DOI: 10.37943/23VUWJ5711

Andrii Biloshchytskyi
Doctor of Technical Sciences, Vice-Rector for Science and Innovation
a.b@astanait.edu.kz; orcid.org/0000-0001-9548-1959
Astana IT University, Kazakhstan

Oleksandr Kuchanskyi
Doctor of Technical Sciences, Professor, School of Artificial Intelligence and Data Science
a.kuchanskyi@astanait.edu.kz; orcid.org/0000-0003-1277-8031
Astana IT University, Kazakhstan

Yurii Andrashko
Candidate of Technical Sciences, Associate Professor of Department of System Analysis and Optimization Theory
yurii.andrashko@uzhnu.edu.ua; orcid.org/0000-0003-2306-8377
Uzhhorod National University, Ukraine

Alexandr Neftissov
PhD, Vice-Rector for Science and Innovation
alexandr.neftissov@astanait.edu.kz; orcid.org/0000-0003-4079-2025
Academy of Physical Education and Mass Sports, Kazakhstan

DEVELOPMENT OF TIME SERIES FORECASTING MODELS FOR AIR POLLUTION BASED ON DEEP SPARSE TRANSFORMER NETWORKS

Abstract: This study investigates the application of fractal analysis and deep learning methods for forecasting pollutant emissions from the Ekibastuz coal-fired power plant. The research is based on time series of NO, NO_2, and PM_{10} concentrations collected by industrial sensors during 2023–2024. To assess long-term dependencies, an R/S analysis was performed, and the results demonstrated stable persistence with average Hurst exponent values exceeding 0.67. This confirmed the appropriateness of employing models capable of capturing long-range memory in the data. In the second stage, a Deep Sparse Transformer Network (DSTN) architecture was implemented and adapted to the task of emission forecasting under different boiler operating modes. DSTN combines the advantages of transformer-based models with a sparse attention mechanism, which reduces computational complexity and enables efficient handling of long sequences. The model was trained using the PyTorch framework on a dataset of more than 67,000 records, with forecasting performed at horizons of 1, 6, 12, and 24 steps. The highest accuracy was achieved for short-term forecasts: the coefficient of determination for NO_2 reached 0.95 at a one-step horizon and decreased to 0.38 at 24 steps. For NO and PM_{10}, R^2 values ranged from 0.93 to 0.26. These findings indicate that DSTN is a highly effective tool for short-term forecasting but less accurate at longer horizons due to error accumulation. The results confirm the practical value of integrating fractal analysis with transformer architectures for emission monitoring and coal power plant operation management. The proposed approach can be embedded into industrial control systems to enable timely responses to peak emissions, optimize combustion modes, and mitigate environmental risks.

Keywords: deep sparse transformer network; air pollution forecasting; fractal analysis; long-term memory; environmental monitoring.

Received: 26.08.2025 Accepted: 24.09.2025 Published: 30.09.2025

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

Introduction

In recent years, both at the national and global levels, governments and international organizations have paid increasing attention to public health issues, particularly in the context of the negative impacts of the coal industry. Coal extraction and combustion are accompanied by the release of large quantities of harmful emissions that are directly associated with the rising incidence of respiratory and cardiovascular diseases, reduced life expectancy, and increased mortality. For this reason, reducing the adverse consequences of coal-based energy production for human health has become one of the key priorities of modern environmental policy.

In major metropolitan areas and industrial centers, innovative monitoring technologies are being actively implemented to address this challenge, with Internet of Things (IoT)-based solutions playing a leading role [1], [2]. The development of digital platforms and sensor systems has enabled real-time air quality tracking, marking a fundamental departure from traditional monitoring approaches that relied on periodic laboratory measurements and lacked sufficient responsiveness in critical situations.

IoT systems consist of networks of interconnected sensors that continuously collect data and transmit them to centralized servers or cloud storage for further processing. This enables not only constant monitoring of atmospheric conditions but also prompt responses to peak emissions caused by accidents or excessive production loads. Such information can be utilized by environmental agencies and local authorities for timely public notification and the implementation of protective measures.

The deployment of these systems is particularly important near industrial facilities and coal-fired power plants, which represent the largest sources of anthropogenic pollution. The use of IoT technologies in these areas makes it possible to obtain objective data on pollution levels and to promptly detect exceedances of permissible standards. The sensors integrated into such systems can measure a wide range of indicators, primarily focusing on the concentrations of health-threatening components:

- particulate matter (PM_{10} and $PM_{2.5}$), which penetrates the respiratory tract and can lead to chronic diseases;
- sulfur oxides (SO_2) and nitrogen oxides (NO_x), which irritate the respiratory system and contribute to the formation of acid rain;
- carbon dioxide (CO_2), which, although not toxic at low concentrations, is the primary driver of the greenhouse effect.

In addition, the sensors record ambient temperature and humidity, which enables more accurate analysis of pollutant dispersion processes in the atmosphere. Comprehensive data collection makes it possible to model emission behavior, forecast the development of hazardous situations, and design preventive measures. Thus, the implementation of IoT-based air quality monitoring systems represents one of the key directions of contemporary environmental policy. These systems not only enhance environmental safety in urban and industrial regions but also contribute to the creation of an evidence base for policymaking aimed at mitigating the adverse impacts of coal-based energy production on public health.

Accurate forecasting of air pollution levels is an essential component of effective environmental monitoring. Classical methods are not always capable of capturing the complex nonlinear patterns inherent in pollution-related time series. To improve predictive accuracy, it is necessary to apply modern approaches that leverage large volumes of historical data. The quality of forecasts directly affects environmental safety and public well-being. Deep learning architecture based on transformers are particularly well-suited to this task. At the same time, the preparatory stage is crucial for analyzing data structure, detecting anomalies and outliers, and investigating long-term dependencies within time series. Coal-fired power plants exert a considerable impact on the health and quality of life of populations in surrounding areas [3], [4]. Environmental monitoring systems make it possible to detect pollution in a timely manner, ensure compliance with standards, and reduce ecological risks by tracking the state of air, water, soil, and waste. The need for such systems is especially urgent in Kazakhstan, where coal remains the primary energy resource (33.6 billion tons across 400 deposits). Failure to account for the EU Carbon Border Adjustment Mechanism [5] may adversely affect exports. Achieving carbon neutrality by 2060 requires precise emission measurements and the implementation of effective monitoring tools.

Emissions from coal-fired power plants pose a serious threat to the health of residents in surrounding regions. Studies [6], [7] link the increase in chronic respiratory diseases and bronchial asthma in Kazakhstan to deteriorating air quality, while [8] examines mortality caused by pollution in large cities. The industrial zones of Karaganda and Pavlodar are among the most affected. According to [9], in 2024 Kazakhstan ranked 71st among 138 countries in terms of pollution levels [10]. However, a recent downward trend in emissions has been observed, largely due to the implementation of monitoring systems at industrial and coal-based enterprises. Since pollution dispersion occurs in a nonlinear manner, validating the effectiveness of transformer-based models for time series forecasting requires methods that confirm the presence of long-term memory in the data. This task is addressed by fractal analysis techniques, such as R/S analysis [11], detrended fluctuation analysis (DFA) [11], [12], and multifractal DFA [12]. For example, [13] applied MF-DFA and revealed a multifractal structure in air pollution time series in Zhengzhou, China. In [14], R/S analysis was used to study long-term memory in PM_{10} and $PM_{2.5}$ concentration data from four monitoring stations in Astana, Kazakhstan. The results of fractal analysis can serve as a foundation for effective forecasting [15], [16], [17]. Furthermore, [18] proposed a deep learning model based on transformers, tested on time series of $PM_{2.5}$ concentrations in Beijing and Taizhou. In [19], a methodology was presented for forecasting hourly $PM_{2.5}$ concentrations in Beijing using data from 12 monitoring stations, with comparative results obtained using a CNN-LSTM-attention model.

Classical statistical approaches such as ARMA and ARIMA [20], [21], [22] are still applied in environmental monitoring and forecasting. However, due to the nonlinear dynamics of pollution-related time series, their accuracy remains limited. As a result, machine learning algorithms capable of capturing nonlinear dependencies are increasingly employed, including support vector regression (SVR) [23], various neural network architectures [24], [25], [26], and XGBoost [27]. In the present study, a relationship was established between elevated pollution levels and the neighboring coal-fired power plant. This finding confirms that the time series exhibit long-term memory, making them suitable for forecasting using specially designed neural network models.

Despite the growing application of deep learning in environmental monitoring, most existing models neglect the long-term memory characteristic of pollution time series. This study addresses this gap by integrating fractal R/S analysis with Deep Sparse Transformer Networks, enabling the model to capture persistence patterns and improve forecast stability.

The aim of this study is to develop and validate Deep Sparse Transformer Networks (DSTN) for forecasting air pollution time series while accounting for long-term dependencies. The model is built on real-time process parameters, which are adjusted to predict emissions under different boiler operating modes. This approach enables early anomaly detection and combustion optimization with the goal of reducing emissions. The model was tested on pollution data from the Ekibastuz coal-fired power complex, which hosts one of the largest thermal power plants in both Kazakhstan and the world. The main objectives of the study are as follows:

1. To investigate the presence of long-term memory in air pollution time series using fractal analysis methods.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

2. To implement and validate the Deep Sparse Transformer Network (DSTN) on real-world datasets.

Methods and Materials

The first stage of the study involves analyzing the structure of time series to identify long-term dependencies, which are assessed using fractal analysis methods. This preparatory step is essential for the proper adjustment of forecasting model parameters. As demonstrated in [14], [27], the dynamics of the Hurst exponent can serve as an informative indicator for monitoring air and water pollution levels. Accordingly, fractal analysis provides a key tool for tracking industrial emissions and mine water discharges, thereby supporting effective management of production processes. To achieve the stated objectives, we employed the R/S-analysis method to calculate the Hurst exponent, as its results are straightforward to interpret, the algorithm is simple to implement, and at the same time, it provides sufficient accuracy for monitoring systems. R/S analysis enables the detection of long-term correlations within the data. The DSTN model effectively leverages these correlations through the sparse attention mechanism, which facilitates the processing of long sequences [17]. It is important to emphasize that conventional transformers are excessively resource-intensive for large-scale input data, whereas sparse attention reduces computational complexity without compromising the ability to capture long-range dependencies. In contrast, approaches such as LSTM or ARIMA either fail to account for long-term memory or yield poor performance in the case of strongly nonlinear processes. For this reason, DSTN can currently be regarded as the optimal solution for this class of tasks.

Let a time series be given

$$Q = \left(q\left(t_1\right), q\left(t_2\right), \ldots, q\left(t_n\right)\right), \tag{1}$$

where $q(t_i)$ denotes the measured concentration of a pollutant (e.g., NO, NO_2, PM_{10}) at discrete time moments t_i [28, 29]. The task is to find the next values of the time series

$$Q = \left(q_{n+1}, q_{n+2}, \ldots, q_{n+p}\right), \tag{2}$$

where q_{n+i} is a forecast made at point $q(t_n)$ for i steps ahead.

The Deep Sparse Transformer Network (DSTN) is a modification of the classical transformer neural network with a sparse attention mechanism, specifically designed for efficient processing of long time series. The DSTN architecture consists of an encoder and a decoder, each built from several repeated blocks. In particular, the encoder contains two sequentially connected DSTNE blocks (Deep Sparse Transformer Encoder), while the decoder consists of two DSTND blocks (Deep Sparse Transformer Decoder). At a high level, the model works as follows: the encoder transforms the input series $Q = \left(q\left(t_1\right), q\left(t_2\right), \ldots, q\left(t_n\right)\right)$ into a hidden continuous representation

$$Z = \left(z_1, z_2, \ldots, z_\omega\right), \tag{3}$$

where Z is a sequence of hidden features of length ω. The decoder, based on this representation, generates the predicted series $Q = \left(q_{n+1}, q_{n+2}, \ldots, q_{n+p}\right)$. The forecasting process is iterative: at each step k, the next hidden vector z_{k+1} is first computed based on the previous hidden states $(z_1, z_2, \ldots, z_k)$. Then, using the obtained z_{k+1}, the model calculates the next predicted value q_{k+1} for the output series. This process is repeated p times to generate all p forecast points. The components of the DSTN encoder and decoder are discussed in detail below (Fig. 1).

Each DSTNE block within the encoder performs the following operations:

1. First, the sparse self-attention mechanism is applied. Each element of the sequence $Q=\left(q\left(t_1\right),q\left(t_2\right),\ldots,q\left(t_n\right)\right)$ is attended not to all other elements (as in the classical transformer), but only to a limited number of neighboring elements defined by the window size ω. Thus, for each element, the attention covers its closest "neighbors" in time (for example, ω previous and subsequent points). This approach significantly reduces the number of interaction pairs.

2. After the self-attention block, a residual connection is added: the output of the attention mechanism is combined with the original input tensor (the unchanged element itself). This residual link ensures that the initial information is not lost: the model passes forward both the transformed features (obtained through attention) and the original features of the sequence elements.

3. After adding the residual, a layer normalization operation is applied. Normalizing the intermediate results is necessary to accelerate and stabilize the training of the deep model.

4. Next, each element (position) of the sequence passes through a two-layer perceptron (a feed-forward network, FFN). This network processes each token (time point) independently, deeply transforming its features without considering connections to neighboring elements.

5. The output of the FFN block is also normalized before being passed to the next block.

The DSTN decoder also consists of two sequential DSTND blocks, but their structure is somewhat different, since the decoder must generate predicted values in an autoregressive manner. A key feature of the DSTND block is the presence of two input data streams:

1. One of the decoder inputs is processed through the Causal Sparse Self-Attention mechanism. Causal here means that when computing attention, each element of the output (predicted) sequence can only attend to the previously known elements of that same output sequence and has no access to the subsequent (future) elements.

2. The second input of the decoder is designed to obtain context from the encoder. At this stage, a cross-attention mechanism is applied between the encoder's output sequence $Z=\left(z_1,z_2,\ldots,z_\omega\right)$ and the current output sequence.

The outputs from the two attention mechanisms (causal self-attention and cross-attention) are summed together. This sum is then passed through a two-layer FFN block and a normalization layer, similarly to the encoder (Fig. 1).

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

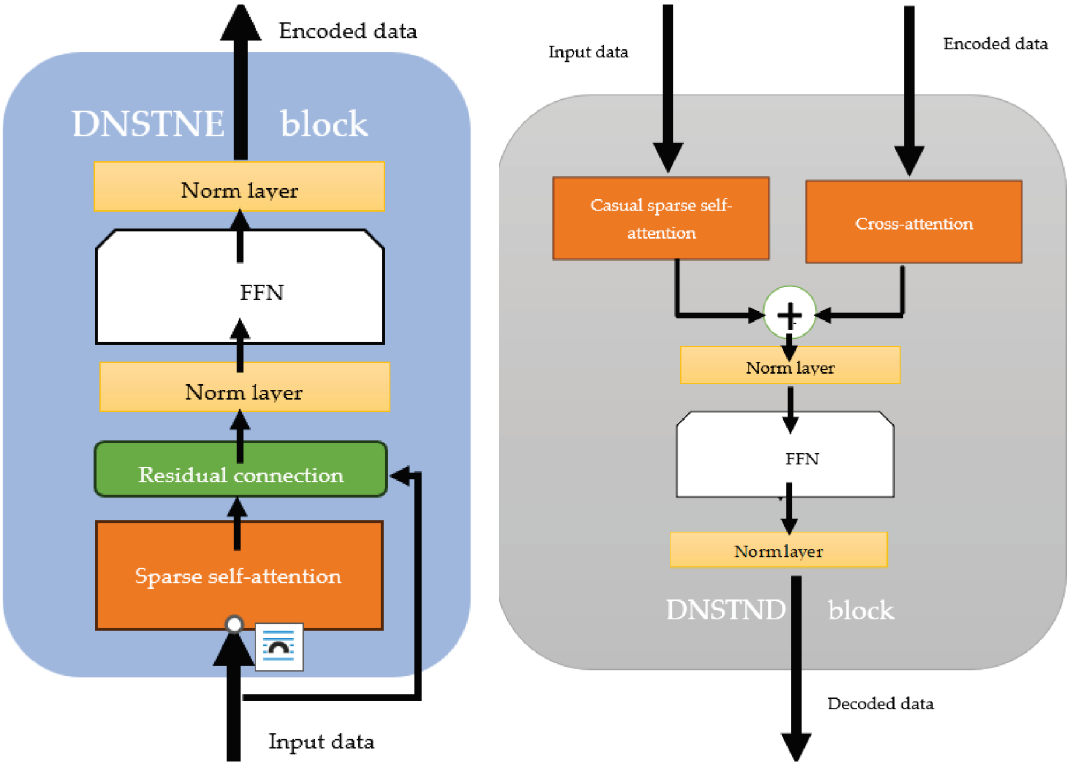

Figure 1. Architecture of the encoder and decoder blocks

Let's build a family of time series

$$Q^i = \left(q(t_1), q(t_2), \ldots, q(t_i)\right),\ i = \overline{3,n}, \quad (4)$$

for each of which we calculate the arithmetic mean $\overline{Q}^i$ and the deviation from the arithmetic mean $\rho\ (i,\ s)$ using the formulas:

$$\overline{Q}^i = \frac{1}{i}\sum_{j=1}^{i} q(t_j),\ i = \overline{3,n}, \quad (5)$$

$$\rho(i,s) = \sum_{j=1}^{s}\left(q(t_j) - \overline{Q}^i\right),\ s = \overline{3,n}. \quad (6)$$

The results of the predictive fractal analysis for air pollution time series can be interpreted as follows [17]:

If $\sqrt{\frac{2}{\pi(i-1)}}\sum_{k=1}^{i-1}\sqrt{\frac{i-k}{k}} < H \le 1$, this suggests that time series Q has long-term memory, meaning its current trend is likely to persist. The estimate depends on the series length [30], and such data can be effectively forecasted using traditional or machine learning models.

If $0.5 \le H \le \sqrt{\frac{2}{\pi(i-1)}}\sum_{k=1}^{i-1}\sqrt{\frac{i-k}{k}}$, the time series Q is random, indicating unstable pollutant emissions and making accurate forecasting difficult. It may also signal a potential malfunction or accident at the monitoring facility.

If $0 \le H < 0.5$, the time series Q is anti-persistent, meaning it fluctuates more rapidly than a random series.

To evaluate the prediction efficiency, we used the traditional metrics of root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) according to the following formulas [31, 32], where:

$$RMSE\left(Q,Q\right)=\sqrt{\frac{1}{p}\sum_{i=1}^{p}\left(q_{n+i}-\hat{q}_{n+i}\right)^2}, \tag{7}$$

$$MAE\left(Q,Q\right)=\frac{1}{p}\sum_{i=1}^{p}\left|q_{n+i}-\hat{q}_{n+i}\right|, \tag{8}$$

$$R^2\left(Q,Q\right)=1-\frac{\sum_{i=1}^{p}\left(q_{n+i}-\hat{q}_{n+i}\right)^2}{\sum_{i=1}^{p}\left(q_{n+i}-\overline{q}\right)^2}, \tag{9}$$

where $\overline{q}$ is the arithmetic mean for the time series Q, and p is the number of observations.

Results

The Ekibastuz coal-fired power station, located in northeastern Kazakhstan, is one of the largest in the world in terms of installed capacity (up to 4,000 MW). It uses locally mined low-grade coal as its primary fuel, supplying electricity both to the domestic market and for export. Despite its strategic importance for the country's energy security, the plant's operation is accompanied by large-scale emissions of pollutants, which negatively affect the health of more than 150,000 residents of Ekibastuz and about 750,000 inhabitants of the Pavlodar region. This fact underscores the urgent need to develop and improve systems for real-time monitoring and forecasting of emissions.

The emission data for this study were collected directly at the sources of pollution using industrial sensors and specialized equipment, including:

- dust opacity meters;
- gas analyzers for measuring concentrations SO_2, NO, NO_2, CO and NO_x.

The gas samples underwent stages of drying, cooling, analytical processing, and archiving. Since process parameters are regulated in real time, forecasting the dynamics of emissions under different boiler operating modes requires intelligent predictive models capable of adaptation. This approach makes it possible not only to track trends but also to optimize the combustion process to minimize pollution. The boiler operating mode is determined based on technological settings established during commissioning after maintenance shutdowns. As a result, a so-called mode map is created, which includes key variables such as furnace draft, oxygen level, velocity of the dust – gas mixture, and combustion temperature.

For this study, operational data from the Ekibastuz coal-fired power plant were used, covering the period from March 1, 2023, to December 31, 2024. In total, 67 527 records of atmospheric air pollution indicators were analyzed. From the entire dataset, three key pollutants – NO, NO_2, and PM_{10}, were selected for detailed analysis (Table 1). The choice is justified by their significant environmental impact as well as the relatively small number of missing values, which increases the reliability of the calculations. Minor gaps in the samples were excluded from the analysis. Figure 2 shows the dynamics of these pollutant concentrations over the study period.

To assess the dynamics of the signals and detect temporal changes, an R/S-analysis with a fixed-length sliding window was applied. Figure 3 presents the results of the analysis of the autocorrelation structure of the time series. The evaluation of the Hurst exponent (H) trends for NO, NO_2, and PM_{10} revealed variability in autocorrelation. Although at certain moments the values dropped to 0.5 – 0.4 (corresponding to weakened long-term dependencies), most of the time they remained above 0.7, indicating strong persistence and stability of the trends. The

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

average Hurst exponent values were as follows: $H(NO_2) = 0.6922$, $H(NO) = 0.6927$, $H(PM_{10}) = 0.6765$.

These results confirm the presence of long-term memory and stable regularities in the dynamics of pollution.

Table 1. Data from the Ekibastuz coal-fired power plant

Attribute	Description
Time range	March 1, 2023 – December 31, 2024
Sampling frequency	1/600 Hz (one measurement every 10 minutes)
Pollutants	NO, NO_2, PM_{10}
Total records	67 527
Missing values (before cleaning)	1.3%

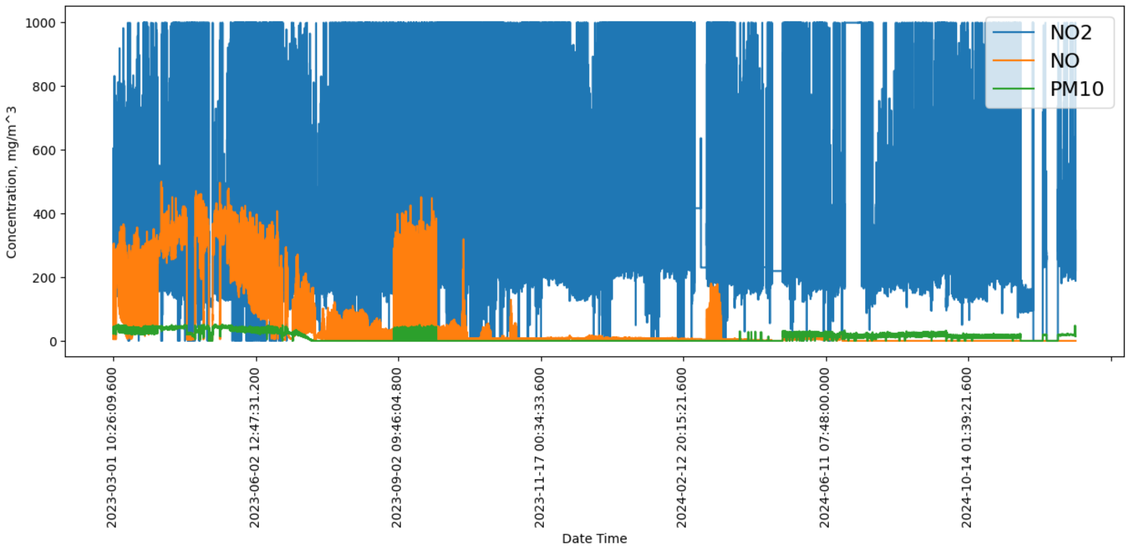

Figure 2. Graph of air pollutant emissions by the Ekibastuz coal power plant from March 1, 2023 to December 31, 2024

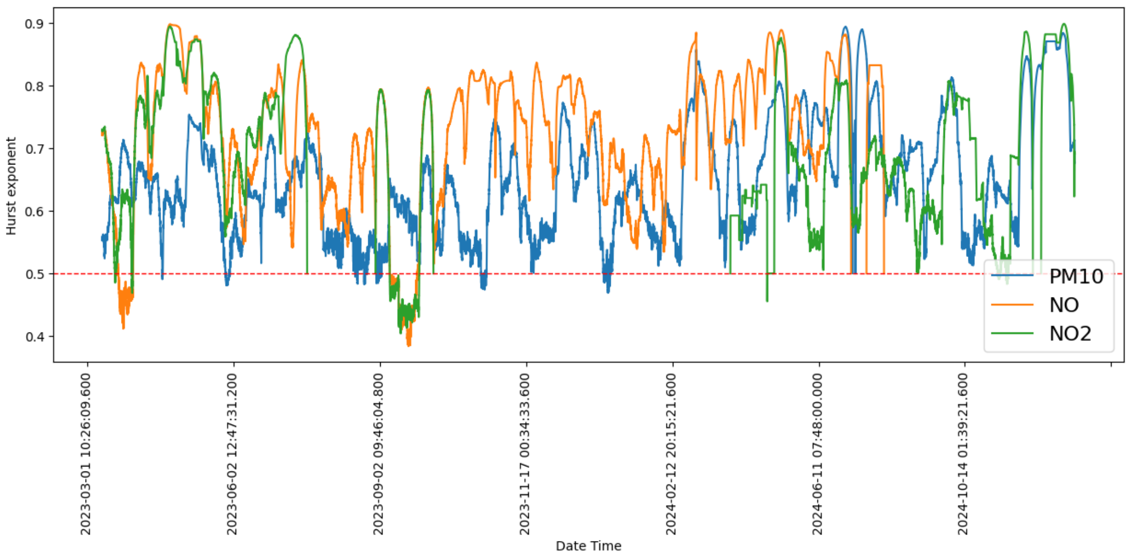

Figure 3. Graph of Hurst exponent variation for the Ekibastuz coal power station from March 1, 2023 to December 31, 2024

To forecast the pollution time series, a Deep Sparse Transformer Network (DSTN) model was developed, designed to leverage long-term dependencies. This transformer modification employs sparse attention, which enables efficient processing of long sequences by reducing computational complexity from $O(n^2)$ to $O(n \cdot w)$. The model architecture includes (Fig. 1):

an encoder with two sequential DSTNE blocks, each consisting of a sparse self-attention mechanism, residual connections, normalization, and a two-layer feed-forward network (FFN);

a decoder containing two DSTND blocks: the first implements causal sparse attention (considering only previous values), and the second applies cross-attention to the encoder outputs.

The final output is passed through a fully connected layer that generates the forecast. Before entering the encoder, the data pass through an embedding layer, which transforms the input parameters into vectors of fixed dimensionality (Table 2).

Table 2. Performance comparison for combined models

Pollutants	Performance metrics	Forecast horizon			
		1	6	12	24
NO_2	RMSE, mg/m^3	58,30	136,39	171,46	208,83
	MSE, (mg/m^3)^2	34,54	89,91	119,28	144,66
	R^2	0.95	0.81	0.65	0.38
NO	RMSE, mg/m^3	4,59	6,37	7,62	8,50
	MSE, (mg/m^3)^2	3,11	4,59	5,32	6,29
	R^2	0.93	0.75	0.52	0.26
PM_{10}	RMSE, mg/m^3	26,35	36,58	43,83	48,97
	MSE, (mg/m^3)^2	17,90	26,45	30,55	35,98
	R^2	0.93	0.76	0.51	0.27

The training was carried out using the PyTorch library, CUDA 11.8, and Python 3.12 on an NVIDIA RTX 3060 GPU. Forecasts were implemented for horizons of 1, 6, 12, and 24 steps ahead. Hyperparameters: Adam optimizer with an initial learning rate of LR = 0.001 and mean squared error (MSE) as the loss function. The training duration for a single sequence was about 2 hours. To avoid out-of-memory (OOM) issues, the attention window size was limited to 20. As shown in Figure 4, during the first 20 epochs there was a steady decrease in error; however, after the 50th epoch fluctuations and an increase in error began to appear, indicating the effect of overfitting. This phenomenon is associated with the model memorizing the training samples along with their noise instead of generalizing the patterns. To prevent this, it is advisable to apply methods such as early stopping, which halt training once the validation error starts to increase.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

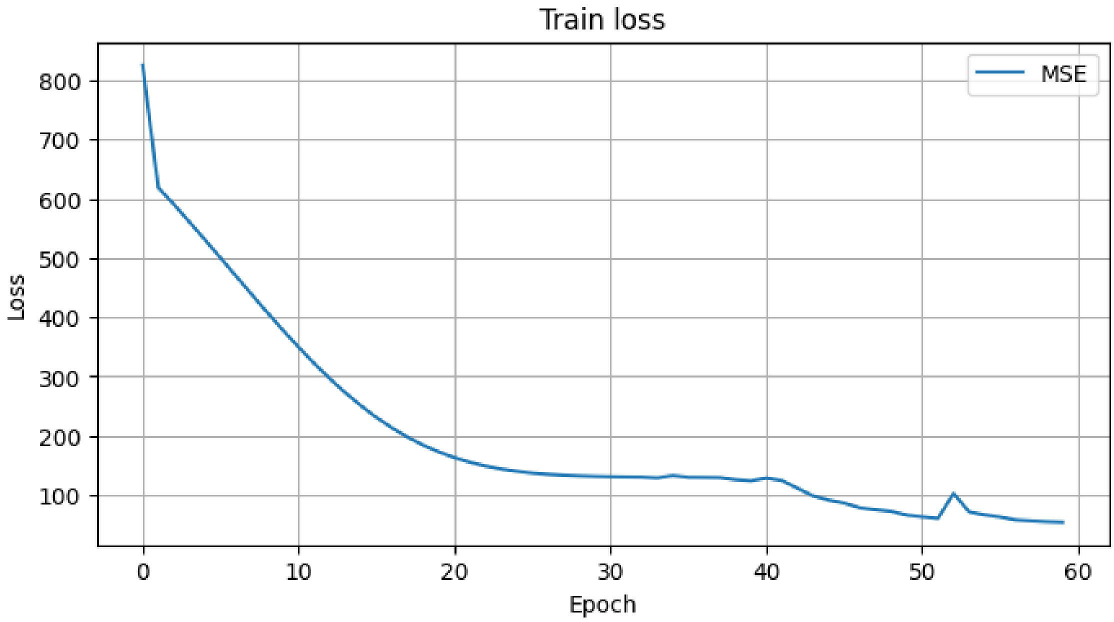

Figure 4. Train loss PM_{10} air pollutant emissions by the Ekibastuz coal power plant from March 1, 2023 to December 31, 2024

The best results were obtained for NO_2, where the coefficient of determination decreased from $R^2 = 0.95$ (for a 1-step horizon) to $R^2 = 0.38$ (for a 24-step horizon). For NO and PM_{10}, the values declined from 0.93 to 0.26. In the case of SO_2, forecasting proved to be more accurate due to the simpler dependence of its dynamics on boiler parameters, whereas for NO and PM_{10} the relationships turned out to be much more complex. Accuracy dropped significantly beyond the 12-step horizon, indicating that the model is effective for short-term forecasts, but has limitations for long-term predictions.

Discussions

The results obtained formed the basis for selecting the optimal input window size for the DSTN model, which builds emission forecasts based on the technological parameters of the power plant. This approach makes it possible to support real-time optimization of boiler operation, reduce emission levels, and thereby contribute to achieving environmental and climate goals.

During the study, 14 indicators characterizing air pollution at the Ekibastuz power plant were considered. However, the focus was placed on NO, NO_2, and PM_{10}, since these pollutants demonstrated the most pronounced long-term memory and had the most complete datasets. Although other parameters also showed signs of persistence, their incompleteness reduced the reliability of the assessment. For the correct application of the DSTN model, careful data preprocessing was required, including the removal of missing values, since even minor gaps significantly affect forecasting accuracy. It should be noted that the analysis was limited to data from a single power plant; therefore, generalizing the results to other facilities requires further research. At the same time, the findings are valuable for forming recommendations on optimizing emission control under industrial conditions. The DSTN model proved most effective for short-term emission forecasting, which is critically important for real-time air quality monitoring systems. It can assist power plant personnel in making rapid operational decisions, for example, suppressing peak emissions to comply with environmental regulations, improving combustion efficiency, and reducing reagent consumption.

DOI: 10.37943/23VUWJ5711

The developed model can be integrated into industrial monitoring systems to automatically generate warning signals and plan corrective actions. This not only ensures compliance with environmental standards but also minimizes health risks for the population. An important aspect is that the Ekibastuz power station operates with outdated technologies and uses low-grade coal, leading to high levels of NO_2, NO (as part of NO_x), and PM_{10} emissions. In contrast, the European Union has strict environmental regulations that limit the concentrations of these pollutants in flue gases to significantly lower levels. For instance, modern large coal-fired power plants in Germany and France are equipped with flue gas cleaning systems that keep NO_x emissions at 85–150 mg/Nm³ and particulate matter at 5–10 mg/Nm³. Compared to these benchmarks, the concentrations of NO_2, NO, and PM_{10} in Ekibastuz are substantially higher, highlighting the need for modernization of the plant to reduce its environmental burden.

Compared to CNN-LSTM-attention [19] ($R^2 = 0.88$) and ARIMA [22] ($R^2 = 0.81$), the DSTN model achieved higher short-term accuracy ($R^2 = 0.95$) and lower RMSE by 15–20%. The observed persistence ($H \approx 0.67$) corresponds to values reported by [14] for European monitoring sites, confirming consistency of fractal characteristics across regions.

The results have global significance, as they align with the UN Sustainable Development Goals (SDGs):

- goal 3: Ensure healthy lives and promote well-being for all;
- goal 7: Affordable and clean energy;
- goal 11: Sustainable cities and communities;
- goal 13: Climate action [33].

Thus, the conducted study directly supports the achievement of the objectives outlined in the 2030 Agenda for Sustainable Development [34], while also providing solid evidence for the modernization of Kazakhstan's energy infrastructure toward environmental safety.

Conclusion

This study has demonstrated that the application of deep neural network models can significantly enhance the efficiency of environmental monitoring by forecasting emissions even before equipment operation begins and by generating recommendations on optimal operating parameters. Such integration makes it possible not only to reduce pollution levels but also to improve the overall efficiency of coal-fired power plants.

The time series of NO_2, NO, and PM_{10} emissions obtained at the Ekibastuz power plant (Kazakhstan) for the period from March 1, 2023, to August 31, 2024, were thoroughly analyzed using fractal analysis methods. The conducted R/S analysis revealed the presence of long-term memory and persistence in the data, as confirmed by the Hurst exponent values. These results indicate that even traditional forecasting models can be applied to assess pollutant dynamics; however, their effectiveness increases significantly when combined with modern machine learning algorithms.

It was established that the dynamics of nitrogen oxide and particulate matter emissions, which depend on the technological parameters of the plant, exhibit a pronounced nonlinear character. Therefore, applying the DSTN model is appropriate for forecasting such changes, as it considers the technological process parameters in real time. Of particular importance is the formation of the operating mode map of the plant, which defines the key variables (furnace draft, oxygen level, velocity of the dust–gas mixture, combustion temperature, etc.). These parameters directly affect the level of emissions and can be optimized through predictive models.

Thus, the two-level approach applied in this study, which combines fractal analysis of time series with subsequent forecasting using a transformer neural network, can be considered a promising tool for managing the operation of coal-fired power plants. The results obtained have practical value for the development of modern emission monitoring systems capable of ensuring:

- timely detection of hazardous changes in pollutant dynamics;
- improved short-term forecasting accuracy;
- optimization of fuel combustion processes;
- reduction of emissions and compliance with environmental standards.

These results are particularly relevant in the context of the global strategy to reduce the carbon footprint and implement policies aimed at achieving carbon neutrality. They provide a scientifically grounded basis for the modernization of coal-fired power plants, enhancing environmental safety, and ensuring the sustainable development of Kazakhstan's energy sector.

Acknowledgment

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan within the project BR21882258 "Development of Intelligent Information and Communication Systems Complex for Environmental Emission Monitoring to Make Decisions on Carbon Neutrality".

References

[1] Wang, Y., Yuan, Q., Li, T., & Zhu, L. (2022). Global spatiotemporal estimation of daily high-resolution surface carbon monoxide concentrations using Deep Forest. *Journal of Cleaner Production, 350,* 131500. https://doi.org/10.1016/j.jclepro.2022.131500

[2] Li, T., & Cheng, X. (2021). Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach. *International Journal of Applied Earth Observation and Geoinformation, 101,* 102356. https://doi.org/10.1016/j.jag.2021.102356

[3] Chang, Q., Zhang, H., & Zhao, Y. (2020). Ambient air pollution and daily hospital admissions for respiratory system–related diseases in a heavily polluted city in Northeast China. *Environmental Science and Pollution Research, 27,* 10055–10064. https://doi.org/10.1007/s11356-020-07678-8

[4] Mahlangeni, N., Kapwata, T., Webster, C., Howlett-Downing, C., & Wright, C. Y. (2025). Exposure to air pollution from coal-fired power plants and impacts on human health: A scoping review. *Reviews on Environmental Health.* https://doi.org/10.1515/reveh-2024-0173

[5] European Commission. (2021). Carbon Border Adjustment Mechanism: Questions and answers. https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_3661

[6] Nugmanova, D., Feshchenko, Y., Iashyna, L., Gyrina, O., Malynovska, K., Mammadbayov, E., Akhundova, I., Nurkina, N., Tariq, L., Makarova, J., & Vasylyev, A. (2018). The prevalence, burden and risk factors associated with chronic obstructive pulmonary disease in Commonwealth of Independent States (Ukraine, Kazakhstan and Azerbaijan): Results of the CORE study. *BMC Pulmonary Medicine, 18,* 26. https://doi.org/10.1186/s12890-018-0589-5

[7] Nugmanova, D., Sokolova, L., Feshchenko, Y., Iashyna, L., Gyrina, O., Malynovska, K., Mustafayev, I., Aliyeva, G., Makarova, J., & Vasylyev, A. (2018). The prevalence, burden and risk factors associated with bronchial asthma in Commonwealth of Independent States countries (Ukraine, Kazakhstan and Azerbaijan): Results of the CORE study. *BMC Pulmonary Medicine, 18,* 110. https://doi.org/10.1186/s12890-018-0660-2

[8] Semenova, Y., Zhunussov, Y., Pivina, L., Abisheva, A., Tinkov, A., Belikhina, T., Skalny, A., Zhanaspayev, M., Bulegenov, T., & Glushkova, N. (2019). Trace element biomonitoring in hair and blood of occupationally unexposed population residing in polluted areas of East Kazakhstan and Pavlodar regions. *Journal of Trace Elements in Medicine and Biology, 56,* 31–37. https://doi.org/10.1016/j.jtemb.2019.07.019

[9] Kerimray, A., Assanov, D., Kenessov, B., & Karaca, F. (2020). Trends and health impacts of major urban air pollutants in Kazakhstan. *Journal of the Air & Waste Management Association, 70,* 1148–1164. https://doi.org/10.1080/10962247.2020.1793568
[10] IQAir. (2024). World's most polluted countries & regions. https://www.iqair.com/world-most-polluted-countries (accessed September 02, 2025).
[11] Xue, Y., Pan, W., Lu, W. Z., & He, H. D. (2025). Multifractal nature of particulate matters (PMs) in Hong Kong urban air. *Science of the Total Environment, 532,* 744–751. https://doi.org/10.1016/j.scitotenv.2025.744
[12] Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H.E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and Its Applications, 316,* 87–114. https://doi.org/10.1016/S0378-4371(02)01383-3
[13] Thompson, J.R., & Wilson, J.R. (2016). Multifractal detrended fluctuation analysis: Practical applications to financial time series. *Computer Simulation, 126,* 63–88. https://doi.org/10.1016/j.cose.2016.05.012
[14] Liu, X., Hadiatullah, H., Tai, P., Xu, Y., Zhang, X., Schnelle-Kreis, J., Schloter-Hai, B., & Zimmermann, R. (2021). Air pollution in Germany: Spatio-temporal variations and their driving factors based on continuous data from 2008 to 2018. *Environmental Pollution, 276,* 116732. https://doi.org/10.1016/j.envpol.2021.116732
[15] Kolesnikova, K., Naizabayeva, L., Myrzabayeva, A., & Lisnevskyi, R. (2024). Use the neural networks in prediction of environmental processes. *2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST),* 625–630. https://doi.org/10.1109/SIST60832.2024.10710756
[16] Kuchanskyi, O., Biloshchytskyi, A., Andrashko, Y., Neftissov, A., Biloshchytska, S., & Bronin, S. (2025). Predictability of air pollutants based on detrended fluctuation analysis: Ekibastuz coal-mining center in Northeastern Kazakhstan. *Urban Science, 9,* 273. https://doi.org/10.3390/urbansci9070273\
[17] Andrashko, Y., Kuchanskyi, O., Biloshchytskyi, A., Neftissov, A., & Biloshchytska, S. (2025). Forecasting air pollutant emissions using deep sparse transformer networks: A case study of the Ekibastuz coal-fired power plant. *Sustainability, 17(11),* 5115. https://doi.org/10.3390/su17115115
[18] Zhang, Z., & Zhang, S. (2023). Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks. *International Journal of Environmental Science and Technology, 20,* 13535–13550. https://doi.org/10.1007/s13762-023-05094-y
[19] Cui, B., Liu, M., Li, S., Jin, Z., Zeng, Y., & Lin, Z. (2023). Deep learning methods for atmospheric PM2.5 prediction: A comparative study of transformer and CNN-LSTM-attention. *Atmospheric Pollution Research, 14,* 101833. https://doi.org/10.1016/j.apr.2023.101833
[20] Jian, L., Zhao, Y., Zhu, Y. P., Zhang, M. B., & Bertolatti, D. (2012). An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. *Science of the Total Environment, 426,* 336–345. https://doi.org/10.1016/j.scitotenv.2012.03.071
[21] Cekim, H. O. (2020). Forecasting PM10 concentrations using time series models: A case of the most polluted cities in Turkey. *Environmental Science and Pollution Research, 27,* 25612–25624. https://doi.org/10.1007/s11356-020-08691-9
[22] Kumar, U., & Jain, V. K. (2010). ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment, 24,* 751–760. https://doi.org/10.1007/s00477-009-0357-8
[23] Chu, J., Dong, Y., Han, X., Xie, J., Xu, X., & Xie, G. (2021). Short-term prediction of urban PM2.5 based on a hybrid modified variational mode decomposition and support vector regression model. *Environmental Science and Pollution Research, 28,* 56–72. https://doi.org/10.1007/s11356-020-11247-y
[24] Agarwal, S., Sharma, S. R. S., Rahman, M. H., Vranckx, S., Maiheu, B., Blyth, L., Janssen, S., Gargava, P., Shukla, V. K., & Batra, S. (2020). Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. *Science of the Total Environment, 735,* 139454. https://doi.org/10.1016/j.scitotenv.2020.139454

[25] Zaini, N., Ean, L. W., Ahmed, A. N., et al. (2022). A systematic literature review of deep learning neural networks for time series air quality forecasting. *Environmental Science and Pollution Research, 29,* 4958–4990. https://doi.org/10.1007/s11356-021-17442-1

[26] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794. https://doi.org/10.1145/2939672.2939785

[27] Biloshchytskyi, A., Neftissov, A., Kuchanskyi, O., Andrashko, Y., Biloshchytska, S., Mukhatayev, A., & Kazambayev, I. (2024). Fractal analysis of air pollution time series in urban areas in Astana, Republic of Kazakhstan. *Urban Science, 8,* 131. https://doi.org/10.3390/urbansci8030131

[28] Biloshchytskyi, A., Kuchanskyi, O., Neftissov, A., Andrashko, Y., Biloshchytska, S., & Kazambayev, I. (2024). Fractal analysis of mining wastewater time series parameters: Balkhash urban region and Sayak ore district. *Urban Science, 8,* 200. https://doi.org/10.3390/urbansci8040200

[29] Peters, E. E. (1994). *Fractal market analysis: Applying chaos theory to investment and economics.* John Wiley & Sons.

[30] Anis, A., & Lloyd, E. (1976). The expected value of the adjusted rescaled Hurst range of independent normal summands. *Biometrika, 63,* 111–116. https://doi.org/10.1093/biomet/63.1.111

[31] Sohrabi, Z., & Maleki, J. (2025). Fusing satellite imagery and ground-based observations for PM2.5 air pollution modeling in Iran using a deep learning approach. *Scientific Reports*, 15, 1, Art. no. 21449. https://doi.org/10.1038/s41598-025-05332-2

[32] Zhang, Z., Zhang, S., Zhao, X., Chen, L., & Yao, J. (2022). Temporal difference-based graph transformer networks for air quality PM2.5 prediction: A case study in China. *Frontiers in Environmental Science, 10,* 924986. https://doi.org/10.3389/fenvs.2022.924986

[33] The 17 Goals, Department of Economic and Social Affairs. (2025). Sustainable development. https://sdgs.un.org/goals (accessed May 23, 2025).

[34] United Nations. (2025). *Transforming our world: The 2030 agenda for sustainable development.* https://sdgs.un.org/2030agenda (accessed May 23, 2025).