DE GRUYTER
OPEN

Ekológia (Bratislava) Vol. 36, No. 3, p. 226–234, 2017
DOI:10.1515/eko-2017-0019

SELECTION OF PARAMETERS AND ARCHITECTURE OF MULTILAYER PERCEPTRONS FOR PREDICTING ICE COVERAGE OF LAKES

NIKOLAEVICH VYACHESLAV BAKLAGIN

Northern Water Problems Institute, Karelian Research Centre, Russian Academy of Sciences, Republic of Karelia, Aleksander Nevsky st., 50, 185030 Petrozavodsk, Russia; e-mail: slava.baklagin@mail.ru

Abstract

Baklagin V.N.: Selection of parameters and architecture of multilayer perceptrons for predicting ice coverage of lakes. Ekológia (Bratislava), Vol. 36, No. 3, p. 226–234, 2017.

The ice cover on lakes is one of the most influential factors in the lakes' winter aquatic ecosystem. The paper presents a method for predicting ice coverage of lakes by means of multilayer perceptrons. This approach is based on historical data on the ice coverage of lakes taking Lake Onega as an example. The daily time series of ice coverage of Lake Onega for 2004–2017 was collected by means of satellite data analysis of snow and ice cover of the Northern Hemisphere. Input signals parameters for the multilayer perceptrons aimed at predicting ice coverage of lakes are based on the correlation analysis of this time series. The results of training of multilayer perceptrons showed that perceptrons with architectures of 3-2-1 within the Freeze-up phase (arithmetic mean of the mean square errors for training epoch $\overline{MSE}$ = 0.0155) and 3-6-1 within the Break-up phase ($\overline{MSE}$ = 0.0105) have the least mean-squared error for the last training epoch. Tests within the holdout samples prove that multilayer perceptrons give more adequate and reliable prediction of the ice coverage of Lake Onega (mean-squared prediction error *MSPE* = 0.0076) comparing with statistical methods such as linear regression, moving average and autoregressive analyses of the first and second order.

Key words: Freeze-up phase, Break-up phase, Lake Onega, training sample, statistical methods.

Introduction

Ice cover on lakes impacts theweather, fishing industry, oxygenation and penetration of sunlight needed for photosynthesis within the lake (Karetnikov, Naumenko, 2008). Prediction of its spatial characteristics such as ice coverage of lakes is of considerable interest for solution of environmental and economic problems related with the lake (Atapaththu et al., 2017). Nowadays certain mathematical models have been developed to be widely used to reproduce the thermohydrodynamics of the lakes, such as POM, NEMO, ECOM, ELCOM and others (Blumberg, Mellor, 1986; Madec, NEMO team, 2015; Quamrul Ahsan, Blumberg, 1999; Dallimore et al., 2003; Menshytkin et al., 2013; Sharif et al., 2013), including ice modelling. However, the listed models are complex and do not have sufficient flexibility in adaptation to specific lakes. Besides, they require collection of a large amount of diverse input information, which makes it difficult to use them in modelling and predicting ice coverage of lakes.

Statistical methods based on the use of probabilistic models, such as regression, autoregressive (ARIMAX, ARHCH), exponential smoothing, and others (Box et al., 2015) can be used to predict daily time series, including one for the ice coverage of lakes. These methods are simple enough, because they do not need a mathematical description of the physical processes in the lake. However, the abovementioned methods are not intended for modelling non-linear processes, since they give large errors. Therefore, they are unsuitable for predicting ice coverage of lakes.

In accordance with a certain training sample, artificial neural networks show good results in solving problems related to the prediction of nonlinear processes (Hagan et al., 2014). The advantages of this method are flexibility, adaptability, ease of use and high accuracy of the forecast.

In this regard, the purpose of this study is to select the parameters of multilayer perceptrons (a specific instance of an artificial feedforward neural network) for predicting ice coverage of lakes taking Lake Onega as an example.

The main objectives of the study are:

1. Formation of training and holdout samples on the basis of daily time series of ice coverage of Lake Onega for 2004–2016 and 2016–2017 correspondingly.
2. Selection of optimal architecture and parameters of multilayer perceptrons for predicting ice coverage of Lake Onega and their training.
3. Testing of the trained perceptrons as well as statistical prediction methods within holdout samples in comparison with the actual values of ice coverage.

Material and methods

As a rule, the ice regime of lakes includes three phases: Freeze-up, Complete Ice cover, Break-up. Each of the phases is characterized by its own peculiarities of the ice coverage dynamics (growth, decrease or stagnation of values). This study considers two phases of the ice regime: Freeze-up and Break-up, since within the Complete Ice cover phase, the area of ice cover remains practically unchanged. Moreover, the perceptron parameters for predicting ice coverage were selected separately for each phase of the ice regime under consideration. This procedure facilitated the optimization of the training process of each perceptron substantially.

The database of the satellite observations (AQUA, TERRA, NOAA-14,15,16,17,18, GOES-9,10,11,13, etc.) on the ice cover of the Northern Hemisphere with a spatial resolution of 4 km was used as a basis for obtaining daily time series of ice coverage of Lake Onega. The observations are conducted from 2004 to the present with a temporal resolution of 1 day (http://nsidc.org/data/G02156). Daily indicators of ice coverage for the period of 2004–2017 were calculated (Fig. 1) and allocated into phases of the ice regime as a result of the automated analysis of used satellite data where the aquatic area coordinates of Lake Onega are identified.

In accordance with the Fig. 1, it can be concluded that Lake Onega appears to be fully ice-covered and ice-free on an annual basis. Therefore, ice coverage behaviour of function of time $f_{ice}(t)$ is cyclical and achieves limited values from 0 to 1. This behaviour of function $f_{ice}(t)$ makes it possible to avoid the normalizing of input and output signals of perceptrons preconditioned by ice coverage of lake.

General processing algorithm of perceptrons for predicting ice coverage of lakes

In this study, multilayer perceptrons with one hidden layer were used (Fig. 2). The architecture of such perceptrons is represented by three layers: the input layer with number of neurons (m), the hidden layer with number of neurons (n), and the output layer with number of neurons (p).

Multilayer perceptrons transmit signals from neurons of one layer to another by means of synaptic connections that have definite weights different for each pair of neurons. The general processing algorithm for a neural network of this type is as follows (Fig. 2):

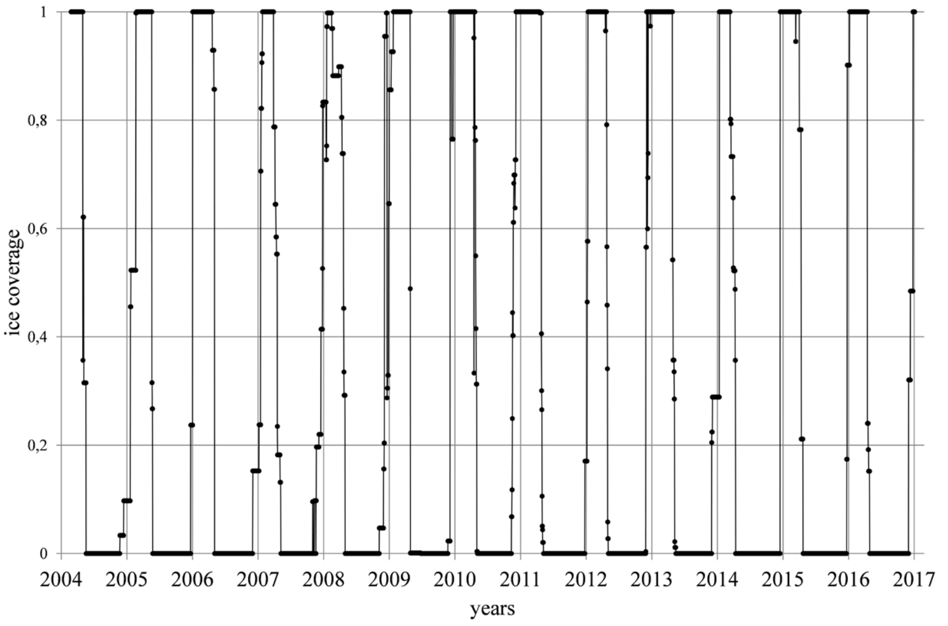

Fig. 1. Ice coverage of Lake Onega according to the satellite observations for the period of 2004–2017.

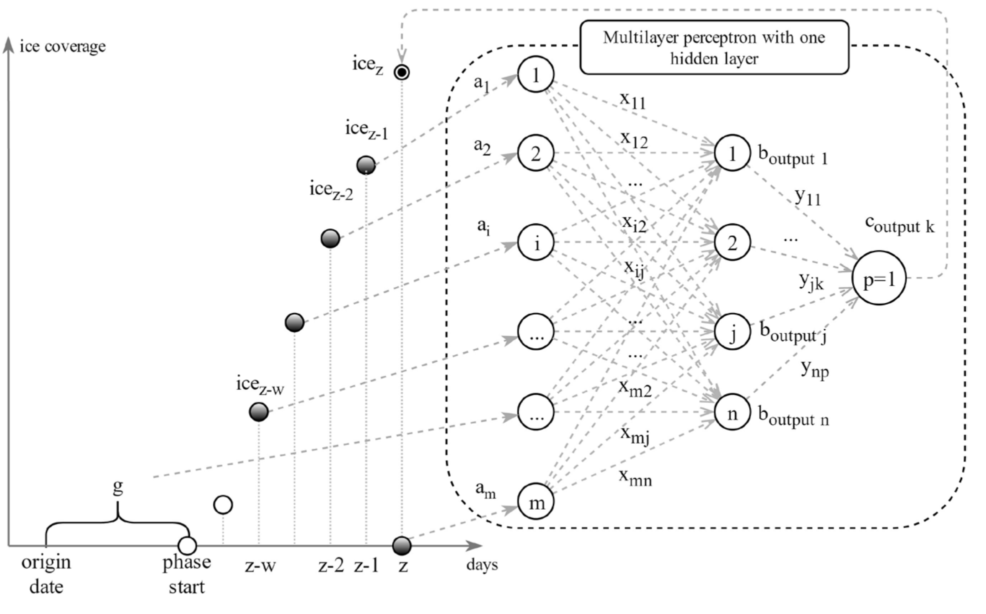

Fig. 2. Scheme of applying of a multilayer perceptron for predicting ice coverage of lakes on the basis of previous values.

1. The input neurons receive a signal $a = [a_1, a_2, ... a_i, ... a_m]$. By means of established weights of the synaptic connections, signals from each input neuron are transformed, summed and fed forward into the neurons of the hidden layer as an input signal: $b_{input\ j} = \sum_{i=1}^{m} a_i \cdot x_{ij}$, where x_{ij} is the weight of synaptic connection between neuron of the input layer and neuron of the hidden layer, $i = 1..m$, $j = 1..n$.
2. Within each neuron of the hidden layer, the signal is transformed with an activation function $b_{output\ j} = f_a\ (b_{input\ j})$. Similarly, signals from the neurons of the hidden layer are transformed, summed and fed forward into the neurons of the next (output) layer: $c_{input\ k} = \sum_{j=1}^{n} b_{output\ j} \cdot y_{jk}$, where y_{jk} is the weight of synaptic connection between neuron of the hidden layer and neuron of the output layer, $k = 1..p$..
3. Within each neuron of the output layer, the signal is transformed with an activation function: $c_{output\ k} = f_a(c_{input\ k})$. Then, the result is interpreted.

Algorithm and parameters of training of multilayer perceptrons for predicting ice coverage of lakes

Within this study, the training of multilayer perceptrons was carried out with the backpropagation method (with teacher) (Haykin, 1999). The backpropagation method assumes the use of some training samples consisting of training sets (input and corresponding output signals). For training purposes, training epochs are formed. They include training sets of the given sample, which are randomly mixed for each epoch. The process of training of perceptrons is carried out for several training epochs with a number of iterations for each. The number of iterations corresponds to the number of training sets included in the training epoch. Each iteration includes the following procedures:

1. The training set is selected out of the training epoch, which consists of an array of input signals $a = [a_1, a_2, ... a_i, ... a_m]$ and an array of target values (teacher) $t = [t_1, t_2, ... t_k, ... t_p]$.
2. The values of the output signal $c_{output\ k}$ are calculated for each neuron of the output layer, as well as the local gradient according to the formula: $\sigma_k = e_k \cdot f_a'(c_{input\ k})$, where e_k is the error signal calculated according to the formula: $e_k = t_k - c_{output\ k}$. Along with that, the value of weight adjustment of the synaptic connection is calculated: $\delta y_{jk} = \alpha \cdot \sigma_k \cdot b_{output\ j}$, where α is the training speed of perceptrons.
3. The local gradient is calculated for each neuron of the hidden layer according to the formula: $\sigma_j = f_a'(b_{input\ j}) \cdot \sum_{k=1}^{p} \sigma_k \cdot y_{jk}$, as well as the value of the weight adjustment of the synaptic connection: $\delta x_{ij} = \alpha \cdot \sigma_j \cdot b_{output\ i}$.
4. The weight of each synaptic connection of the perceptron changes due to the addition of a corresponding corrective value: $x_{ij}^{new} = x_{ij}^{old} + \sigma x_{ij}$, $y_{ij}^{new} = y_{ij}^{old} + \sigma y_{ij}$.

In each iteration, within the process of training of the perceptron, the mean square error (*MSE*) of the output signal with respect to the target values (teacher) was calculated. The arithmetic mean of the mean square errors $\overline{MSE}$ was calculated for each training epoch. The criterion for stopping of training was the restriction (Hagan et al., 2014): $\frac{|\overline{MSE}_{ep} - \overline{MSE}_{ep-1}|}{\overline{MSE}_{ep-1}} \cdot 100\% \leq 0.01\%$, where *ep* is the number of the current training epoch.

The speed of training α is chosen by experiment: $\alpha = 0.001$. The initial weights of synaptic connections of neurons were selected randomly from a uniform distribution with a mathematical expectation equal to 0, and the variance, which is determined by the formula (Hagan et al., 2014): $D = \frac{1}{\sqrt{s}}$, where s is the number of synaptic connections of the neuron. Considering the variance formula for the uniform distribution: $D = \frac{(u-v)^2}{12}$, where u, v are the upper and lower limits of the interval respectively, and also taking into account the fact that the mathematical expectation is 0, the interval is as follows: $\left[-\sqrt{\frac{3}{\sqrt{s}}}; +\sqrt{\frac{3}{\sqrt{s}}}\right]$.

Formation of training and holdout samples

Training sets in the form of input and corresponding output signals were formed into training samples separately for the Freeze-up and Break-up phases on the basis of a daily time series of ice coverage of Lake Onega for the period 2004–2016.

Along with training samples, holdout samples were formed separately for the Freeze-up and Break-up phases and included input and corresponding output signals based on a daily time series of ice coverage of Lake Onega for the period 2016–2017.

Within this study, the ice coverage values preceding the predicted value ice_z were taken as input signals, where z is the day number when the considered phase of the ice regime started (for which the prediction is performed) (Fig. 2). Correlation analysis of the daily time series of ice coverage of Lake Onega for the period 2004–2016 showed a strong correlation (pair correlation coefficient $r = 0.7..0.92$) between the values of ice_{z-d} and ice_z, where is the number of days preceding the predicted day ($d = 1 — 9$). It was also found out that the day number when the considered phase of the ice regime started is also a significant factor (degree of correlation is: within the Freeze-up phase $r = 0.32..0.9$, within the Break-up phase $r = —0.55..—0.96$). In addition, it was found out that the duration of time interval g between an origin date and phase start (Fig. 2) has a significant correlation with the duration of the phase: for the Freeze-up phase, r = —0.56 andfor the Break-up phase, r = —0.94. In this study, the origin date corresponds to the earliest for the period 2004-2016, start of the ice regime phase of Lake Onega: for the Freeze-up phase it is November 9th, for the Break-up phase it is February 28th.

For the optimal performance of perceptrons, all input signals should be normalized within the interval [0;1].

The ice coverage values satisfy the normalization condition and the input signal corresponding to the day number when the considered phase of the ice regime started, z was normalized within the study as follows:

$$h_z = \begin{cases} \frac{z}{l}, & z \le l \\ 1, & z > l \end{cases},$$

where l is the maximum value of duration of the considered phase of ice regime for the time period, measured in days (for the Freeze-up and Break-up phases of Lake Onega for the period 2004–2016, it is 80 and 71 days correspondingly).

The input signal corresponding to the duration of the time interval was also normalized as follows:

$$h_g = \begin{cases} \frac{g}{l}, & g \le q \\ 1, & g > q \\ 0, & g < 0 \end{cases},$$

where q is the maximum value of time interval g for the considered phase for the time period (for the Freeze-up and Break-up phases of Lake Onega for the period 2004–2016, it is 56 and 81 days correspondingly).

Thus, the array of input signals is as follows: $a_z = [ice_{z-w}, \ldots, ice_{z-2}, ice_{z-1}, h_z, h_g]$, where ice is ice coverage value; w is the number of ice coverage values preceding the predicted value, and fed forward to the input of an artificial neural network (Fig. 2). In this study, w was assumed to be 1, since according to heuristic recommendations for improving the network performance, input variables should not be correlated (Haykin, 1999). In the case of preceding values, there is a strong correlation between the values ice_{z-d-1} and ice_{z-d} ($r > 0.9$).

The array of output signals contains the predicted value of ice coverage for the corresponding day when phase started and has the following form: $c_z = [ice_z]$.

Selecting parameters and architecture of multilayer perceptrons

The activation function was selected by experiment. The hyperbolic tangent function $f_a = \varphi \cdot tanh(\beta x)$, where $\varphi = 1.7159$, $\beta = 2/3$, demonstrated the best results in perceptrons training.

Within the study, the number of neurons of the input layer was determined by relation: $m = w + 2 = 3$. The number of neurons of the output layer resulted from one predicted value of ice coverage: $p = 1$. The number of neurons of the hidden layer was settled by experiment within several steps. At the first step, a perceptron with the simplest architecture (3-2-1) was trained. At each subsequent step, a perceptron for training was the one with one more hidden neuron than the perceptron in the previous step. At each step, $\overline{MSE}$ was calculated for the last training epoch. If at the current step, the value was less than in the previous step, the next step was performed; otherwise the procedure of selecting the optimal perceptron architecture was completed and the perceptron trained in the previous step was selected.

Results

Figures 3, 4 present the graphs of dependence of $\overline{MSE}$ on the number of training epochs for perceptrons with different numbers of neurons in the hidden layer.

Perceptrons with architectures: 3-2-1 – in the Freeze-up phase ($\overline{MSE}$ = 0.0.155), 3-6-1 – in the Break-up phase ($\overline{MSE}$ = 0.0.105) demonstrated the least $\overline{MSE}$ in the last training epoch (within the training). These perceptrons were selected for testing on holdout samples.

The results of testing of perceptrons with architectures of 3-2-1 and 3-6-1 on the holdout samples in comparison with the actual values are shown in Figs 5, 6. The Figs 5, 6 also demonstrate the results of predictionby means of various statistical methods: linear regression, moving average and autoregressions of the first and second orders.

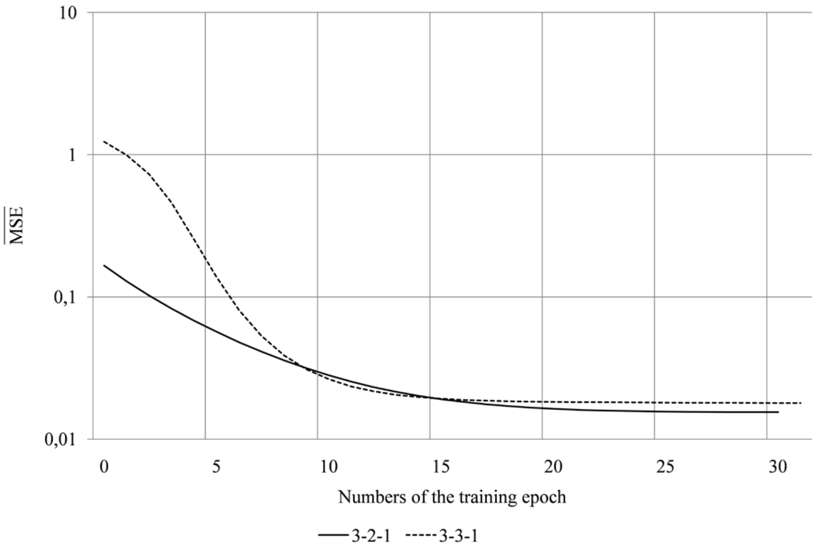

Fig. 3. Graph of dependence of $\overline{MSE}$ on the number of the training epochs when training perceptrons with different number of neurons in the hidden layer within the Freeze-up phase.

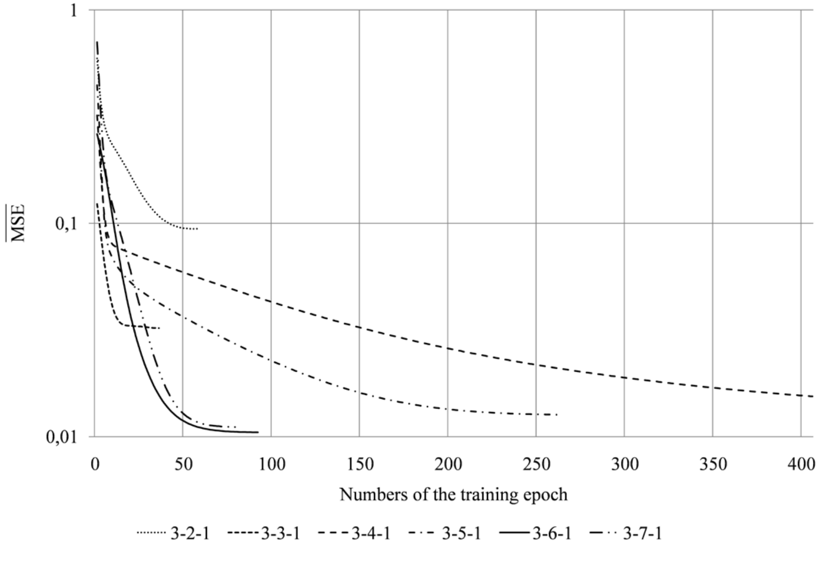

Fig. 4. Graph of dependence of $\overline{MSE}$ on the number of the training epoch when training perceptrons with different number of neurons in the hidden layer within the Break-up phase.

The statistical analysis of test results (the mean-square error of the predicted and actual values *MSPE*, the mean value of absolute deviations of the predicted and actual values *MAD*, the maximum absolute deviation of the predicted and actual values Δ_{max}) of different methods for predicting ice coverage on the basis of holdout samples is presented in Table 1.

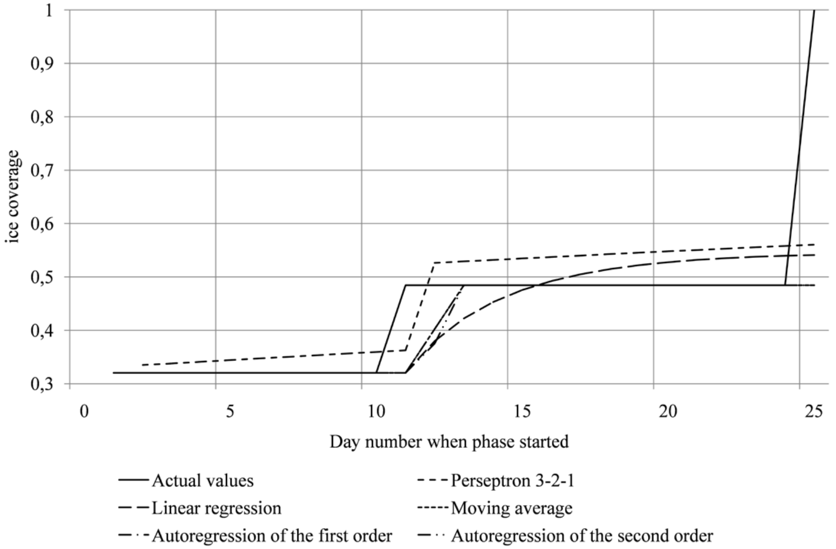

Fig. 5. Predicted and actual ice coverage of Lake Onega within the Freeze-up phase for the period 2015–2016 (holdout samples).

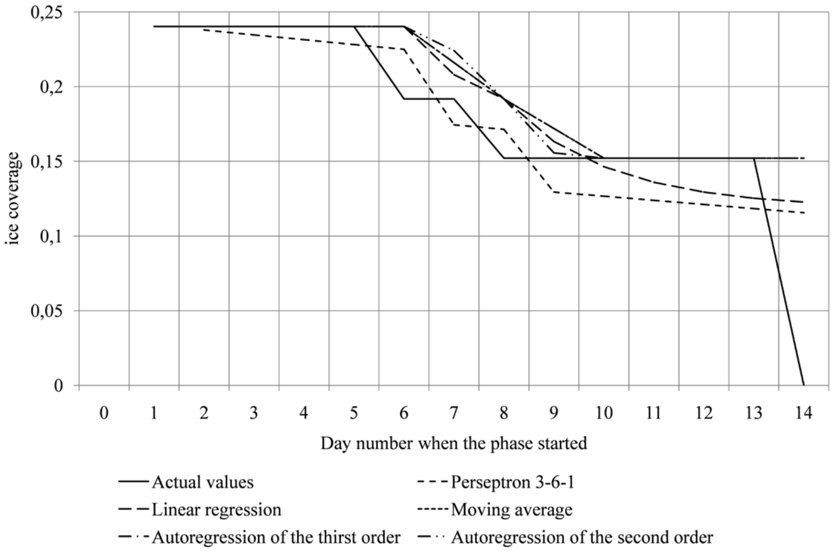

Fig. 6. Predicted and actual ice coverage of Lake Onega within the Break-up phase for the period 2015–2016 (holdout samples).

T a b l e 1. Results of testing of methods for predicting ice coverage on the basis of holdout samples.

Prediction methods	**Freeze-up phase**			**Break-up phase**			**Total *MSPE***
	MSPE	***MAD***	Δ_{max}	***MSPE***	Δ_{max}	***MAD***	
Perceptrons	0.0108	0.065	0.439	0.0015	0.027	0.116	0.0076
Linear regression	0.0117	0.051	0.459	0.0017	0.026	0.123	0.0083
Moving average	0.0130	0.033	0.516	0.0023	0.024	0.152	0.0094
Autoregression of the first order	0.0136	0.035	0.516	0.0025	0.026	0.152	0.0099
Autoregression of the second order	0.0145	0.038	0.516	0.0028	0.028	0.152	0.0107

Discussion

Perceptrons allow preforming more reliable prediction of the ice coverage than statistical methods. This is proved by the smallest value of *MSPE* in holdout samples (Table 1).However, it should be noted that in both holdout samples, the values of *MAD* of statistical methods were lower than values of perceptrons. This results from the fact that at the intervals of linear behaviour of function of the ice coverage of time $f_{ice}(t)$, statistical methods can give a more accurate prediction than perceptrons (absolute deviations predicted by statistical methods from actual values $\Delta \approx 0$). In particular, this is relevant in the case when the area of ice formation remains constant. However, at intervals of nonlinear behaviour of the function $f_{ice}(t)$ within both holdout samples, the statistical methods gave much more errors than perceptrons. This is proved by the maximum values of the absolute deviations of predicted and actual values of ice coverage Δ_{max} (Table 1).

All in all, it could be concluded that multilayer perceptrons allow giving more adequate and reliable prediction of ice coverage of lake than statistical methods, especially when the function $f_{ice}(t)$ behaves nonlinearly, which is a typical situation for lakes in most cases. Furthermore, at the intervals of the nonlinear behaviour of the function $f_{ice}(t)$, statistical methods give unacceptable errors in prediction ($\Delta > 0.3$).

It should also be noted that the training sample for multi-layer perceptrons is formed on the basis of 12 years. This is an insignificant period for revealing all the patterns of ice cover formation; therefore, errors can occur in cases of abnormal situations. Therefore, the coverage changes significantly within a short period of time (> 0.1/*day*); the absolute deviations predicted by perceptrons from the actual values of ice coverage are not always satisfactory as well ($\Delta > 0.2$). Data on the ice cover of lakes with necessary fractionality and spatial resolution for the formation of a training sample can be obtained only using satellite observations. Satellite observations in the necessary mode are conducted not so long ago (2000–2006); therefore, the increase in the volume of training sample is not available. To solve this problem, it is necessary to study additional factors affecting the ice coverage of lakes, such as air and water temperature and wind speed. The inclusion of these factors in the number of input signals of perceptrons greatly extends their functionality.

Conclusion

The parameters and architecture of multilayer perceptrons were selected within the study for predicting ice coverage of Lake Onega during the Freeze-up and Break-up phases. It has been found through experiment that for predicting ice coverage of Lake Onega, the optimal architectures of perceptrons are 3-3-1 within the Freeze-up phase and 3-6-1 within the Break-up phase. These perceptrons showed sufficient results when tested on holdout samples: the mean-squared errors are 0.0108 and 0.0015 within the Freeze-up and Break phases respectively. As a result of testing in accordance with all the holdout samples, it was found that in comparison with the statistical methods, perceptrons have a lower mean-squared prediction error ($MSPE$ = 0.0076), as well as lower maximum absolute deviation (Δ_{max} = 0.439). In this connection, it can be concluded that perceptrons give a more adequate prediction of ice coverage of lakes than statistical methods. However, it is required to improve the quality of perceptron prediction in cases of anomalous (nonlinear) behaviour of the function $f_{ice}(t)$ due to increase in the number of input neurons in the architecture of perceptrons.

Acknowledgements

This work was supported by the Russian Science Foundation, project No. 14-17-00740.

References

Atapaththu, K.S.S., Asaeda, T., Yamamuro, M. & Kamiya H. (2017). Effects of water turbulence on plant, sediment and water quality in reed (*Phragmites australis*) community. *Ekológia (Bratislava)*, 36(1), 1–10. DOI: 10.1515/eko-2017-0001.

Blumberg, A.F. & Mellor G.L. (1986). A description of a three-dimensional coastal ocean circulation model. In N.S. Heaps (Ed.), *Three-dimensional coastal ocean models* (pp. 1–16). Washington, DC: American Geophysical Union.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. & Ljung G.M. (2015). *Time series analysis: Forecasting and control.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Dallimore, C., Hodges, B.R. & Imberger J. (2003). Coupling an underflow model to a three-dimensional hydrodynamic model. *Journal of Hydraulic Engineering,* 129, 748–757. DOI: 10.1061/(ASCE)0733-9429(2003)129:10(748).

Hagan, M.T., Demuth, H.B., Beale, M.H. & De Jesus O. (2014). *Neural network design*. New Castle, DE: International Edition.

Haykin, S.(1999). *Neural network: A comprehensive foundation*. Hamilton, Ontario: Pearson Education, Inc.

Karetnikov, S.G. & Naumenko A.M. (2008). Recent trends in Lake Ladoga ice cover. *Hydrobiologia,* 599, 41–48. DOI: 10.1007/s10750-007-9211-1.

Madec, G. & NEMO team (2015). *NEMO ocean engine*. Paris: Insitute Pierre Simon Laplace.

Menshytkin, V.V., Rukhovets, L.A. & Filatov N.N. (2013). Ecosystem modeling of freshwater lakes (Review): 1. Hydrodynamics of lakes. *Water Resources*, 40, 606–620. DOI: 10.1134/S0097807813060080.

Sharif, A.J., Elias, Z.R. & Omar M.F. (2013). Water flow model for the Harrier basin, Kurdistan of Iraq. *Ekológia (Bratislava),* 32(2), 242–247. DOI: 10.2478/eko-2013-0020.

Quamrul Ahsan, A.K.M. & Blumberg A.F. (1999). Three-dimensional hydrothermal model of Onondaga Lake. *Journal of Hydraulic Engineering,* 125, 912–923. DOI: 10.1061/(ASCE)0733-9429(1999)125:9(912).