UDC 528.852+845 : 519.237.8 : 004.93

Implementation of linear regression and neural network models for local climate change assessment and forecasting based on time series data

Ya.I. Zyelyk[1]*, **L.V. Pidgorodetska**[1], **S.V. Chornyy**[1], **L.M. Kolos**[1], **Yu.R. Dykach**[2]

[1]Space Research Institute NASU & SSAU, Hlushkova Avenue 40, building 4/1, 03187 Kyiv, Ukraine
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Avenue 37, 03056 Kyiv, Ukraine

Climate change studies are based on the data processing of the following time series for number of Ukrainian cities: daily precipitation and air temperature from the site of the European Climate Assessment & Dataset (ECA&D) project and monthly carbon dioxide emissions from fossil fuels in the atmosphere from the area of one geographical degree size from the site of the Carbon Dioxide Information Analysis Center (CDIAC). Assessment of synchronous dynamics and forecasting of the air temperature, precipitation according to the period 1950–2016 and the carbon dioxide concentration in atmosphere according to the period 1951–2013 was performed using two types of models: linear regression and neural network model in the form of perceptron with one hidden neuron layer. These models implementation is based on modern approaches to the Big Data intellectual analysis: Data Mining and Knowledge Discovery in Databases. The program scenarios for processing, intelligent analysis and forecasting of the above mentioned data time series using the constructed forecasting models have been developed and implemented in the unified analytical platform Deductor. The monthly time series of the emission in atmosphere of carbon dioxide from fossil fuels to area of 1 geographical degree in 1951–2013 for number of Ukrainian cities shows clear trend, which is well approximated by the cubic polynomial and does not contain a periodic seasonal component. There is the tendency to increasing CO_2 emissions before 1991 and the declining tendency after 1991 with some "plateau" in the period 2000–2009 that are years of relative stability in the economic development of Ukraine. However, there is no correlation between the trend of CO_2 emissions and the trend of average monthly temperature in 1951–2013. The predictive model of linear regression was the most acceptable for time series of average monthly temperature in 1950–2016, which are characterized by strong seasonal component with the periodicity of 1 year. Using the predictive model of linear regression $[240 \times 1]$, trained on the dataset, constructed by the sliding window method with retrospective depth of 240 months, the consistent forecast of the average monthly temperature time series for Kyiv after 12.01.2016 with the forecast horizon of 60 months was derived. However, the application of the predictive model of linear regression is characterized by a relatively short horizon of consistent forecasting. For the time series of the average monthly temperature for Kyiv city after 12.01.2016, the predictive neural network model $[360 \times 5 \times 1]$ (number of inputs is equal to the selected prehistory depth of the time series samples that is 360; number of outputs is 1; number of neurons in the hidden layer is 5; activation function type is sigmoid with the given slope 1) provided the consistent forecast with the horizon of 120 months. Low values of the maximum and average relative errors of the neural network model were achieved on the training set ($4,45 \cdot 10^{-2}$ and $2,99 \cdot 10^{-4}$, respectively) and on the test set ($3,60 \cdot 10^{-2}$ and $5,69 \cdot 10^{-3}$, respectively). Similarly, for the time series of monthly CO_2 emissions for the Kyiv city after December 1, 2013, the predictive neural network model $[240 \times 5 \times 1]$ provided consistent forecast with the horizon of 60 months. In general, the time series of monthly CO_2 emission values are characterized by much smaller values of the consistent forecast horizon in comparison with the time series of the average monthly temperature, at least when using predictive neural network models.

Keywords: climate change; carbon dioxide emission; time series prediction; Data Mining; linear regression; neural network model; single-layer perceptron.

1. INTRODUCTION

One of the most important challenges in the 21st century is the negative effects of global and local climate change minimization. This problem is of particular importance for the sustainable society development and is closely related to the need of assessment and forecasting of the key climate indicators dynamics.

This study focuses on air temperature and precipitation forecasting depending from the carbon dioxide concentration dynamics in the atmosphere for number of Ukrainian cities. These three factors are essential climate variables. An essential climate variable (ECV) is a physical, chemical or biological variable or group of related variables that critically affects the Earth's climate. ECV datasets provide the empirical evidence needed to understand and predict climate evolution, to plan for climate change mitigation and risk assessment, and to identify the causes and effects of climate change relationship.

Two model types for assessment and forecasting of local climate change based on time series data have been studied in this article: linear regression model and neural network model in the form of perceptron with one hidden neuron layer using modern approaches to intellectual analysis of Big Data: Data Mining (DM)

* Yarema Zyelyk; ✉ adapt09@gmail.com

and Knowledge Discovery in Databases (KDD) [1–5]. Data Mining is the detection of previously unknown, non-trivial, practically useful and of the accessible for the interpretation patterns in the source data. Knowledge Discovery in Databases is the analytical process of studying of the large information volumes to identify hidden structures, relationships, and interconnections. KDD is a synthetic information technology based on the combination of models and methods of artificial intelligence, numerical methods, statistics and heuristic approaches. The KDD process involves the following stages: forming of input data sets, i.e. samples; preliminary processing and clearing of data, i.e. detection and processing of gaps, overshoots, extreme and abnormal values; data conversion and reconciliation; Data Mining, i.e. construction of descriptive and predictive models and implementation of appropriate methods based on regression, classification, etc.; data post-processing, i.e. testing of models and methods and interpretation of the results obtained concerning to certain subject area. The indicated approaches and stages of the intellectual data analysis performing regarding the assessment and forecasting of local climate change based on the time data series are implemented in the unificated program analytical Deductor platform [3–5].

Climate change studies are based on the processing of daily precipitation and air temperature data from meteorological stations in the number of cities in Ukraine derived from the European Climate Assessment and Data Analysis (ECA & D) project [6]. This resource contains information on changes in weather and climatic conditions and contains the daily dataset which is necessary for monitoring and climate analysis, including the extreme situations. In addition, was processed the time series of the monthly emission of carbon dioxide from fossil fuels in the atmosphere to the area of the one geographic degree for number of Ukrainian cities, obtained from the electronic resource of the Carbon Dioxide Information Analysis Center (CDIAC) [7]. The above mentioned time series data cover the common time interval from 1951 to 2013.

2. THE TIME SERIES DATA PROCESSING TECHNOLOGY AND LINEAR REGRESSION AND NEURAL NETWORK PREDICTIVE MODELS CONSTRUCTING

The program scenarios for processing, intellectual analysis and forecasting of the specified time data series using the constructed prediction models has been developed and implemented in the environment of the unificated analytical Deductor platform. The program scenarios nodes implement the basic procedures described below using the special masters for data processing.

Import data in the text format with the delimiters and conversion of calendar dates to month and year data to aggregate daily data to total monthly precipitation and average monthly temperature values. Visualization the results of data conversion and data aggregation using OLAP (Online Analytical Processing) cube construction [2–5] of data, displaying the necessary OLAP cube data slices and constructing of corresponding diagrams for aggregated data. The cross tables constructing for further analysis and forecasting of the selected time series type of aggregated climatic variables for further intellectual analysis and forecasting. The data quality evaluating with the allocation of gaps, overshoots, extreme and abnormal values, the choice and implementation of ways to eliminate these anomalies and forming of the aggregated climatic variables purified time series. Autocorrelation analysis of the transformed aggregated data time series in order to identify of seasonality. Decomposition of the aggregated data time series to the trend with the choice of its approximating function, to regular periodic component, random remainder and the decomposition results visualization. Forming by the sliding window method the input dataset with the specified time series samples prehistory depth for the time series prediction models constructing.

Implementation of the linear regression predictive model [2–5] with the number of inputs equal to the experimentally selected time series samples prehistory depth, and one output equals to the time series current sample value. The indicators of the adequacy and quality of the model are determined: the determination coefficient, Fisher's criterion, the residues dispersion, the Student's criterion for the regression coefficients statistical significance. The scatter diagram is constructed with displaying on the plane of the model etalon output, the actual model output and the limits of the permissible training error. The input dataset table for the prediction model of linear regression is also displayed, supplemented by field of the actual model output on the training set and by field of model error. The mean square error value is calculated by the linear regression model approximation of the current time series samples on the training data set. The linear regression predictive model is corrected by the time series samples prehistory depth value changing if necessary.

The predictive linear regression model identification is that to find the model coefficients estimates from the condition of the mean square model error minimizing on the training dataset.

The predictive neural network model in the form of perceptron with one hidden neuron layer [3–5, 8, 9]. The inputs number is equal to the experimentally selected time series samples prehistory depth; the outputs number is 1 that is current time series sample value; number of hidden neuron layers is 1; experimentally is selected number of neurons in the hidden layer; the type of the activation function in our case is the sigmoid with the specified slope value 1. The algorithm for training the neural network is chosen, in our case it is Resilient Propagation (RPROP) [4, 5] according to which training is conducted in off-line mode that is

correction of the connections weight of all neurons is carried out after presentation of all examples of the training set and taking into account the gradient sign for each weight. The number of training epochs is specified that is the amount of training cycles on the training dataset and tests cycles on the test dataset; these sets are in the given percentage relation to the number of elements, but in each epoch these sets are randomly formed from the input data set of the neural network model. The acceptable error is also set for the neural network to recognize the training or test example that is a pair of data "vector input — scalar output". It is also an acceptable error to recognize a training or test example — a pair of data "vector input — scalar output" of the neural network. With the implementation of the above RPROP algorithm and additional set parameters, the neural network training process is started. The training results of the prediction neural network model are displayed: the graph of the neural network, the scatter diagram with the displaying on the plane of the etalon model output, the actual model output and the limits of the permissible training error. The input data set for the neural network training is also displayed, supplemented by field of values of the actual model output on the training set and by field of model error values. The average value of the approximation error by the neural network model of current samples on the test data set is calculated on the base of the specified depth of the time series samples prehistory.

The identification of the neural network model with the defined structure is that to find estimates of the weight coefficients of the neurons links in the hidden layer with the model inputs by implementing of the training algorithm of the neural network, in our case of Resilient Propagation algorithm, on the training data set constructed by the sliding window method with the experimentally determined depth of the time series samples prehistory.

For the neural network predictive model it is impossible to evaluate its adequacy and quality, as in the case of the linear regression model, based only on the input data set, which is completely used for the linear regression model training. In the case of the neural network prediction model, a ten-fold cross-validation procedure is implemented to assess of adequacy and quality. In order to this procedure implementation, the input data set for training (which is constructed on the base of the sliding window method) multiplies by 10 times by combining 10 data blocks, each of which enters the data from the input data set of the model, mixed randomly. In this way, ten blocks validation data set is generated. To carry out such validation procedure, the same sequence of training and testing procedures is performed, but for all 10 blocks. The obtained 10 values of the average error of the model are averaged over the number of blocks and the result is considered as average validated model error. To predict the time series values, the optimal neural network model is selected with parameters that are estimated from the data of the test subset with the block number, which achieves the minimum validated average error of the model.

The time series forecasting of the aggregated climatic variables at the given forecast horizon on the base of constructed trained and validated predictive linear regression model and neural network model in the form of perceptron with one hidden neurons layer with the given depth of the time series samples prehistory. The forecasting procedure is implemented in the following way. Let the resulting time series data, for example, by the sliding window method, consist the sequence of time samples: $x(-n), ..., x(-2), x(-1), x$, where x is the current sample value. The forecast in the next step $x(+1)$ is based on the constructed sequence using the predictive model. To construct the forecast for the value of $x(+2)$, it is necessary to shift the whole sequence of time samples to one sample to the right, so that the previously made forecast $x(+1)$ also became part of the input data of the prediction model. Then the calculation of the predicted value of $x(+2)$ is implemented using the predictive model and so on in accordance with the specified forecast horizon.

3. THE RESEARCH RESULTS ANALYSIS

Some of the research results were reported [11, 12]. In Fig. 1 and 2 the decomposition diagrams of the time series of total monthly precipitation and the average monthly temperature values in 1950–2016 for Kyiv city are represented on which the initial time series, cubic trend and seasonal component are shown. Both time series have significant seasonal components with the period of 1 year (12 months). Time series of total monthly precipitation in Fig. 1 contains the small cubic trend in the small dynamic range of 460–560 mm in 66 years, and the time series of average monthly temperature contains the slight cubic rising trend in the small dynamic range of 7.4–10.1 degrees, that is, the growth of the temperature trend is 2.7 degrees per 66 years.

Fig. 3 represents the decomposition diagram of the time series of the monthly emission of carbon dioxide from fossil fuels in the atmosphere to the area of the one geographic degree in 1951–2013 for Kyiv city with the show of the initial time series, the clearly expressed cubic trend with an analytical form of its function and the zero seasonal component. There is clearly non-periodic tendency in this time series, which is confirmed by zero seasonal component. There is the tendency to increase CO_2 emissions before 1991 and the tendency of their decline after 1991 with some "plateau" for the period 2000–2009 that are the years of relative stability in economic development in Ukraine. All these tendencies are illustrated by the corresponding change in the cubic trend of the time series of CO_2 emission. From the comparison of the time series trends, shown in
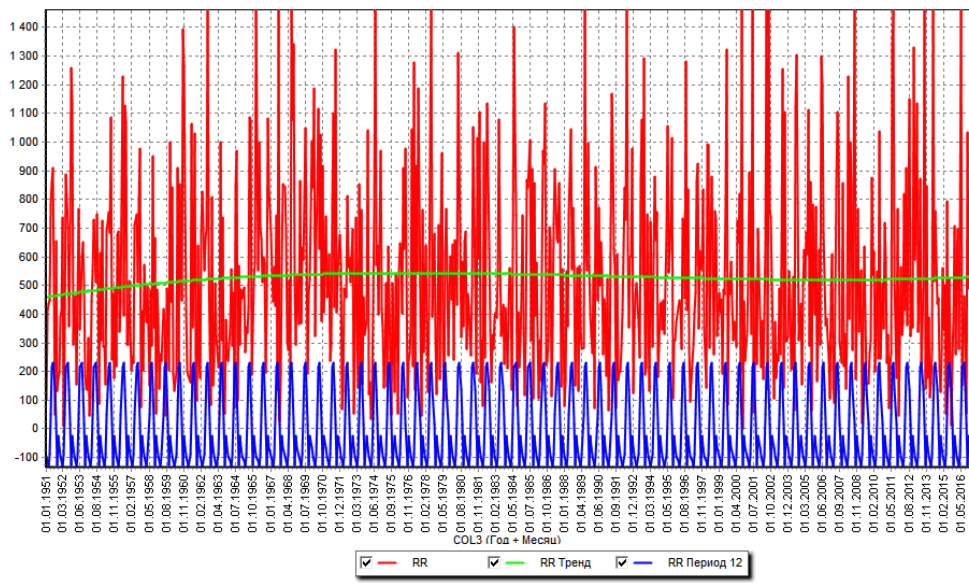
Fig. 1. Decomposition diagram of the time series of the total monthly precipitation values in 1950–2016 for Kyiv city with the displaying of the initial time series, cubic trend and seasonal component
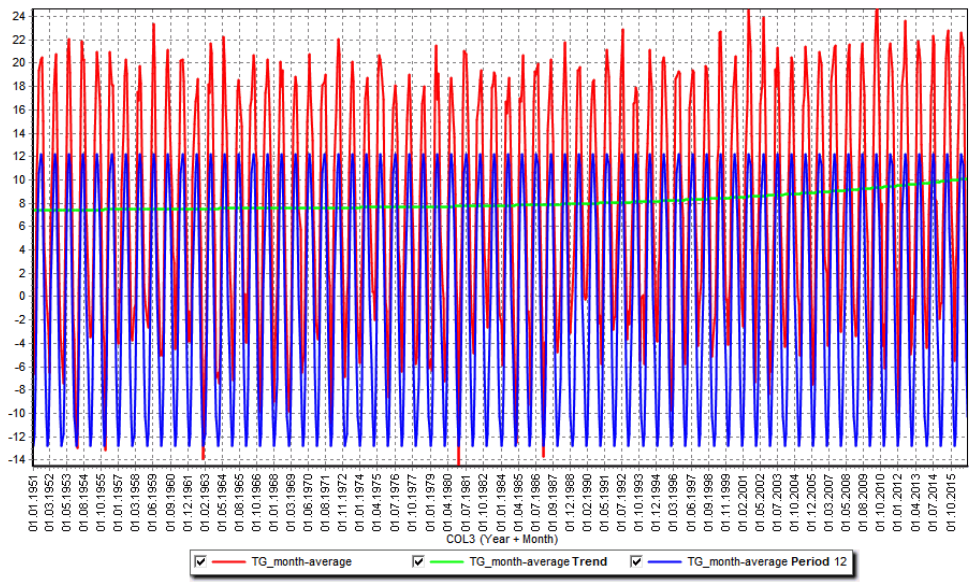

Fig. 2. Decomposition diagram of the time series of the average monthly temperature in 1950–2016 for Kyiv city with the displaying of the initial time series, cubic trend and seasonal component
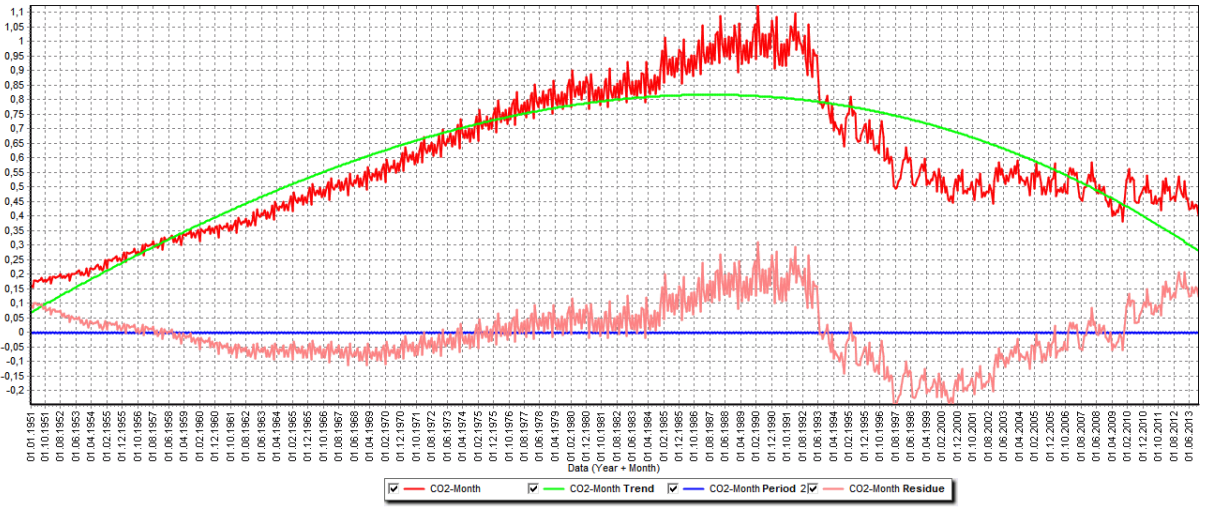

Fig. 3. Decomposition diagram of the time series of the monthly CO_2 emission, in 1951–2013 for Kyiv city with the displaying of the initial time series, cubic trend with the analytical form of its function and the zero seasonal component
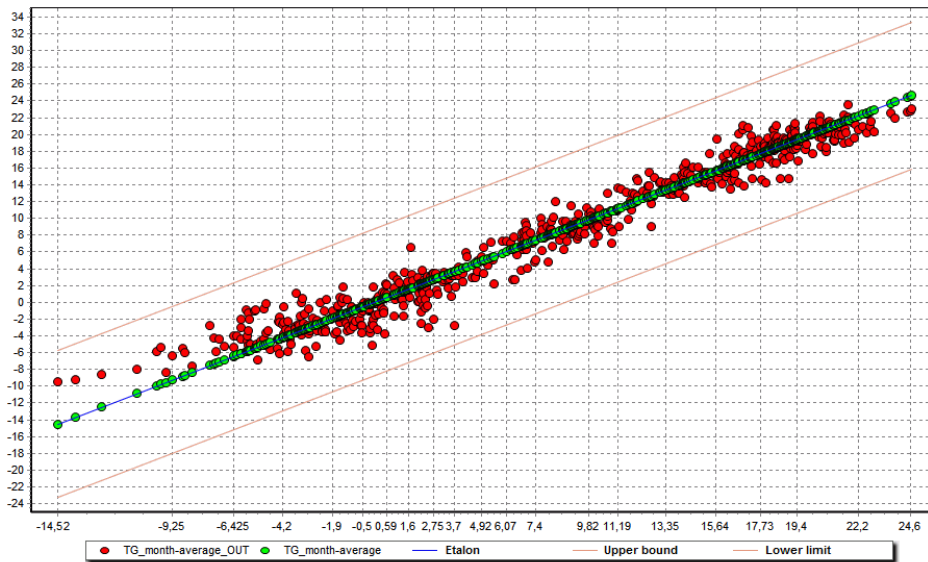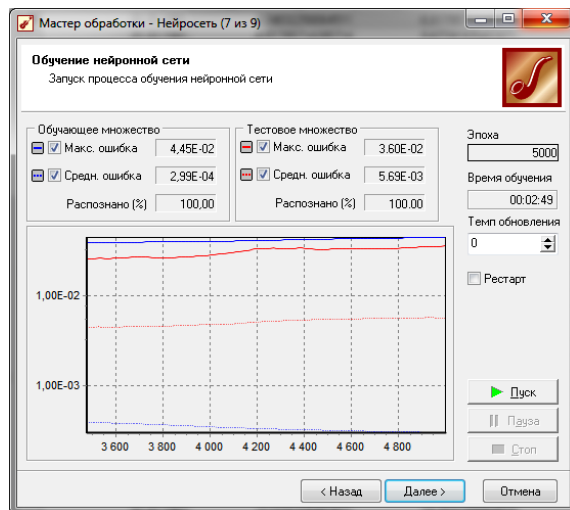
Fig. 4. Scattering diagram for the predictive linear regression model $[240 \times 1]$ on the training data set constructed by the sliding window method with the prehistory depth of 240 months for the time series of average monthly temperature in 1950–2016 for Kyiv city
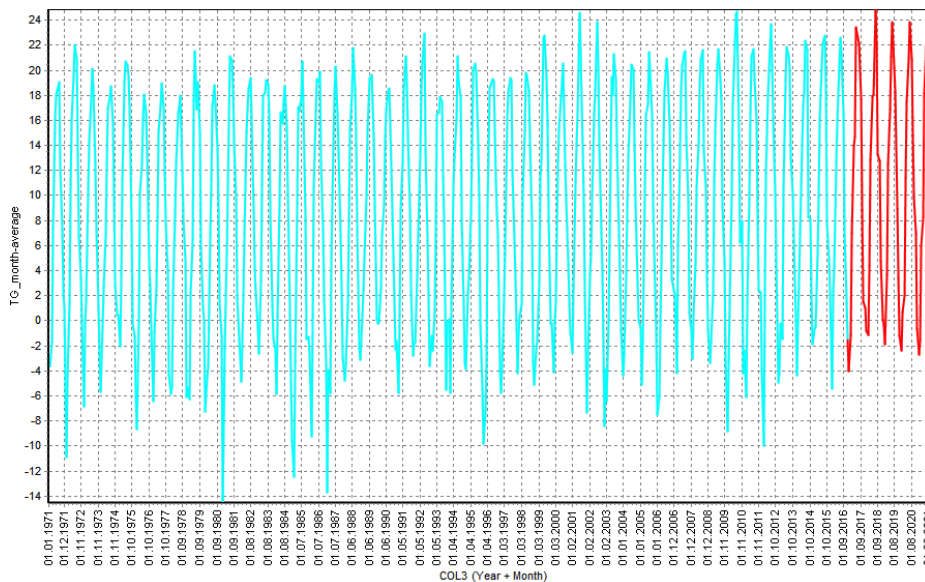

Fig. 5. Forecast diagram of the time series of average monthly temperature for Kyiv city after 12.01.2016 with the forecast horizon of 60 months using the predictive model of linear regression $[240 \times 1]$ with the prehistory samples depth of 240 months
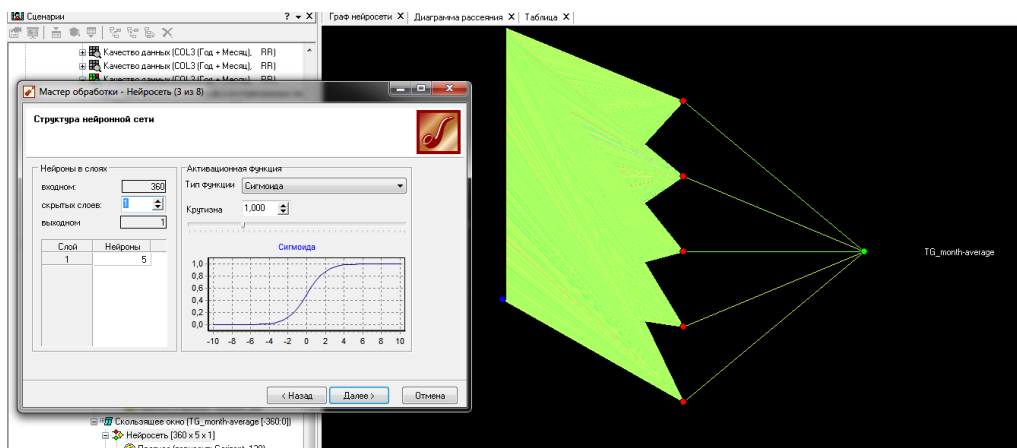
Fig. 6. Interface for specifying of the predictive neural network model $[360 \times 5 \times 1]$ parameters for the average monthly temperature in Kyiv city in 1951–2016 with the displaying of the neural network graph

Zyelyk Ya.I., Pidgorodetska L.V., Chornyy S.V., Kolos L.M., Dykach Yu.R.

Fig. 7. Result of the training process by the RPROP algorithm after 5000 epochs for the predictive neural network model $[360 \times 5 \times 1]$ for average monthly temperature in Kyiv city in 1950–2016
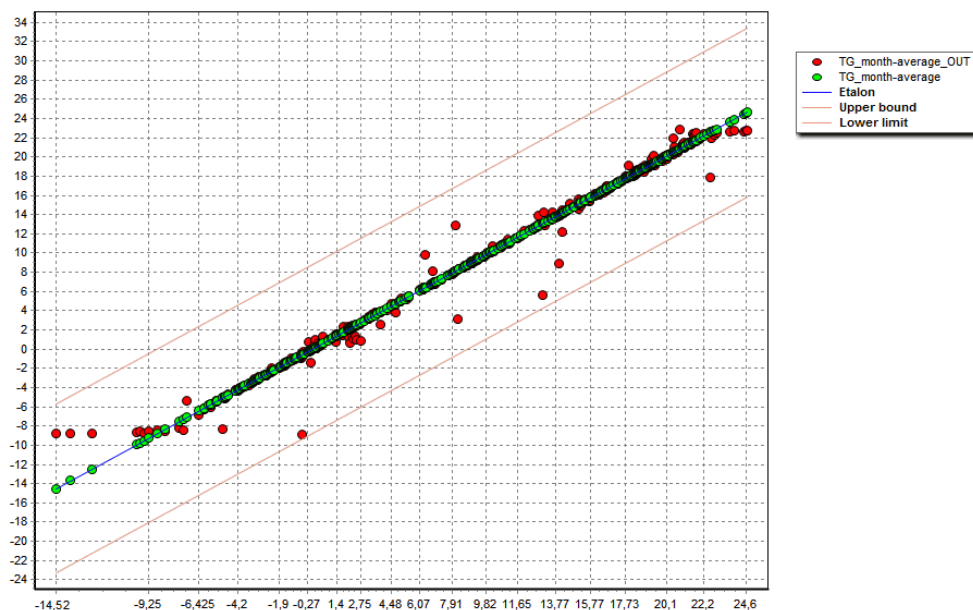
Fig. 8. Scattering diagram for the predictive neural network model $[360 \times 5 \times 1]$ for average monthly temperature in Kyiv city in 1950–2016 on the training and test sets jointly

Fig. 2 and Fig. 3 respectively it follows that there is no correlation between the trend of CO_2 emission and the average monthly temperature trend for Kyiv city in 1951–2013.

Fig. 4 shows the scattering diagram for the predictive linear regression model $[240 \times 1]$ on the training dataset constructed by the sliding window method with the prehistory depth of 240 months for the time series of the average monthly temperature in 1950–2016 for Kyiv city with the representation on the plane of the etalon model output and actual model output within the tolerable training error with the reliability of 0.95.
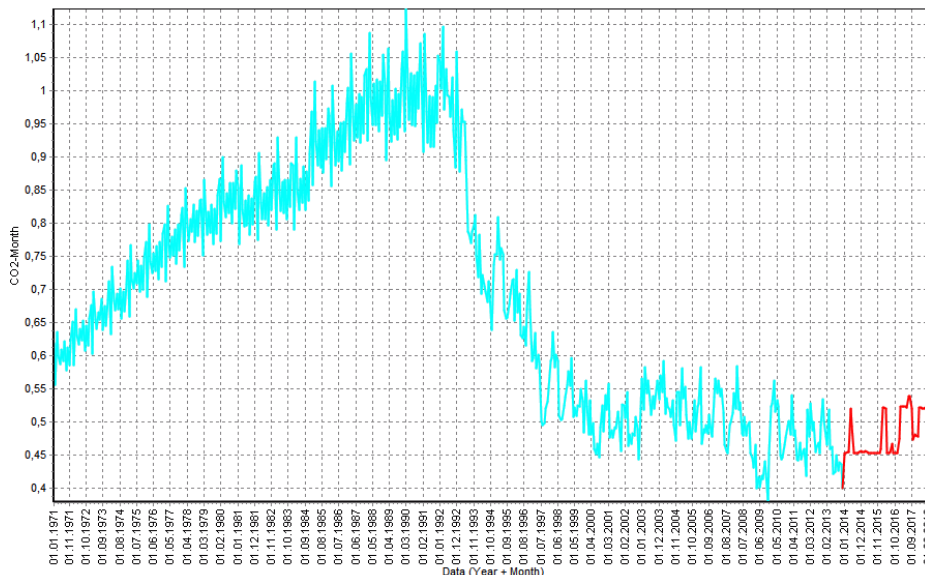
Fig. 5 shows the forecast diagram of the average monthly temperature time series for Kyiv city after 12.01.2016 with the forecast horizon of 60 months using the linear regression $[240 \times 1]$ predictive model trained on the dataset constructed by sliding window method with the prehistory samples depth of 240 months.

Fig. 6 shows the interface for specifying the parameters of the predictive neural network model $[360 \times 5 \times 1]$ with the show of the neural network graph and the following parameters: the number of inputs, which is equal to the experimentally selected prehistory depth of the time series samples that is 360; number of outputs is 1; number of hidden neuron layers is 1; experimentally selected neuron number in the hidden layer is 5; type of activation function is sigmoid with given slope equal 1.

Fig. 7 shows the result of the training process by the RPROP algorithm for the predictive neural network model $[360 \times 5 \times 1]$ for average monthly temperature in Kyiv city in 1950–2016 (the structure of the model is
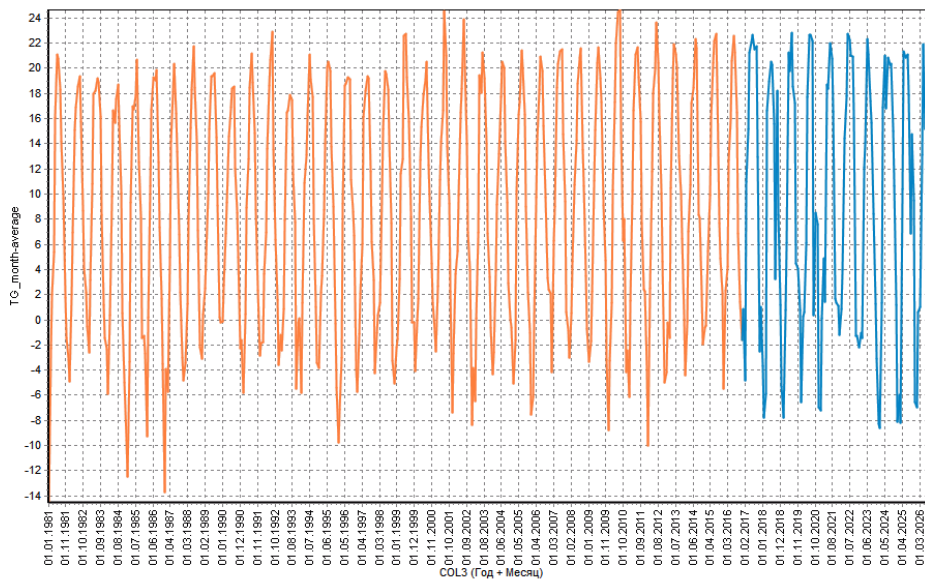
Fig. 9. Forecast diagram of the time series of the average monthly temperature for Kyiv city after 12.01.2016 with the forecast horizon of 120 months using the neural network model [360 × 5 × 1] with the prehistory depth of 360 months

Fig. 10. Forecast diagram of the time series of the monthly CO_2 emission values for Kyiv city after 12.01.2016 with the forecast horizon of 60 months using the neural network model [240 × 5 × 1] with the prehistory depth of 240 months

shown in Fig. 6) after the implementation of 5000 epochs, that is of training cycles on the training set and test cycles on the test set. The values of the maximum and average relative errors of the neural network model on the training set ($4,45 \cdot 10^{-2}$ and $2,99 \cdot 10^{-4}$ respectively) and on the test set ($3,60 \cdot 10^{-2}$ and $5,69 \cdot 10^{-3}$ respectively) are shown, that is rather low errors values in the permissible limits achieved as a result of training, indicating the adequacy of the model.

Fig. 8 shows the scattering diagram for the predictive neural network model [360 × 5 × 1] of average monthly temperature for the years 1950–2016 in Kyiv on the training and test sets jointly, for which, in a given percentage, is randomly separated the input data set constructed by the sliding window method with the retrospective depth of 360 months. Fig. 8 shows the etalon model output and the actual model output within the tolerable training error with the reliability of 0.95.

Fig. 9 shows the forecast diagram of the time series of average monthly temperature for Kyiv city after 12.01.2016 with the forecast horizon of 120 months using the neural network model [360 × 5 × 1] trained on the dataset constructed by sliding window method with the prehistory depth of 360 months.

Fig. 10 shows the forecast diagram of the time series of the monthly CO_2 emission values for Kyiv city after 12.01.2016 with the forecast horizon of 60 months using the neural network model [240 × 5 × 1] trained on the dataset constructed by sliding window method with the prehistory depth of 240 months.

4. CONCLUSIONS

Two types of models for assessment and forecasting of local climate change based on time series data have been studied: linear regression model and neural network model in the form of perceptron with one hidden neuron layer, using modern approaches to intellectual analysis of Big Data: Data Mining and Knowledge Discovery in Databases.

The study is based on the processing of time series of daily precipitation and temperature data in the period 1950–2016 and the monthly emission of carbon dioxide from fossil fuels in the atmosphere to the area of 1 geographic degree in the period 1951–2013 for number Ukraine cities. The program scenarios for processing, intellectual analysis and forecasting of the specified time data series using the constructed prediction models has been developed and implemented in the environment of the unificated analytical Deductor platform.

The decomposition of the time series of the total monthly precipitation and the average monthly temperature in 1950–2016 on the trend, seasonal component and random remainder for number of Ukrainian cities shows that there are significant seasonal components with the period of 1 year in these series, and the time series of average monthly temperatures have the very strong periodic component. For Kyiv city, the time series of total monthly precipitation contains the small cubic trend in the small dynamic range of 460–560 mm in 66 years, and the time series of average monthly temperature contains the slight cubic growing trend in the small dynamic range of 7.4–10.1 degrees, i.e. the growth of the temperature trend is 2.7 degrees per 66 years.

The monthly time series of the emission in atmosphere of carbon dioxide from fossil fuels to area of 1 geographical degree in 1951–2013 for number of Ukrainian cities shows clear trend, which is well approximated by the cubic polynomial and does not contain a periodic seasonal component. There is the tendency to increasing CO_2 emissions before 1991 and the declining tendency after 1991 with some "plateau" in the period 2000–2009 that are years of relative stability in the economic development of Ukraine. However, there is no correlation between the trend of CO_2 emissions and the trend of average monthly temperature in 1951–2013.

The predictive model of linear regression was the most acceptable for time series of average monthly temperature in 1950–2016, which are characterized by strong seasonal component with the periodicity of 1 year. Using the predictive model of linear regression $[240 \times 1]$, trained on the dataset, constructed by the sliding window method with prehistory depth of 240 months, the consistent forecast of the time series of average monthly temperature for Kyiv city after 12.01.2016 has been obtained with the forecast horizon of 60 months. However, the implementation of the predictive model of linear regression is characterized by a relatively short horizon of consistent forecasting. The predictive model of linear regression was the most acceptable for time series of average monthly temperature in 1950–2016, which are characterized by strong seasonal component with the periodicity of 1 year.

Concerning the time series of total monthly precipitation and the time series of the monthly emission in the atmosphere of carbon dioxide from fossil fuels, as well as to provide of prolonged consistent forecast horizon, the predictive neural network model was significantly more effective.

For the time series of the average monthly temperature for Kyiv after 12.01.2016 the predictive neural network model $[360 \times 5 \times 1]$ (number of inputs is equal to the selected prehistory depth of the time series samples is 360; number of outputs is 1; number of neurons in the hidden layer is 5; type activation function is sigmoid with the given slope 1) provide the consistent forecast with the horizon of 120 months. Low values of the maximum and average relative errors of the neural network model were achieved on the training set ($4,45 \cdot 10^{-2}$ and $2,99 \cdot 10^{-4}$, respectively) and on the test set ($3,60 \cdot 10^{-2}$ and $5,69 \cdot 10^{-3}$, respectively).

Similarly, for the time series of monthly CO_2 emissions for the Kyiv city after December 1, 2013, the predictive neural network model $[240 \times 5 \times 1]$ provided consistent forecast with the horizon of 60 months. In general, the time series of monthly CO_2 emission values are characterized by much smaller values of the consistent forecast horizon in comparison with the time series of the average monthly temperature, at least when using predictive neural network models.

1. *Fayyad U., Piatetsky-Shapiro G., Smyth P.* From data mining to knowledge discovery in databases // AI Magazine. — 1996. — Vol. 17, No. 3. — P.37–54. https://doi.org/10.1609/aimag.v17i3.1230
2. *Barsegian A.A., Kuprianov M.S., Kholod I.I., Tess M.D., Yelisarov S.I.* Analysis of data and processes: handbook. — St. Petersburg: BHV-Petersburg, 2009. — 512 p. [in Russian]
3. *Gladun A.Ya., Rogushyna Yu.V.* Data Mining: search for knowledge in data. — Kyiv: Ltd "PH ADEF-Ukraine". — 452 p. [in Ukrainian]
4. Deductor [in Russian] [Electronic resource]. https://basegroup.ru/deductor/description (01.11.2020)
5. Deductor. Analitic Guide. Version 5.3. — BaseGroup Labs Companie, 1995–2013. — 219 p. [in Russian]

6. European Climate Assessment & Dataset (ECA&D) project. Daily data [Electronic resource]. http://eca.knmi.nl/dailydata/index.php (01.11.2020)

7. Carbon Dioxid Information Analysis Center (CDIAC). Monthly Fossil-Fuel CO_2 Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude [Electronic resource]. https://cdiac.ess-dive.lbl.gov/epubs/fossil_fuel_CO2_emissions_gridded_monthly_v2016.html# (01.11.2020)

8. *Diakonov V., Kruglov V.* Mathematical packages of Matlab extension. Special reference book. — St. Petersburg: Peter, 2001. — 480 p. [in Russian]

9. *Kussul N.M., Shelestov A.Yu., Lavreniuk A.M.* Intelligent computing: handbook. — Kyiv: Naukova dumka, 2006. — 186 p. [in Ukrainian]

10. *Pidgorodetska L., Yatsenko V., Zyelyk Ya.* Development of mathematical models to predict the impact of the greenhouse effect on climate change at the regional level // Abstracts of the 17th Ukrainian Conference on Space Research, Odessa, August, 21–25, 2017. — Kyiv: SRI, NAS and SSA of Ukraine, 2017. — P.189.

11. *Pidgorodetska L., Yatsenko V., Zyelyk Ya.* Development of nonlinear dynamic models of the climate change and its applications for carbon dioxide emission time series data over Ukraine's territory // Proceedings of Conference GEO-UA 2018, 18–19 September 2018, Kyiv, Ukraine. — P.56–57

Застосування лінійних регресійних та нейромережевих моделей для оцінки та прогнозування локальних змін клімату на основі часових рядів даних

Зєлик Я.І.[1], *Підгородецька Л.В.*[1], *Чорный С.В.*[1], *Колос Л.Н.*[1], *Дикач Ю.Р.*[2]

[1]Інститут космічних досліджень НАН України і ДКА України,
03187, м. Київ, пр. Академіка Глушкова, 40, корпус 4/1

[2]Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»,
03056, м. Київ, пр. Перемоги, 37

Дослідження змін клімату базуються на оброблені даних таких часових рядів для низки міст України: щоденних атмосферних опадів та температури повітря з електронного ресурсу Європейського проекту з оцінки клімату та даних про клімат (European Climate Assessment & Dataset (ECA&D) Project) та щомісячної емісії двоокису вуглецю від викопного палива в атмосфері на ділянку розміром в 1 географічний градус з електронного ресурсу Carbon Dioxide Information Analysis Center (CDIAC). Оцінювання синхронної динаміки та прогнозування температури повітря, атмосферних опадів за даними періоду 1950–2016 рр. та концентрації двоокису вуглецю в атмосфері за даними періоду 1951–2013 рр. проведено із застосуванням двох типів моделей: лінійної регресії та нейромережевої моделі у формі персептрона з одним прихованим шаром нейронів на основі сучасних підходів до інтелектуального аналізу великих обсягів даних: Data Mining (DM) — видобування даних і Knowledge Discovery in Databases (KDD) — виявлення знань у базах даних. Розроблено та реалізовано у середовищі уніфікованої аналітичної платформи Deductor програмні сценарії оброблення, інтелектуального аналізу та прогнозування вказаних часових рядів даних із застосуванням побудованих прогнозуючих моделей. Декомпозиція часових рядів щомісячної емісії в атмосфері двоокису вуглецю від викопного палива на ділянку розміром в 1 географічний градус у 1951–2013 рр. для низки міст України демонструє явно виражений тренд, що добре апроксимується кубічним поліномом, та відсутність періодичної сезонної складової. Спостерігається тенденція зростання викидів CO_2 до 1991 р. і тенденція їх спадання після 1991 р. з деяким «плато» на відрізку 2000–2009 рр. — роки відносної стабільності в економічному розвитку України. Однак не має місця жодної кореляції між трендом викидів CO_2 та трендом середньомісячної температури у 1951–2013 рр. Прогнозуюча модель лінійної регресії виявилась найбільш прийнятною для часових рядів середньомісячної температури у 1950–2016 рр., для яких властива потужна сезонна складова з періодичністю 1 рік. З використанням прогнозуючої моделі лінійної регресії $[240 \times 1]$, навченої на множині даних, побудованій за методом ковзного вікна з глибиною передісторії відліків 240 місяців, отримано спроможний прогноз часового ряду значень середньомісячної температури для м. Києва після 01.12.2016 р. з горизонтом прогнозу 60 місяців. Однак для застосування прогнозуючої моделі лінійної регресії характерний відносно короткий горизонт спроможного прогнозу. Для часового ряду значень середньомісячної температури для м. Києва після 01.12.2016 р. прогнозуюча нейромережева модель $[360 \times 5 \times 1]$ (кількість входів рівна вибраній глибині передісторії відліків часового ряду — 360; кількість виходів — 1; кількість нейронів у прихованому шарі — 5; тип активаційної функції — сігмоїда з заданою крутизною — 1) забезпечила спроможний прогноз з горизонтом 120 місяців. Були досягнуті низькі значення максимальної і середньої відносних помилок нейромережевої моделі на навчальній множині ($4{,}45 \cdot 10^{-2}$ і $2{,}99 \cdot 10^{-4}$ відповідно) та на тестовій множині ($3{,}60 \cdot 10^{-2}$ і $5{,}69 \cdot 10^{-3}$ відповідно). Аналогічним чином, для часового ряду значень щомісячної емісії CO_2 для м. Києва після 01.12.2013 р. прогнозуюча нейромережева модель $[240 \times 5 \times 1]$ забезпечила спроможний прогноз з горизонтом 60 місяців. Взагалі, часові ряди значень щомісячної емісії CO_2 характеризуються значно меншими значеннями горизонту спроможного прогнозу порівняно з часовими рядами значень середньомісячної температури, принаймні, при використанні прогнозуючих нейромережевих моделей.

Ключові слова: зміни клімату; емісія двоокису вуглецю; прогнозування часових рядів; Data Mining; лінійна регресія; нейромережева модель; одношаровий персептрон.

Применение линейных регрессионных и нейросетевых моделей для оценки и прогнозирования локальных изменений климата на основе временных рядов данных

Зелык Я.И., Подгородецкая Л.В., Черный С.В., Колос Л.Н., Дыкач Ю.Р.

[1]Институт космических исследований НАН Украины и ГКА Украины,

03187, г. Киев, пр. Академика Глушкова, 40, корпус 4/1

[2]Национальный технический университет Украины «Киевский политехнический институт имени Игоря

Сикорского», 03056, г. Киев, пр. Победы, 37

Исследование изменений климата базируются на обработке данных таких временных рядов для ряда городов Украины: ежедневных атмосферных осадков и температуры воздуха из электронного ресурса Европейского проекта по оценке климата и данным о климате (European Climate Assessment & Dataset (ECA&D) Project) и ежемесячной эмиссии двуокиси углерода от ископаемого топлива в атмосфере на участок размером в один географический градус из электронного ресурса Carbon Dioxide Information Analysis Center (CDIAC). Оценивание синхронной динамики и прогнозирование температуры воздуха, атмосферных осадков по данным периода 1950–2016 гг. и концентрации двуокиси углерода в атмосфере по данным периода 1951–2013 гг. проведено с применением двух типов моделей: линейной регрессии и нейросетевой модели в форме персептрона с одним скрытым слоем нейронов на основе современных подходов к интеллектуальному анализу больших объемов данных: Data Mining (DM) — добыча данных и Knowledge Discovery in Databases (KDD) — выявление знаний в базах данных. Разработаны и реализованы в среде унифицированной аналитической платформы Deductor программные сценарии обработки, интеллектуального анализа и прогнозирования указанных временных рядов данных с применением построенных прогнозирующих моделей. Декомпозиция временных рядов ежемесячной эмиссии в атмосфере двуокиси углерода от ископаемого топлива на участок размером в 1 географический градус в 1951–2013 гг. для ряда городов Украины демонстрирует явно выраженный тренд, хорошо аппроксимирующийся кубическим полиномом, и отсутствие периодической сезонной составляющей. Наблюдается тенденция роста выбросов CO_2 до 1991 г. и тенденция их падения после 1991 г. с некоторым «плато» на отрезке 2000–2009 гг. — годы относительной стабильности в экономическом развитии Украины. Однако отсутствует корреляция между трендом выбросов CO_2 и трендом среднемесячной температуры в 1951–2013 гг. Прогнозирующая модель линейной регрессии оказалась наиболее приемлемой для временных рядов среднемесячной температуры в 1950–2016 гг., для которых свойственна мощная сезонная составляющая с периодичностью 1 год. С использованием прогнозирующей модели линейной регрессии $[240 \times 1]$, обученной на множестве данных, построенном по методу скользящего окна с глубиной предыстории отсчетов 240 месяцев, получен состоятельный прогноз временного ряда значений среднемесячной температуры для г. Киева после 01.12.2016 г. с горизонтом прогноза 60 месяцев. Однако для применения прогнозирующей модели линейной регрессии характерен относительно короткий горизонт способного прогноза. Для временного ряда значений среднемесячной температуры для г. Киева после 01.12.2016 г. прогнозирующая нейросетевая модель $[360 \times 5 \times 1]$ (количество входов равно выбранной глубине предыстории отсчетов временного ряда — 360, количество выходов — 1, количество нейронов в скрытом слое — 5, тип активационной функции — сигмоида с заданной крутизной — 1) обеспечила состоятельный прогноз с горизонтом 120 месяцев. Были достигнуты низкие значения максимальной и средней относительных ошибок нейросетевой модели на обучающем множестве ($4,45 \cdot 10^{-2}$ и $2,99 \cdot 10^{-4}$ соответственно) и на тестовом множестве ($3,60 \cdot 10^{-2}$ и $5,69 \cdot 10^{-3}$ соответственно). Аналогичным образом, для временного ряда значений ежемесячной эмиссии CO_2 для г. Киев после 01.12.2013 г. прогнозирующая нейросетевая модель $[240 \times 5 \times 1]$ обеспечила состоятельный прогноз с горизонтом 60 месяцев. Вообще, временные ряды значений ежемесячной эмиссии CO_2 характеризуются значительно меньшими значениями горизонта состоятельного прогноза по сравнению с временными рядами значений среднемесячной температуры, по крайней мере, при использовании прогнозирующих нейросетевых моделей.

Ключевые слова: изменения климата; эмиссия двуокиси углерода; прогнозирования временных рядов; Data Mining; линейная регрессия; нейросетевая модель; однослойный персептрон.