V. E. Kopniak; V. B. Mokin, Doctor of Science (Eng.), Professor;
S. O. Zhukov, Cand. Sc. (Eng.), Associate Professor;
I. V. Varchuk, Cand. Sc. (Eng.), Associate Professor; T. V. Skrypnyk

BOOSTING METHOD OF HETEROSKEDASTIC MODELS FOR THE PREDICTION OF THE SAHARA DUST CONCENTRATION IN THE ATMOSPERIC AIR OF UKRAINE

The paper presents new method of heteroskedastic models boosting and its applied usage on the example of the Sahara dust concentration in the atmospheric air of Ukraine. Recently the increased frequency of fine dust transfer from the Sahara Desert across the Mediterranean sea on the territory of Europe, in particular – in Ukraine, is observed. This phenomenon complicates the prediction of atmospheric air quality as a result of destruction of the stable regularities of pollution, as new factors are added, they need special models for adequate description. Special meteorological mode of the Sahara Desert dust spread enables to assume that the dispersion of the remains of ARIMA model may be random process and for its description it is expedient to use heteroskedastic models, such as GARCH. However, the conventional GARCH-models are efficient only if one dominating random process is available. If several such processes are added, conventional models loose their efficiency.

In the given study the application of boosting approach for the construction of the assembly models is suggested, these models, unlike the available, comprise the cooperation of the decision trees and heteroskedastic models for modeling complex heteroskedastic processes. The suggested method, as it is accepted for the boosting models, is based on the iterative process of models selection, where the next model takes into account the errors of the previous model. For the verification of the efficiency of the method the data of civil monitoring of atmospheric air EcoCity were used, in particular, data regarding Vinnytsia region by PM1 index, which indicate the periods, when the concentration of the Sahara fine dust in the atmospheric air of the region reached especially abnormal values.

It was proved that the process of the Saharan dust spreading in Vinnytsia region is heteroskedastic. SARIMAX models and typical GARCH-models using Python-libraries statsmodels and arch are constructed. It is revealed that the model ARIMA demonstrates far better results as compared with classic GARCH-models with different parameters, this shows insufficient efficiency of these GARCH-models. The suggested method of boosting heteroskedastic models allows to reach far greater accuracy than all these models in the whole range of values, except the value of the largest anomaly, this value is impossible to foresee. Thus, the forecasting method, developed in this research is an efficient approach for the solution of complex forecasting problems, the example of which is forecasting of the atmospheric air quality during the Sahara dust spreading in Ukraine.

***Key words**: artificial intelligence, time series, heteroskedastic models, GARCH, ARIMA, SARIMAX, boosting of machine learning models, the Saharan dust, atmospheric air quality, ecological monitoring, pollution forecasting, fine dust, PM1, EcoCity, assembly models, machine learning.*

Introduction

Recently, characteristic features of metrological conditions cause the motion of the fine dust from the Sahara Desert across the Mediterranean Sea and its spread on the territory of Europe, in particular – in Ukraine. This fact is recorded by the systems of both state and civil monitoring of atmospheric air quality.

According to the observation data of 2024 such phenomenon is observed in different periods from March 17 till April 28 (Fig. 1, 2) [1 – 3]. This phenomenon has great impact on ecosystems, including increase of harmful substances concentration in the air, influencing human health, plants, animals. High concentration of fine dust may inhibit plants growth, reduce visibility and contribute to emergence respiratory diseases among people. Besides, dust from the Sahara may transfer pathogens and other polluting substances, worsening the state of ecosystems [4 – 6].

Another negative result of this phenomenon is making forecasting of the atmospheric air more difficult. Dominating polluting impact destroys the established regularities. Other models and approaches are needed for its forecasting. For the description of the time series ARIMA models and its variations are traditionally used [7]. But special meteorological mode of the Sahara dust spreading allows to assume that the dispersion of the remains of ARIMA model may also be random process which is described by a certain model. For the description of such processes heteroskedastic models are most suitable [8, 9]. However, the process of modeling is complicated by the fact that traditional polluting factors do not disappear but are added and, in this case, more complex assembly models and methods of their identification are needed.

Objective of the paper is to increase the accuracy of time series prediction with dominant impact and usage of heteroskedastic models on the example of dust concentration modeling in the atmospheric air of Ukraine during the active spreading of the Saharan dust. This enables to evaluate more accurately the range of values of short time prediction of atmospheric air quality during this peculiar meteorological phenomenon.

Idea of the method

It is noted in the paper [10] that the specific class of time series is heteroskedastic process with the dispersion, which changes in time according to certain stochastic law. For its prediction special heteroskedastic models (GARCH-models) are used. In the same study [10] the review of such models is done , identification of which is automated in Python-library arch.

However, it is quite clear, that GARCH-model is efficient when one basic random process with variable in time dispersion at certain model takes place. If the sum of two such processes takes place, this model will not be efficient. One law will overlap another. Such an example is studies in the research [10], where it is proved, that the process of index «PM10» change at the station № 650 («Turbiv») of EcoCity network according to the data of 2020 – 2023 is heteroskedastic but it is shown that classic models ARCH, GARCH, EGARCH, APARCH, HARCH are less adequate that the models ARIMA, which also gives important error. One of the reasons of this phenomenon may be the fact that the station records not one heteroskedastic process but several processes simultaneously. For such complex cases it is suggested to construct the assembly model.

In the work [11] the review of the main types of assembly models of machine learning is presented: beting, boosting, stacking, voting. If this complex process is considered as the set of simpler processes, then the best variant is boosting. This is an iterative process of model selection. At each iteration the prediction discrepancy is analyzed by the previous model. New model tries to take into account this discrepancy. The result is a complex of models, applied in series, each of the models takes into consideration certain characteristics of the data and decreases the resultant error. Usually, this method is applied for combining simple models (for instance, Decision Tree). In the study [12] the attempt was made to construct the assembly of heteroskedastic models on the base of boosting (method Wild Boost GARCH), but this approach did not gain wide application. Authors think that the reason is that the process of boosting is efficient if simple models are used but not only complex heteroskedastic models.

Main idea of the suggested method is the construction of boosting, based on cooperation of the decision trees and heteroskedastic models. This method is worth applying only for heteroskedastic data series.

Let us considered the most generalized case when the problem is regressive and not classificational, i. e., when the value of the series is, for instance, fractional numbers. In this case, for boosting the regression variant of the decision tree «Decision Tree Regressor» (DTR) of Sklearn Python library [11] should be used. As heteroskedastic model one of the most generalized its types should be used, this type enables to take into account asymmetric changes of the dispersion and takes into consideration residuals sign, that allows to describe differently the impact of positive and negative deviations on the volatility of a series: «Exponential GARCH» (EGARCH) [10]. For launching boosting algorithm on zero iteration ($t = 0$, where t – is the number of iteration) one of the most efficient assembly boosting models – «Random Forest Regressor» (RFR) [11], is suggested to use, further pairs of the models «DTR- EGARCH» will work .

Algorithm of the method application

In general case input data X for modeling may be the table with numerous features (and may be either time series and time intersection), and target feature y – it is separate time series of the same length. In the first approximation most widely spread in practice variant will be considered, when the input dataset X contains only one series of values at certain moment of time, which are simultaneously the input data and the target feature y. To select optimal variant of the model classic method of the division of the input data X into the training X_T and validation X_V is used, for instance, randomly in the ratio of 80 % to 20 % [11]. And as the testing data we use data $y = X$, to analyze how the optimal model will describe the total series of the observations. This is more the problem of modeling than forecasting, when the adequacy of the model to all input data is studied. If it is adequate it can be also be used for the forecasting. The following method of heteroskedastic data series X and target feature y on the base of classic boosting algorithm is suggested but taking into account the decision trees DTR cooperation and one of heteroskedastic models EGARCH:

1. Initialization of the remainders at zero iteration ($t = 0$):

$$r[0] = y. \tag{1}$$

2. Initial iteration:
   - Learning of the model $f_0(X_T)$ Random Forest Regressor:

$$f_0(X_T) = RFR(X_T, r[0]). \tag{2}$$

   - Calculation of the predictions on zero iteration (t = 0) $y_{pred}[0]$ by means of substitution in the model (2) validation data X_V:

$$y_{pred}[0] = f_0(X_V). \tag{3}$$

   - Calculation of the remainders on the next iteration where, as it is accepted in boosting models [12], from remainders $r[0]$ the predictions are subtracted (3) with the account of L coefficient, that is often called «learning_rate»:

$$r[1] = r[0] - Ly_{pred}[0] = \theta(r[0], y_{pred}). \tag{4}$$

   - Learning of the model EGARCH (EG), where remainders $r[1]$ are substituted [1]:

$$\sigma^2[1] = EG(r[1]). \tag{5}$$

   - Correction of the remainders with the account of very small constant ξ (for instance 10^{-6}), to avoid division into zero in the denominator of the fraction – this is a typical method from [12]:

$$r^*[1] = \frac{r[1]}{\sqrt{\sigma^2[1] + \xi}} = \Omega_1(r[1], \sigma^2[1], \xi). \tag{6}$$

3. Next iterations $t = 1 \ldots N-1$:
   - Learning of the model Decision Tree Regressor:

$$f_t(X_T) = DTR(X_T, r^*[t]). \tag{7}$$

   - Calculations of the predictions:

$$y_{pred}[t] = f_t(X_V). \tag{8}$$

   - Calculations of the remainders:

$$r[t+1] = r^*[t] - Ly_{pred}[t]. \tag{9}$$

- Learning of the model EGARCH:

$$\sigma^2[t+1] = EG\left(r[t+1]\right). \tag{10}$$

- Correction of the remainders:

$$r^*[t+1] = \frac{r[t+1]}{\sqrt{\sigma^2[t+1]+\xi}} = \Omega_{t+1}(\cdot). \tag{11}$$

4. Final prediction – weighted sum of predictions at all iterations:

$$y_\Sigma = \sum_{t=0}^{N-1} Ly_{pred}[t]. \tag{12}$$

5. Assessment of the decision accuracy by the matrix MSE («Mean Square Error»):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{\Sigma,i}\right)^2. \tag{13}$$

Fig. 1 shows block-diagram of the suggested algorithm.

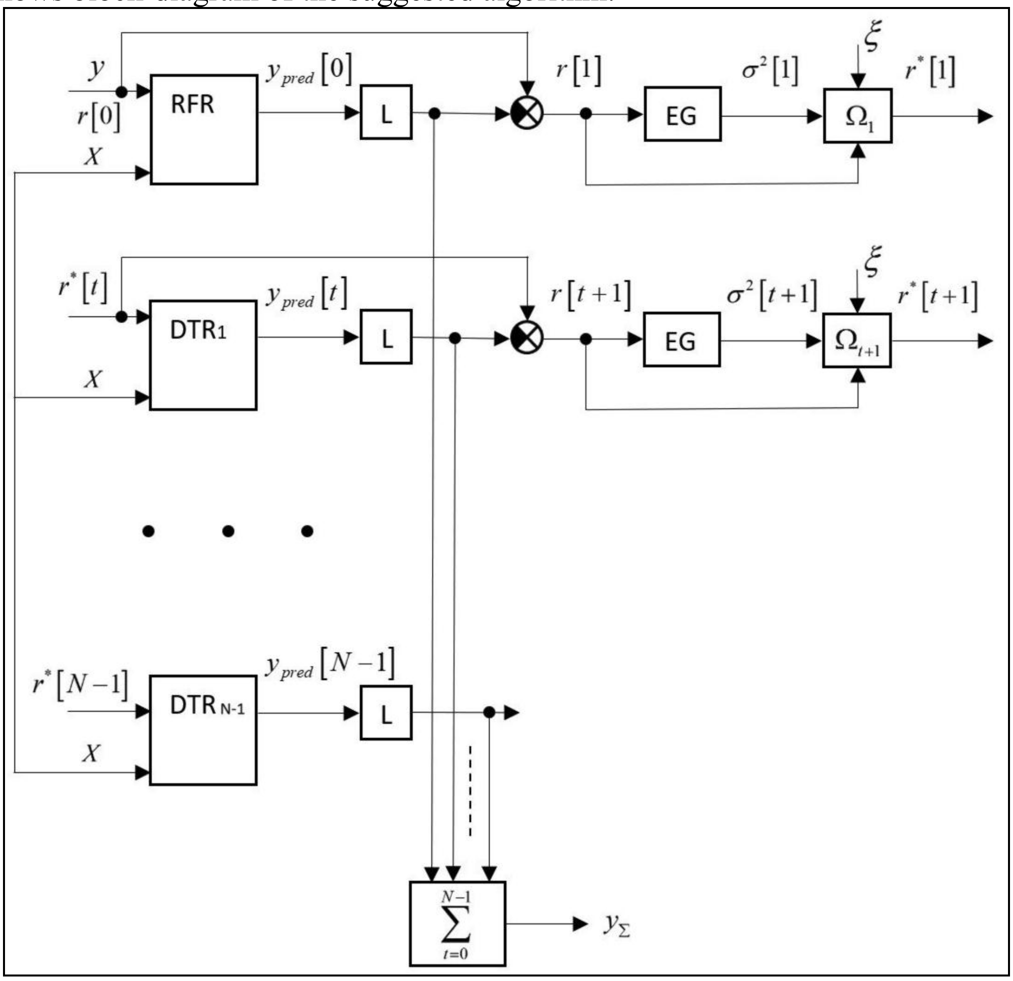


Fig. 1. Block-diagram of the suggested algorithm of the method of heteroskedastic data series prediction , on the base of boosting and cooperation of decision trees and heteroskedastic models

For the application of the suggested algorithm in practice, the following supplement of the operations is proposed at the beginning:

1. According to the data of observations, taking into consideration the sources of information about the specific feature of the subject area and object of study, reveal the period(s) of time, when dominant process potentially took place. Make a sample of data during this period.

2. Verify if the sample is heteroskedastic process, for instance, using Shapiro-Wilks [9] criterion. If the condition is proved, then pass to p. 3, otherwise – at the end of the algorithm.
3. If necessary, apply typical operation of preprocessing of data (inputing, standardization). Divide data into learning and validational , for instance, 80 % and 20 %, correspondingly.
4. Adjust boosting heteroskedastic model to this sample (Fig. 1).
5. Realize prediction of the data and calculate the matrices and errors.

This algorithm can be applied to different indices of the air quality, connected with the spreading of air pollution. Most brightly it can be demonstrated for the prediction of the greatest concentrations, connected with the impact of the Sahara dust on the quality of the atmospheric air in Ukraine. This can be proved.

Example of the application of the suggested method for the problem of the Saharan dust spread in the atmospheric air of Ukraine

Analysis of the sources showed that the Saharan dust reached Ukraine at the end of March 2024, spreading from the western regions to the rest of the territory. Dust cloud from the Sahara covered Germany, France and Switzerland and moved in the direction of Ukraine. This phenomenon was noticed, in particular, in Vinnytsia region, where white curtain appeared in the sky. Satellite images also proved the spread of the Sahara dust above Ukraine (Fig. 2, 3) [13, 14]. Dust storms in the Sahara usually contain microscopic particles, which can travel thousands of kilometers and influence the quality of air and visibility.

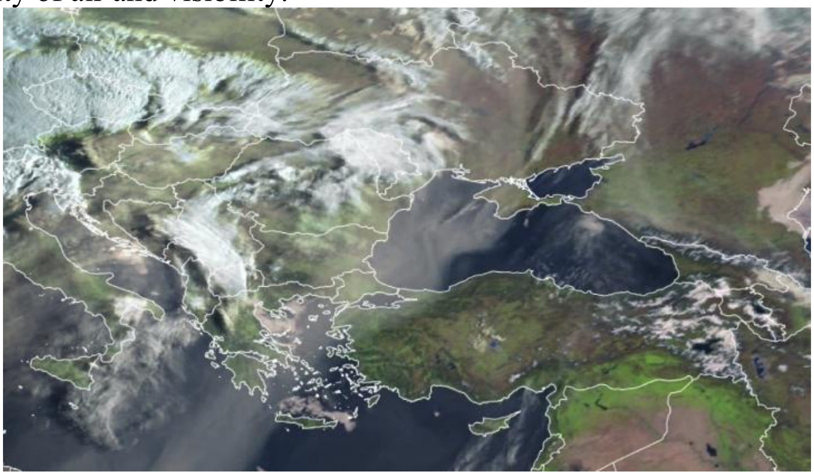


Fig. 2. Satellite image as of the 1st of April 2024

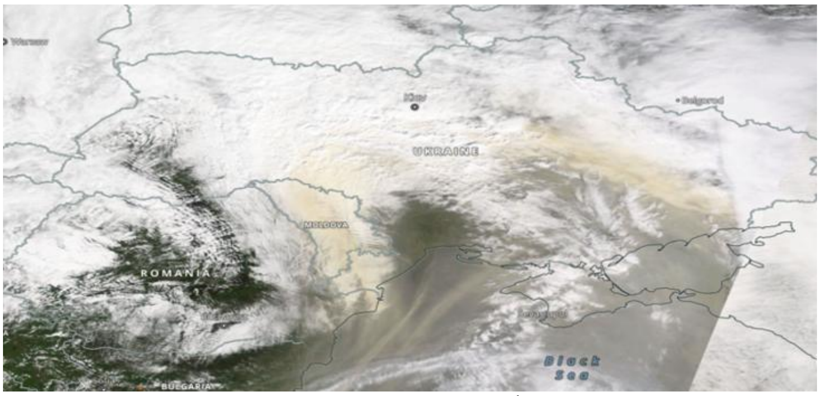


Fig. 3. Sattelite image as of the 24th of April 2024

Data of «Cabinet of the explorer of the air quality in Ukraine» (https://archive.eco-city.org.ua) of the network of civil monitoring of the atmospheric air EcoCity was used, Vinnytsia National Technical University has an authorized access to this resource. Post graduate student of VNTU Kopniak V. E. performed the import of data of the observations in Vinnytsia region by «PM1» index (concentration of the fine dust, transferred on large distances) during the period from 25.03.2024 till 10.05.2024 from this Cabinet and loaded the data into Kaggle-dataset «Air Quality Monitoring from Eco City» [15]. For instance, according to the data from the station № 1315, installed in scientific-research laboratory of the ecological studies and ecological monitoring of the Chair of System analysis and information technologies (SAIT) of VNTU, it is seen at the end of March and at the beginning of April the period of time with abnormally large values (Fig. 4), this period corresponds to the spreading of the Saharan dust in Vinnytsia region (Fig. 2).

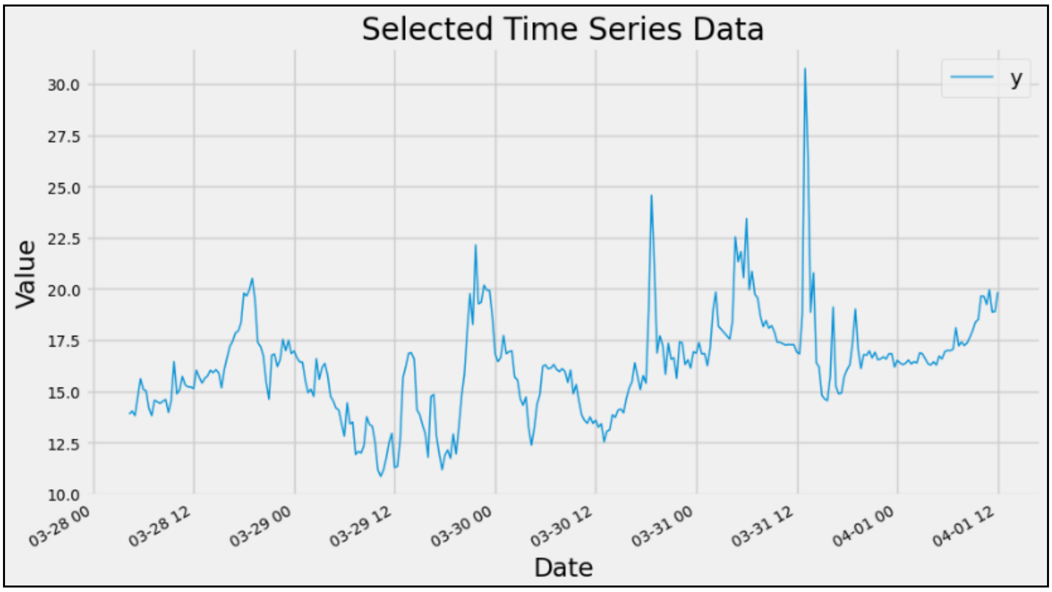


Fig. 4. Graph of «PM1» index data observations at the station № 1315 (Vinnytsia, VNTU) of Eco City network during the dust cloud motion from the Sahara in the atmospheric air of Vinnytsia region

Analysis of series data from Fig. 4 on heteroskedasticity is shown in Fig. 5 [16].

```
# Test for heteroskedasticity
arch_test = het_arch(resid, maxlag=lags_max)
shapiro_test = shapiro(st_resid)

print(f'Lagrange mulitplier p-value: {arch_test[1]}')
print(f'Shapiro-Wilks p-value: {shapiro_test[1]}')

if arch_test[1]>0.05:
    print("Time series is heteroskedasticity")
else: print("Time series is homoscedastic")
```

```
Lagrange mulitplier p-value: 0.9262081876457032
Shapiro-Wilks p-value: 1.0658690596798692e-10
Time series is heteroskedasticity
```

Fig. 5. Result of the verification of zero hypothesis regarding the heteroskedasticity of the observation data of «PM1» index at the stations of Eco City network during sand cloud passage from the Sahara Desert in the atmospheric air of Vinnytsia region (from the author`s Kaggle-notebook [16])

Standardization of the data is carried out, SARIMAX model from the library statsmodels.tsa.arima_model and typical models from the library arch, using author`s program-notebook on the base of the platform Kaggle [16] (Fig. 6) was constructed.

model_name	AIC	BIC	params
ARIMA	1060.904542	1079.571248	[1, 0, 2]
EGARCH	1961.595627	1980.246126	[1, 2]
EGARCH	1963.386028	1985.766627	[2, 2]
EGARCH	1970.596960	1989.247459	[2, 1]
EGARCH	1978.098144	1993.018543	[1, 1]
APARCH	1999.483736	2021.864335	[1, 2]
APARCH	2001.205673	2027.316371	[2, 2]
APARCH	2002.252056	2020.902555	[1, 1]
GARCH	2002.445240	2021.095739	[1, 2]
APARCH	2004.252056	2026.632654	[2, 1]
GARCH	2004.445241	2026.825839	[2, 2]
GARCH	2005.004308	2019.924707	[1, 1]
ARCH	2005.652019	2020.572418	[2, 1]
ARCH	2005.652019	2020.572418	[2, 2]
HARCH	2005.652019	2020.572418	[2, 1]
HARCH	2005.652019	2020.572418	[2, 2]
GARCH	2007.004308	2025.654807	[2, 1]
ARCH	2012.279722	2023.470021	[1, 2]
ARCH	2012.279722	2023.470021	[1, 1]
HARCH	2012.279722	2023.470021	[1, 1]
HARCH	2012.279722	2023.470021	[1, 2]

Fig. 6. Accuracy of typical models of heteroskedastic data series

As it is seen from Fig. 6, although the series is heteroskedastic, but the model ARIMA demonstrates better results but not the best results. Among GARCH-models the best is EGARCH-model. Similar conclusion was substantiated in the paper [10].

Fig. 7 contains the result of the identification of different parameters of models from Fig. 1, it gives rather high accuracy by the matrices and r2_score (takes into account the correctness of the direction of the values growth), and by the matrix «MAPE» («Mean Absolute Percentage Error»)which is the analog of the relative error in machine learning of the models.

n_estimators	learning_rate	max_depth	rmse	r2_score	mape
30	0.038	6	0.525978	0.969829	0.015924
30	0.038	5	0.530082	0.969356	0.015963
30	0.04	5	0.564941	0.965193	0.021762
25	0.04	5	0.619013	0.958211	0.0064
30	0.04	7	0.619418	0.958157	0.02373
25	0.04	7	0.632898	0.956316	0.008254
50	0.02	6	0.638552	0.955532	0.007497
30	0.038	7	0.640419	0.955271	0.019301
25	0.038	6	0.693737	0.947513	0.009052
25	0.038	5	0.697322	0.94697	0.008905
25	0.038	7	0.711547	0.944784	0.01001
25	0.04	6	0.722393	0.943088	0.010866
30	0.04	6	0.742197	0.939925	0.026406
50	0.02	7	0.925277	0.906631	0.01973
50	0.02	5	1.060899	0.877254	0.022432
30	0.02	6	1.492676	0.75701	0.046486
30	0.02	7	1.49818	0.755214	0.046474
30	0.02	5	1.576136	0.729077	0.044712
25	0.02	6	1.74112	0.66939	0.057347
25	0.02	5	1.743243	0.668584	0.057425
25	0.02	7	1.751656	0.665377	0.057316
50	0.038	6	1.867448	0.619675	0.091782
50	0.04	7	2.085526	0.525661	0.107327
50	0.038	5	2.147649	0.496981	0.094501
50	0.04	6	2.355412	0.39495	0.128006
50	0.04	5	2.408582	0.367325	0.105215
50	0.038	7	40.412197	-177.107771	1.096049

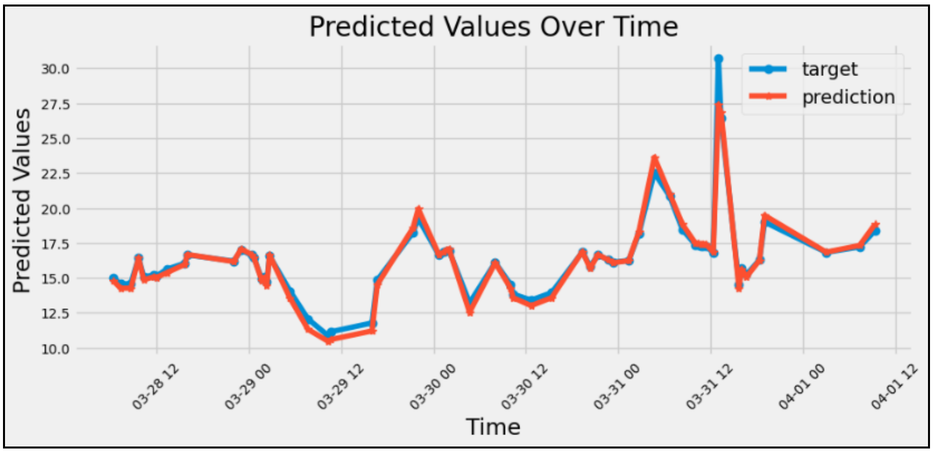
a)

n_estimators	learning_rate	max_depth	rmse	r2_score	mape
25	0.04	5	0.619013	0.958211	0.0064
50	0.02	6	0.638552	0.955532	0.007497
25	0.04	7	0.632898	0.956316	0.008254
25	0.038	5	0.697322	0.94697	0.008905
25	0.038	6	0.693737	0.947513	0.009052
25	0.038	7	0.711547	0.944784	0.01001
25	0.04	6	0.722393	0.943088	0.010866
30	0.038	6	0.525978	0.969829	0.015924
30	0.038	5	0.530082	0.969356	0.015963
30	0.038	7	0.640419	0.955271	0.019301
50	0.02	7	0.925277	0.906631	0.01973
30	0.04	5	0.564941	0.965193	0.021762
50	0.02	5	1.060899	0.877254	0.022432
30	0.04	7	0.619418	0.958157	0.02373
30	0.04	6	0.742197	0.939925	0.026406
30	0.02	5	1.576136	0.729077	0.044712
30	0.02	7	1.49818	0.755214	0.046474
30	0.02	6	1.492676	0.75701	0.046486
25	0.02	7	1.751656	0.665377	0.057316
25	0.02	6	1.74112	0.66939	0.057347
25	0.02	5	1.743243	0.668584	0.057425
50	0.038	6	1.867448	0.619675	0.091782
50	0.038	5	2.147649	0.496981	0.094501
50	0.04	5	2.408582	0.367325	0.105215
50	0.04	7	2.085526	0.525661	0.107327
50	0.04	6	2.355412	0.39495	0.128006
50	0.038	7	40.412197	-177.107771	1.096049

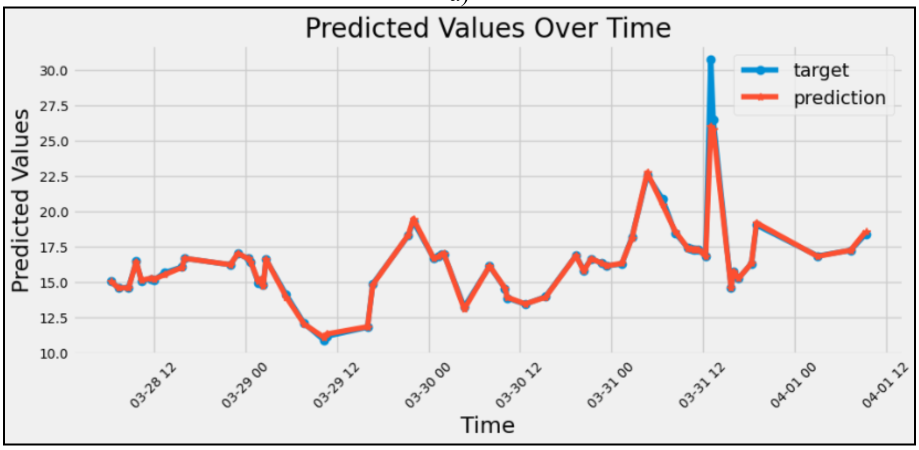
b)

Fig. 7. Accuracy of the typical models of heteroskedastic data series: a) sorted by the matrix r2_score; b) sorted by the matrix MAPE [16]

Figs. 8 a, b contain graphs, illustrating the accuracy of prediction of all the data of the sample from Fig. 1, which are optimal by two basic matrices with the parameters, indicated in the first row of the tables in Figs. 7 a, b, correspondingly.



a)



b)

Fig. 8. Result of the data prediction by the set and predicted values: a) model, optimal by matrix r2_score; b) model, optimal by matrix MAPE

As it is seen in Fig. 8, high accuracy was achieved. Main reason of the deviation is especially high abnormal value of the main peak of atmospheric air pollution with the Saharan dust on 31st of March 2024 at 12:00, it can not be predicted by this model.

Thus, the suggested method proved its efficiency during the solution of complex applied problem.

Conclusions

The paper considered the method of modeling and prediction of the Saharan dust concentration in the atmospheric air of Ukraine by means of boosting heteroskedastic models. Research, carried out, showed that dominanting impact of the Saharan dust on the quality of the atmospheric air complicates the prediction, using the conventional models as a result of the destruction of the established pollution regularities. That is why, it was suggested to use more complex assembly models, they take into consideration more complex dynamic of the process under the impact of various polluting factors simultaneously.

Main idea of the suggested method is the construction of the boosting model, based on the cooperation of decision trees and heteroskedastic models. Boosting approach enables to reduce gradually prediction errors at the expense of iteration sampling of the models, which take into account the errors of the models, constructed on the previous iterations. The analysis, performed, of the data of civil monitoring Eco City by the index of fine dust concentration «PM1» in Vinnytsia proved the efficiency of the proposed method.

The proposed method comprises several important steps: revealing of the periods, when the impact of dominant polluting process takes place, verification of the sample on heteroskedasticity and, in case of positive result of verification: preprocessing of the data, construction of the boosting model, based on the cooperation of the decision trees and heteroskedastic models, assessment of the accuracy metrices.

Efficiency of the method is proved by high adequacy of the model on the whole sample of data, where the impact of the Saharan dust in the atmospheric air of Vinnytsia region took place, as this is important for reliable prediction and, on its base – decision making in the sphere of ecological monitoring and health protection of the population. Proposed approach enables to take into account adequally the impact of complex meteorological conditions and additional polluting factors, that is critical for the development of the efficient management strategies of atmospheric air quality.

Thus, boosting approach with the cooperation of the decision trees and heteroskedastic models opens new possibilities for modeling and prediction of complex heteroskedastic processes. This method can be applied for various indices of the air quality, it allows to improve the accuracy of prediction and correspondingly, efficiency of managerial measures, aimed at the protection of people`s health, in the conditions of active spreading of the Saharan dust and other similar phenomena.

REFERENCES

1. The Copernicus Atmosphere Monitoring Service [Electronic resource]. Access mode : https://atmosphere.copernicus.eu/climate-atmosphere-podcast-understanding-impact-saharan-dust-storms.

2. Portal of the system of civil monitoring of the air quality EcoCity [Electronic resource]. Access mode : https://eco-city.org.ua/. (Ukr).

3. WorldView [Electronic resource]. Access mode : https://worldview.earthdata.nasa.gov.

4. The Saharan dust plume : Current knowledge on the impact on health, human activities, and the ecosystem, with comments on research gaps [Electronic resource] / Olaoluwa Omoniyi Olarewaju, Olufunke Omowumi Fajinmi, Theophilus Clavell Davies // Authorea Preprints. – 2023. – Access mode : https://d197for5662m48.cloudfront.net/documents/publicationstatus/131361/preprint_pdf/058d51d7471809a08d290dcf8a5e0950.pdf. DOI:10.22541/essoar.167979167.72582095/v1.

5. Wang Q. The impact of Sahara dust on air quality and public health in European countries / Q. Wang, J. Gu, X. Wang // Atmospheric Environment. – November 2020. – Volume 241. – P. 117771. DOI : https://doi.org/10.1016/j.atmosenv.2020.117771.

6. Knight J. Dust from the Central Sahara : Environmental and Cultural Impacts / J. Knight, S. Merlo, A. Zerboni // Landscapes and Landforms of the Central Sahara. World Geomorphological Landscapes. Springer, Cham, 2024. – P. 131 – 141. DOI: https://doi.org/10.1007/978-3-031-47160-5_14.

7. Methodology and organization of scientific investigations : Manual – 3*rd* Edition, revised and expanded / [B. I. Mokin, O. B. Моkіn, V. B. Моkіn]. – Vinnytsia : VNTU, 2023. – 230 p. (Ukr).

8. How to Model Volatility with ARCH and GARCH for Time Series Forecasting in Python [Electronic resource] / Jason Brownlee // Machine Learning Mastery. Time Series. – 24 August 2018. – Access mode : https://machinelearningmastery.com/develop-arch-and-garch-models-for-time-series-forecasting-in-python.

9. Volatility Clustering and GARCH / Nicholas Holloway // Kaggle notebook. – 18 March 2019. – Access mode : https://www.kaggle.com/code/nholloway/volatility-clustering-and-garch.

10. Study of the problems with heteroskedastic data of atmospheric air quality monitoring [Electronic resource] / V. E. Kopniak, V. B. Моkіn // Materials of LII Scientific technical conference of the Department of intelligent information technologies and automation of Vinnytsia National Technical University, Vinnytsia, June 21 – 23, 2023. – Access mode : https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2023/paper/view/18962/15728. (Ukr).

11. Science about data : machine learning and intelligent analysis of the data : electronic manual of the combined (local and network) usage [Electronic resource] / V. B. Моkіn, M. V. Dratovanyi – Vinnytsia: VNTU, 2024. – 258 p. – Access mode : https://iq.vntu.edu.ua/repository/getfile.php/8163.pdf. (Ukr).

12. Boosting GARCH and neural networks for the prediction of heteroskedastic time series / J. M. Matías, M. Febrero-Bande, W. González-Manteiga [et al.] // Mathematical and Computer Modelling. – 2010. – Volume 51, Issues 3 – 4. – P. 256 – 271. – DOI : https://doi.org/10.1016/j.mcm.2009.08.013.

13. Satellite images: Dust from the Sahara reached Ukraine [Electronic resource] / Access mode : https://universemagazine.com/en/satellite-images-dust-from-the-sahara-reached-ukraine/.

14. Why is Sahara dust blowing into Europe? [Electronic resource] / Access mode : https://www.dw.com/en/why-is-sahara-dust-blowing-into-europe/a-68927274.

15. Air Quality Monitoring from EcoCity [Electronic resource] / V. Mokin, D. Shmundiak, V. Kopniak // Kaggle Dataset. – May 2024. – Access mode : https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity.

16. Sahara's Dust in UA: Boosting GARCH models [Electronic resource] / V. Mokin. V. Kopniak // Kaggle Notebook. – June 2024. – Access mode : https://www.kaggle.com/code/vbmokin/sahara-s-dust-in-ua-boosting-garch-models/settings.

Editorial office received the paper 20.06.2024.
The paper was reviewed 24.06.2024.

Kopniak Volodymyr – Post Graduate with the Chair of SystemAanalysis and Information Technologies, e-mail: vkopnyak@gmail.com.

Mokin Vitaliy – Dr. Sc. (Eng.), Professor, Head of the Department of System Analysis and Information Technologies, e-mail: vbmokin@gmail.com.

Zhukov Sergiy – Cand. Sc. (Eng.), Associate Professor, Associate Professor with the Chair of System Analysis and Information Technologies, e-mail: sazhukov@gmail.com.

Varchuk Ilona – Cand. Sc. (Eng.), Associate Professor, Associate Professor with the Chair of System Analysis and Information Technologies, e-mail: ilona.varchuk@gmail.com.

Skrynyk Taras – Student of the group 2ICT-20б, Chair of System Analysis and Information Technologies, e-mail: tarasskrynnyk1@gmail.com.

Vinnytsia National Technical University.