamics with a similar noise level better than the network trained on a signal with $D_0 = 0.05$.

Figure 2b shows the time series of the dynamics of the original FitzHugh–Nagumo neuron and the predictions of the reservoir computing network with $D_0 = 0.2$ and N_{RC} equal to 100, 500, and 1000, respectively. As one can see, even with $N_{RC} = 100$, the network is able to model the general dynamics of the stochastic neuron, but there is a noticeable difference in the generation time of some spikes from the original system.

After that, each of the three trained networks, the results of which were shown in Fig. 2b, was used to predict the FitzHugh–Nagumo time series with a range of D from 0.05 to 1 with a step size of 0.05 over the range $t = 3000–53000$ s. The coefficient of variation R is used to evaluate the fit of the statistical characteristics of the predicted signals to the original model:

$$R = \frac{\sigma}{\mu}, \tag{7}$$

where σ is the standard deviation of the interspike intervals of the signal, μ is the interspike intervals mean value.

The results of stochastic neuron modeling under different noise exposure are shown in Fig. 3a. The network with $N_{RC} = 500$ showed the most similar dynamics of the dependence of R on D compared to the original model and also demonstrated the best results in predicting a single signal with $D = 0.2$. The network with $N_{RC} = 1000$ also showed similar dynamics, but at a high noise level it deviates significantly from the target curve. At $N_{RC} = 100$ the network is

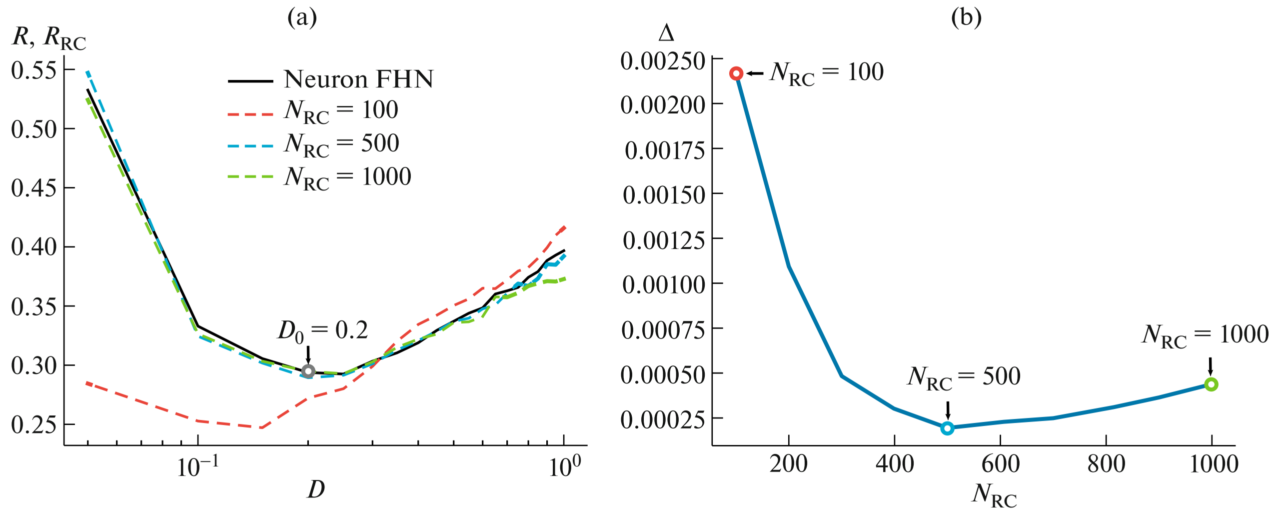

Fig. 3. Dependence of the coefficients of variation of the FitzHugh–Nagumo stochastic neuron R_{RC} and predicted signals of the reservoir computing network R on the amplitude of the applied white noise D (a) and dependence of the prediction performance Δ on the number of reservoir neurons N_{RC} (b).

unable to correctly reproduce the dependence of the stochastic neuron on the noise amplitude; the coefficient of variation close to the original system was achieved only at $D = 0.3$.

Next, to evaluate the dependence of the overall modeling ability of the network on the reservoir size N_{RC}, signal prediction data of the FitzHugh–Nagumo neuron were obtained for the entire range of N_{RC} from 100 to 1000. To compare the dependence of R on D of the trained network and the stochastic neuron, the value Δ is used to characterize the integral difference between the predicted dependence curve and the target curve:

$$\Delta = \sqrt{\int (R(D) - R_{RC}(D))^2 \, dD}. \quad (8)$$

The values of the Δ metric for the reservoir computing networks with each of the considered N_{RC} values are presented in Fig. 3b. Consistent with previous results, the network with $N_{RC} = 500$ demonstrated the highest predictive ability, and the network with $N_{RC} = 100$ showed the lowest. Like Fig. 2a, a drastic improvement in modeling quality occurs when N_{RC} increases from 100 to 500, after which the prediction error starts to increase gradually.

CONCLUSIONS

In this study, a reservoir computing network was successfully trained to reproduce the dynamics of an excitable FitzHugh–Nagumo model with varying white noise amplitude by training it on a single, average, noise level. An increase in the modeling quality of the FitzHugh–Nagumo neuron signal was observed when the reservoir size was increased up to an average value, after which a gradual deterioration of the prediction was observed when further recurrent neurons were added to the hidden layer.

At the minimum reservoir size, a sufficiently correct prediction of the stochastic neuron signal was achieved at only one noise amplitude, in other cases, a significant deviation from the statistical data of the original model is observed. At medium and large hidden layer sizes, high modeling quality was observed over the entire range of white noise intensity, but at the highest noise level, the medium-sized network showed the best results.

These results indicate the existence of an optimal hidden layer size for the reservoir computing network in predicting the stochastic signal, deviations from which cause the quality of the original system modeling to deteriorate.

FUNDING

This work was supported by the Russian Science Foundation (grant no. 24-71-10072).

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Hewamalage, H., Bergmeir, C., and Bandara, K., *Int. J. Forecast.*, 2021, vol. 37, no. 1, p. 388.
2. Salmela, L., Tsipinakis, N., Foi, A., et al., *Nat. Mach. Intell.*, 2021, vol. 3, no. 4, p. 344.

3. Vlachas, P.R., Pathak, J., Hunt, B.R., et al., *Neuron Networks*, 2020, vol. 126, p. 191.
4. Sangiorgio, M. and Dercole, F., *Chaos, Solitons Fractals*, 2020, vol. 139, p. 110045.
5. Batmanova, A., Kuc, A., Maksimenko, V., et al., *Mathematics*, 2022, vol. 10, no. 17, p. 3153.
6. Hodgkin, A.L. and Huxley, A.F., *J. Physiol.*, 1952, vol. 117, no. 4, p. 500.
7. FitzHugh, R., *Biophys. J.*, 1961, vol. 1, no. 6, p. 445.
8. Hindmarsh, J.L. and Rose, R.M., *Proc. R. Soc. London, Ser. B*, 1984, vol. 221, no. 1222, p. 87.
9. Lee, S., Neiman, A., and Kim, S., *Phys. Rev. E*, 1998, vol. 57, no. 3, p. 3292.
10. Pikovsky, A.S. and Kurths, J., *Phys. Rev. Lett.*, 1997, vol. 78, no. 5, p. 775.
11. Wu, S., Ren, W., He, K., and Huang, Z., *Phys. Lett. A*, 2001, vol. 279, nos 5–6, p. 347.
12. Hramov, A.E., Kulagin, N., Andreev, A.V., and Pisarchik, A.N., *Chaos, Solitons Fractals*, 2024, vol. 178, p. 114354.
13. Andreev, A.V., Badarin, A.A., Maximenko, V.A., and Hramov, A.E., *Chaos*, 2022, vol. 32, no. 10, p. 103126.