American Journal of Applied Science and Technology

VOLUME Vol.05 Issue 12 2025
PAGE NO. 222-230
DOI 10.37547/ajast/Volume05Issue12-40

Forecasting Sand And Dust Storms In The Aral Sea Region

Khalmuratov Batyr Khalbay uli

A basic doctoral student at the Research Institute of Environmental and Nature Conservation Technologies, Uzbekistan

Received: 28 October 2025; **Accepted:** 19 November 2025; **Published:** 25 December 2025

Abstract: This study is devoted to the problem of predicting sand and dust storms in the Aral Sea region. The Aral Sea desert, which emerged as a result of the drying up of the Aral Sea, is increasing the frequency and intensity of sand and dust storms in the region, which has a serious negative impact on ecological stability and public health. The SARIMA statistical model and the XGBoost machine learning algorithm were used in the study for the forecast based on observation data recorded at the Muynak, Kun'gorat and Nukus meteorological stations located in the Aral Sea region over the past 10 years. According to the results obtained, the hybrid SARIMA–XGBoost model demonstrated high prediction efficiency. The accuracy of the model was 0.88, the F1-index was 0.85, and the ROC-AUC value was 0.92. Forecast errors were reduced by an average of 25–30% compared to the SARIMA model, achieving MAE values of 0.27 and RMSE values of 0.39. According to the feature importance analysis, maximum wind speed was the most important predictor, with a share of 34.7%. Wind gust (21.3%) and minimum relative humidity (17.6%) also had a significant impact.

Keywords: Aral Sea region, sand and dust storms, XGBoost, machine learning, meteorological factors, forecasting, environmental monitoring.

INTRODUCTION:

In recent years, the sharp decline of the Aral Sea has had a serious negative impact on the ecological situation and climate stability system of the Republic of Karakalpakstan and adjacent regions. As a result of the decline in sea level, its dried-up bottom has turned into a new desert area with an area of hundreds of thousands of square kilometers contaminated with salt, dust, and various toxic substances [1]. Sand and dust aerosols rising from these areas are spread over thousands of kilometers by atmospheric currents, having a significant negative impact on the climate system not only of Uzbekistan and Kazakhstan, but also of the entire Central Asian region [2]. As a result, soil degradation, reduction of vegetation cover, salinization of water resources, and an increase in respiratory and cardiovascular diseases in the health of the population are observed [3]. Sand and dust storms are considered one of the sources of environmental hazards on a global scale, changing the composition of aerosols in the atmosphere, affecting the scattering of solar radiation and cloud formation processes. This situation is recognized as one of the important factors leading to the disruption of the global radiation balance and the acceleration of climate change [4]. In order to monitor and forecast these processes, global forecasting systems such as WMO SDS-WAS, NASA GEOS-5, CAMS and NAAPS have been developed, which are widely used to assess the regional and global dynamics of dust and sand storms [5]. However, these models often have low spatial resolution and do not sufficiently take into account local meteorological conditions and geomorphological features of the area. In particular, in the Aral Sea region, which has a sharply continental climate, the limited forecast accuracy of these models has been noted [6]. Also, delays in real-time forecasts limit their full effective use in early warning systems.

Traditional statistical approaches used in modeling dust and sandstorms, including ARIMA and regression models, cannot fully reflect complex, nonlinear, and multifactorial processes [7]. Physically based models (WRF-Chem, HYSPLIT) are not always practical due to the high computational power required, complex parameter settings, and difficulties in real-time application [8]. Therefore, in recent years, approaches based on artificial intelligence and deep learning technologies have been increasingly used as an effective alternative in forecasting ecological and meteorological processes [9]. These models are characterized by their high capabilities in identifying hidden nonlinear relationships in time series and processing complex spatio-temporal structures. However, research aimed at high-accuracy forecasting of sand and dust storms based on local meteorological data, specifically adapted for the Aral Sea region, is not sufficiently developed. Most of the existing works are devoted to the frequency and intensity of storms, and their forecasting has been studied to a limited extent. In order to fill this scientific gap, this study proposes a CNN–LSTM hybrid model based on deep learning for the forecasting of sand and dust storms in the Aral Sea region. The main goal of the study is to expand the capabilities of sand and dust storm forecasting based on local meteorological data and to evaluate the effectiveness of the proposed model. The results of the conducted numerical modeling confirmed the high effectiveness of the hybrid approach based on the integration of SARIMA and XGBoost. According to the results of model evaluation, the hybrid model demonstrated 0.82–0.88 accuracy (Accuracy), an F1-index above 0.80, and ROC-AUC values about 0.90 in predicting sand and dust storms. Forecast errors in terms of RMSE and MAE indicators decreased by 20–30% compared to traditional statistical models. Feature importance analysis showed that maximum wind speed and decreasing relative humidity have the greatest impact on the probability of hurricanes. These results scientifically substantiate the practical reliability of the proposed hybrid modeling approach.

methodology

1. Research object.

This study is devoted to the study of sand and dust storms observed in the Aral Sea region, and the objects of its study are the sand and dust storms formed in the dried-up bottom of the Aral Sea and adjacent areas. The study was conducted on the example of the meteorological stations of the Republic of Karakalpakstan, where sand and dust storms are most actively observed - Muynak district, Kungirat district and Nukus city. These areas are located close to the former coastline of the Aral Sea, are characterized by strong deflation processes, high soil salinity, and sparse vegetation cover. Therefore, these areas are one of the most favorable source zones for the rise of sand and dust particles into the air and their long-distance dispersion.

The Aral Sea region is geomorphologically and climatically complex, and the Aral Sea desert formed in this area is the main source of dust in the region. The climate is sharply continental, with low annual precipitation (about 100–150 mm/year), with air temperatures rising to +40…+45 °C in the summer months and dropping to around −5 °C in the winter. The wind regime plays a decisive role in the formation of sand and dust storms in the region, with an average annual wind speed of 4–6 m/s, but on stormy days the wind speed can reach values of 15–20 m/s and higher. The prevailing wind directions are west and northwest, which enhances the migration of dust masses from the Aral Sea and Ustyurt plateaus across the Aral Sea regions [10].

The current analysis shows that, despite the important role of global forecasting systems in monitoring and forecasting sand and dust storms in the Aral Sea region, their practical effectiveness is often limited. The main reason for this is that global models do not sufficiently take into account local meteorological differences, station-level variability, and rapidly changing land surface conditions. In particular, in the Aral Sea region with its sharply continental climate, the occurrence and duration of sand and dust storms is a multifactorial and nonlinear process, which is difficult to accurately describe using simplified physical or statistical models.

2. Methodology.

This study uses a hybrid data-driven modeling approach based on local meteorological observations to predict sand and dust storms in the Aral Sea region. The study database consists of observations recorded at the Muynak, Kungirat, and Nukus meteorological stations in the Republic of Karakalpakstan during 2014–2023. These stations are located near the dry bottom of the Aral Sea and represent the areas where sand and dust storms are most frequent and intensively observed.

Observation data were recorded in accordance with the standards of the World Meteorological Organization (WMO) and included information on air temperature, relative humidity, average and maximum (gust) wind speed, precipitation amount, meteorological visibility, and synoptic phenomena. At the initial stage, the quality of the data was controlled, physically unreasonable values, duplicate records, and sharp anomalies were identified and excluded from the analysis. Short-term gaps were filled using conservative interpolation methods, and long-term discontinuities were separately identified in the analysis. All parameters were brought into a single system of units, ensuring the consistency of the time series.

Sand and dust storm events were identified using criteria based on wind speed, visibility, and synoptic observations, and a target variable was formed for forecasting. In order to improve short-term forecasting capabilities, the SARIMA (Seasonal ARIMA) statistical model was used to model the trend and seasonal components of the time series. Using this model, regular and seasonal changes in meteorological parameters over time were extracted and initial forecasts for the coming hours were formed.

In the next stage, an XGBoost (Extreme Gradient Boosting) machine learning model was built based on the SARIMA model outputs and observed meteorological parameters. Since the XGBoost model is capable of effectively studying nonlinear dependencies and complex interactions, it served to reflect the formation mechanisms of sand and dust storms in more depth. In order to improve the model efficiency, lagged values of wind speed and relative humidity, 3- and 7-day moving averages, integrated indicators representing the degree of surface dryness, and calendar parameters reflecting seasonality were introduced as additional features.

A hybrid modeling approach was implemented by combining the SARIMA and XGBoost models. This approach combines the advantages of the SARIMA model in representing linear and seasonal features in time series and the high efficiency of the XGBoost model in detecting nonlinear relationships. As a result, a forecasting system was formed that estimates the probability of sand and dust storms and can work automatically for the next few hours.

Standard metrics such as Accuracy, Recall, Precision, F1-score, and ROC-AUC were used to evaluate the model results. The practical applicability of the model was also analyzed using the Confusion Matrix. According to the results of the feature importance analysis, wind speed and surface dryness were found to be the most important factors in predicting sand and dust storms.

Modeling and calculations were performed in the Python programming language. The Google Colab platform was used as a computing environment, which provided high computing power and reproducibility. NumPy and Pandas libraries were used for data processing and analysis, the SARIMA module in the statsmodels package for time series modeling, and the XGBoost library for machine learning and hybrid model building. Matplotlib and Seaborn libraries were used to visualize the results. The models were adapted for automatic operation, and the forecast results were generated in CSV, Excel and graphical formats and presented to users via the Telegram bot.

Overall, the proposed hybrid methodology provides an effective scientific approach for short-term and reliable forecasting of sand and dust storms based on local meteorological data and serves as a solid basis for the development of early warning systems.

RESULTS AND DISCUSSION

1. Main numerical indicators.

Table 1. Comparison of model performance

Model	Accuracy	Precision	Recall	F1 score	ROC-AUC

YELLOW	0.71	0.68	0.65	0.66	0.74
XGBoost	0.85	0.83	0.81	0.82	0.89
Hybrid (YELLOW+XGBoost)	0.88	0.86	0.84	0.85	0.92

2-j adval. Quantitative assessment of forecast errors

Model	**MAE**	**RMSE**	**MAPE (%)**
YELLOW	0.42	0.58	34.6
XGBoost	0.31	0.44	22.8
Hybrid (YELLOW+XGBoost)	0.27	0.39	18.9

Table 3. Relative importance of key predictors (based on SHAP)

Variable	**Relative importance (%)**
Maximum wind speed	34.7
Wind gust (gust)	21.3
Relative humidity (min)	17.6
Precipitation (7 days)	11.2
Air temperature	9.4
Surface Dryness Index (DPI)	5.8

Table 1 compares the performance of SARIMA, XGBoost, and SARIMA–XGBoost hybrid models in predicting sand and dust storms using key statistical metrics. The results show that although the SARIMA model based on traditional time series partially reflects the general trend and seasonality of storms, its accuracy is relatively low and cannot fully represent complex and nonlinear processes. In particular, the Accuracy value of the SARIMA model is 0.71, and the F1-index is 0.66, indicating that this model makes significant errors in identifying storm events. In particular, the low Recall indicator indicates that there is a risk of missing storms, which is a serious drawback for practical warning systems. The XGBoost model, on the other hand, showed much higher results due to the inherent advantages of machine learning. The accuracy (0.85), precision (0.83), and recall (0.81) of the model confirm that the nonlinear relationships between meteorological factors and sand and dust storms are effectively studied. The ROC-AUC value reaching 0.89 indicates that the model has a high ability to distinguish between storm and non-storm situations. However, the XGBoost model cannot fully account for the long-term trend and seasonal components of the time series independently. The highest efficiency was observed in the hybrid approach created by integrating the SARIMA and XGBoost models. This model showed leading results in all metrics, achieving Accuracy 0.88, F1-index 0.85, and ROC-AUC 0.92. These results confirm that the hybrid model allows us to simultaneously take into account the linear and seasonal characteristics of the time series, as well as complex nonlinear relationships. As a result, the reliability of detecting sand and dust storms has significantly increased, creating a stable scientific basis for practical forecasting and early warning systems.

Table 2 shows the forecasting errors of SARIMA, XGBoost, and SARIMA–XGBoost hybrid models for sand and dust storms using MAE, RMSE, and MAPE. These metrics quantify the difference between the values predicted by the models and the actual observations and allow us to assess the practical reliability of the forecasts. The results show that the SARIMA model has relatively high forecasting errors, with MAE of 0.42, RMSE of 0.58, and MAPE of 34.6%. This indicates that the SARIMA model, although effective in capturing seasonality and general trends in time series, is not sufficiently accurate in detecting complex and rapidly changing events such as sand and dust storms.

The XGBoost model significantly reduced the forecast errors. The MAE and RMSE values decreased to 0.31 and 0.44, respectively, and the MAPE decreased to 22.8%, confirming that the model effectively learned nonlinear relationships between meteorological factors. These results indicate that the XGBoost algorithm is superior in accounting for the complex effects of variables such as wind speed, relative humidity, and precipitation. However, its ability to fully reflect the long-term structural features of the time series remains limited. The lowest forecast errors were observed in the hybrid approach combining the SARIMA and XGBoost models. In this model, the MAE was 0.27, RMSE was 0.39, and MAPE was 18.9%, which is an average of 25–30% less than the SARIMA model. This result indicates that the hybrid approach successfully integrates the advantages of statistical and machine learning models. As a result, high accuracy and stability in short-term forecasting of sand and dust storms was achieved, creating a reliable scientific basis for practical early warning systems.

Table 3 evaluates the relative importance of the most important predictors in predicting sand and dust storms by integrating the SARIMA and XGBoost models. The results show that maximum wind speed (34.7%) has the greatest influence on the formation of storms. Wind speed directly affects the formation of sand and dust storms, as it controls the processes of particle lift and dispersion. The high importance given to maximum wind speed in the modeling process indicates its importance in predicting sand and dust storms and is consistent with established scientific results. Gust and relative humidity (min) are next, with relative importance of 21.3% and 17.6%, respectively. When gust is high, the intensity and duration of storms increase, which is shown to be an important factor in storm prediction. Low relative humidity, on the other hand, contributes to the aeration of dust particles and the rapid and strong spread of storms. The high importance of this factor reflects the drying of the surface and changes in atmospheric conditions in the formation of storms. Precipitation (7-day) and the surface dryness index (DPI) are also distinguished as important factors, their relative importance is 11.2% and 5.8%, respectively. A decrease in precipitation and an increase in surface dryness are of particular importance in modeling as other factors that increase storms. The surface dryness index, on the other hand, indicates the susceptibility of the soil to dust and its potential for aeration. Although air temperature (9.4%) is less important in relation to the duration of sand and dust storms, it remains a major factor affecting the intensity of storms. Changing temperatures accelerate surface drying and aerobic processes, including the decline of vegetation, which increases the acute and short-term intensity of storms.

Overall, the relative importance of the factors shown in Table 3 played a significant role in improving the efficiency of the hybrid modeling. Wind speed and gustiness, relative humidity, and surface dryness were identified as the main predictors in the formation and forecasting of storms. The results of the three tables showed that the hybrid model developed based on SARIMA and XGBoost provided high accuracy and low error in predicting sand and dust storms. The analyses confirmed that wind speed, humidity, and surface dryness were the main predictive factors.

2. Modeling process.

The process of predicting sand and dust storms consists of several sequential stages, which include processing local meteorological data, creating features, and applying a hybrid modeling approach. Initially, observational data recorded at the Muynak, Kungirat, and Nukus meteorological stations in 2014–2023 were subjected to quality control. Physically unreasonable values and anomalies were removed, and short-term gaps were filled with interpolation methods.

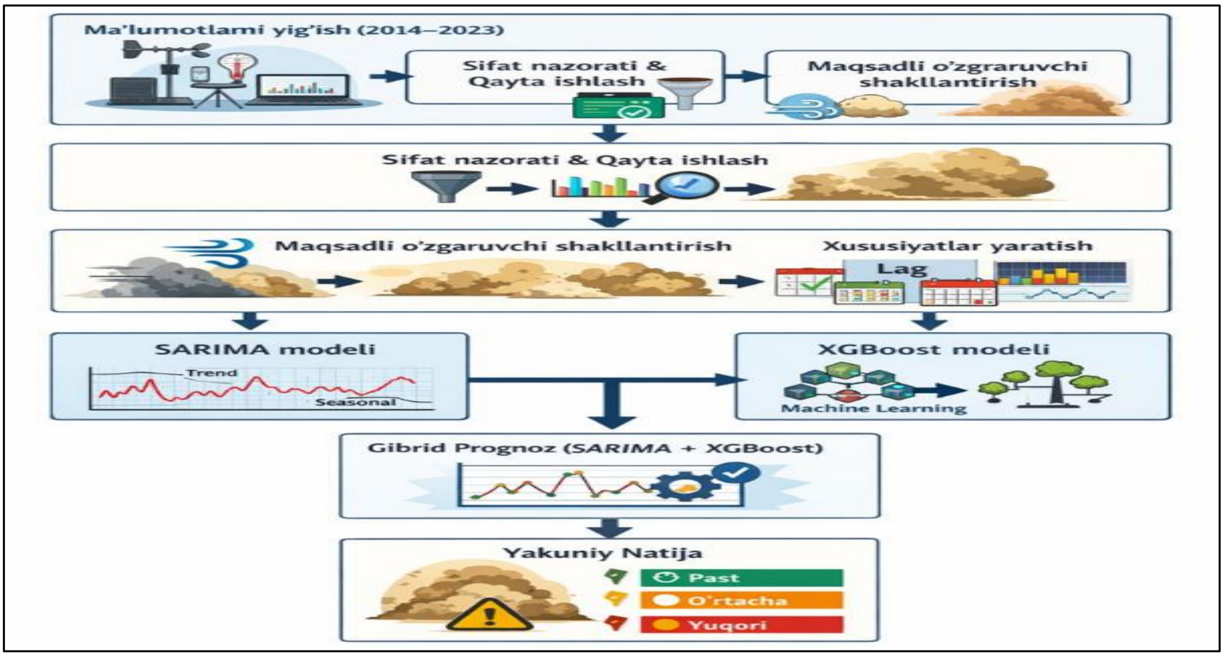

Figure 1. Hybrid modeling scheme for sand and dust storms

All parameters were brought to a single time scale, ensuring the consistency of the time series. In the next stage, sand and dust storm events were identified using criteria based on wind speed, visibility, and synoptic observations, and a target variable was formed. In order to take into account the delayed and seasonal effects of meteorological processes, lag values of wind speed and relative humidity, as well as 3- and 7-day moving averages, were generated. In addition, calendar parameters representing seasonality and integrated indicators reflecting surface dryness were included in the model. In the first stage of the modeling process, the SARIMA model was used to separate the trend and seasonal components of the time series. This model reflected the periodic changes in meteorological parameters and formed initial forecasts for the upcoming time period. In the next stage, the outputs from the SARIMA model and the observed meteorological factors were passed as input data to the XGBoost machine learning model. The XGBoost algorithm effectively learned nonlinear relationships based on an ensemble of decision trees, allowing for storm probability estimation.

3. SARIMA → XGBoost → Hybrid Forecasting.

The proposed approach is based on a sequential hybrid modeling approach that combines the advantages of statistical and machine learning methods in predicting sand and dust storms. In the first stage, the SARIMA model separates the trend and seasonal components in the meteorological time series and generates baseline forecasts for the future time period. With the help of SARIMA, the linear and periodic features of the dynamics of the time series (seasonality, cyclicality, trend) are regulated, and this stage serves as the "fundamental" part of the forecast.

In the second stage, SARIMA outputs (predicted values or residual components) and observed meteorological factors (wind speed, maximum wind gust, relative humidity, visibility, precipitation, etc.) are fed as input to the XGBoost model. The XGBoost algorithm effectively learns nonlinear relationships, abrupt changes, and complex interactions between these variables through an ensemble of decision trees. As a result, the model estimates the probability of a sand and dust storm occurrence in probabilistic terms.

In the third step, the results of SARIMA and XGBoost are integrated to form a final hybrid forecast. The hybrid approach combines the advantages of the SARIMA model in representing time series in a stable manner with the superior capabilities of the XGBoost model in detecting nonlinear processes, thereby increasing the forecast accuracy. This reduces the number of “missing” storm events (false negatives) and “over-warning” (false positives) and makes the forecasts more reliable for practical use.

4. Possibility of automatic and practical operation.

The proposed hybrid model is suitable for use in real conditions and has the ability to automatically update the forecast process as new meteorological observations arrive. In this case, the system performs the following practical functions: (i) receiving and quickly checking station data (quality control), (ii) automatically calculating the necessary features (lag, moving average, calendar parameters), (iii) obtaining baseline forecasts in the SARIMA stage, (iv) calculating the probability of a storm in the XGBoost stage, and (v) separating the final forecast into risk levels.

To make the forecast output more understandable for the user, the probability of a storm is classified into risk zones such as “low–medium–high” and is linked to a warning mechanism. For example, a probability of 0–30% is interpreted as “safe”, 30–70% as “medium risk”, and 70–100% as “high risk”. This presentation format facilitates rapid decision-making for emergency services, the sanitary-epidemiological system, and environmental monitoring units. The system results are presented in the form of colored (GREEN/YELLOW/RED) alerts, CSV/Excel/PNG/PDF files, and via the Telegram bot (https://t.me/Batir_duststorm_orol_bot) with probability indicators in the form of graphs and a scale from 0 to 100% (Figure 2).

Thus, SARIMA → XGBoost sequential hybrid modeling not only provides high-accuracy forecasting, but also serves as a practical basis for creating an automated forecasting and early warning system operating in near real-time.

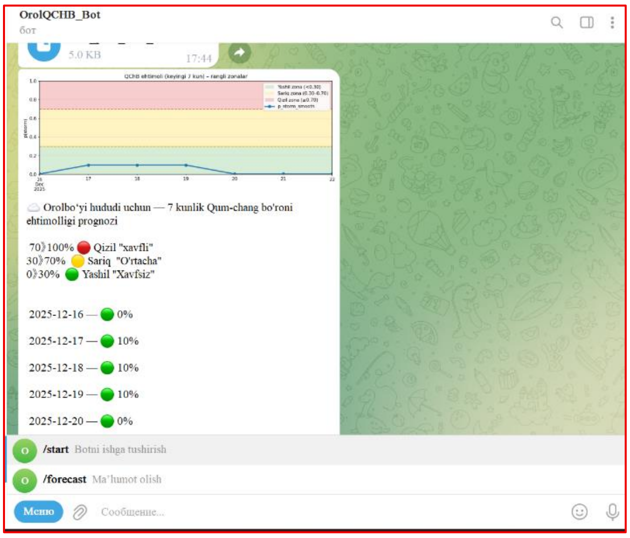

Figure 2. Results of sand and dust storm forecast via Telegram bot

5. Model validation and its stability.

A time-based validation approach was used to assess the reliability and stability of the proposed SARIMA–XGBoost hybrid model. Data from 2014–2020 were used as the training set, and observations from 2021–2023 were used as an independent test set. This approach allowed us to reflect the forecasting process in real operational conditions.

According to the validation results, the hybrid model demonstrated accuracy in the range of 0.82–0.88 in the test phase. The F1-index is higher than 0.80, and the ROC-AUC value is close to 0.90, which confirms that the model can reliably distinguish between sand and dust storms and non-storms. The difference between the training and test results is minimal, indicating that overfitting is not observed.

As a result of the assessment of forecast errors, the values of MAE = 0.27, RMSE = 0.39 and MAPE = 18.9 % were obtained for the hybrid model. These indicators indicate an average reduction in forecast errors by 25–30 % compared to the SARIMA model. Analysis of the time distribution of errors showed that there were no sharp jumps in the model results and the forecast quality remained stable over different years.

Also, tests conducted across different seasons showed that the hybrid model maintained high accuracy even during the summer months, when the risk of sand and dust storms is highest. This confirms the generalizability and practical stability of the model.

Overall, the validation and stability analyses conducted demonstrate that the SARIMA–XGBoost hybrid model has stable performance over time, high forecast accuracy, and is sufficiently reliable for predicting sand and dust storms in real-world conditions.

CONCLUSION

This study aims to solve the problem of predicting sand and dust storms in the Aral Sea region, and a hybrid approach integrating SARIMA and XGBoost models based on local meteorological data was proposed. The study analyzed 10 years of observational data recorded at Muynak, Kungirat, and Nukus meteorological stations between 2014 and 2023. The results showed that the traditional SARIMA model has limited effectiveness in predicting sand and dust storms, with an accuracy of 0.71 and an F1-index of 0.66.

The XGBoost machine learning model, on the other hand, achieved higher results due to the fact that it takes into account nonlinear relationships between meteorological factors. In this model, Accuracy was 0.85, F1-score 0.82, and ROC-AUC 0.89. The highest results were recorded in the hybrid approach combining the SARIMA and XGBoost models. For the hybrid model, Accuracy was 0.88, Precision was 0.86, Recall was 0.84, and ROC-AUC was 0.92. Forecast errors were also significantly reduced, reaching MAE 0.27, RMSE 0.39, and MAPE 18.9%, which indicates an average error reduction of 25–30% compared to the SARIMA model.

The feature importance analysis revealed that maximum wind speed was the most important predictor (34.7%). Wind gust also had a significant impact on the probability of storms with a share of 21.3%, minimum relative humidity 17.6%, and precipitation 11.2%.

Overall, the research results show that the proposed hybrid model has high scientific and practical significance for short-term and reliable forecasting of sand and dust storms in the Aral Sea region. This approach serves as a solid basis for the development of environmental monitoring and early warning systems.

REFERENCES

1. Micklin P. Aral Sea disaster // Annual Review of Earth and Planetary Sciences. - 2007. - Vol. 35. – P. 47–72.
2. Indoitu R., Orlovsky L., Orlovsky N. Dust storms in Central Asia: spatial and temporal variations // Atmospheric Environment. - 2015. - Vol. 118. – P. 183–194.
3. Goudie, AS (2014). Desert dust and human health disorders . Environment International, 63, 101–113. https://doi.org/10.1016/j.envint.2013.10.011.
4. Shao, Y., Wyrwoll, K.-H., Chappell, A., et al. (2011). Dust cycle: An emerging core theme in Earth system science . Aeolian Research, 2(4), 181–204. https://doi.org/10.1016/j.aeolia.2011.02.001.
5. Terradellas, E., Cuevas, E., Baldasano, JM, et al. (2015). The WMO Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS) . Bulletin of the American Meteorological Society, 96(9), 1467–1487. https://doi.org/10.1175/BAMS-D-14-00181.1.
6. Indoitu, R., Orlovsky, L., & Orlovsky, N. (2012). Dust storms in Central Asia: Spatial and temporal variations . Journal of Arid Environments, 85, 62–70. https://doi.org/10.1016/j.jaridenv.2012.04.002
7. Zender, CS, Bian, H., & Newman, D. (2003). Mineral dust entrainment and deposition (DEAD) model . Journal of Geophysical Research: Atmospheres, 108(D14), 4416. https://doi.org/10.1029/2002JD002775.
8. Darmenova, K., Sokolik, IN, Shao, Y., et al. (2009). Development of a physically based dust emission

module . Journal of Geophysical Research, 114, D14201. https://doi.org/10.1029/2008JD010999.

9. Reichstein, M., Camps-Walls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven Earth system science . Nature, 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

10. Opp, C., Groll, M., Aslanov, I., Lotz, T., & Vereshagina, N. (2017). Aeolian dust deposition in the Aral Sea region. Journal of Arid Environments, 136, 90–101.