CYBERNETICS AND PHYSICS, VOL. 13, NO. 1, 2024, 20–31

PREDICTION OF EXTREME EVENTS AND CHAOTIC DYNAMICS USING WAVENET

Nikolay V. Gromov
Control Theory Department
Lobachevsky University
Nyzhny Novgorod, Russia
gromov@itmm.unn.ru

Lev A. Smirnov
Control Theory Department
Lobachevsky University
Nyzhny Novgorod, Russia
lev.smirnov@itmm.unn.ru

Tatiana A. Levanova
Control Theory Department
Lobachevsky University
Nyzhny Novgorod, Russia
tatiana.levanova@itmm.unn.ru

Article history:
Received 05.04.2024, Accepted 13.05.2024

Abstract

In this paper we propose an approach for chaotic time series and extreme events prediction based on WaveNet model, which is a deep neural network that directly synthesizes speech waveforms from acoustic features. We test our approach on artificial data obtained from long time series with extreme events generated by two coupled bursting Hindmarsh-Rose neurons, and on two real-life data sets (local field potentials recorded in mice and electroencephalogram recorded in humans) containing patterns of epileptiform activity, which also can be viewed as extreme events.

Key words

extreme events, epileptiform activity, LFP, EEG, WaveNet, LSTM.

1 Introduction

Extreme events are rare, repetitive and strong deviations from the typical behavior of an observed variable in biological and engineered systems that greatly influence their dynamics [Lehnertz, 2006]. Examples of such behavior include oceanic rogue waves [Dysthe et al., 2008; Kharif et al., 2008], events at financial markets [Longin, 2016], cell dynamics [Kashiwagi et al., 2006], failure of mechanical parts [Zio and Pedroni, 2009], events in laser dynamics [Pisarchik et al., 2012; Kumarasamy and Pisarchik, 2018] and epileptic seizures [Lehnertz, 2006; Pisarchik et al., 2018; Frolov et al., 2019].

Understanding and accurate prediction of such phenomena still remains an urgent problem in applied mathematics [Qi and Majda, 2016; Majda and Tong, 2015; Viotti and Dias, 2014; Cai et al., 2001]. Extreme events can be both isolated rare events [Guth and Sapsis, 2019; Cousins and Sapsis, 2015] and frequent in space and time [Majda and Lee, 2014; Grooms and Majda, 2014]. The curse of dimensionality is one of the important obstacles for accurately predicting extreme events in large complex systems, where both new models and efficient numerical algorithms are required [Qi and Majda, 2018; Mohamad and Sapsis, 2018; Chen and Majda, 2017].

Several approaches have been developed that successfully deal with non-Gaussian statistics of extreme events, see [Varadhan, 1984; Weinan and Vanden-Eijnden, 2010; Qi and Majda, 2016; Mohamad and Sapsis, 2015; Majda et al., 2019; Farazmand and Sapsis, 2017]. Such approaches often use a specific problem structure, which is applicable in cases where the dynamics of the system are well known. For systems in which our understanding of the governing laws is partial, or in situations where there are serious errors in model construction, it has been shown that combining existing models (incomplete or with errors) with a data-driven approach improves the efficiency of prediction compared to each of the methods separately [Wan et al., 2018].

However, there are situations when the only available information characterizing the system is observations. In this case the task can be solved using machine learning algorithms, especially deep neural networks, which have been effectively applied to a wide range of problems, in particular, those related to extreme events, see, e.g. [Jordan and Mitchell, 2015; Hramov et al., 2017; Maksimenko et al., 2018; Qi and Majda, 2018]. Namely, different deep learning architectures, such as long short-term memory (LSTM) [Ding et al., 2019; Wambura et al., 2020; Li et al., 2023] and gated recurrent unit (GRU) [Durairaj et al., 2023; Zhang et al., 2024] networks, reservoir computings [Pyragas and Pyragas, 2020] and transformers [Feng and Fox, 2022; Gao et al., 2023], were applied to a problem of extreme events prediction. In this literature, LSTM is the most commonly used architecture because of its ability to process long-term dependences in the temporal sequences and deal with the vanishing gradient problem.

It should be noted that in some cases machine learning algorithms developed for specific tasks also show good performance in the task of predicting complex dynamics and extreme events. For instance, deep neural networks developed for image processing have been proposed for data-driven prediction of chaotic dynamical systems [Ma and Wang, 2019; Pathak et al., 2018; Raissi and Karniadakis, 2018], climate and weather forecasting [Bolton and Zanna, 2019; Weyn et al., 2019], and parameterization of unsolved processes [Han et al., 2018; Han and Jentzen, 2017; Brenowitz and Bretherton, 2018]. Nevertheless, building optimal deep learning algorithms for predicting extreme events is still an actively developing topic [Durairaj et al., 2023].

In this paper, we develop a deep machine learning framework to predict chaotic dynamics and extreme events using WaveNet, which is promising neural speech generation model [Oord et al., 2016]. In speech technologies, WaveNet is a widely used fully probabilistic generative deep learning model for creating sound [Oord et al., 2016]. It can be considered as a causal filter that does not look into the future, therefore it can be applicable for time series prediction. Speech time series are obviously irregular, with mode switching and large-amplitude spikes (events), which resemble extreme events. We test our approach on artificial data obtained from long time series with extreme events generated by two coupled bursting Hindmarsh-Rose neurons, and on two real-life data sets (local field potentials recorded in mice and electroencephalogram recorded in humans) containing patterns of epileptiform activity, which also can be viewed as extreme events.

The paper is organized as follows. In Section 2 we describe data sets based on which we demonstrate the performance of our approach. In Section 3 we describe the architecture of WaveNet, which is a type of convolutional deep neural networks. Here we also briefly describe an architecture of LSTM network, which is a type of recurrent deep neural networks, and compare two these architectures. In Section 4 we formulate the problem of chaotic time series prediction and introduce evaluation metrics. In Section 5 we present results of prediction of chaotic time series and time series containing extreme events using WaveNet. In Section 6 we discuss our findings and draw our conclusions.

2 Data: time series with extreme events

2.1 Data Set 1: artificial data

2.1.1 Data source In order to create data set containing sufficient number of extreme events, we calculated long time series using well-known dynamical system that is able to generate extreme events. Similar approach to creating data set with extreme events was used in [Gromov et al., 2022]. It can help to overcome issues related to data imbalance that naturally arise in tasks of extreme events prediction: the time series often contains very few examples of extreme events comparing to other events. Such a strong data imbalance can potentially lead any machine learning model into one of two problematic situations: either it has difficulties learning extreme events patterns and simply recognizes all samples as belonging to the most popular non-standard dynamics, or it faces the overfitting problem, when the algorithm perfectly remembers train samples, but does not actually learns complex nonlinear dependencies and fails on test set. Deep neural networks widely used for classical time series prediction, are observed to have troubles in dealing with data imbalance [Wang et al., 2019].

In order to generate artificial data we used the system of two bursting Hindmarsh-Rose neurons with reciprocal chemical couplings [Mishra et al., 2018], which is described by the following equations:

$$\begin{cases} \dot{x}_i = y_i + bx_i^2 - ax_i^3 - z_i + I - k_i(x_i - v_s)\Gamma(x_j) \\ \dot{y}_i = c - dx_i^2 - y_i \\ \dot{z}_i = r[s(x_i - x_R) - z_i] \\ i, j = 1, 2\ (i \neq j). \end{cases} \tag{1}$$

Here the variable x_i describes the membrane potential of the i-th neuron, the variables y_i and z_i correspond to fast and slow ion currents flowing through the membrane of the i-th neuron. The parameter $r << 1$ determines the ratio of the characteristic time scales of these currents. In our study, we fix $r = 0.001$. The parameter I describes the external current applied to the neuron. In the study, we fix its value $I = 4$. Other parameters describe the nonlinearity of the membrane conductivity: $a = 1,\ b = 3,\ c = 1,\ d = 5, x_R = -1.6,\ s = 5$, which are typical values for bursting regime in an isolated element.

Chemical synaptic couplings are described by the sigmoid function

$$\Gamma(x) = \frac{1}{1 + exp^{-\lambda(x-\Theta)}} \tag{2}$$

with parameters $\lambda = 10,\ \Theta = -0.25,\ v_s = 2$. Parameters $k_{1,2}$ correspond to the strength of chemical couplings and are control parameters in the system. Depending on its values, we can simulate different types of impact: inhibitory ($k_{1,2} < 0$), excitatory (($k_{1,2} > 0$) and mixed ($k_1 > 0$ and $k_2 < 0$ or vice versa). Note that such coupling function takes into account the basic principles of the chemical interaction of neurons: (i) the presence or absence of activity of the postsynaptic element depends on the level of activity of the presynaptic element; (ii) all interactions between neuronal cells are inertial due to the fact that signal transmission is not instantaneous.

We choose the variable $x_{II} = x_1 + x_2$ as the observable variable based on biological background of this system. It was shown in [Mishra et al., 2018] that described system demonstrates extreme events for a wide range of governing parameters k_1 and k_2. We took $k_1 = k_2 = -0.17$ in our numerical experiments. The example of corresponding time series containing extreme

events is shown in Fig. 1. Probability density function of all events demonstrates dragon-king-type hump for the tailed part of the distribution [Mishra et al., 2018].

Here we employ the simplest framework consists only of two elements. For more biologically plausible modelling one can use ensembles with greater number of elements, see, e.g., [Plotnikov, 2021]. Also more sophisticated single neuron models, such as Hodgkin–Huxley model [Borisenok, 2022], can be used in tasks of simulating epileptiform activity.

2.1.2 Definition of extreme events and numerical criterion As the term "extreme event" is used in various disciplines, a precise definition of extreme events is not available. Rather, it depends on the particular discipline where this term is being used. In this study, we select the extreme events based on their magnitude. Therefore, it is crucial to set a threshold height so that we can call an event "extreme" when it exceeds the threshold.

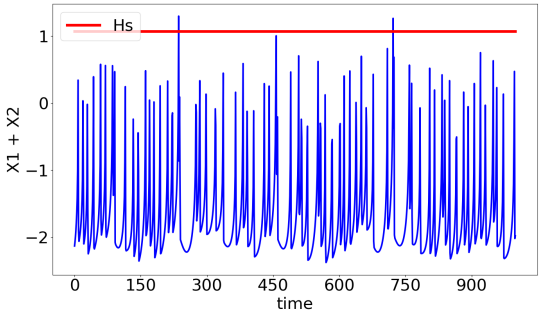


Figure 1. Time series for the two coupled Hindmarsh-Rose bursting neurons (1). Parameters are specified in the main text. Red line marks H_s level. Here $H_s = \mu + 6\sigma$.

Here to define extreme events we adopt a numerical criterion first introduced in [Dysthe et al., 2008] and then widely used in numerous studies, see e.g. [Chowdhury et al., 2019; Mishra et al., 2020]. Namely, we calculate the level

$$H_s = \mu + N\sigma, \tag{3}$$

where μ is mean value and σ is standard deviation of all events in temporal diagram of the observable variable. Constant N depends on particular system or data source; in this case $N = 6$. In order to do this we took a long runs ($\approx 2 \cdot 10^6$ iterations) of observable variable, measured the local maximum values and estimate their mean $\mu = \langle P_n \rangle$ and standard deviation $\sigma = \langle P_n^2 \rangle - \mu^2$. An event is then classified as extreme when it crosses a threshold, H_s. This criterion reflects that the amplitude of extreme events exceeds several times the standard deviation of the observed variable from its mean value.

Mathematically, extreme events are connected with highly irregular, chaotic (or even hyperchaotic [Olenin and Levanova, 2023]) dynamics of the system under study. The prediction horizon, which is the maximum number of steps ahead that one can make a prognosis, is finite for chaotic time series. The prediction horizon is attributed to an exponential divergence of initially close trajectories due to the Lyapunov instability of chaotic time series [Pyragas and Pyragas, 2020], which results in exponential error growth for multi-step prediction. The highest Lyapunov exponent serves here as the exponent coefficient. This explains the fact that most papers dealing with chaotic time series prediction, discuss results for a single step prediction only, whereas the problem of multi-step prediction for chaotic time series is still unresolved.

2.2 Data Set 2: recordings of epileptiform activity in mice

2.2.1 Data source A dataset of recordings of chronic neuronal activity containing patterns of epileptiform activity was obtained at the Experimental Animal Clinic of the Institute of Theoretical and Experimental Biophysics of the Russian Academy of Sciences (Pushchino, Russia). Local field potentials (LFPs) were recorded in the hippocampus and medial entorhinal cortex of mice. Young healthy outbred CD1 mice were used, divided into two groups: control (n=6) and with induced chronic epileptiform activity (n=6). Epileptiform activity in mice was modeled using intraperitoneal administration of pilocarpine (280 mg/kg). Registration of DILI in awake mice was carried out one month after the induction of epileptiform activity. Detailed protocol of biological experiment can be found at [Gerasimova et al., 2023].

In this study, recordings of chronic neuronal activity from the hippocampus and MEC III were used, while data on the study of behavioral patterns at the time of recording of neuronal activity were not used.

The data were pre-processed before carrying out numerical experiments. In particular, artifacts were removed and a Gaussian filter was applied in order to remove noise. Then the data was normalized so that the mean value was zero and the spread was equal to one. Next, the data were converted to a time sequence–response format. For example, to predict one step, 20 time samples were fed to the model input, and the 21st was used as the response. To predict several steps, the model's response to the previous step was iteratively added to the sequence at the input to the network.

2.2.2 Definition of extreme events and numerical criterion Here we define extreme events as spike-wave discharges [Dolinina et al., 2022] and other epileptiform activity temporal patterns. The marking of patterns of epileptiform activity was carried out by specialists in [Gerasimova et al., 2023]. Fig. 2 shows time series for the local field potentials recorded in mice *in vivo*.

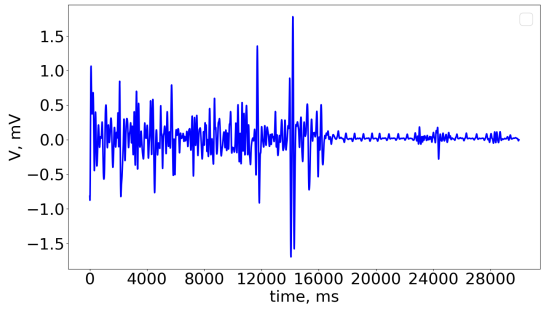


Figure 2. Time series for the local field potentials recorded in mice *in vivo*.

2.3 Data Set 3: EEG data with epileptic seizures

2.3.1 Data source As a source of EEG data containing epileptic seizures we used CHB-MIT Scalp EEG Database freely available at Physionet [Goldberger et al., 2000]. The data is collected from Children's Hospital Boston (CHB), composed of EEG recordings observed for several days. This is the only open database that contains continuous long-term scalp EEG recordings suitable for the purpose of our study.

The recordings are grouped into 23 cases collected from 22 subjects (17 females of age 1.5–19 years old and five males aged 3–22 years old [Shoeb, 2009]. The signals are sampled at 256 samples/sec of 16-bit resolution, which includes 23 EEG signals. EEG electrode position and nomenclature (International 10–20 systems) are also recorded. In some cases, other signals are also recorded (ECG signal). The EEG signals are in easy to read format where the dummy signals are ignored. The files are composed of 664 .edf files, and the records (seizures) list the files with more seizures [Shoeb, 2009]. The records include 198 seizures (182 original sets with 23 cases). The long sequences of EEG samples contain labeled seizures.

In total, 18 channels are consistent across all 24 cases: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ.

We perform single-channel prediction, for which data only from FP1-F7 channel were used. On the one hand, this approach (single-channel prediction) allows one to reduce the workload to the total number of electrode points. On the other hand, it still allows to test our deep learning architectures and compare its performance.

For numerical experiments, we normalized the data by converting them to zero mean and unit variance. After that, the noise was filtered with a Gaussian filter with $\sigma = 3$ and with $radius = 5$. The data was then separated into inputs and outputs for the model ("time sequence-response" format). The input data consisted of 20 time samples, and the output was the 21-st sample. The optimal number of time samples in the input data was determined experimentally during the optimization process. After normalizing and splitting the data, we split it into train and test sets in a 4:1 ratio.

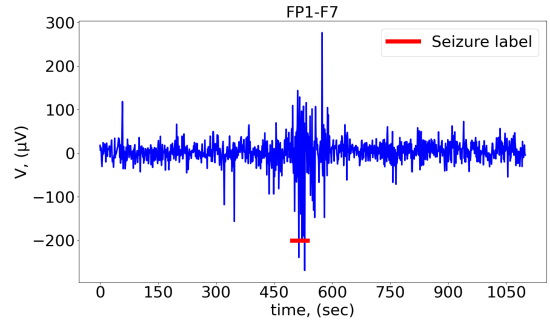


Figure 3. Example of human epileptic EEG signals from CHB-MIT Scalp EEG Database (single channel, FP1-F7, frontal area). Red line marks seizure according to given annotation file.

2.3.2 Definition of extreme events and numerical criterion Epileptic seizures can be viewed as extreme events in this case [Frolov et al., 2019]. The long sequence of EEG samples contains seizure intervals, the beginning time and the end time of which are annotated by doctors for each subject with seizures. These seizure intervals are termed as ictal stages. The example of EEG data that contains recorded seizures is shown in Fig. 3.

3 Methods

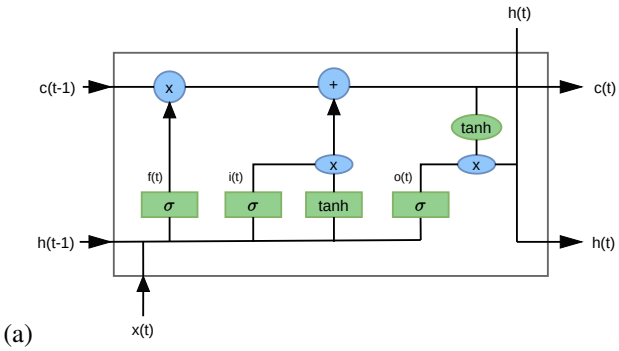
In this Section we compare the performance of WaveNet deep learning architecture with the performance of long short-term memory (LSTM) network, which is widely used in time series prediction tasks.

3.1 Long short-term memory (LSTM) network

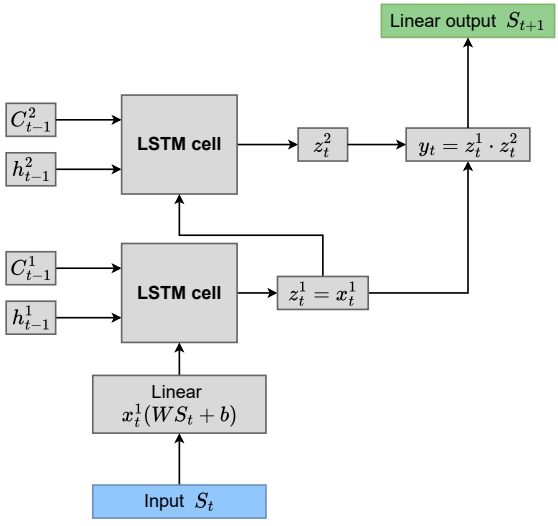
LSTM networks are one of the most well-known deep learning architectures that belongs to a class of recurrent neural networks. LSTM network consists of LSTM cells.

An LSTM cell is schematically presented in Fig. 4(a). The cell has two hidden states, one representing short-term memory(h_t) and the other representing long-term memory(c_t), which interact through filters. The idea of long-term memory is that, based on the totality of information from short-term memory and input, it is regulated which information should be forgotten from it and which should be remembered. First, let us consider the information that needs to be forgotten, a forgetting filter with a sigmoid function is responsible for this, it returns values from 0 to 1 and they are multiplied component-by-component by the state of long-term memory, this can be interpreted as follows: if the output is 0, then this information is forgotten, 1 – left .

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_t + b_f) \tag{4}$$



(a)



(b)

Figure 4. LSTM architecture. (a) LSTM cell. (b) LSTM network.

For information that needs to be remembered, a sigmoid input filter is used to understand which components of the long-term memory state to insert the information into

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_t + b_i), \tag{5}$$

and a hyperbolic tangent filter is used to understand what information needs to be added

$$g_t = \tanh(W_{xg}x_t + W_{hg}h_t + b_g). \tag{6}$$

Then the final formula for changing long-term memory is:

$$c_t = f_t * c_{t-1} + i_t * g_t. \tag{7}$$

To obtain an output from a cell, an output filter and information from long-term memory are used:

$$\begin{gathered} h_t = o_t * \tanh(c_t), \\ o_t = \sigma(W_{xo}x_t + W_{ho}h_t + b_o). \end{gathered} \tag{8}$$

Weight matrices W and bias b are trained in the network using the backpropagation method, optimizing the loss function [LeCun et al., 2015].

The deep neural network LSTM used in this study has the following architecture, see Fig. 4(b). The first is the input linear layer, which translates the input information into a feature space of dimension 100. Then follow two layers of LSTM cells. The result is projected by the linear output layer. Weight matrices W and bias b are trained using backpropagation using the mean square loss function. This LSTM network architecture was tested in our previous studies [Gromov et al., 2022; Gromov et al., 2023; Gerasimova et al., 2023; Beltyukova et al., 2023].

3.2 WaveNet

WaveNet allows to process and generate audio signals [Oord et al., 2016]. Namely, it models time series with conditional probabilities, where the probability of each value x_t in the series is conditioned on r previous values: $p(x_t \| x_{t-r}, \ldots, x_{t-1})$. It models these relationships between past and future values with multiple causal convolutional layers, where the causal property ensures that no future values are considered when making the prediction for the next time step. Therefore, the network receives the last r time samples of a time series as input, and predicts the next $t+1$ sample.

The architecture of WaveNet is schematically depicted in Fig. 5. Here each layer is displayed as a block of its own color. All blocks of the same color represent the same layer, i.e. they have the same weights. As one may see, the structure of a network is a tree-like. Each block has a fairly simple architecture. First comes the dilated convolution, which allows to exponentially increase the distances between the inputs to the next layer.

$$DilatedConvOut_k = W_{conv_k}[x_i]. \tag{9}$$

where i in $[t-k, t-2k, t-3k \ldots]$, W_{conv_k} is a trainable weight matrix for block k and x_t is a sample in the moment t. The core of this convolution contains gaps, the size of which can be specified. This convolution allows to consider a larger receptive field with fewer parameters. Two such convolutions are used, one for the sigmoid output, the other for the hyperbolic tangent, i.e. a typical example of filters is implemented, as in recurrent neural networks, such as LSTM and GRU.

$$\begin{gathered} Gating_k = tanh(DilatedConvOut_k) * \\ * \sigma(DilatedConvOut_k) \end{gathered} \tag{10}$$

The output is passed through fully connected layers, one for the next block:

$$x_t^{k+1} = x_t^k + W_k Gating_k \tag{11}$$

and another for the predictor:

$$x_{t+1} = Softmax(W_{out_2} Relu(W_{out_1} Relu(\sum_{k=1}^{K} x_t^k))), \tag{12}$$

where k is a depth of the network (number of blocks), W_k is another trainable weight matrix for block k, W_{out_1}, W_{out_2} are output trainable weight matrices.

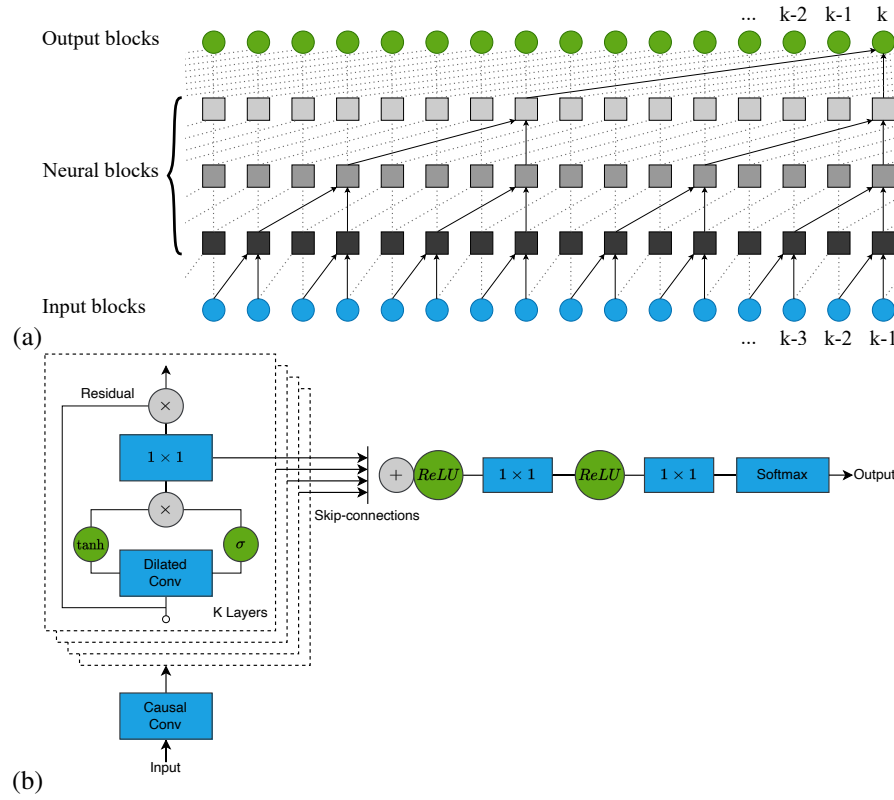


Figure 5. WaveNet vocoder architecture.(a) Stack of dilated causal convolution layers. (b) Residual block and the entire architecture.

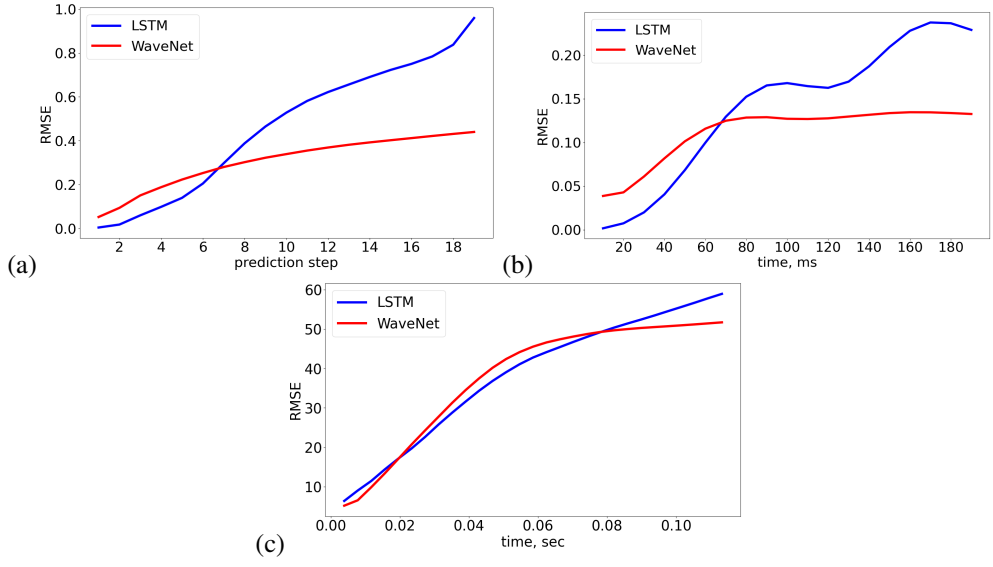


Figure 6. The dependence of the value of evaluation metrics on the prediction horizon. (a) Data Set 1. (b) Data Set 2. (c) Data Set 3.

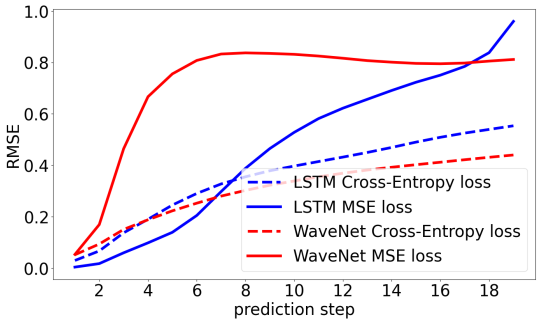


Figure 7. Comparison of predictions of both of our networks for Data Set 1 with a different formulation of the task, as classification with cross-entropy loss function (dashed lines) and as regression task with MSE loss function (continuous lines).

The network predicts values autoregressively, i.e. when the t-th sample is received, it is added to the list, the $(t-r)$-th sample is removed from the list, so the time series is synthesized step by step.

Note that WaveNet is in fact a convolutional deep network, which has a valuable advantage in time series prediction compared with recurrent deep networks. Namely, recurrent networks tends to predict a very similar sample to the last one seen to optimize the loss function, and thus the network tends to converge to predict the mode. Convolutional deep networks converge to mode much more slowly due to a large receptive field, and the network does not give preference to any sample.

4 Problem statement and evaluation metrics

The task of predicting chaotic time series and time series containing extreme events can be formulated both as a regression task and as a classification task.

Regression task implies predicting the next time step. In this case one should optimise mean squared error (MSE) loss function, which is

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad (13)$$

where n is number of training sequences, y_i is correct sample and $\hat{y}_i$ is predicted sample.

Classification task implies dividing samples into 256 classes, representing uniform intervals from the minimum value of the time series to the maximum. In this task one should optimise multiclass cross-entropy loss function, which is

$$CrossEntropy = -\frac{1}{n}\sum_{i=1}^{n}(y_i log(\hat{y}_i)). \qquad (14)$$

In previous studies researchers used different evaluation metrics to examine the quality of a deep learning models in similar tasks by quantitatively measuring the accuracy in terms of RMSE, co-efficient of determination (R^2), MSE and MAE, see [Hussain et al., 2019]. In our study we used as an evaluation metric a root-mean-square error (RMSE), which has very intuitive interpretation in terms of relative error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \qquad (15)$$

Here the meaning of parameters is the same as in case of MSE, see Eq.(13). The RMSE metric shows how far the predicted value is from the true value in absolute measure [LeCun et al., 2015].

We have compared the performance of both architectures in two task settings on artificial data (Data Set 1), see Fig. 7. It was shown that for single step prediction LSTM network demonstrates better performance with MSE loss function (regression task) than with Cross-Entropy loss function (classification task). This is also true for predicting a small number of steps ahead. WaveNet showed better performance with Cross-Entropy loss function (classification task) in all cases. Therefore, in all further numerical experiments we will compare the performance of LSTM with MSE loss and WaveNet with Cross-Entropy loss.

5 Results

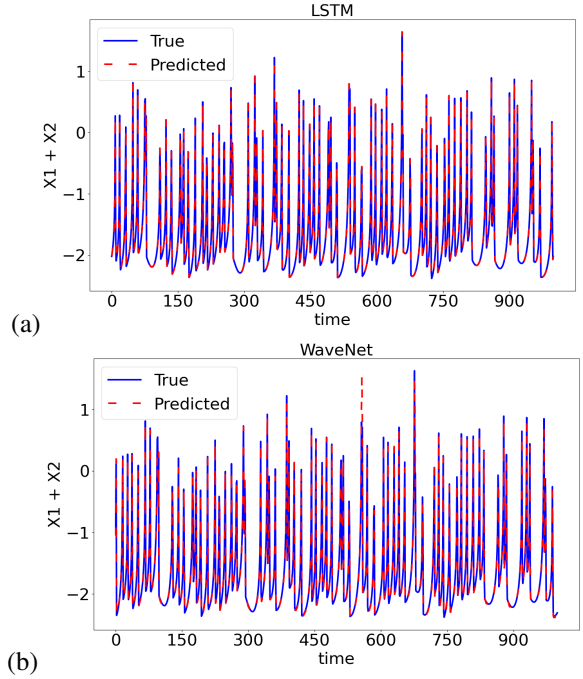


Figure 8. True (blue solid line) and predicted (red dashed line) value for 1 step of prediction for Data Set 1. (a) LSTM. (b) WaveNet.

Described in Section 3 LSTM and WaveNet deep architectures were trained on sequences ”time sequence –

response". Here the response of the model to the previous step was iteratively added to the sequence to the network input. We compared the performance of two neural networks using performance metrics on the example of artificial and real data.

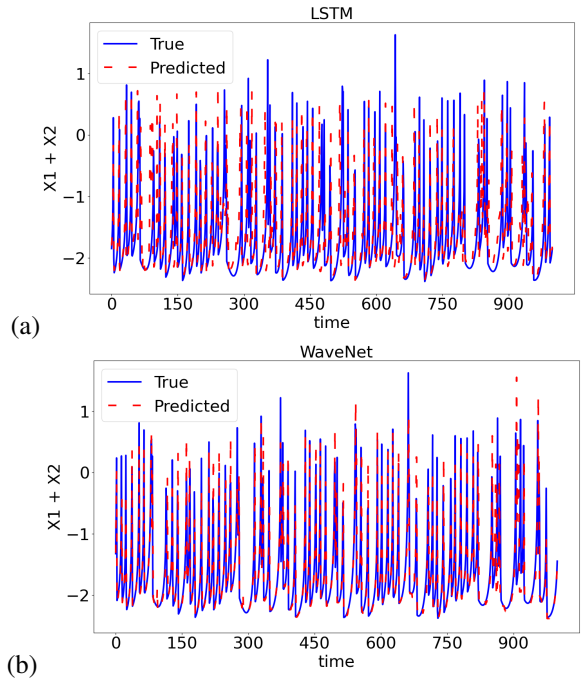


Figure 9. True (blue solid line) and predicted (red dashed line) value for 15 steps of prediction for Data Set 1. (a) LSTM. (b) WaveNet.

First, let us study how these two models deal with artificial data containing chaos and extreme events (Data Set 1). In Fig. 8 examples of true and predicted artificial signal are shown for one-step prediction. In Fig. 9 examples of true and predicted artificial signal are shown for multi-step prediction (namely, for 15 steps). The dependence of the value of evaluation metrics on the prediction step length (prediction horizon) is presented in Fig. 6(a). It can be clearly seen that for a small number of prediction steps LSTM perform better. Nevertheless, in case of LSTM prediction error roughly increases with the increase in the number of prediction steps. Vice versa, WaveNet shows better performance for multistep prediction. Note that prediction error for WaveNet increases much slower than for LSTM.

Second, let us study how these two models deal with real-life data that contain records with epileptiform activity in mice (Data Set 2). The sampling frequency for recording local field potentials was 100 Hz. Accordingly, the time between adjacent points (i.e., 1 time step) was 10 ms. In Fig. 10 examples of true and predicted signal with epileptiform activity are shown for 10 ms prediction. In Fig. 11 examples of true and predicted LFP signal in mice are shown for multi-step prediction (namely, for 8 steps, which are 80 ms). The dependence of the value of evaluation metrics on the prediction step length (prediction horizon) is presented in Fig. 6(b). As in the case of artificial data (Data Set 1), WaveNet shows better performance for multistep prediction and much slower growth in the value of prediction error with the increase in prediction step length.

Third, let us study how these two models will perform on another real-life data, namely, on EEG data with epileptic seizures (Data Set 3). Note that the sampling frequency of the EEG recording was 256 Hz, respectively, the time between adjacent points (i.e., 1 time step) is $1/256$ s. In Fig. 12 examples of true and predicted EEG signal with epileptic seizures are shown for 1 step prediction. In Fig. 13 examples of true and predicted EEG signal with epileptic seizures are shown for multistep prediction (namely, for 10 steps). From Fig. 6(c) one can see that for this data set Wavenet shows worse performance, than the LSTM network.

6 Conclusion

We proposed WaveNet deep learning architecture, which is usually used for speech generation, to be employed in tasks of predicting chaotic time series and time series containing extreme events. To the best of our knowledge, this is a first attempt to use vocoders in the task of extreme events prediction.

We train WaveNet network on artificial and real-life data sets and compared its performance with the performance of LSTM deep network, which is a first-choice architecture in time series and sequences prediction. We showed that for chosen prediction horizon prediction LSTM network demonstrates better performance with MSE loss function than with Cross-Entropy loss function. We also showed that WaveNet is able to capture complex relationships within the signals of all tested types and in several cases is consistently superior in performance compared to the LSTM model.

This study can make a contribution to improving real-time estimation of neural activity in different protocols, e.g. in neuroprosthesis [Beltyukova et al., 2023] and brain-computer interfaces [Chholak et al., 2019], [Rasheed, 2021], providing higher performance in real-time applications compared to previous state-of-the-art methods, such as LSTM.

Acknowledgements

We thank A. Emelin and S. Olenin for their help with Figs. 4-5.

Numerical experiments with Data Sets 1 and 3 were supported by Russian Science Foundation (Project No. 19-72-10128), numerical results with Data Set 2 was supported by Ministry of Science and Education of Russian Federation (Contract FSWR-2024-0005).

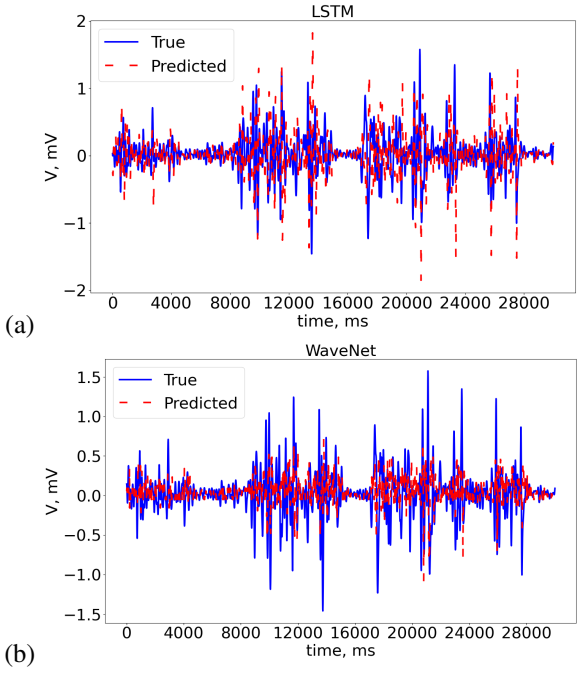


Figure 11. True (blue solid line) and predicted (red dashed line) value for 80 ms (8 steps) of prediction for Data Set 2. (a) LSTM. (b) WaveNet.

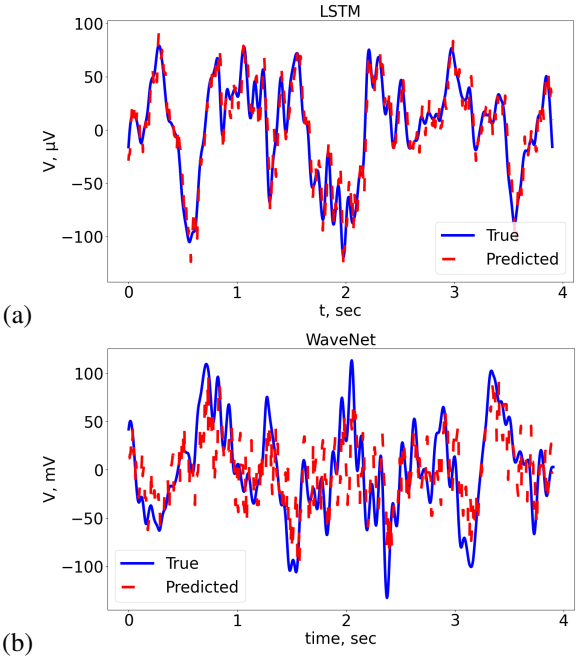


Figure 13. True (blue solid line) and predicted (red dashed line) value for 10 step prediction (sample rate is 256 Hz) of prediction for Data Set 3. (a) LSTM. (b) WaveNet.

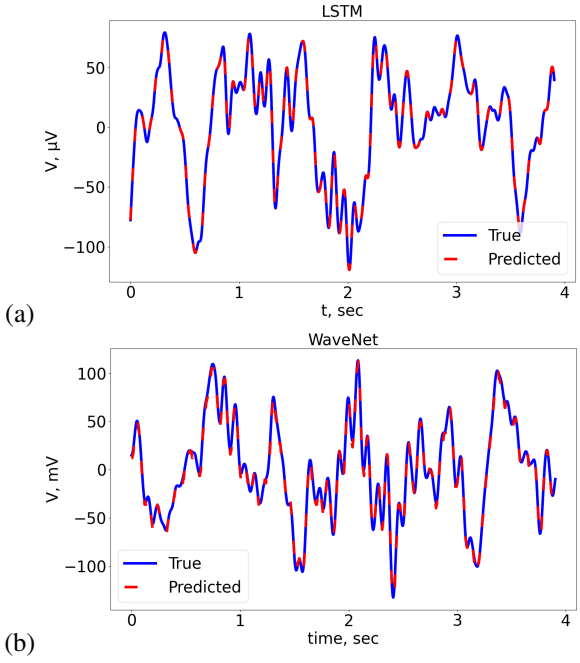


Figure 12. True (blue solid line) and predicted (red dashed line) value for 1 step prediction (sample rate is 256 Hz) of prediction for Data Set 3. (a) LSTM. (b) WaveNet.

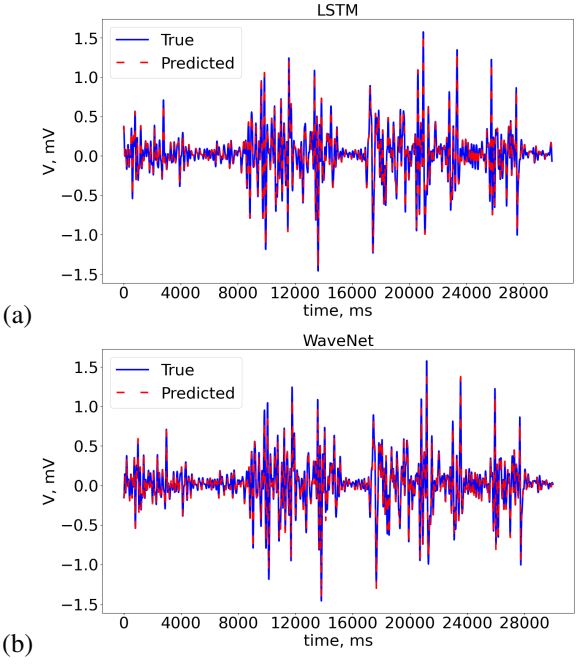


Figure 10. True (blue solid line) and predicted (red dashed line) value for 10 ms of prediction for Data Set 2. (a) LSTM. (b) WaveNet.

References

Beltyukova, A. V., Razin, V. V., Gromov, N. V., Samburova, M. I., Mishchenko, M. A., Kipelkin, I. M., Malkov, A. E., Smirnov, L. A., Levanova, T. A., Gerasimova, S. A., and Lebedeva, A. (2023). The concept of hippocampal activity restoration using artificial intelligence technologies. In *International Conference on Mathematical Modeling and Supercomputer Technologies*, Springer, pp. 240–252.

Bolton, T. and Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, **11** (1), pp. 376–399.

Borisenok, S. (2022). Detection and control of epileptiform regime in the hodgkin–huxley artificial neural networks via quantum algorithms. *Cybernetics and physics*.

Brenowitz, N. D. and Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, **45** (12), pp. 6289–6298.

Cai, D., Majda, A. J., McLaughlin, D. W., and Tabak, E. G. (2001). Dispersive wave turbulence in one dimension. *Physica D: Nonlinear Phenomena*, **152**, pp. 551–572.

Chen, N. and Majda, A. J. (2017). Beating the curse of dimension with accurate statistics for the fokker–planck equation in complex turbulent systems. *Proceedings of the National Academy of Sciences*, **114** (49), pp. 12864–12869.

Chholak, P., Pisarchik, A. N., Kurkin, S., Maksimenko, V. A., and Hramov, A. E. (2019). Neuronal pathway and signal modulation for motor communication. *Cybernetics and physics*, **8** (3), pp. 106–113.

Chowdhury, S. N., Majhi, S., Ozer, M., Ghosh, D., and Perc, M. (2019). Synchronization to extreme events in moving agents. *New Journal of Physics*, **21** (7), pp. 073048.

Cousins, W. and Sapsis, T. P. (2015). Unsteady evolution of localized unidirectional deep-water wave groups. *Physical Review E*, **91** (6), pp. 063204.

Ding, D., Zhang, M., Pan, X., Yang, M., and He, X. (2019). Modeling extreme events in time series prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1114–1122.

Dolinina, A. Y., Sysoeva, M. V., van Rijn, C. M., and Sysoev, I. V. (2022). Detection of spike-wave discharge restarts in genetic rat model based on frequency dynamics. *Cybernetics and physics*, **11**, pp. 121–130.

Durairaj, P., Sundararam, G. K., Kanagaraj, S., and Rajagopal, K. (2023). Prediction of dragon king extreme events using machine learning approaches and its characterizations. *Physics Letters A*, **489**, pp. 129158.

Dysthe, K., Krogstad, H. E., and Müller, P. (2008). Oceanic rogue waves. *Annual Review of Fluid Mechanics*, **40**, pp. 287–310.

Farazmand, M. and Sapsis, T. P. (2017). A variational approach to probing extreme events in turbulent dynamical systems. *Science advances*, **3** (9), pp. e1701533.

Feng, B. and Fox, G. (2022). Gtrans: Spatiotemporal autoregressive transformer with graph embeddings for nowcasting extreme events. *arXiv preprint arXiv:2201.06717*.

Frolov, N. S., Grubov, V. V., Maksimenko, V. A., Lüttjohann, A., Makarov, V. V., Pavlov, A. N., Sitnikova, E., Pisarchik, A. N., Kurths, J., and Hramov, A. E. (2019). Statistical properties and predictability of extreme epileptic events. *Scientific Reports*, **9** (1), pp. 1–8.

Gao, C., Zhou, L., and Zhang, R.-H. (2023). A transformer-based deep learning model for successful predictions of the 2021 second-year la niña condition. *Geophysical Research Letters*, **50** (12), pp. e2023GL104034.

Gerasimova, S. A., Lebedeva, A. V., Gromov, N. V., Malkov, A. E., Fedulina, , Levanova, T. A., and Pisarchik, A. N. (2023). Memristive neural networks for predicting seizure activity. , **15** (4 (eng)), pp. 30–37.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, E. H. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, **101** (23), pp. e215–e220.

Gromov, N., Gubina, E., and Levanova, T. (2022). Loss functions in the prediction of extreme events and chaotic dynamics using machine learning approach. In *2022 Fourth International Conference Neurotechnologies and Neurointerfaces (CNN)*, IEEE, pp. 46–50.

Gromov, N., Lebedeva, A., Kipelkin, I., Elshina, O., Yashin, K., Smirnov, L., Levanova, T., and Gerasimova, S. (2023). The choice of evaluation metrics in the prediction of epileptiform activity. In *International Conference on Mathematical Modeling and Supercomputer Technologies*, Springer, pp. 280–293.

Grooms, I. G. and Majda, A. J. (2014). Stochastic superparameterization in a one-dimensional model for wave turbulence. *Communications in Mathematical Sciences*, **12** (3), pp. 509–525.

Guth, S. and Sapsis, T. P. (2019). Machine learning predictors of extreme events occurring in complex dynamical systems. *Entropy*, **21** (10), pp. 925.

Han, J. and Jentzen, A. (2017). Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, **5** (4), pp. 349–380.

Han, J., Jentzen, A., and E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, **115** (34), pp. 8505–8510.

Hramov, A. E., Maksimenko, V. A., Pchelintseva, S. V., Runnova, A. E., Grubov, V. V., Musatov, V. Y., Zhu-

ravlev, M. O., Koronovskii, A. A., and Pisarchik, A. N. (2017). Classifying the perceptual interpretations of a bistable image using eeg and artificial neural networks. *Frontiers in neuroscience*, **11**, pp. 674.

Hussain, L., Saeed, S., Idris, A., Awan, I. A., Shah, S. A., Majid, A., Ahmed, B., and Chaudhary, Q.-A. (2019). Regression analysis for detecting epileptic seizure with different feature extracting strategies. *Biomedical Engineering/Biomedizinische Technik*, **64** (6), pp. 619–642.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, **349** (6245), pp. 255–260.

Kashiwagi, A., Urabe, I., Kaneko, K., and Yomo, T. (2006). Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. *PloS one*, **1** (1), pp. e49.

Kharif, C., Pelinovsky, E., and Slunyaev, A. (2008). *Rogue waves in the ocean*. Springer Science & Business Media.

Kumarasamy, S. and Pisarchik, A. N. (2018). Extreme events in systems with discontinuous boundaries. *Physical Review E*, **98** (3), pp. 032203.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, **521** (7553), pp. 436–444.

Lehnertz, K. (2006). Epilepsy: Extreme events in the human brain. In *Extreme events in nature and society*, pp. 123–143. Springer.

Li, Y., Xu, J., and Anastasiu, D. C. (2023). An extreme-adaptive time series prediction model based on probability-enhanced lstm neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 8684–8691.

Longin, F. (2016). *Extreme events in finance: A handbook of extreme value theory and its applications*. John Wiley & Sons.

Ma, C. and Wang, J. (2019). W. e, model reduction with memory and the machine learning of dynamical systems, commun. *Journal of Computational Physics*, **25** (4), pp. 947–962.

Majda, A. J. and Lee, Y. (2014). Conceptual dynamical models for turbulence. *Proceedings of the National Academy of Sciences*, **111** (18), pp. 6548–6553.

Majda, A. J., Moore, M., and Qi, D. (2019). Statistical dynamical model to predict extreme events and anomalous features in shallow water waves with abrupt depth change. *Proceedings of the National Academy of Sciences*, **116** (10), pp. 3982–3987.

Majda, A. J. and Tong, X. T. (2015). Intermittency in turbulent diffusion models with a mean gradient. *Nonlinearity*, **28** (11), pp. 4171.

Maksimenko, V. A., Kurkin, S. A., Pitsik, E. N., Musatov, V. Y., Runnova, A. E., Efremova, T. Y., Hramov, A. E., and Pisarchik, A. N. (2018). Artificial neural network classification of motor-related eeg: An increase in classification accuracy by reducing signal complexity. *Complexity*, **2018**.

Mishra, A., Leo Kingston, S., Hens, C., Kapitaniak, T., Feudel, U., and Dana, S. K. (2020). Routes to extreme events in dynamical systems: Dynamical and statistical characteristics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **30** (6), pp. 063114.

Mishra, A., Saha, S., Vigneshwaran, M., Pal, P., Kapitaniak, T., and Dana, S. K. (2018). Dragon-king-like extreme events in coupled bursting neurons. *Physical Review E*, **97** (6), pp. 062311.

Mohamad, M. A. and Sapsis, T. P. (2015). Probabilistic description of extreme events in intermittently unstable dynamical systems excited by correlated stochastic processes. *SIAM/ASA Journal on Uncertainty Quantification*, **3** (1), pp. 709–736.

Mohamad, M. A. and Sapsis, T. P. (2018). Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, **115** (44), pp. 11138–11143.

Olenin, S. M. and Levanova, T. A. (2023). Extreme events in small ensemble of bursting neurons with chemical and electrical couplings. In *2023 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp. 1–6.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E. (2018). Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical Review Letters*, **120** (2), pp. 024102.

Pisarchik, A. N., Grubov, V. V., Maksimenko, V. A., Lüttjohann, A., Frolov, N. S., Marqués-Pascual, C., Gonzalez-Nieto, D., Khramova, M. V., and Hramov, A. E. (2018). Extreme events in epileptic eeg of rodents after ischemic stroke. *The European Physical Journal Special Topics*, **227** (7), pp. 921–932.

Pisarchik, A. N., Jaimes-Reátegui, R., Sevilla-Escoboza, R., and Huerta-Cuellar, G. (2012). Multistate intermittency and extreme pulses in a fiber laser. *Physical Review E*, **86** (5), pp. 056219.

Plotnikov, S. (2021). Synchronization conditions in networks of hindmarsh–rose systems. *Cybernetics and physics*, **10**, pp. 254–259.

Pyragas, V. and Pyragas, K. (2020). Using reservoir computer to predict and prevent extreme events. *Physics Letters A*, **384** (24), pp. 126591.

Qi, D. and Majda, A. J. (2016). Predicting fat-tailed intermittent probability distributions in passive scalar turbulence with imperfect models through empirical information theory. *Communications in Mathematical Sciences*, **14** (6), pp. 1687–1722.

Qi, D. and Majda, A. J. (2018). Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models. *Communications in Mathematical Sciences*, **16** (1),

pp. 17–51.

Raissi, M. and Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, **357**, pp. 125–141.

Rasheed, S. (2021). A review of the role of machine learning techniques towards brain–computer interface applications. *Machine Learning and Knowledge Extraction*, **3** (4), pp. 835–862.

Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology.

Varadhan, S. S. (1984). *Large deviations and applications*. Society for Industrial and Applied Mathematics.

Viotti, C. and Dias, F. (2014). Extreme waves induced by strong depth transitions: Fully nonlinear results. *Physics of Fluids*, **26** (5), pp. 051705.

Wambura, S., Li, H., and Nigussie, A. (2020). Fast memory-efficient extreme events prediction in complex time series. In *Proceedings of the 2020 3rd International Conference on Robot Systems and Applications*, pp. 60–69.

Wan, Z. Y., Vlachas, P., Koumoutsakos, P., and Sapsis, T. (2018). Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PloS one*, **13** (5), pp. e0197704.

Wang, X., He, X., Wang, M., Feng, F., and Chua, T.-S. (2019). Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 165–174.

Weinan, E. and Vanden-Eijnden, E. (2010). Transition-path theory and path-finding algorithms for the study of rare events. *Annual review of physical chemistry*, **61**, pp. 391–420.

Weyn, J. A., Durran, D. R., and Caruana, R. (2019). Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, **11** (8), pp. 2680–2693.

Zhang, Y., Wu, R., Dascalu, S. M., and Harris Jr, F. C. (2024). A novel extreme adaptive gru for multivariate time series forecasting. *Scientific Reports*, **14** (1), pp. 2991.

Zio, E. and Pedroni, N. (2009). Estimation of the functional failure probability of a thermal–hydraulic passive system by subset simulation. *Nuclear Engineering and Design*, **239** (3), pp. 580–599.