LANDSCAPE APPROACH TO NORMALIZED DIFFERENCE VEGETATION INDEX FORECAST BY ARTIFICIAL NEURAL NETWORK: EXAMPLE OF DIYALA RIVER BASIN

A.S. Alhumaima, engineer_alisubhi@yahoo.com,
S.M. Abdullaev, abdullaevsm@susu.ru
South Ural State University, Chelyabinsk, Russian Federation

This study examines the perspective of artificial neural networks for forecast Normalized Differential Vegetation Index (NDVI) on Diyala River basin and also how information about of bioclimatic landscapes will affect to forecasting performance. To do this, in the first stage of the experiment, a total of 20 perceptrons with different one hidden layer architectures were trained with site-specific variables (latitude, longitude, minimal, maximal and mean height, landcover type) and seasonal meteorological variables (precipitation sum, and minimal, maximal and average daily temperatures) by error back propagation algorithm on the data of 2000–2010 years and tested on data for 2011–2016 years. It has been shown that the best performance, with determination coefficient $R^2$ of 0.78, was achieved by perceptron model with 12 hidden neurons the activated by logistic activation function and hyperbolic tangential activation of output value of NDVI. The large spatial heterogeneity of forecasting performance of the best perceptron was detected: in upper part of basin characterized according to Köppen–Trewartha bioclimatic classification, as landscapes of temperate mountain climate and the subtropical climate with dry summers, $R^2$ was 0.76–0.80, whereas in dry steppe landscapes and semi-desert landscapes of Diyala downstream $R^2$ was 0.6–0.7. The second stage of experiments with 20 models of perceptrons where the type of landscape was added as input variable or where 150 individual perceptrons were selected for each landscape, have shown that these approaches allows to $R^2$ increase up to 0.73–0.85. However, the strong contrast between characteristics of individual models complicates their use in the practice and requires to finding of new forecasting approaches.

Keywords: remote sensing, NDVI forecast, perceptron, bioclimatic landscapes, precipitation, temperature, climatic response.

Introduction

Modern Earth sciences are not conceivable without the analysis of multispectral satellite data. The Normalized Difference Vegetation Index (NDVI) and other proxies of primary biological productivity are important products of this analysis [1]. The values of NDVI are highly depended on environmental conditions, so that NDVI is one of principal indicators for evaluating climate impact onto terrestrial ecosystems [2–12]. Particularly, extent and evolution NDVI are often used to estimate climate changes global and vegetation activity [2, 3] – net primary production and vegetation dynamics overlarge arid regions such as Sahel [4, 5], arid regions of Central Asia and Kazakhstan [6, 7], Mongolia and arid area of China [8, 9], Tibetan Plateau [10]. In addition to monitoring arid zones in the works [11, 12] explores the long-period changes in forest-steppe, forest and tundra vegetation of the Russian Federation.

Mapping of NDVI dynamics is one of the main instruments for evaluation and prediction of agricultural productivity [13–18]. First, as in the case of natural biomes, some work explore the impact of climate change on productivity of rain-fed zones [13] and other use NDVI data to model of ecological regimes of rural territory [14]. The changes in vegetation indices of rural areas allows to separate healthy vegetation crops from weak developed fields in irrigated agriculture [15]; to monitor droughts [16, 17] and, with availability of additional surface data, to implement crop forecast [18]. From this point, the capabilities to predict vegetation index under an appropriate spatiotemporal scale [13–18] are critical for decision making to adapt agricultural techniques or to limit socio-economic losses associated with urbanization [19, 20].
In the analysis of remote sensing data are used by all kinds of statistical analysis and machine learning techniques [21]. For example, the linear regressions [22–24] are used in the analysis of NDVI time series; the stepwise cluster analysis is used to NDVI simulation [25] and artificial neural network (ANN) to short-range NDVI forecast [26]. Recently began studies related to the application of the ensemble approach. So in the forecast NDVI model [27] uses deep stacking network, consisting of a stack of multilayer perceptron, each of which models the spatial feature of the associated region at a particular time instant. The study [28] estimates vegetation health on the basis of trained gradient boosted machine models, which combine gradient-based optimization and boosting of base trees models that divide predictor variable into distinct geographic regions.

The main objective of this work is to develop a ANN based prediction model of NDVI, which (i) would take into account changes in rainfall and temperature in the basin of the river Diyala; and (ii) at the same time could be useful in other regions. The relevance of first question is related with unknown reaction natural-human systems the Tigris River basin to climate change [29] and that can be explored by simple multilayer perceptron model. The relevance of second question is related to success of ensembles stacking and boosting [27, 28] where individual model of ensemble member is constructed by specific geographical information. At the same time, simple and ensemble networking approaches possess a general inability to understand cause-and-effect relationships between the input and output of such networks.

The main idea of present study is to associate all possible geographical predictors on input layer of ANN with some form of geographical landscapes. The geographical landscape can be defined as a homogeneous geosystem (in origin and history of development) with a specific uniform of indivisible by zonal factors terrain, geological base, regional climate and hydrothermal conditions, types of soil and ecological communities [30]. The landscape can be introduced as a collection of smaller geosystems or “natural boundaries” – specific morphological elements of the landscape that contains natural communities merged by unique physical and geographical processes and developed on one form of the landscape terrains and homogeneous substrate. The concept of the natural boundaries is quite useful in the ecological assessment studies where it allows introduction of the anthropogenic landscapes [31]. Perceiving this ideological content as a basis, we nonetheless adopt the technical definition of landscape classification [32]: the degree of climate continentality; belonging to morphological structures of the highest order; the splitting of the terrain; bioclimatic differences and geochemical type (the simplification of “geochemical catena” position to the maximum, mean and minimum elevation of smallest landscape unit is used in our study).

On the basis of the above, the objectives of the work consisted of 1) processing the digital terrain data and comparing two climate data sets for classification of bioclimatic landscapes of Diyala; 2) finding a set of best predictors described hydrothermal regime which influence to NDVI value; and finally 3) to determine the best ANN architecture working in all types of landscapes.

The rest of paper is organized onto 3 sections: 1) datasets and methods with brief overview of the study area, evaluation of climate datasets and landscape classification; 2) ANN forecasting results, and 3) short conclusion.

1. Datasets and methods
Four stages of geo-climatic information processing have been performed in our research work in order to get the required results. All the results presented in this paper have been calculated using MATLAB programming language version R2018b, whereas the ArcGIS version 10.5 has been used to simulate the results as geographical maps.

1.1. Land cover and topography of study area
Diyala river basin shapefile used within the ArcGIS has been projected using the maps presented in the inventory of shared water resources in western Asia by the United Nations Economic and Social Commission for Western Asia [33]. The Diyala river basin with a total area of 32,600 km², located approximately between 33.216°–35.833°N and 44.500°–46.833°E and distributed between Iraq (43%) and Iran (57%).

The data from digital elevation model (DEM) are used in this paper both as input variable and to landscape classification. These data were extracted from the Advanced Spaceborne Thermal Emission
Алхумайма А.С., Абдулаев С.М.

Применение искусственных нейронных сетей для прогноза нормализованного вегетационного индекса (NDVI)...

and Reflection Radiometer (ASTER) Global DEM which is distributed from ASTER website (https://asterweb.jpl.nasa.gov/gdem.asp) as Geo-referenced Tagged Image File Format file containing $1^\circ \times 1^\circ$ tiles of earth surface with height and latitude, longitude of 30-meter grid and other additional information [34]. Overlaying boundaries of the basin of Diyala to DEM data (Fig. 1a), it is easy to see that the nature of the relief from the beginnings of Diyala to your mouth varies significantly: large portion of terrain in Iran part of basin is mountainous with peak of 3356 m and becomes abruptly flat on Iraqi part of basin.

![Fig. 1 Study area (a) elevation level according to the ASTER Global DEM; (b) cover types based on the GLC-SHARE classification](image)

Land cover map of basin (Fig. 1b) was obtained from the Global Land Cover-SHARE (GLC-SHARE [35]) database, version 2014, with a spatial resolution of 30" arc-seconds (http://www.fao.org/geonetwork). The GLC-SHARE is a new land cover database created by the Land and Water Division of the Food and Agriculture Organization of the United Nations in partnership with contributions from various institutions by a combination of “best available” land cover database.

According to GLC-SHARE dataset the main land cover categories of the study area are: bare soil ~ 12%, croplands ~ 18%, grasslands ~ 16%, shrub covered areas ~ 26%, tree covered areas ~ 9%, sparse vegetation ~ 14%, and herbaceous cover. Thus, about 34% of Diyala basin occupy by arable fields and pastures, and 35% can be attributed to natural vegetation (shrubs and forest). Therefore, should be expected that the vegetation of these areas will react differently to the thermal regime and the excess or deficiency of the accumulated moisture. All this promises that the basin of Diyala will be difficult place for the prediction of the response of vegetation to changes in weather and climate.

**1.2. NDVI and other data preprocessing**

Terra and Aqua are two earth observation satellites that were launched by NASA in 1999 and 2002, respectively. One of the instruments carried by both the Terra and Aqua satellites is the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS acts a significant role in meeting a very wide range of scientific research objectives like the monitoring of vegetation cover change. In this study, MODIS NDVI dataset (MOD13Q1) [36] of the finest available spatial resolution of 250 m and 16-day composites has been directly obtained from USGS data-center (https://www.usgs.gov). The 34 of 16-day MODIS NDVI composite images centered on March, 7 and March, 23 for period of 17 years (2000–2016) were chosen in order to reduce cloud impacts and to ensure a high chance of having the best quality of pixels representing the NDVI cover.

The preprocessing stage includes: first: extract data that only related to or within the boundaries of our study area, second: from the 2 maps of each March month, construct one map (i.e. 17 monthly maps for our study period) using maximum value composite method to reduce cloud disturbance and increase the overall quality of the dataset [37], and third: remove missing data and NDVI values which are less than 0.1 to reduce unwanted signals coming from potentially non-vegetated pixels (bare soil and ice cover) [38]. For the precipitation and temperatures, the preprocessing stage includes only data extraction.
that related to our study area for the period (1981–2016). The same situation is for the land cover and digital elevation datasets, where the preprocessing stage includes only extracting maps that related to our study area. It is clear that our downloaded data of Diyala river basin are not in the same spatial resolution, therefore we have converted all of them into a $0.05^\circ \times 0.05^\circ$ grid resolution, resulting in up to 1520 grid sample. We combined these datasets and reshaped them into a singular matrix where each row ($1520 \times 17 = 25840$ row) corresponds to a grid sample at one time and each column (11 input-output variables) is a measured factor or variable.

### 1.3. Climate Datasets evaluation

Two climate datasets have been used in this paper, the first one obtained from the University of East Anglia (UEA)/ Climatic Research Unit (CRU) (Version 4.01) [39], which provides monthly total precipitation and monthly mean, minimum, and maximum surface air temperatures for the period 1901–2016. The second dataset used in this paper which are the observations of monthly total precipitation and monthly mean surface air temperatures for the period 1900–2017 are obtained from the University of Delaware (Version 5.01) [40]. The two datasets provide a monthly global gridded data at spatial resolution of $0.5^\circ \times 0.5^\circ$.

For the temporal correlation analysis, the nonparametric Spearman correlation method, which it is characterized by its robustness against the effect of outliers, has been performed in order to investigate the effects of essential environmental variables, namely precipitation and temperatures time series data on the variability of NDVI. This process is very important if we want to obtain the best prediction performance, where it has been used to explore in which accumulated amounts, we must use our meteorological predictors before using them with the ANN models.

To determine the best-accumulated amounts of the precipitation and temperature that have the most influence on the NDVI vegetation index of March growing month within our study area and that could enhance the results of the ANN models, the nonparametric Spearman correlation (SR) analysis has been used. The correlation coefficients SR and their significance levels (p) for CRU and UD datasets are presented in Table 1. Thus, we can see from the table that the SR between March NDVI with CRU precipitation and temperatures are positive and generally have major absolute values comparing to the correlation coefficients of UD dataset.

<table>
<thead>
<tr>
<th>Months</th>
<th>CRU dataset</th>
<th></th>
<th>CRU dataset</th>
<th></th>
<th>UD dataset</th>
<th></th>
<th>UD dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>p</td>
<td>SR</td>
<td>p</td>
<td>SR</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>Mar.</td>
<td>0.47</td>
<td>0.06</td>
<td>0.29</td>
<td>0.26</td>
<td>0.53</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>Feb.–Mar.</td>
<td>0.15</td>
<td>0.57</td>
<td>0.53</td>
<td>0.03</td>
<td><strong>0.66</strong></td>
<td>0.01</td>
<td>0.37</td>
</tr>
<tr>
<td>Jan.–Mar.</td>
<td>0.12</td>
<td>0.65</td>
<td><strong>0.65</strong></td>
<td>0.01</td>
<td><strong>0.66</strong></td>
<td>0.01</td>
<td><strong>0.65</strong></td>
</tr>
<tr>
<td>Dec.–Mar.</td>
<td>0.21</td>
<td>0.42</td>
<td>0.40</td>
<td>0.11</td>
<td>0.51</td>
<td>0.04</td>
<td>0.27</td>
</tr>
<tr>
<td>Nov.–Mar.</td>
<td>0.40</td>
<td>0.11</td>
<td>0.43</td>
<td>0.08</td>
<td>0.59</td>
<td>0.01</td>
<td>0.24</td>
</tr>
<tr>
<td>Oct.–Mar.</td>
<td><strong>0.68</strong></td>
<td>0.01</td>
<td>0.41</td>
<td>0.11</td>
<td>0.60</td>
<td>0.01</td>
<td>0.17</td>
</tr>
</tbody>
</table>

The main results of correlation analysis depicted by bold letters: the vegetation index is moderately correlated (SR = 0.68) with the CRU total precipitation of winter season from October of last year to March and NDVI is correlated (SR = 0.65–0.66) with accumulated CRU temperatures during January to March. Despite of the moderate degree of CRU correlation coefficients, their significance level is 99%, and bearing in mind that this database provides optional minimum and maximum temperatures, we will only use CRU data.

### 1.4. Bioclimatic landscape classification

The bioclimatic classification of Köppen – Trewartha (K–T) [41–43] can be obtained by applying their K–T criteria to at least 30 years of observations of mean monthly precipitation and air temperature. The general groups of bioclimatic classifications established by the K–T are A: tropical humid climates, B: dry climates, C: subtropical climates, D: temperate climates, E: boreal climates, and F: polar climates.
One of the main features of the K–T classification is to delineate dry climates BS (steppes) and BW (deserts) by dryness threshold R. The dryness threshold R (mm) is defined in [42] as $R = 23T - 6.4Pw + 410$, where $T$: mean annual temperature (°C) and $Pw$ is the percentage of annual precipitation occurring in winter.

The general climate groups are further sub-divided based on temperature and precipitation seasonality. Thus, Cs climate is subtropical climate with dry summer (more 70% of precipitation during winter). A third and fourth letter can be added to include information about the warmest and coldest months for every climate class. For example, the temperate continental climates DC with summer temperatures 18–22 °C and winter −9–0 °C, will be described as DCbo, and if an elevation is higher than 1000 m are changed to mountains climates GDCbo. See [41–43] for more details about K–T bio-climate classification rules and descriptions.

In this study, applying classification rules [41–43] to the 30-years-based CRU monthly precipitation and temperature gridded data in Diyala basin we obtain five (5) bioclimatic landscapes types BWil, BSil, BShk, CShk and GDCbo. Fig. 2a–c illustrates the annual hydrothermal regimes of BSil, CShk and GDCbo landscapes and Fig. 2d present spatial distribution of these landscapes. It is evident from Fig. 2a–c that precipitation in steppes, subtropical and temperate climates rainy season lasts from November to April, and May is transition months to dry summer which lasts from June to September in dry and subtropical climates, and with only some convective precipitation in temperate mountains climates.

The end of the rainy season was the main starting point for the selection of the March or April for the characteristics of the vegetation period. If you look at the climate map (Fig. 2d) and temperature charts (Fig. 2a–c), it becomes clear why the month of March was chosen for the characteristics of the vegetation period. Firstly, desert, steppe and subtropical climates occupy a large portion of the basin. For dry climates (Fig. 2a) do not have temperature limits for the growing season, here, is almost always, the temperature is above +10 °C, and conditions for the growth of winter are only limited by soil moisture. In these regions, the irrigated year-round farming is applied. In subtropical landscapes of Diyala (Fig. 2b), the mean temperatures of 3 month, from December to February, are below +10 °C, but March temperatures generally above +10 °C, and taking into account the truth that grass communities start vegetation at +5 °C, the March is months of active vegetation period for the subtropical and dry climate biomes. It is clear from Fig. 2c that the temperature of GDCbo forest landscapes becomes above +10 °C only at April. This means that our study is mainly describe the growing season and consequently the green mass productivity of desert, steppes and subtropical landscapes. On the other hand, we can see

**Fig. 2.** Annual course of temperature and rainfall of steppes (a), subtropical (b) and temperate climates (c) in Diyala basin (1981–2016). The Köppen – Trewartha classification of climates is presented on Fig. 2d
2. Application of ANN to NDVI forecast

In the Earth sciences, the popularity of ANNs has grown mainly to solve prediction problems with nonlinear, stochastic nature, or unknown variations of variables. [21, 26, 27, 44]. Although we know a priori [45], that lack of winter rains causes adverse arid conditions in Tiger basin, i.e. reducing the NDVI, we nevertheless try to prove that besides the meteorological parameters and land cover category, an important role in the response vegetation plays a landscape type itself. On the other hand, there is some reason to assume that the response of the vegetation of different landscapes on precipitation and temperature will be non-linear. Therefore, as a starting ANN model we have selected perceptron with one hidden layer. To answer if the landscape types are relevant to explanation of NDVI distribution, in section 2.1 we trained various ANN models without information about landscape and find model having best performance for entire basin and then, in section 2.2, we observe visual results and rated the quality of the forecast of this model for each of the landscapes. Using the results of last procedure as base to comparison we construct new ANN models by using landscape information.

2.1. ANN models without landscape information

To design of the model without landscape information, initially, the ten predictors are used as input neurons: 1) 5 neurons represented the geographical site-specific characteristics: latitude, longitude and minimal, maximal and mean altitude; 2) 4 neurons represent of the meteorological properties derived from CRU data (winter precipitation, and minimal, maximal and mean temperature of January to March); 3) 1 neuron correspond to land cover types. The value NDVI of corresponding March months is considered as output neuron (forecasting value). The period from 2000 to 2010 with 16720 observations or \( \approx \) 65% of total number of 25 840 input vectors was used to form the training dataset. The period from 2011 to 2016 with 9120 observations or \( \approx \) 35% of total input vectors was used for the testing dataset.

One of the challenges of this work is that choosing of one of four group approaches [46]: from the most primitive trial and error search to the heuristic method using knowledge gained from previous experiments where a near-optimal ANN topology achieved and from exhaustive search through all possible topologies to pruning and constructive algorithms devising an efficient network structure by incrementally adding or removing links. By virtue of the practical prediction NDVI, we have chosen heuristic approach using some preliminary estimates. The review of [47] present dozens equations to calculate number of hidden nodes depending to (i) the number of input and output nodes, or their combinations with (ii) number of samples in training data. We spent 25 calculations on various equations given in [46, 47] and got a different number of nodes from 4 up to 1000, with the highest frequency of the number of nodes between 4 and 9. For these reasons we decided to vary the number of hidden nodes from 4 to 8, 12, 16 and 20 nodes and tracking the model performance during training and testing procedure by using three common metrics: Root-Mean Square Error (RMSE); Mean Absolute Error (MAE); and coefficient of determination (\( R^2 \)). We also understand that the optimal structure ANN will depend also on the functions of neuronal activation and training algorithm. Therefore, all of 20 different models presented in Table 2 were well-trained with the same Levenberg – Marquardt back propagation algorithm. The difference between models with the same number of hidden nodes was that these hidden nodes were activated by one of two functions, logistic or hyperbolic tangent (tanh) function, and signals from hidden layer were transmitted to the output layer by linear or tangential function. Recall that the coefficient \( R^2 \) is the proportion of variance of the dependent variable NDVI which can be explained by variance of 10 predictors, so that the best model is choosing from 20 models as the model with the biggest value of \( R^2 \) obtained after testing procedure, and observing also if this value converges with value of \( R^2 \) obtaining after training procedures.

As shown in Table 2, the best ANNs are depended both activation function and number of hidden nodes. Thus, model #18 with logistic sigmoid for the hidden layer and hyperbolic tangent functions for output and 12 hidden nodes exhibits the highest \( R^2 \) of 0.776 during testing process with little difference with training value of \( R^2 \) of 0.814.
Training and testing performance of ANN models with 10 input variables

<table>
<thead>
<tr>
<th>Model #</th>
<th>Transfer function</th>
<th>Hidden nodes</th>
<th>Training performance</th>
<th>Testing performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>1</td>
<td>Tanh Linear</td>
<td>4</td>
<td>0.082</td>
<td>0.063</td>
</tr>
<tr>
<td>2</td>
<td>Tanh Linear</td>
<td>8</td>
<td>0.067</td>
<td>0.052</td>
</tr>
<tr>
<td>3</td>
<td>Tanh Linear</td>
<td>12</td>
<td>0.062</td>
<td>0.047</td>
</tr>
<tr>
<td>4</td>
<td>Tanh Linear</td>
<td>16</td>
<td>0.058</td>
<td>0.045</td>
</tr>
<tr>
<td>5</td>
<td>Tanh Linear</td>
<td>20</td>
<td>0.053</td>
<td>0.041</td>
</tr>
<tr>
<td>6</td>
<td>Logistic Linear</td>
<td>4</td>
<td>0.078</td>
<td>0.059</td>
</tr>
<tr>
<td>7</td>
<td>Logistic Linear</td>
<td>8</td>
<td>0.068</td>
<td>0.052</td>
</tr>
<tr>
<td>8</td>
<td>Logistic Linear</td>
<td>12</td>
<td>0.063</td>
<td>0.048</td>
</tr>
<tr>
<td>9</td>
<td>Logistic Linear</td>
<td>16</td>
<td>0.058</td>
<td>0.045</td>
</tr>
<tr>
<td>10</td>
<td>Logistic Linear</td>
<td>20</td>
<td>0.054</td>
<td>0.041</td>
</tr>
<tr>
<td>11</td>
<td>Tanh Tanh</td>
<td>4</td>
<td>0.078</td>
<td>0.060</td>
</tr>
<tr>
<td>12</td>
<td>Tanh Tanh</td>
<td>8</td>
<td>0.067</td>
<td>0.052</td>
</tr>
<tr>
<td>13</td>
<td>Tanh Tanh</td>
<td>12</td>
<td>0.060</td>
<td>0.046</td>
</tr>
<tr>
<td>14</td>
<td>Tanh Tanh</td>
<td>16</td>
<td>0.057</td>
<td>0.043</td>
</tr>
<tr>
<td>15</td>
<td>Tanh Tanh</td>
<td>20</td>
<td>0.053</td>
<td>0.040</td>
</tr>
<tr>
<td>16</td>
<td>Logistic Tanh</td>
<td>4</td>
<td>0.062</td>
<td>0.047</td>
</tr>
<tr>
<td>17</td>
<td>Logistic Tanh</td>
<td>8</td>
<td>0.066</td>
<td>0.050</td>
</tr>
<tr>
<td>18</td>
<td>Logistic Tanh</td>
<td>12</td>
<td>0.077</td>
<td>0.059</td>
</tr>
<tr>
<td>19</td>
<td>Logistic Tanh</td>
<td>16</td>
<td>0.055</td>
<td>0.042</td>
</tr>
<tr>
<td>20</td>
<td>Logistic Tanh</td>
<td>20</td>
<td>0.052</td>
<td>0.040</td>
</tr>
</tbody>
</table>

From the Table 2 it is also obvious that model #1, 6 and 12 with other functions and a smaller number of hidden nodes give similar results with $R^2 = 0.74–0.77$. Thus, in these cases, the equivalent coefficient of linear correlation would be equal to 0.85. Note also that difference in performances of the best and worst model is 18%, but if we reject these two cases, the difference of $R^2$ would be only 11%.

2.2. ANN with landscape information

Visual comparison of model #18 forecast with actual NDVI maps (Fig. 3) shows that the quality of forecast varies from year to year, and also varies in space.

For example, in the year 2011, while maintaining the overall structure of the active vegetation NDVI > 0.3, the model visibly minimized NDVI values in BWil and BSil landscapes, was close to observed values in BShk and subtropical and temperate landscapes. Comparing the quality of predictions in other years, we can see that in most cases, the model generally underestimates the value of NDVI dry landscapes, with one exception of BShk landscape which almost always stands out a strip of mountain vegetation.

Testing performance of model #18 estimated separately by landscapes (Table 3) depict that performance of the desert and steppe landscapes had markedly lower values of $R^2$ than the average value for the entire model, but model #18 good predicted NDVI temperate landscapes.

Given the differences of predictability on landscape level we, ceteris paribus conditions of models #1–20, added landscape types to previous ten predictors and further trained models #21–40 around the entire basin. The best forecast result of #21–40 with shared $R^2 = 0.789$ present model #28 with 12 hidden neurons and logistic to linear transfer function. Comparing model performance in the landscapes (Table 3), we observe that new best perceptron has slightly lower capability in temperate and subtropical landscapes but is advanced in very hot desert and steppe landscapes where $R^2$ rise to 0.66–0.76 or to 11–14% more relative to model #18. Despite the importance of such an increase, you must say that could not predict which of the models in Table 2 when you add a new input variable will give the best results. For example, the model #8, corresponding to model #28, had demonstrated previously very low potential (Table 3).
On the other hand, comparison of all 20 models without information about landscapes and 20 models with the addition of landscape types showed average growth of determination coefficient of 5%. Note that $R^2$ rise was observed for 19 models, but for five model this growth was less than 2%.

Replying to a question whether it is possible to improve the quality of prediction if develop individual models for each of the landscapes, we initially thought that the subtraction of one or more predictors from input layer will reduce the coefficient of determination affecting negatively to testing performances. But as previous experience with adding data showed that it could be the opposite we began to study the impact of removal of the input data. On the other hand, it is obvious that when you rebuild the synapses in new models, past ways of neuron activation need not necessarily be optimal. Therefore, we fix the number of hidden nodes as 12 (as in best model #18, and #28) and then by varying the activation function train and test of 100 ANN models where one input variable (precipitation, minimal temperature $T_{min}$, maximal $T_{max}$, mean temperature) was removed from input; and 25 models with absence of $T_{min}$ and $T_{max}$. Totally we create 150 individual ANN models (including 25 models using all ten input variables) or 30 models per landscape and then choose from these 30 models the models which present the best performance in prediction of NDVI of certain landscape (Table 3).

Looking at Table 3, we see that only one best model created for desert landscape need all meteorological data. For other landscapes, the best models require only one or two temperatures. In general, two models for steppe (BShk) and subtropical landscape (Cshk) inherit the logistic and hyperbolic tangent

### Table 3

<table>
<thead>
<tr>
<th>Type</th>
<th>Testing performance, $R^2 \times 10^3$</th>
<th>Best individual model description</th>
<th>Input variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#18</td>
<td>#21–40</td>
<td>#41–190</td>
</tr>
<tr>
<td>BWil</td>
<td>618</td>
<td>656</td>
<td>748</td>
</tr>
<tr>
<td>BSil</td>
<td>663</td>
<td>760</td>
<td>783</td>
</tr>
<tr>
<td>BShk</td>
<td>717</td>
<td>739</td>
<td>859</td>
</tr>
<tr>
<td>CShk</td>
<td>757</td>
<td>712</td>
<td>728</td>
</tr>
<tr>
<td>GDCbo</td>
<td>801</td>
<td>800</td>
<td>838</td>
</tr>
</tbody>
</table>

Fig. 3. Vegetation index maps predicted by model #18 against their actual maps for 2011–2016 years.
Landscapes are highlighted by thin line.
activation functions of the model #18, and model for desert landscape was the heiress of model #28, the remaining two best individual models reproduced the other probable combinations of hidden and output transfer functions. Comparing the quality of individual models with performance of ANN #28, it can be seen that the improvements in 14–16% concern only models of two dry very hot landscapes, but the performance of models developed for other landscapes rises only to 2–3%.

Conclusion
This work investigated the possibility of ANN trained on conventional geographic and specially adapted meteorological data to predict of NDVI values over Diyala river basin with spatial resolution of ~ 5 km. The principal feature of our work was that we initially assumed that delineation of large basin by geographical landscapes and using landscape type as additional input variable or as spatial boundary for individual models construction can improve the performance of ANN forecasting. The analysis of climate data reveals that Diyala intersects four main Köppen – Trewartha climates from temperate and subtropical climates to steppe with two subtypes and desert climates. For the purity of the experiment we construct ANN models representing perceptrons with one hidden layer and different number of hidden nodes and four combination of layer activation function, and after that all 20 ANN were trained by errors backpropagation algorithm over dataset without landscape data. This was done in order to find the best model configuration which has a biggest coefficient of determination $R^2$ during testing procedure. Despite the moderate level of $R^2 \approx 0.78$ of the best model, it turned out that the NDVI value of desert and steppe landscapes was predicted by this model noticeably worse with $R^2 = 0.59–0.66$. We found that adding the type of landscape to input layer of the previous models improves the mean performance for 5%, especially to desert and steppe landscapes, where the proportion of explained variance grew by 11–14%. However, the best performances with $R^2 = 0.73–0.86$ were obtained when models were training and testing individually within the boundaries of one of five landscapes types. Analysis of performance of 150 ANN individual models, reveal that the optimal configuration of model developed to certain landscape can completely be different from the configuration of optimal models of other landscapes. This fact and the evidence of the climate and landscape changing complicate the prospects of using neural networks to forecast the nature of vegetation. On the other hand the analysis of application of neural network to forecast of complex environmental systems allows us to create some preliminary requirements when new types of forecasting will appear.

References
8. Tulokhonov A.K., Tsydypov B.Z., Voloshin A.L., Batueva D.Zh., Chimedorrj Ts. Spatio-temporal Characteristics of Vegetation Cover of Arid and Semiarid Climatic Zones in Mongolia on


33. UN-ESCW and BGR (United Nations Economic and Social Commission for Western Asia; Bundesanstalt und Rohstoffe). *Inventory of Shared Water Resources in Western Asia*. Beirut, 2013.


ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗА НОРМАЛИЗОВАННОГО ВЕГЕТАЦИОННОГО ИНДЕКСА (NDVI) БИОКЛИМАТИЧЕСКИХ ЛАНДШАФТОВ БАССЕЙНА РЕКИ ДИЯЛА

А.С. Алхумайма, С.М. Абдуллаев
Южно-Уральский государственный университет, г. Челябинск, Россия

Данное исследование касается перспектив использования искусственных нейронных сетей для прогнозирования распределений Normalized Differential Vegetation Index (NDVI) в бассейне реки Дияла и главным образом того, каким образом информация о типах биоклиматических ландшафтов повлияет на прогнозируемость NDVI. Для этого в первом этапе эксперимента на вход персептронов с одним скрытым слоем и различными функциями активации подавались только общегеографические характеристики одного из 25 000 участков бассейна размером 0,05° × 0,05° (широта и долгота, минимальная, средняя и максимальная высота над уровнем моря, тип земного покрова) и сезонные метеорологические факторы (сумма осадков и средние температуры, минимальные и максимальные температуры) и прогнозировалось значение NDVI в начале вегетационного периода. Все 20 персептронов с 4–20 скрытыми узлами обучались на данных 2000–2010 гг. с помощью алгоритма обратного распространения ошибки и тестировались на данных за 2011–2016 гг. Было показано, что лучшее соответствие между прогнозируемым и фактическими NDVI с коэффициентом детерминации (КД), равным 0,78, достигается персепtronом с логистической функцией активации 12 скрытых нейронов и гиперболической тангенциальной активацией выходного нейрона. При этом обнаружена пространственная неоднородность качества прогноза: в верховьях реки, характеризуемых согласно Кеппену – Треварта, как ландшафты умеренного горного климата и субтропического климата с сухим летом, КД = 0,76–0,80, тогда как в сухих степных и полупустынных ландшафтах низовий реки КД = 0,59–0,66. Эксперименты с 20 моделями с добавлением типа ландшафтов на вход персептронов показали возможное улучшение КД на 5 %, а индивидуальный подбор модели персептрона для каждого ландшафта (всего 150 моделей) позволил увеличить КД до 0,73–0,85. Тем не менее сильное отличие характеристик индивидуальных моделей осложняет перспективы их использования в практических целях и требует поиска новых подходов.

Ключевые слова: дистанционное зондирование, прогноз NDVI, персептрон, биоклиматические ландшафты, гидротермический режим, вегетационный период.

Литература


18. Прогнозирование урожайности зерновых и зернобобовых культур в Центральных Черноземных областях на основе комплексирования наземных и спутниковых данных / А.И. Страшная,
Информатика и вычислительная техника


22. Exploring spatially variable relationships between NDVI and climatic factors in a transition zone using geographically weighted regression / Zhiqiang Zhao, Jianbo Gao, Yanglin Wang et al. // Theoretical and Applied Climatology. – 2015. – Vol. 120, no. 3–4. – P. 507–519. DOI: 10.1007/s00704-014-1188-x


33. UN-ESCWA and BGR (United Nations Economic and Social Commission for Western Asia; Bundesanstalt für Geowissenschaften und Rohstoffe), Inventory of Shared Water Resources in Western Asia. Beirut, 2013.


Алхумайма Али Субхи, аспирант кафедры системного программирования, Южно-Уральский государственный университет, г. Челябинск; engineer_alisubhi@yahoo.com.

Абдуллаев Санжар Муталович, д-р геогр. наук, профессор кафедры системного программирования, Южно-Уральский государственный университет, г. Челябинск; abdullaevsm@susu.ru.

Поступила в редакцию 9 апреля 2019 г.