UDC 004.8: 551.509.32
DOI: 10.15587/1729-4061.2025.321664

DEVISING A FORWARD PROPAGATION ARTIFICIAL NEURAL NETWORK APPLICATION TECHNOLOGY FOR NOWCASTING WEATHER ELEMENTS

Boris Perelygin
Corresponding author
PhD, Associate Professor*
E-mail: b.perelygin@gmail.com
Halyna Borovska
PhD, Associate Professor
Department of Meteorology and Climatology**
Hanna Hnatovska
PhD, Associate Professor*
Andrey Sergienko
PhD, Associate Professor*
Tetiana Tkach
PhD***
Nataliia Shtefan***
*Department of Information Technologies**
**Odesa I. I. Mechnikov National University
Vsevoloda Zmiienka str., 2, Odesa, Ukraine, 65082
***Department of Software Engineering
Odesa Polytechnic National University
Shevchenko ave., 1, Odesa, Ukraine, 65044

The object of this paper is the procedure of applying a forward-propagation artificial neural network of surface learning for the purpose of short-term forecasting one of the weather elements – the temperature of the near-surface air layer. Known methods to forecast weather elements, namely, hydrodynamic, physical-statistical, and synoptic, have been successfully supplemented in recent years by forecasting using artificial neural networks. It has become possible to build large networks for a large amount of training data for deep learning. However, the level of development of the theory of artificial neural networks does not make it possible to build the required network. Therefore, when solving applied tasks such as the one reported in this paper, the developer has to build a forecasting system blindly or based on some heuristic considerations, experimenting with neural networks. At the same time, the path of network complexity often does not lead to a qualitative improvement in forecasting results. Therefore, during the research, the main problem to be solved was to optimize the use of a simple neural network with a well-developed training algorithm for the purpose of nowcasting meteorological elements. The optimization criterion adopted was the surface air temperature short-term forecast accuracy at different time intervals. The parameters enabling the achievement of optimality are the parameters of the data that train the network and the parameters of the network itself. By selecting these parameters, a high accuracy rate for short-term forecasts of different timeliness has been achieved. The accuracy of a three-hour forecast reaches 100 percent. The same value is achieved for the forecast accuracy with a one-day lead time. Predicting temperature values for three days has an accuracy rate exceeding 90 percent

Keywords: forward propagation artificial neural network, nowcasting of weather elements

Received 06.11.2024
Received in revised form 18.12.2024
Accepted date 10.01.2025
Published date 27.02.2025

How to Cite: *Perelygin, B., Borovska, H., Hnatovska, H., Sergienko, A., Tkach, T., Shtefan, N. (2025). Devising a forward propagation artificial neural network application technology for nowcasting weather elements. Eastern-European Journal of Enterprise Technologies, 1 (2 (133)), 18–32. https://doi.org/10.15587/1729-4061.2025.321664*

1. Introduction

Weather is a set of weather phenomena and values of weather elements over a certain area at a certain time. There are not many primary weather elements: air temperature, atmospheric pressure, air humidity, visibility, insolation. Forecasting of these weather elements is carried out in three well-known main ways – hydrodynamic, physical-statistical, and synoptic. Obtaining a forecast by the hydrodynamic method involves solving a system of nonlinear aerohydrodynamic equations for a point or for a region. The physical-statistical method provides a forecast based on the analysis of a fairly long series of observations of weather elements at the point of observation or in the region. The synoptic forecast is obtained by analyzing the fields of weather elements.

Weather forecasting is carried out with different lead times – ultra-short-term, short-term, medium-term, long-term, circular.

Forecasting is one of the most important tasks that require prompt solutions in almost all areas of science and life. Forecasting weather elements is one of the oldest forecasting tasks due to their great impact on all aspects of human life. Meteorological weather forecasts are scientifically based assumptions about the future state of the weather. Studies and development of weather forecasting methods have been carried out in many countries. The success of modern short-term weather forecasts is quite high but there are also those that did not come true, especially in the cases of turbulent atmospheric conditions (presence of atmospheric fronts, cyclones, etc.). Therefore, research in this area is relevant.

2. Literature review and problem statement

In various fields of science, artificial neural networks (ANNs) are widely used for forecasting purposes. For

Copyright © 2025 Authors. This is an open access article under the Creative Commons CC BY license

example, in work [1], the application of ANN for forecasting impurities in atmospheric air is considered. Along with the conventional methods for weather forecasting mentioned above, the use of ANNs for prediction is also considered as a promising area of research. For example, it is proposed to use fully connected ANNs of forward propagation to forecast time series of moisture in mountain soils. To forecast geophysical time series, it is suggested to use multi-wavelet polymorphic ANNs. However, the work does not address the issue of selecting network parameters and optimizing output data because the networks used are complex and require deep learning on a large amount of data.

In paper [2], it is proposed to use extreme learning machines for forecasting long-term series, and in [3] the use of convolutional ANNs is considered for forecasting the energy of time series. These works consider general issues of construction and training of networks but the issues of optimization of constructed networks and, even more so, optimization of initial data, are not touched upon.

Work [4] deals with the forecasting of electric load using an ensemble of weather forecasts, which definitely complicates the artificial neural network and requires an increase in computing resources.

Paper [5] reports the results of combining different ANNs and their learning paradigms as an ensemble for weather forecasting and evaluates the weighted sum of their combination. But the questions related to the method of preparation of initial data, numerical evaluation of the quality of forecasting and influence on these estimates of ANN parameters and selection of initial data remained unresolved.

In [6], the results of a comparative analysis of the use of direct propagation ANNs, radial base networks, and feedback networks for weather forecasting are given. However, issues related to the method of preparation of initial data, numerical assessment of the quality of forecasting and influence on these estimates of ANN parameters and selection of initial data also remained unresolved.

Work [7] reports the results of a comparative analysis of radial-base networks, Hopfield, Elman networks as part of an ensemble for forecasting various weather elements. In the work, the questions related to the methodology for preparation of initial data, numerical assessment of the quality of forecasting and influence on these estimates of ANN parameters and selection of initial data also remained unresolved. The reason for this may be the sufficient amount of resources available to researchers, which makes relevant research impractical.

In [8], a probabilistic rainfall forecasting model based on Bayesian ANN with variational inference is proposed. It includes a combination of epistemic and aleatoric estimates of uncertainty. The model coped well with forecasting. However, this is a very resource-intensive model with a large volume of raw data. In addition, the methodology for preparing raw data is not covered.

In work [9], the problem of short-term, up to 12 hours in advance, precipitation forecasting is considered. The authors claim that the existing grid forecasting of weather elements requires powerful computing resources and there is a time delay in obtaining the result. They consider deep learning ANNs to be a promising alternative. It is proposed to make a hybrid network from the standard MetNet deep learning ANN by supplementing it with additional statistical processing. It is clear that such a system cannot be installed on a personal computer. Since the prediction is made with deep learning of the network on large initial data, the issues of their optimization are not considered.

Paper [10] tackles weather forecasting using deep learning 3D ANNs. The issues of medium-term forecasting are being considered. To avoid overtraining, random rearrangement of the raw data is assumed. Unfortunately, the issues of optimization of network parameters and output data are not considered.

In [11], to increase the probability of forecasting, a methodology of weather forecasting was devised based on the combination of ANNs. It has been proven that this approach to forecasting increases the quality of forecasts. However, the WeatherBench training data set is more than 300 gigabytes. It is planned to reduce this volume by reducing the number of combined ANNs at the same time. But all the same, this approach to forecasting remains very resource intensive.

In [12], the authors proposed a new ANN paradigm devised by them – a multimodal graph neural network. It is intended to improve weather forecasting in areas without specialized weather stations. These areas remain without meteorological coverage and do not have the detailed meteorological data necessary for accurate forecasts. Therefore, this network is designed for area forecasting and requires satellite data for forecasting. That is, it is a very resource-intensive system both from the point of view of computers and in terms of raw data.

Work [13] is the closest to our research area. It deals with the definition of an appropriate ANN model with a proper architecture (multilayer perceptron type) and its application for visibility (fog) prediction. Of the set of forecasting quality parameters, only visibility is investigated, and only for a set of fixed parameters of this ANN. The methodology of preparing initial data, numerical evaluations of the quality of forecasting, and the influence of initial data parameters on these evaluations are not covered.

Therefore, the following conclusions can be drawn from our review of available literature related to the subject of the current research.

Previously, due to the resource-intensive nature of the learning task, the modest success of ANNs was explained by the small size of the trained ANNs. Now it is possible to build a significantly larger ANN, but this often does not lead to a qualitative improvement in forecasting results. The level of development of the ANN theory currently does not provide a constructive method for building the desired network. Therefore, when solving applied tasks, it is necessary to experiment with different ANN paradigms, the number of hidden layers, the number of neurons in each layer, and the topology of connections between them, practically blindly or on the basis of some heuristics.

In scientific publications, papers on the optimization of ANN meta-parameters when using them for forecasting periodically appear. At the same time, making the proposed changes to the ANN topology most often entails its complete retraining, which, in turn, leads to a significant increase in the necessary computing resources. Moreover, the modern tendency to increase the size of training samples and the search for increasingly complex dependences also lead to the need to increase the size of ANNs and to the same need for computing resources.

There are many methods for weather forecasting using ANNs based on extrapolation but all of them have limited abilities to capture non-linear dependences of changes in weather elements. Therefore, the vector of research turned towards deep learning methods, which are considered to better represent nonlinear regularities. The way of using complex computer systems for forecasting units of the hydrometeorological service is problematic because it is resource intensive. Therefore, there is an urgent need to use ordinary personal computers for forecasting with not very complex ANNs, with a well-developed learning algorithm, and with optimized ANN parameters and output data. This combination of methods and tools for data processing make up forecasting technology.

Therefore, in our opinion, the problem related to the development of technology for forecasting weather elements using ANN of forward propagation on personal computers remains unsolved. As part of this task, it is necessary to determine the optimal parameters for the ANN itself and the parameters of the output data.

A likely option to overcome these challenges is the approach used in work [14]. This approach gives reasons to assert that it is appropriate to conduct a study aimed at devising a technology for the use of a not very complex ANN to forecast meteorological elements in order to meet the need for such technology by the production and forecasting units at the hydrometeorological service organizations.

3. The aim and objectives of the study

The purpose of our study is to devise a technology for the application of ANN with a well-developed learning algorithm to forecast weather elements with optimized ANN parameters and output data. This will allow the production and forecasting divisions at the hydrometeorological service organizations to forecast meteorological elements with different short-term lead times. This lead time is 3 hours, 1 day, and 3 days.

To achieve the goal, the following tasks must be solved:

– to determine the influence of the parameters of the data used to train the ANN on the accuracy of forecasts with the specified advance (length of training vectors, number of training vectors, types of output data) and to determine the values that ensure the best accuracy of forecasts;

– to establish the effect of ANN parameters on the accuracy of forecasts (number of hidden layers, presence of limitations in the activation functions of neurons of hidden layers) and optimal parameters of ANN, which ensure the best accuracy of forecasts with the specified lead time.

4. The study materials and methods

The object of our study is the process of using forward propagation ANN for forecasting air temperature values.

The subject of the study is a forward propagation ANN intended for short-term forecasting of air temperature values, and the parameters of the output data and the network.

The basic hypothesis assumes the possibility of building, using a personal computer, the technology of applying forward propagation artificial neural ANN for short-term forecasting of weather elements with optimization of the parameters of the initial data and network.

A well-known network of the following architecture was used as an ANN – a forward propagation ANN with surface learning using the error backpropagation procedure. The Levenberg-Marquardt algorithm was used for training.

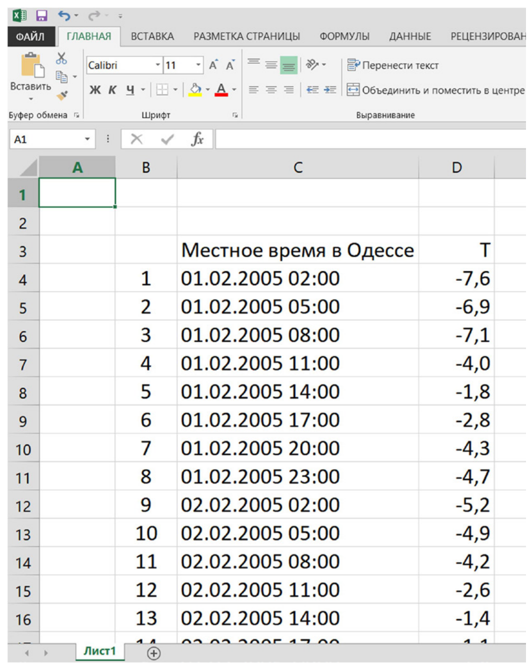
Based on our technical capabilities, short-term forecasting was carried out.

As an example of a weather element, the results of a study on forecasting air temperature values are given. This parameter is selected as the main one among all weather parameters, although another weather element could be selected.

The previously described types of data should be understood as the actual series of observations, the so-called "raw" data, and its transformation: the centered series is obtained by subtracting the arithmetic mean value of the series from all the values of the series. This transformation of the series was performed in order to assess the quality of forecasting depending on the type of data and their variable characteristics.

Our research was carried out by the method of forecasting modeling using ANN. Appropriate input data were used to train and operate the network. The initial data for weather forecasting are, as a rule, the results of regular measurements of weather elements in the form of numerical series.

The value of air temperature was chosen as the data for research because of the continuity of these data and the visibility of the results obtained. The data represent a long 15-year series of air temperature values obtained during regular eight-time observations at weather station 33837, Odesa, from February 1, 2005, to December 31, 2019 (Fig. 1). Graphically, this series is shown in Fig. 2.



	B	C	D
3		Местное время в Одессе	T
4	1	01.02.2005 02:00	-7,6
5	2	01.02.2005 05:00	-6,9
6	3	01.02.2005 08:00	-7,1
7	4	01.02.2005 11:00	-4,0
8	5	01.02.2005 14:00	-1,8
9	6	01.02.2005 17:00	-2,8
10	7	01.02.2005 20:00	-4,3
11	8	01.02.2005 23:00	-4,7
12	9	02.02.2005 02:00	-5,2
13	10	02.02.2005 05:00	-4,9
14	11	02.02.2005 08:00	-4,2
15	12	02.02.2005 11:00	-2,6
16	13	02.02.2005 14:00	-1,4

Fig. 1. A fragment of the used data on the air temperature of weather station 33837, Odesa

The data processing methodology for obtaining the results was as follows. Due to the large variability of meteorological values in space and time, the specific value of any value specified in the forecast should be considered as the most likely value that this value would take during the forecast period. After the expiration of the short-term forecast, an assessment of its precision is performed, which is based on accuracy. Accuracy is the degree of correspondence between forecast and actual meteorological values, phenomena with certain established tolerances. The accuracy of the temperature forecast is evaluated alternatively [15]. If the forecast temperature differed from the actual temperature by no more than 2.0 °C, then the accuracy of the forecast is 100 %, if the difference is 3.0 °C, then the accuracy of the forecast is 50 %, if the difference is greater than or equal to 4.0 °C, then the accuracy of the forecast is 0 %. During the research, the accuracy was calculated similarly to the above methodology. However, alternative scores (50 %, 0 %) were not used. Accuracy was quantified as the ratio of the number of forecasts that were accurate (which fell within the range of ±2 °C) to the total number of forecasts at a given lead time. That is, stricter conditions were used to determine accuracy.

The entire series of observations was divided into two large groups: the first group of data – for training, and the second group of data – for forecasting (Fig. 3).

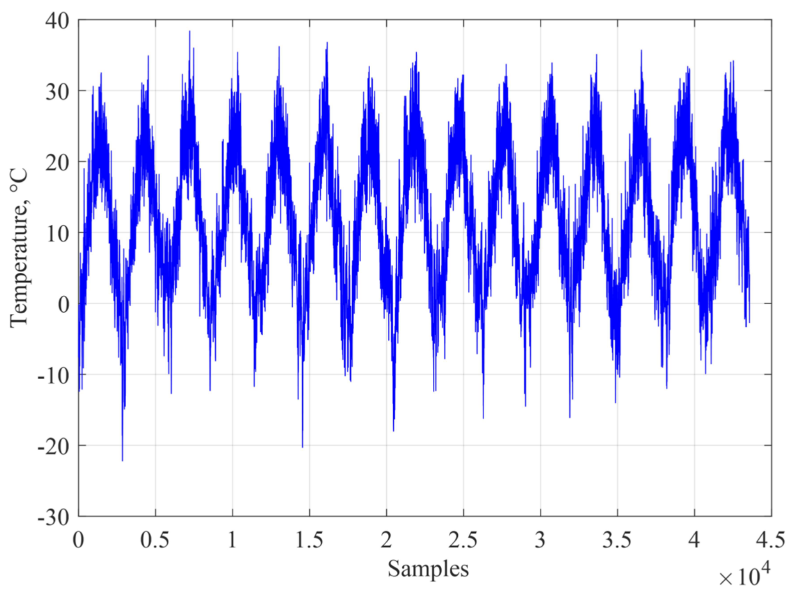


Fig. 2. A series of temperature values used to forecast air temperature values applying an artificial neural network (43,569 samples from 02/01/2005 to 12/31/2019)

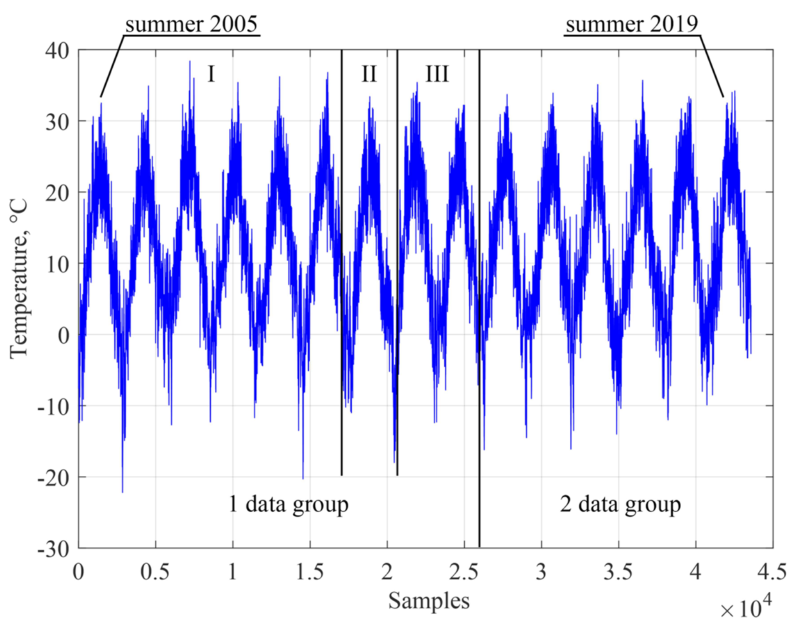


Fig. 3. Dividing the raw data into groups and arrays necessary for training an artificial neural network (data group 1: I — training set, II — control set, III — test set) and forecasting (data group 2)

From the first group of data intended for training, the necessary arrays of initial data were formed in the following way. The entire first group was divided into 3 parts. The first part (I in Fig. 3) was used for training ANN (training set). The second (II in Fig. 3) was used as a verification (control) set to check the quality of training. Multiple repetition of experiments leads to the fact that the control set begins to play a key role in the construction of the model, that is, it becomes part of the learning process. Thus, its role as an independent criterion of model quality is weakened – with a large number of experiments, there is a risk of choosing a network that gives a good result on the control set. In order to give the final model adequate reliability, the third part of the first group of data constituted a reserve (test) set of observations (III in Fig. 3). Therefore, the final model was tested on the data from the last set (III in Fig. 3) to make sure that the results achieved on the training and control sets are real. Based on the obtained data, the learning quality indicator was calculated as the difference between the temperature values of the training array and their approximation by the network. Subsequently, forecasting was carried out based on the second group of data.

From the first group of data, all arrays necessary for training (training, verification, and test) were built by the method illustrated in Fig. 4.

We shift along the original array of "raw" (or centered) data, starting from a given reference (which is chosen for the analysis of a certain season of the year). At each step, the "window", the size of which is equal to the specified length of the training (control, test) vector plus one count, is shifted to the right by one count. The training, verification, and test arrays were formed from the temperature values that fell into the "window", and the target and both reference arrays (for the verification set and for the test set) were formed from additional samples. The length of the "window" varied widely, so the training vectors also varied widely.

For the training of ANN, a training array and a target array corresponding to it were presented. For control over training, verification (or test) and reference arrays were presented. In the course of our research, the length and number of training vectors were changed, as well as the starting point of the array was changed to assess the seasonal impact of the data on the quality of the forecast.

After training the ANN from the second set of data to obtain a forecast in the order shown in Fig. 4, a data array for forecasting and a reference array for assessing the accuracy of the forecast were formed. The network was presented with an array of data for prediction, on the basis of which the network produced an extrapolated temperature value (forecast), which was compared to a reference temperature value (true temperature value). The timeline of the forecast with such array construction was determined by the number of vectors. In the case of forecasting for 3 hours, it is one vector, in the case of forecasting for one day, it is 8 vectors, and in the case of forecasting for three days, it is 24 vectors.

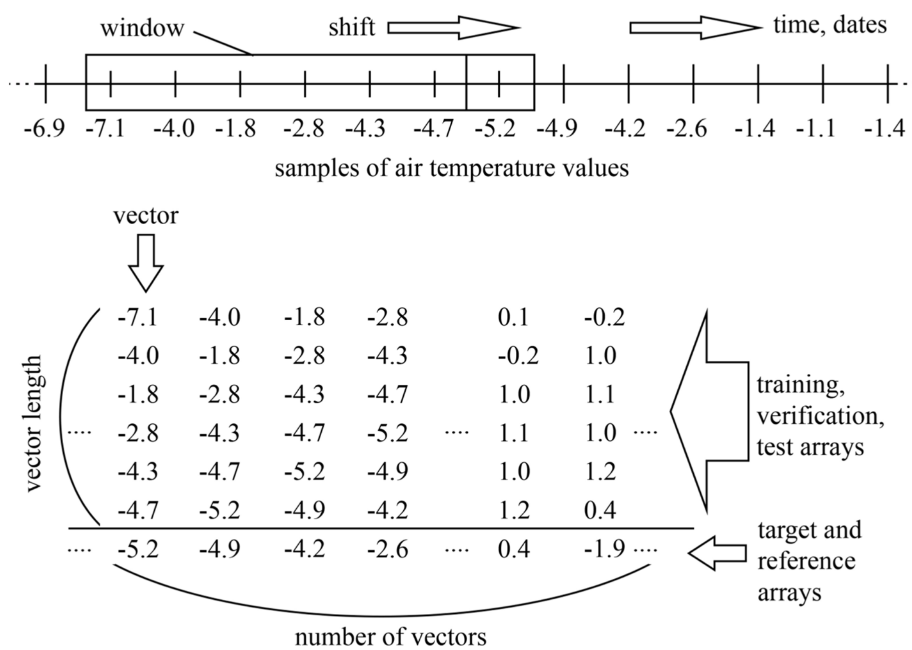


Fig. 4. The principle of forming arrays for training an artificial neural network and for forecasting temperature values with its help

The above procedure for obtaining a forecast was termed the one-time simulation procedure. When performing this procedure, the values of the quality of ANN training and the values of the forecast accuracy of forecasts were obtained in the form of plots. The results are obtained for different types of data, different length of training vectors and their number, different number of hidden layers and different activation functions of neurons, as well as different seasons of the year.

5. Results of investigating the use of an artificial neural network of forward propagation for short-term forecasting of temperature values

5. 1. Determining the influence of data parameters on the quality of forecasting and finding their optimal values

The process of using ANN for forecasting includes several stages, namely:

– data collection for training;
– preparation and normalization of data;
– selection of ANN topology;
– experimental selection of ANN characteristics;
– experimental selection of learning parameters;
– training;
– checking the adequacy of training;
– adjustment of parameters;
– final training;
– verbalization of the ANN for further use.

In the research process, not all stages were completed; no attention was paid to the last three. This is due to the fact that during the research, proving the solution to the problem to the level of developing a software package for forecasting was not considered.

During the formation of the array of initial data, missing air temperature values were interpolated as the arithmetic mean of neighboring temperature values. There were only 6 such omissions in the data, therefore, for the length of the series of 43569 samples, their correction did not have a significant impact on the research findings.

As a result of one procedure of one-time modeling, 72 three-dimensional plots were constructed. Since the scope of the paper does not make it possible to represent them all, two of them, as an example, are shown in Fig. 5. This is an indicator of the quality of training and the accuracy of the forecast for one day with a nonlinear activation function, two hidden layers, and "raw" data.

The length of the vectors is plotted along the abscissa axis, the number of training vectors along the ordinate axis, and the value of the parameter of the quality of training or the value of the parameter of the accuracy of the forecast of the corresponding timeliness are plotted along the axis of the application. Each point of these plots was calculated for the given length of the vector, the number of vectors, the number of hidden layers of the ANN, the type of activation function, and the type of output data. From a mathematical point of view, the plot in Fig. 5, *a* reflects the quality of approximation by the network of the training array, and in Fig. 5, *b* – the quality of data extrapolation by the trained network. The plot in Fig. 5, *b* clearly shows the instability of forecasts (highly variable value of accuracy) for any length of training vectors and for a small number of them.

In addition, the values of the approximation error and the distribution of the value of the approximation depending on the number of the training vectors were estimated for different lengths of the training vectors. As an example, they are shown in Fig. 6, 7 for the conditions described above, namely, with a nonlinear activation function, two hidden layers, "raw" data and with a length of training vectors equal to 15.

Of all the ANN paradigms, forward-propagation ANN using one-shot simulation was chosen for research. All parameters of the ANN (number of inputs corresponds to the length of the training vector, number of layers, type of activation function) were set and changed within the capabilities of the employed personal computer.

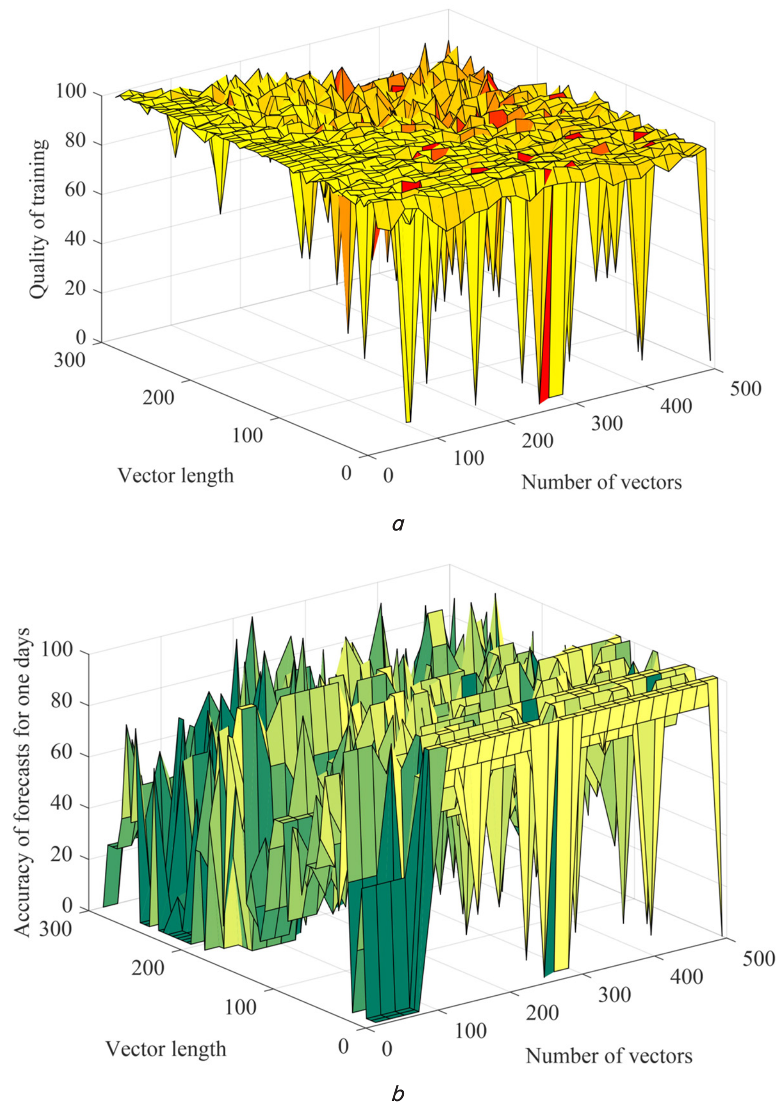


Fig. 5. An example of displaying parameters: *a* – quality of training; *b* – accuracy of forecasts

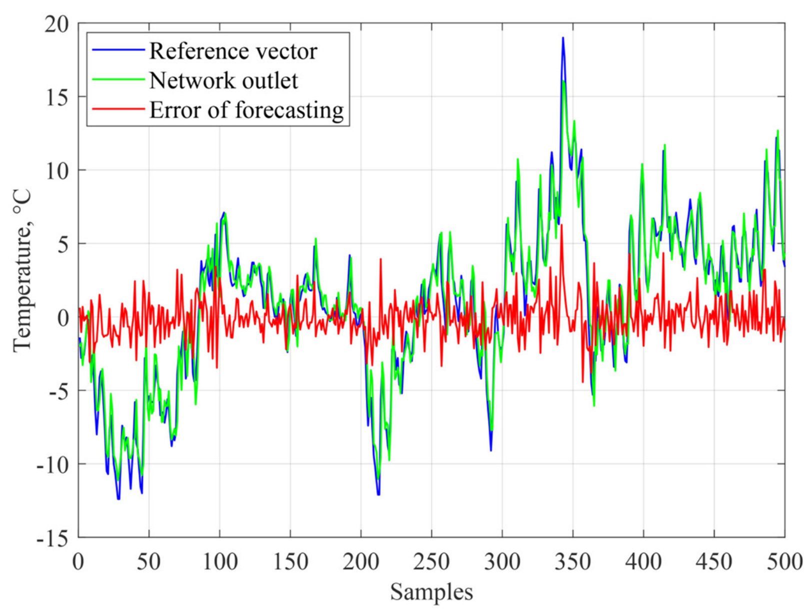


Fig. 6. Approximation error of the temperature values for the training array depending on the number of training vectors for the length of the training vectors equal to 15 and for the nonlinear activation function, two hidden layers, and "raw" data

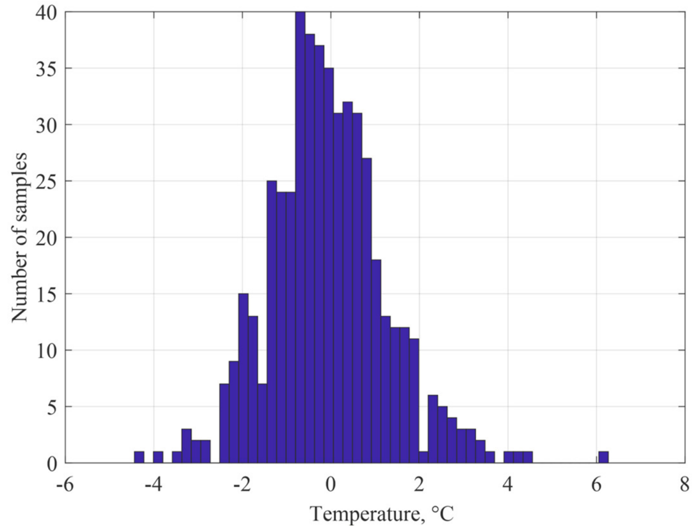


Fig. 7. Distribution of the values of approximation error of the training array temperature for the length of the training vectors equal to 15 and for the nonlinear activation function, two hidden layers, and "raw" data

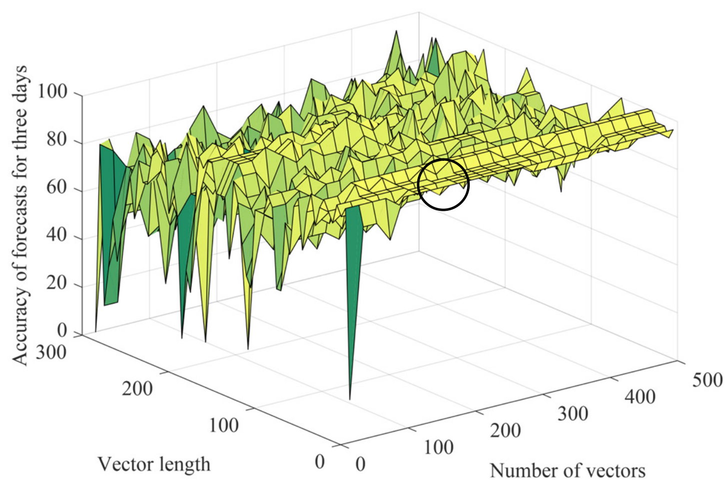
The number of hidden layers was changed from one to three since their further increase is impractical due to the complexity of the ANN and the increase in training time, which, however, did not lead to an improvement in the forecast accuracy of predictions. The activation function of artificial neurons was used in two ways: linear without limitation and sigmoidal type with limitation.

During the determination of the optimal length of the training vector and the optimal number of vectors that train the ANN, from the point of view of the forecast accuracy of the predictions, the simulation of the training procedure and the prediction procedure using the network was performed. As a result, the size of the training array (the length of the vectors and their number) was selected so that the training procedure was completed in no more than one hour. Otherwise, the sense of short-term forecasting was lost since the result could be obtained after the predicted time.

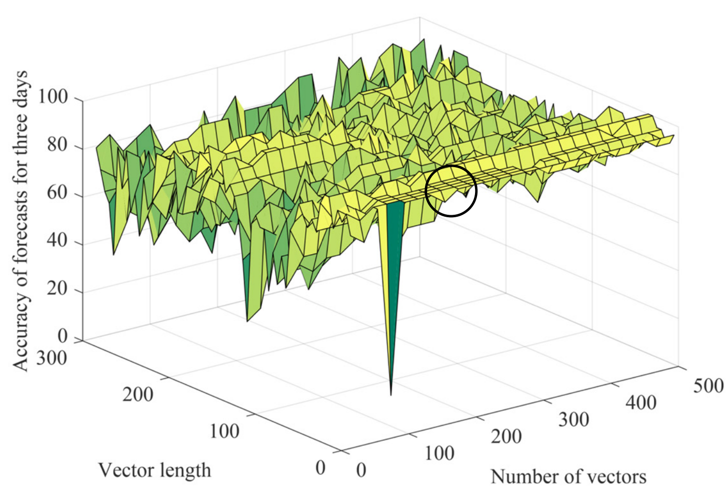
For the number of training vectors up to 500 values and for the length of each of them up to 300 values, as a result of simulation, 150,000 values of the parameters of the quality of training and the accuracy of forecasts appear. Solving a problem of this size on a personal computer takes more than 24 hours, and the results of calculations of neighboring points of the plot are very close. Therefore, after several simulation runs, the modeling step for the number of training vectors was taken equal to 20, and for the size of training vectors it was taken to be equal to 5. As a result, the plot began to contain 1500 points and turned out to be more compact and clearer, which can be seen in Fig. 5.

During the simulation, the quality of training and the accuracy of all three forecasts of different timeline were evaluated. At the same time, the following changes were made: the type of initial data ("raw", centered); number of hidden layers (1, 2, 3); activation function from linear to sigmoidal for hidden and output layers; seasons of the year.

According to the research results, it was found that the type of initial data does not affect the quality of forecasting, the accuracy does not change significantly when replacing "raw" initial data with centered ones. Plots in Fig. 8 illustrate this using the example of a three-day forecast.



a



b

Fig. 8. Accuracy of forecasts for 3 days with one hidden layer and a linear function of neuron activation for different types of input data: *a* – "raw" data, *b* – centered data

Similar results were obtained with other forecast timelines.

5. 2. Determining the impact of artificial neural network parameters on the quality of forecasting and finding their optimal values

Analysis of all three-dimensional plots with simulation results proved that the presence of non-linearity (limitation) in the activation function significantly worsens the indicator of forecasting quality – accuracy. This is well illustrated by Fig. 9.

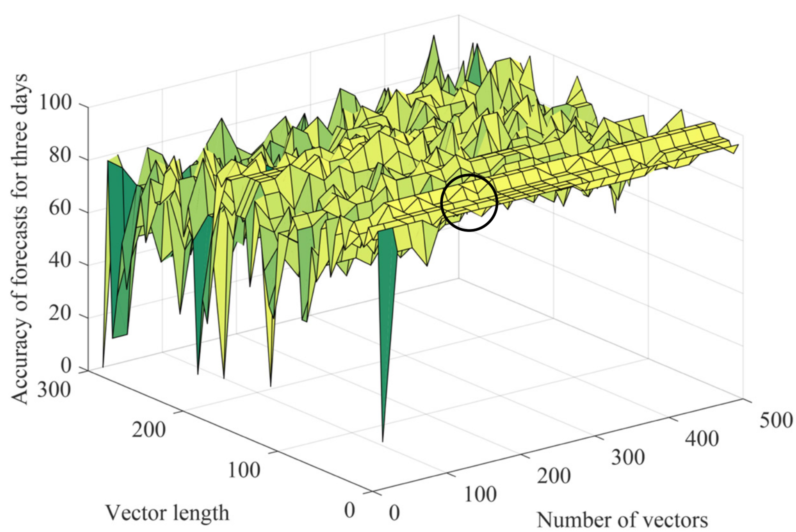
Significant instability of accuracy values is observed for a forecast with an advance of 3 days with a small number of vectors and with an arbitrary length of vectors. For forecasts with a different lead time, the picture is similar.

In addition, the presence of a non-linear activation function increases the training time of the network by more than 3 times. Therefore, when solving such a task, the neurons of the network must have a complete linear activation function.

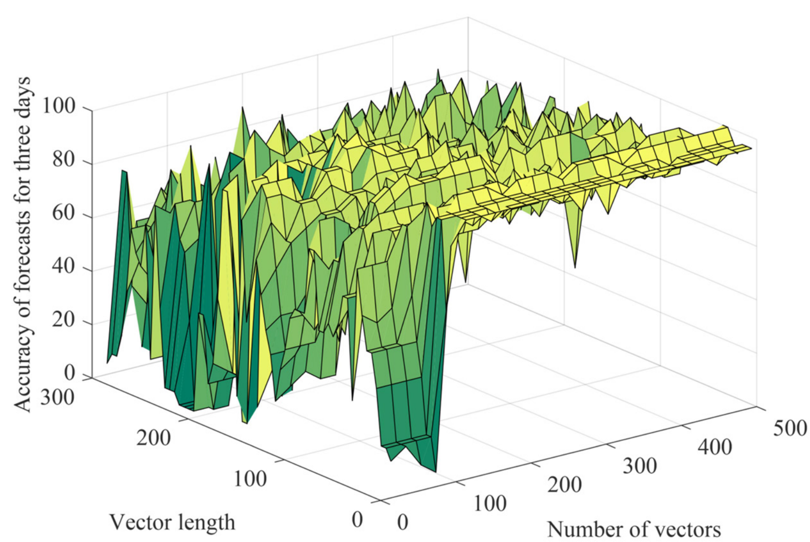
Increasing the number of hidden layers of the network does not improve, but also does not worsen the quality of prediction, the accuracy does not change significantly. In Fig. 10, using the example of a one-day forecast by a network with one, two, and three hidden layers, it is clearly visible; the fluctuations of the accuracy values are very similar.

However, if the number of hidden layers increases, the architecture of the network becomes more complicated, and with the same learning algorithm (the Levenberg-Marquardt algorithm in the procedure of error backpropagation), the learning time increases several times.

In addition, analysis of our data on the quality of network training and the accuracy of predictions reveals that good quality of training does not guarantee obtaining high accuracy of predictions. Moreover, this effect is observed with different power of the network, that is, with different numbers of neurons and layers. This thesis is well illustrated by Fig. 11, 12. Fig. 11 shows that the quality of training is good and slightly decreases with a large number of training vectors.



a



b

Fig. 9. Validity of forecasts for 3 days for the same type of initial data ("raw" data) and the same number of hidden layers (one layer) for different activation functions of neurons: *a* – for a linear activation function, *b* – for a nonlinear activation function

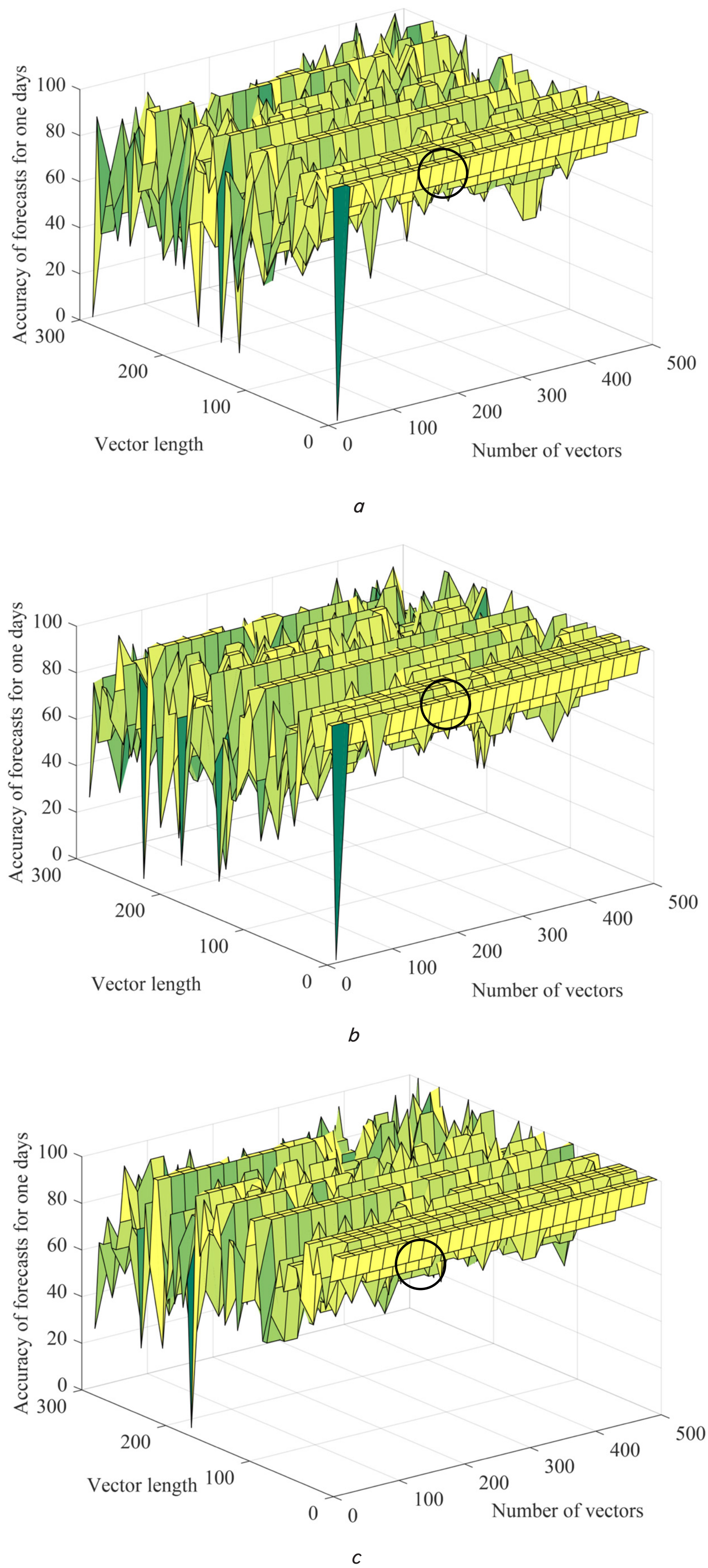


Fig. 10. Accuracy of forecasts for 1 day based on "raw" data and with a linear activation function based on the presence of: *a* – one hidden layer, *b* – two hidden layers, *c* – three hidden layers

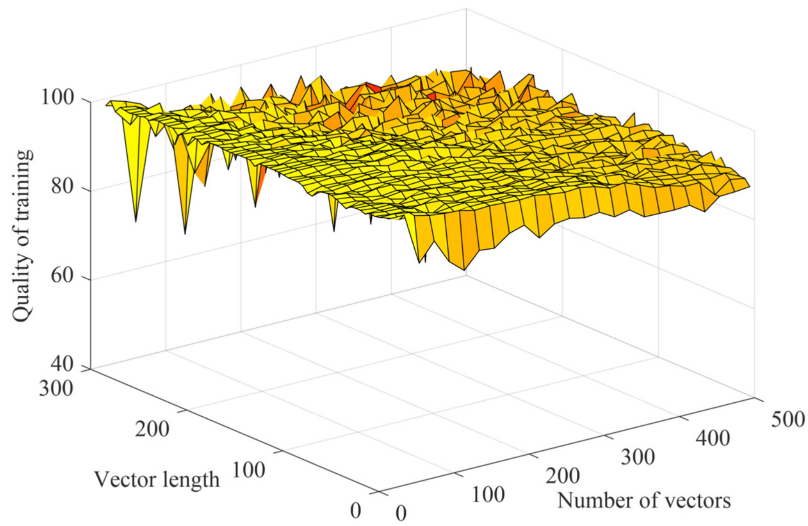


Fig. 11. Illustration of the quality of network training with one hidden layer and with a linear activation function

However, with an acceptable quality of training, the assessment of accuracy is unstable and fluctuates strongly (Fig. 12, *a–c*).

This may mean that the concept of good training quality is very conditional, i.e. a good approximation of the training data network is not synonymous with correct learning. This effect is very similar to human learning.

The studies described above were the first stage at which it became possible to determine the optimal type of input data, amount of output data, and main parameters of forward propagation ANN for forecasting temperature values.

Based on the results of the first stage of research, to obtain further numerical estimates of the parameters of the quality of learning and forecasting, a single-layer forward propagation ANN was used with linear functions of neuron activation, with training based on "raw" data, using the Levenberg-Marquardt algorithm, and with the volume of initial data in in the form of 150 vectors with a length of 15 each (the areas of stable forecasting are outlined in Fig. 8–10). Moreover, during training on such an insignificant amount of data, one mandatory condition must be observed: while forecasting in a certain season (time of the year), earlier data must be taken for training, but necessarily from exactly the same time of year. Otherwise, the prediction error will be unacceptably large.

The second stage of research involved the application of the specified network with the above parameters during short-term forecasting three hours, one day, and three days in advance and obtaining numerical estimates of parameters for the quality of training and forecasting. The results of the second stage of research are shown in Fig. 13–15.

The value of the error during operation, obtained as a result of the first stage of research of the network on the target set, is actually the result of the approximation of the objective function by the network (Fig. 6). The training quality indicator was 72 %, and the root mean square error (RMS) of approximation was 1.57 °C.

Fig. 13 shows the result of a one-time forecast for 3 hours. Since the short-term forecast for 3 hours is a forecast one step ahead in eight-time forecasting, we get one forecast value as a result. Fig. 13 demonstrates the reference temperature, network output value, and prediction error. From these data, it can be seen that the accuracy of the forecast for three hours was 100 %.

The desire to obtain a complete set of reliable forecasting statistics for three hours for plotting error values and error value distribution histograms requires multiple (at least 50 times for statistical reliability) repetition of the one-time simulation procedure.

While with a one-time simulation of the forecast for 3 hours, one can only get an alternative estimate of accuracy of 0 % or 100 %, then with a one-time simulation of the forecast for one day and three days, it is already possible to reasonably construct a histogram of the distribution of the forecast error value, estimate the root mean square deviation, and obtain the value of the evaluation of forecast accuracy.

Fig. 14 shows the result of one-time forecasting for one day. This short-term forecast for one day is a forecast eight steps ahead of an eight-term forecast, so we get eight forecast values as a result. Fig. 14, *a* shows the prediction error, and Fig. 14, *b–a* histogram of distribution of the error value. The accuracy of the forecast for one day is 100 %, and the RMS of the forecast for one day is 1.78 °C.

As in the first case, the desire to obtain a complete set of reliable statistics for constructing a plot of error values and a histogram of the distribution of error values requires multiple (at least 50 times for statistical reliability) repetition of the one-time simulation procedure.

Fig. 15 shows the result of a one-time forecast for three days. This three-day short-term forecast is twenty-four steps ahead of the eight-day forecast, so we end up with 24 forecast values. Fig. 15, *a* shows the prediction error, and Fig. 15, *b* – a histogram of the distribution of the error value. The accuracy of the three-day forecast is 91.7 %, and the RMS of the three-day forecast is 2.28 °C.

As in the first two cases, the desire to obtain a complete set of reliable statistics for constructing a plot of error values and a histogram of the distribution of error values requires multiple (at least 50 times for statistical reliability) repetition of the one-time simulation procedure.

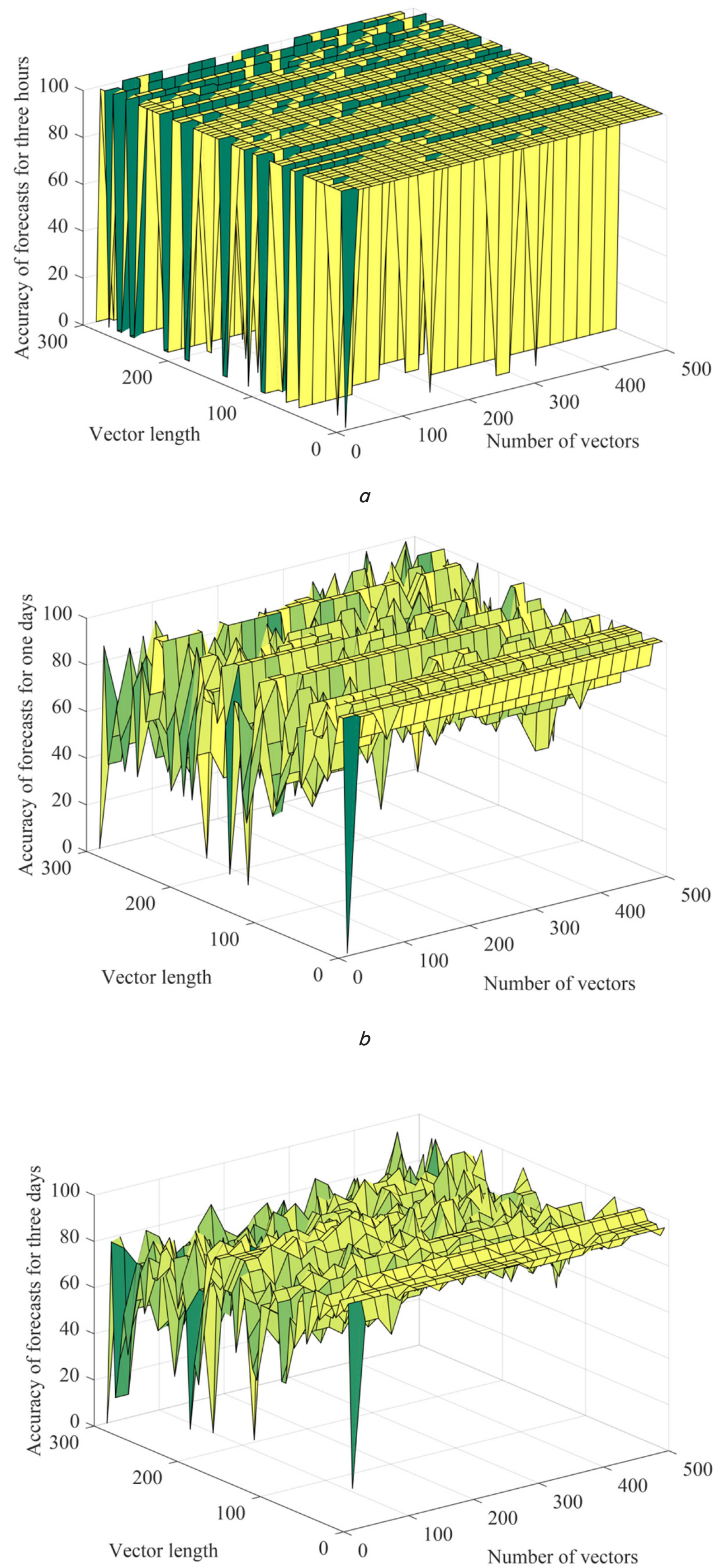


Fig. 12. Illustration of the quality of forecasting with one hidden layer and with a linear activation function: *a* – accuracy of the forecast for 3 hours, *b* – accuracy of the forecast for 1 day, *c* – accuracy of the forecast for 3 days

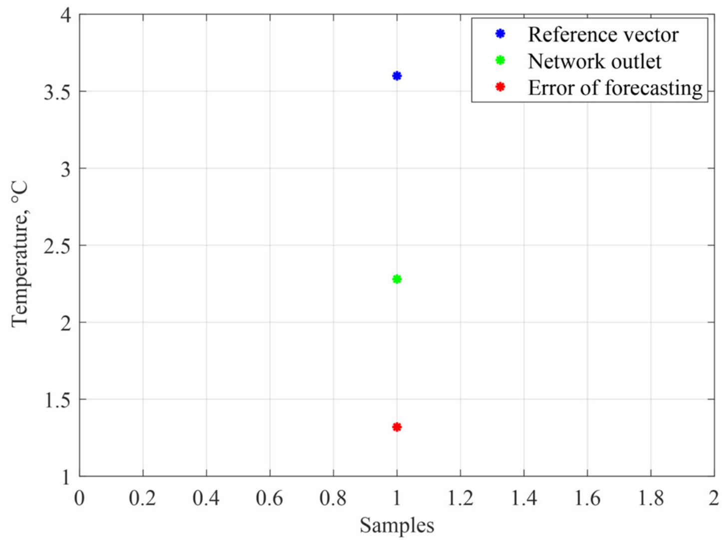
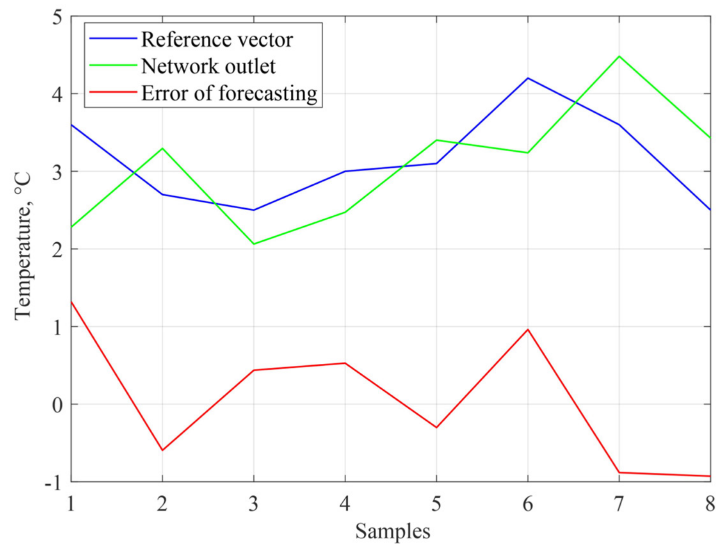
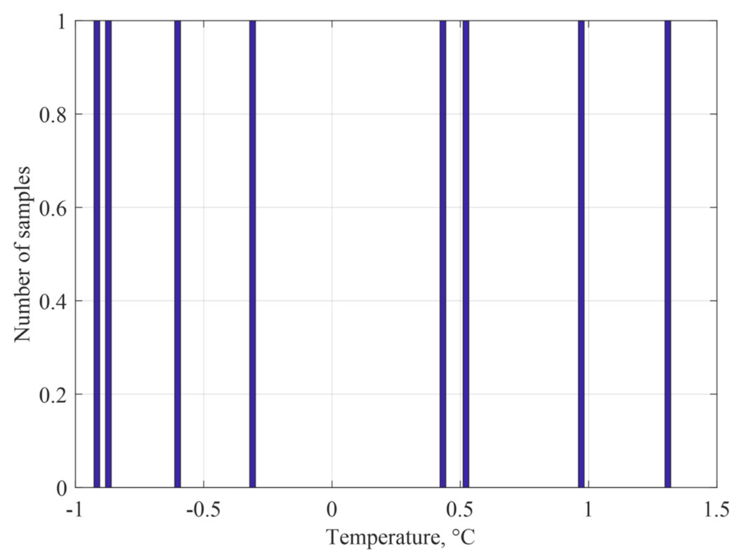


Fig. 13. Result of forecasting for three hours

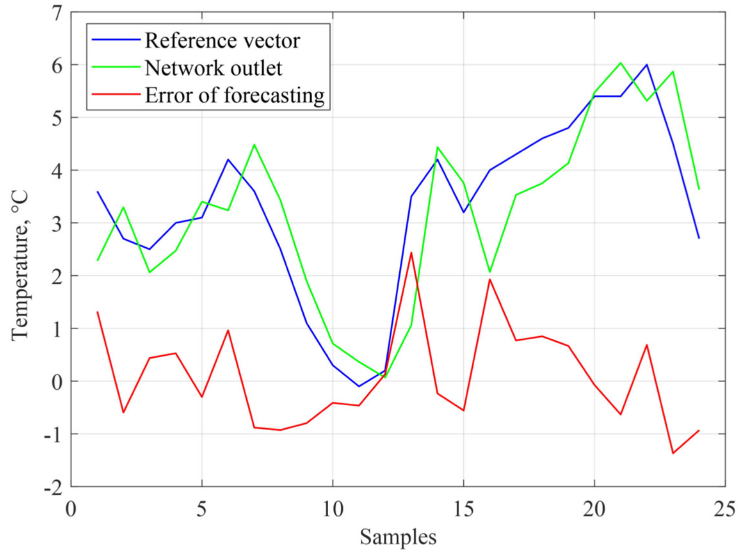


a

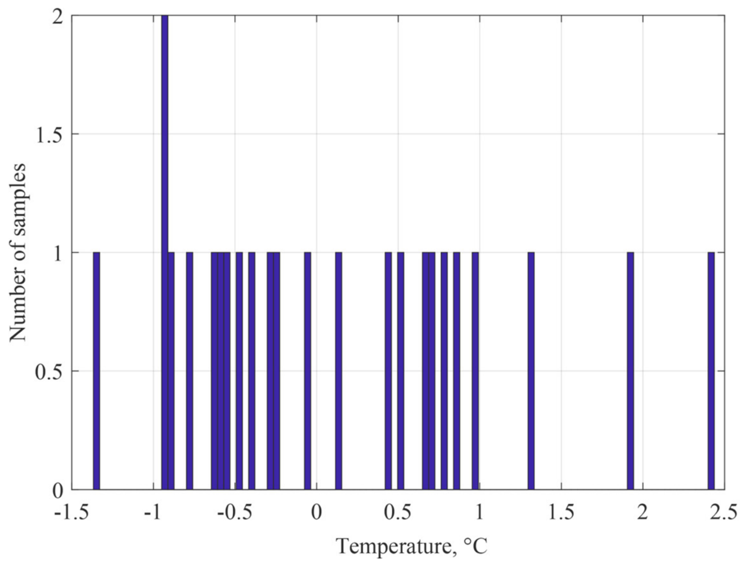


b

Fig. 14. Results of one-time forecasting for one day: *a* – forecasting error; *b* – histogram of the distribution of the error value



a



b

Fig. 15. Results of one-time forecasting for three days:
a — forecasting error; *b* — histogram of the distribution of the error value

6. Discussion of results base investigating the application of artificial neural network of forward propagation for short-term forecasting of temperature values

There are many technologies for the use of artificial neural networks for forecasting weather elements, as well as various ANNs. But there is a need for simple means and methods of short-term forecasting for use in the production and forecasting units at the hydrometeorological service organizations. These means and methods are ordinary personal computers with not very complex ANNs, with a well-developed learning algorithm, and with optimized ANN parameters and output data. The advantage of the proposed forecasting technology is its simplicity. A not very complex ANN is a forward propagation network with training using the method of error backpropagation and with the Levenberg-Marquardt learning algorithm as the fastest. In contrast to [13], in which only the parameters of neural networks considered the influence on the quality of visibility prediction, the influence of the parameters of the initial data was also studied. This prompted the idea to formulate the technology of using a relatively simple artificial neural network for forecasting air temperature values with the help of personal computers. This approach was formulated in [14]. Based on this approach, the technology of using an artificial neural network of forward propagation for forecasting weather elements has been developed. It demanded, among other things, investigating how the parameters of both the ANN and the source data affect the quality of forecasting. As a result, the best (optimized) parameters of the ANN and the best (optimized) parameters of the initial data have been determined.

As a result of our study, it was established that the type of initial data does not significantly affect the quality of forecasting (Fig. 8). The optimal length of training vectors and their number are equal to 15 and 150, respectively (outlined areas in Fig. 8–10). Such ANN and data parameters provide the best accuracy for short-term forecasts (Fig. 13–15).

An unexpected result of the study is that high-quality ANN training does not always ensure the quality of the forecast (Fig. 11, 12). This result requires clarification.

Our research has also revealed that ANN parameters significantly affect the quality of forecasting. The presence of any limitation in the neuron activation function leads to a sharp deterioration in the quality of forecasts, therefore the neuron activation function must be completely linear (Fig. 9). The number of hidden layers does not significantly affect the quality of forecasts (Fig. 10).

Thus, the proposed technology for short-term forecasting of weather elements, in particular, air temperature, will make it possible to achieve this with simple means and methods, without the involvement of powerful computers and cloud systems. Based on the developed technology, short-term forecasts of temperature values were obtained, and their quality was assessed. The quality of the obtained short-term forecasts is high (the accuracy of the forecast for three hours and one day is almost 100 %, and the accuracy of the forecast for three days is at least 91 %).

With the practical application of such technology, there is a need to forecast other weather elements in a similar way. However, significant problems could arise when forecasting atmospheric pressure. The normal pressure value is 1013 hPa, and its changes are less than 10 percent of this value. The behavior of ANNs when trained on such data requires additional research.

The drawback of the research is that all the reported research findings are based on data that reflect a calm atmosphere in some seasons. Both network training and forecasting were carried out on such data. Therefore, in the development of this study, it is of undoubted interest to study the impact on the quality of short-term forecasting of the temperature values of the disturbed atmosphere. It is also necessary to carry out multiple simulations of the forecast, during which the number of obtained estimates increases, the possibility of obtaining statistical estimates of accuracy appears, which increases the reliability of the forecast. The use of ANNs for medium- and long-term forecasting of weather elements is also of particular interest.

7. Conclusions

1. The optimal data sample for training the forward propagation ANN has been determined, which provides the best accuracy of short-term temperature forecasts. This training sample consists of 150 training vectors with a length of 15 counts each. Larger and smaller sizes of vectors and their larger and smaller number lead to a deterioration in the quality of forecasts and to instability of forecasts. The impact on the validity of short-term temperature forecasts of the parameters of the data used to train the neural network and the optimal parameters of the data have been determined. Optimal parameters are "raw" data and the same seasonality of data for training and forecasting. "Raw" data and centered data give close accuracy values, but "raw" is the best because it does not require any additional processing unlike centered. The condition of the same seasonality of the data arose due to the fact that the length of the training vectors turned out to be small.

2. The optimal parameters of the ANN have been determined. First, it is the requirement for neurons to have a complete linear activation function. The presence of any non-linearity (constraint) in the activation function leads to unacceptable accuracy errors in prediction. Secondly, it is the presence of one hidden layer in the network. Increasing the number of hidden layers does not improve the quality of prediction. This is due to the nature of the initial data, namely, a simple harmonic numerical series with seasonal periodicity and an imposed noise component in the form of daily temperature fluctuations. According to the understanding of the operation of such a neural network, two hidden layers would be optimal for the task of temperature forecasting, one to respond to seasonal fluctuations, the second to daily fluctuations. However, the energy of the noise component turned out to be very insignificant and did not manifest itself in the experiment. Thirdly, training was carried out on the basis of the procedure of error backpropagation using the Levenberg-Marquardt algorithm. The error backpropagation procedure is well established, and the Levenberg-Marquardt algorithm is the fastest. Applying other training algorithms more than doubles the training time of the network without significantly improving the accuracy.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

All data are available, either in numerical or graphical form, in the main text of the manuscript.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Moustris, K., Larissi, I., Nastos, P. T., Koukouletsos, K., Paliatsos, A. G. (2012). 24-Hours Ahead Forecasting of PM10 Concentrations Using Artificial Neural Networks in the Greater Athens Area, Greece. Advances in Meteorology, Climatology and Atmospheric Physics, 1121–1126. https://doi.org/10.1007/978-3-642-29172-2_156
2. Lei, Y., Zhao, D., Cai, H. (2015). Prediction of length-of-day using extreme learning machine. Geodesy and Geodynamics, 6 (2), 151–159. https://doi.org/10.1016/j.geog.2014.12.007
3. Koprinska, I., Wu, D., Wang, Z. (2018). Convolutional Neural Networks for Energy Time Series Forecasting. 2018 International Joint Conference on Neural Networks (IJCNN), 1–8. https://doi.org/10.1109/ijcnn.2018.8489399

4. Taylor, J. W., Buizza, R. (2002). Neural network load forecasting with weather ensemble predictions. IEEE Transactions on Power Systems, 17 (3), 626–632. https://doi.org/10.1109/tpwrs.2002.800906
5. Maqsood, I., Khan, M., Abraham, A. (2004). An ensemble of neural networks for weather forecasting. Neural Computing and Applications, 13 (2). https://doi.org/10.1007/s00521-004-0413-4
6. Maqsood, I., Abraham, A. (2007). Weather analysis using ensemble of connectionist learning paradigms. Applied Soft Computing, 7 (3), 995–1004. https://doi.org/10.1016/j.asoc.2006.06.005
7. Maqsood, I., Khan, M. R., Abraham, A. (2003). Weather Forecasting Models Using Ensembles of Neural Networks. Intelligent Systems Design and Applications, 33–42. https://doi.org/10.1007/978-3-540-44999-7_4
8. Harnist, B., Pulkkinen, S., Mäkinen, T. (2024). DEUCE v1.0: a neural network for probabilistic precipitation nowcasting with aleatoric and epistemic uncertainties. Geoscientific Model Development, 17 (9), 3839–3866. https://doi.org/10.5194/gmd-17-3839-2024
9. Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C. et al. (2022). Deep learning for twelve hour precipitation forecasts. Nature Communications, 13 (1). https://doi.org/10.1038/s41467-022-32483-x
10. Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619 (7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3
11. Clare, M. C. A., Jamil, O., Morcrette, C. J. (2021). Combining distribution-based neural networks to predict weather forecast probabilities. Quarterly Journal of the Royal Meteorological Society, 147 (741), 4337–4357. https://doi.org/10.1002/qj.4180
12. Yang, Q., Giezendanner, J., Civitarese, D. S., Jakubik, J., Schmitt, E., Chandra, A. et al. (2024). Multi-modal graph neural networks for localized off-grid weather forecasting. arXiv. https://doi.org/10.48550/arXiv.2410.12938
13. Chaudhuri, S., Das, D., Sarkar, I., Goswami, S. (2015). Multilayer Perceptron Model for Nowcasting Visibility from Surface Observations: Results and Sensitivity to Dissimilar Station Altitudes. Pure and Applied Geophysics, 172 (10), 2813–2829. https://doi.org/10.1007/s00024-015-1065-2
14. Perelygin, B., Tkach, T., Gnatovskaya, A. (2021). Preparing data and determining parameters for a feedforward neural network used for short-term air temperature forecasting. Short Paper Proceedings of the 2nd International Conference on Intellectual Systems and Information Technologies (ISIT 2021) co-located with 1st International Forum "Digital Reality" (DRForum 2021). Odesa, 56–61. Available at: https://ceur-ws.org/Vol-3126/paper8.pdf
15. Nastanova z meteorolohichnoho prohnozuvannia. Ukrainskyi hidrometerolohichnyi tsentr. Kyiv, 35. Available at: https://www.meteo.gov.ua/f/pro_nas/normatuvni_akt/Nastanova%20z%20meteoprognozuvannia.pdf