

Article

Predictability of Air Pollutants Based on Detrended Fluctuation Analysis: Ekibastuz Coal-Mining Center in Northeastern Kazakhstan

Oleksandr Kuchanskyi ^{1,2,3} , Andrii Biloshchytskyi ^{4,5,*}, Yurii Andrashko ⁶ , Alexandr Neftissov ⁷ ,
Svitlana Biloshchytska ^{1,5,*} and Sergiy Bronin ⁸ 

¹ Department of Computational and Data Science, Astana IT University, Astana 010000, Kazakhstan; kuchanskyi.o@gmail.com

² Department of Information Control Systems and Technologies, Uzhhorod National University, 88000 Uzhhorod, Ukraine

³ Department of Biomedical Cybernetics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 03056 Kyiv, Ukraine

⁴ Department of Administration, Astana IT University, Astana 010000, Kazakhstan

⁵ Department of Information Technology, Kyiv National University of Construction and Architecture, 03037 Kyiv, Ukraine

⁶ Department of System Analysis and Optimization Theory, Uzhhorod National University, 88000 Uzhhorod, Ukraine; yurii.andrashko@uzhnu.edu.ua

⁷ Research and Innovation Center “Industry 4.0”, Astana IT University, Astana 010000, Kazakhstan; alexandr.neftissov@astanait.edu.kz

⁸ Institute of Social Sciences and Applied Informatics, Vilnius University, 44280 Kaunas, Lithuania; sbronin@me.com

* Correspondence: a.b@astanait.edu.kz (A.B.); bsv@astanait.edu.kz (S.B.)

Abstract

Environmental comfort and air pollution are among the most important indicators for assessing the population’s quality of life in urban agglomerations. This study aims to explore long-term memory in air pollution time series by analyzing the dynamics of the Hurst exponent and evaluating the predictability index. This type of statistical pre-forecast analysis is essential for developing accurate forecasting models for such time series. The effectiveness of air quality monitoring systems largely depends on the precision of these forecasts. The Ekibastuz coal-mining center, which houses one of the largest coal-fired power stations in Kazakhstan and the world, with a capacity of about 4000 MW, was chosen as an example for the study. Data for the period from 1 March 2023 to 31 December 2024 were collected and analyzed at the Ekibastuz coal-fired power station. During the specified period, 14 indicators (67,527 observations) were collected at 10 min intervals, including mass concentrations of CO, NO, NO₂, SO₂, PM_{2.5}, and PM₁₀, as well as current mass consumption of CO, NO, NO₂, SO₂, dust, and NO_x. The detrended fluctuation analysis of a time series of air pollution indicators was used to calculate the Hurst exponent and identify long-term memory. Changes in the Hurst exponent in regards to dynamics were also investigated, and a predictability index was calculated to monitor emissions of pollutants in the air. Long-term memory is recorded in the structure of all the time series of air pollution indicators. Dynamic analysis of the Hurst exponent confirmed persistent time series characteristics, with an average Hurst exponent of about 0.7. Identifying the time series plots for which the Hurst exponent is falling (analysis of the indicator of dynamics), along with the predictability index, is a sign of an increase in the influence of random factors on the time series. This is a sign of changes in the dynamics of the pollutant release concentrations and may indicate possible excess emissions that need to be controlled. Calculating the dynamic changes in the Hurst exponent for the emission



Academic Editor: Jianming Cai

Received: 2 May 2025

Revised: 6 July 2025

Accepted: 9 July 2025

Published: 16 July 2025

Citation: Kuchanskyi, O.; Biloshchytskyi, A.; Andrashko, Y.; Neftissov, A.; Biloshchytska, S.; Bronin, S. Predictability of Air Pollutants Based on Detrended Fluctuation Analysis: Ekibastuz Coal-Mining Center in Northeastern Kazakhstan. *Urban Sci.* **2025**, *9*, 273. <https://doi.org/10.3390/urbansci9070273>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

time series made it possible to identify two distinct clusters corresponding to periods of persistence and randomness in the operation of the coal-fired power station. The study shows that evaluating the predictability index helps fine-tune the parameters of time series forecasting models, which is crucial for developing reliable air pollution monitoring systems. The results obtained in this study allow us to conclude that the method of trended fluctuation analysis can be the basis for creating an indicator of the level of air pollution, which allows us to quickly respond to possible deviations from the established standards. Environmental services can use the results to build reliable monitoring systems for air pollution from coal combustion emissions, especially near populated areas.

Keywords: detrended fluctuation analysis; Hurst exponent; air pollution; long-term memory; predictability index; coal-fired power station

1. Introduction

It is well known that coal is a relatively inexpensive fuel source for generating electricity. Moreover, the Energy Institute [1] reports that coal accounts for up to 35% of the global electricity balance as of 2023. However, because coal burning harms human health and the environment, coal consumption in many countries is generally declining. In particular, in the United States, coal consumption has decreased by nearly half over the past 12 years, and the number of coal-fired power stations has also declined. A similar trend is observed in Europe. By contrast, estimates indicate growing coal production and use in China, India, Russia, and other parts of Asia. In some cases, coal spas occur at unmodernized stations that are highly polluted by the environment [2]. Also, in these regions, people continue to burn coal to heat private homes, which leads to a significant concentration of pollutants in the air. Coal burning releases high concentrations of solid particles (PM_{10} , $PM_{2.5}$), metal contaminants, and carcinogenic gases [3]. This results in significant health problems for the local population, specifically an increase in respiratory problems, cardiovascular diseases, etc. In particular, Mahlangeni et al. [4] examine the health impact of coal-fired power station operations on the health of the population living near these stations. As a result, it was found that living near coal-fired power stations is often associated with an increased likelihood of respiratory disorders, adverse consequences of childbirth, and problems with a child's development.

The American Lung Association [5] notes that burning coal contaminates the air with solid particles (PM_{10} , $PM_{2.5}$) that can enter the lungs and circulatory system, increasing the risk of cardiovascular and respiratory diseases. Also, when coal burns, sulfur dioxide (SO_2) and nitrogen oxides (NO_2) are released, causing respiratory irritation. Mercury (Hg), lead (Pb), and arsenic (As) particles are also released, which accumulate in the body and can cause neurological and oncological diseases. Research has also examined the impact of burning coal indoors on human health. A study by the International Agency for Research on Cancer (IARC) determined that indoor emissions from coal burning comprise a group 1 carcinogen [6]. At the same time, indoor coal burning remains the main source of air pollution in many developing countries [7]. Despite this, the main factor contributing to air pollution worldwide remains the production of energy based on coal.

Henneman et al. [8] analyzed deaths in the United States associated with exposure to $PM_{2.5}$ from coal-fired power stations. As a result, a large dataset of Medicare deaths dating back to 1999 was analyzed. Then, the way in which air flows carry emissions from 480 separate coal-fired power stations was modeled. This allowed them to assess the impact of $PM_{2.5}$ carbon particles in locations where affected individuals lived

and died. The research team estimated that between 1999 and 2020, approximately 460,000 deaths would not have occurred if there had been no emissions from coal-fired power stations. Another important conclusion from the study [8] is that installing pollution control sensors (scrubbers) reduced emissions-related deaths. For example, the Keystone facility in Pennsylvania was one of the most powerful power stations during the study period. This station was associated with an average of more than 600 deaths per year before the installation of emissions scrubbers. After installing the scrubbers, this number dropped to 80 deaths per year [8,9]. Therefore, it is important to ensure reliable monitoring of pollutant emissions from coal-fired power stations. This is the basis for ensuring the necessary quality of life and health of the population near power stations.

Wesz et al. [10] developed seven measures based on the ISO standard 37120:2018 [11] to assess the population's quality of life in urban agglomerations. These include urban services; culture; economy and recreation; urban mobility; friendliness; safety; and environmental comfort. Environmental comfort includes noise and air pollution, climate comfort, and abundance of water resources. Therefore, studying the impact and characteristics of pollutant emissions from coal-fired power stations is highly relevant for urban studies. This is important for safeguarding the quality of life of the urban population, considering the environmental impact of the urban development of coal mining regions and ensuring the effective use of coal in coal-fired power stations. Ensuring effective monitoring of pollution and considering the characteristics and changes in the concentration of various pollutants should lead to a reduction in air pollution levels and accordingly, a decrease in population morbidity. This, in turn, should improve the population's standard of living and reduce state spending on medical care.

The pollution issue concerns urban agglomerations that are located near coal mining sites. For example, in the Republic of Kazakhstan, the Pavlodar region in northeastern Kazakhstan is one of the regions where the problem of ensuring environmental safety in cities is acute. The Ekibastuz coal-mining center is located in this region. Ekibastuz is a city, as well as a center of open-pit coal mining. Coal has been extracted here since 1876 in small volumes, but large-scale mining began in the mid-1900s [12]. In the 1970s, Ekibastuz was the third largest coal producer in the Soviet Union. However, at that time, this region was mostly a rural area, and the urban center was still in the process of formation. Currently, the population of Ekibastuz exceeds 150 thousand people [13]. The Ekibastuz coal-fired power station is also located in the city, providing electricity to both the domestic market and for export. The Ekibastuz coal-fired power station is one of the largest thermal power stations in Kazakhstan and the world, with a capacity of 4000 MW.

The goal of this study is to investigate long-term memory in air pollution time series by analyzing the dynamic behavior of the Hurst exponent and assessing the predictability index. This type of statistical pre-forecast analysis is essential for building accurate forecasting models for such time series. The accuracy of these forecasts directly impacts the effectiveness of air quality monitoring systems.

2. Literature Review

This section presents a literature review regarding the use of fractal analysis for assessing long-term memory in time series, particularly in the context of the emissions of air pollutants. Conducting a comprehensive study to identify the causes of pollution and explore ways to reduce or mitigate it requires an understanding of various influencing factors, including climatic, economic, technological, and other aspects. Currently, Kazakhstan has adopted the concept of a green economy, along with a national development strategy for Kazakhstan until 2050. These documents serve as the strategic foundation for a consistent transformation of the country's economy, with the focus on sustainable development. The

key factor for achieving carbon neutrality in Kazakhstan's economy is the energy sector. The energy sector is the main source of greenhouse gas emissions in Kazakhstan. Moreover, the main source of emissions is the coal industry. In 1990, the coal industry accounted for 65% of Kazakhstan's greenhouse gas emissions. In 2017, this share was estimated at 59% [14]. Kazakhstan is among the top ten countries in the world when it comes to its coal reserves, estimated at 33.6 billion tons across approximately 400 fields. About 72% of the country's coal is extracted in the Pavlodar and Karaganda regions.

According to the report in Reference [14], the EU's carbon border adjustment mechanism (CBAM) priorities [15] can significantly impact Kazakhstan's industry if the country does not abandon its use of coal. Losses in export revenues can reach USD hundreds of millions of annually. The gradual abandonment of coal in the production and energy sectors will help mitigate the associated problems. In 2018, 239 and 263 cancer cases per 100,000 people were recorded in the Karaganda and Pavlodar regions, respectively, while the national average was 184 cases per 100,000 people. According to recent satellite observations, 27,500 kg/h of methane emissions from the Ekibastuz mines was observed in October 2021 [14]. This corresponds to the emissions of approximately 1.3 million vehicles yearly. Consequently, the region near the city of Ekibastuz experiences significant environmental problems that will continue to accumulate over the years.

In recent years, there has been an improvement in the environmental situation in Kazakhstan due to the fact that industrial enterprises are required to obtain permits for emissions within the established permissible limits [16]. Maintaining these values depends on ensuring air quality standards and sanitary standards in the surrounding localities. For example, in 2019, Kazakhstan ranked 29th in the global ranking of the most polluted countries [17]. In 2024, Kazakhstan ranked 71st out of 138 countries in terms of pollution (based on annual average $\text{PM}_{2.5}$ concentration ($\mu\text{g}/\text{m}^3$)) [18]. Nevertheless, many cities in Kazakhstan exhibit high pollution rates and rank among the most polluted cities, including Karaganda, Astana, Almaty, etc. The climate in most regions of Kazakhstan is sharply continental, with cold winters; thus, the centralized heating season lasts about 6 months. Accordingly, one of the main sources of pollution is the high rate of coal consumption at power stations, along with coal burning by private households. This leads to significant health problems for local residents. In studies [19,20], the reasons for the increased prevalence of chronic obstructive pulmonary disease and bronchial asthma in Kazakhstan were analyzed. Also, the impact of air pollution in Kazakhstan on public health has been evaluated in studies [21,22]. In study [23], the mortality rate from air pollution in the large cities of Kazakhstan was estimated. The highest mortality rate from the potential effects of increased $\text{PM}_{2.5}$ concentrations is observed in industrial cities. Studies [24,25] describe solutions that can improve air quality in regions with high levels of pollution. In particular, the concept of moss-based biotechnological purification filters used in urban environments is described. In study [25], a model was constructed to determine the best locations for biotechnological cleaning filters by solving the discrete optimization problem. Such approaches often overlook the root causes of air pollution and tend to be ineffective without a thorough analysis of time series data from pollution sources, identification of technological operating modes that influence emissions, and determination of optimal parameters for forecasting, planning, and managing emissions.

This study is dedicated to the analysis of emissions from the Ekibastuz coal-fired power station. The Ekibastuz coal-mining center is located in the Pavlodar region, which, in addition to coal extraction, is known for significant man-made pollution caused by mining activities, oil refining, the chemical industry, and both non-ferrous and ferrous metallurgy. However, the main pollution in the region is due to coal mining and the operation of thermal power stations near the city of Ekibastuz. The population of the

region exceeds 750,000 people [13]. The urban agglomerations of Pavlodar, Aksu, and Ekibastuz are located directly near the regions of intensive mining and the location of thermal power stations. Although the city of Ekibastuz is located 16 km away from the Ekibastuz coal-fired power station, local dispersion models for this plant show that the highest concentrations of pollutants are typically observed within an 8 km radius around the station and spread horizontally [26]. Moreover, weather conditions can cause the transport of pollutants beyond 8 km. Similar studies conducted for other power stations have revealed that pollution can travel significantly greater distances. For instance, Levy et al. used the CALPUFF atmospheric dispersion model, determining that high concentrations of particulate matter can be detected at distances ranging from 1 km to 40 km [27]. Goyal et al. observed $PM_{2.5}$ dispersion at distances of 60 km and beyond, noting a substantial impact on air quality in cities located up to 100 km away from coal-fired power plants [28]. The negative effects of locating cities near coal power plants are also discussed in studies [4,29].

Therefore, it can be concluded that emissions from the Ekibastuz coal-fired power station have an impact on air quality in the city of Ekibastuz. Given that this power station contributes a significant portion of the region's annual pollutant load, understanding changes in emissions at the source is critical for predicting their impact on urban areas through the use of dispersion models. Analyzing emissions near the source helps to identify the causes of exceedances and provides a foundation for reducing pollution in urban agglomerations. The complexity of analyzing time series of changes in air pollution levels, especially in urban agglomerations and industrial zones, is associated with the nonlinearity of pollution spread. Therefore, before applying methods for predicting such time series, it is necessary to analyze the structure and investigate the presence of long-term memory in these time series. This can be achieved using fractal analysis methods, specifically, variable range analysis (R/S analysis) or detrended fluctuation analysis (DFA). These methods can also be used to build reliable air quality monitoring systems. In particular, in study [30], long-term memory in a time series of concentrations of air pollutants (solid particles PM_{10} , $PM_{2.5}$) from four stations in Astana (Kazakhstan) was studied using the fractal R/S analysis method. It was found that the reason for the significant increase in air pollutant concentration is the close location of the coal-fired power station to the city. In study [31], The MF-DFA method was constructed, and the multifractal structure of the air pollution time series was revealed. The study was conducted using the city of Zhengzhou (China) as an example. In study [32], data from observations of air pollution levels by solid particles (PM_{10} , $PM_{2.5}$), as well as gaseous materials and meteorological indicators at 412 stations in Germany from 2008 to 2018, were analyzed. According to the Hurst exponent's analysis of air pollution time series, it was determined that air pollution at certain monitoring stations is generally persistent. In study [33], the behavior of the time series of air pollution levels in the United States (PM_{10} , $PM_{2.5}$) was analyzed. The results showed that Western states have a higher persistence rate in their air pollution time series compared to Eastern states, where a general downward trend is observed in the Hurst exponent. In addition, the study of long-term memory in time series has been explored in studies [34–37]. Other studies have used nonparametric methods, such as DFA [38] and multifractal detrended fluctuation analysis (MF-DFA) [39], to estimate the Hurst exponent. If the time series of air pollution are short, the DFA method is used to analyze them [39,40]. The predictability index can be calculated based on the fractal dimension, which depends on the Hurst exponent [41]. This index can be used as an indicator for air quality monitoring systems.

Therefore, the application of fractal analysis to air pollution time series to establish long-term memory and identify the stability of the emissions of pollutants into the air is an urgent task. This is especially true for urban agglomerations near coal-fired power stations,

which are characterized by high levels of air pollution because of the high concentrations of pollutants in the air, which harm the health and quality of life of the population.

This paper aims to investigate long-term memory in the time series of air pollution indicators resulting from coal burning, which impacts environmental safety and the population's quality of life. Another goal of the study is to establish a connection between the calculated estimates of long-term memory and the technological operating modes of the coal-fired power station, as well as the parameters of air pollution forecasting models. This information can be utilized by air quality monitoring systems to enhance their accuracy and reliability. The Ekibastuz coal-mining center, which houses one of Kazakhstan's largest coal-fired power stations, was chosen as an example for the study. To achieve this goal, the following tasks were identified: to apply the DFA method of a time series of air pollution indicators to calculate the Hurst exponent and the predictability index, as well as to investigate the dynamic changes of the Hurst exponent over time, and to justify how the obtained predictability index estimates can be applied to forecast air pollution time series. The analysis also examined how the Hurst exponent values reflect the technological operating modes of the Ekibastuz coal-fired power station and how this information can be used to reduce the emissions of air pollutants.

The long-term stability of pollutant emissions in the air enables the prediction and control of pollution parameters. The results of the dynamic analysis of the Hurst exponent and the evaluation of the predictability index will help improve the accuracy of forecasting models for air pollution time series and provide insights into how the operating conditions of the power station influence these dynamics. This is crucial for the development of effective air quality monitoring systems. Finally, the results are important for developing urbanism in terms of ensuring the environmental safety and sustainable development of urban agglomerations, particularly those located near coal mining areas and coal-fired power stations.

3. Materials and Methods

3.1. Calculation of the Hurst Exponent Based on Detrended Fluctuation Analysis

This section describes the method for calculating the Hurst exponent using DFA. It also provides an interpretation of the obtained estimates in the context of analyzing time series data for the emissions of air pollutants. The DFA of the time series was used to calculate the Hurst exponent for this case. In study [42], a modified method is presented that allows for the elimination of the deviations and is suitable for short sections of the time series. However, in study [43], it is indicated that the use of DFA introduces a bias and is more computationally expensive. On the other hand, a preliminary analysis of the input time series of air pollution indicators at the Ekibastuz coal-fired power station revealed the presence of a significant number of sequential passes. This makes it challenging to use the traditional and less expensive R/S analysis method to calculate the Hurst exponent. The advantage of the DFA method is that it can be used for shorter time series sections. Therefore, it is possible to bypass gaps to study the data structure. This is important for accurately assessing the Hurst exponent and the predictability index, as well as for identifying long-term memory in air pollution time series. These parameters, along with the analysis of changes in the dynamics of the Hurst exponent, are very important for monitoring the air pollution level.

A discrete time series of air pollution indicators is given. In this case, the values of the indicators are recorded by the corresponding sensors at time points t_1, t_2, \dots, t_n , where t_1 is the starting point in time and

$$R = (r(t_1), r(t_2), \dots, r(t_n))$$

Calculate the mean using the following formula:

$$\bar{R} = \frac{1}{n_j} \sum_{j=1}^n r(t_j)$$

where \bar{R} is the mean for a time series R .

Then, calculate the cumulative sum:

$$\rho_s = \sum_{j=1}^s (r(t_j) - \bar{R}^h)$$

where ρ_s is the cumulative sum of the input time series R .

A sequence of integers is defined as $I = (i_1, i_2, \dots, i_k)$, $i_1 < i_2 < \dots < i_k$, $i_b \in \mathbb{Z}$, $i_b > 0$, $b = \overline{1, k}$. Empirically obtained estimates for the minimum and maximum values of the sequence are I : $i_1 \approx 4$, $i_k \approx n$. Moreover, the sequence is evenly distributed on a logarithmic scale, i.e., $\log(i_2) - \log(i_1) \approx \log(i_3) - \log(i_2) \approx \dots$. For each sequence value $i \in I$ and cumulative sum ρ_s , using the least squares method, we will determine the local trend. As a result, we obtain a piecewise-linear fit $(Y_1^i, Y_2^i, \dots, Y_n^i)$ [44].

From the obtained local trends, we will calculate the root mean square deviation σ_{ij} and root-mean-square σ_i :

$$\sigma_{ij} = \sqrt{\frac{1}{i} \sum_{s=ij+1}^{ij+i} (\rho_s - Y_s^i)^2}$$

$$\sigma_i = \sqrt{\frac{1}{\lambda} \sum_{j=1}^{\lambda} \sigma_{ij}^2}, \quad \lambda = \frac{n}{i}$$

and if the time series R long-range power law is correlated, the fluctuation functions σ_i increase by a power law of $\sigma_i \propto i^\alpha$, $\alpha > 0$ [42].

Assessment $\alpha > 0$ is a generalization of the Hurst exponent, i.e., if $\alpha > \frac{1}{2}$, then the data is correlated, and the time series is persistent. If $\alpha < \frac{1}{2}$, then the data is anticorrelated, and the time series is antipersistent. Indicator value $\alpha \simeq \frac{1}{2}$ indicates white noise. In this case, the fractal dimension D of the time series R [45,46] is

$$D = 2 - \alpha$$

Value D indicates how much the fractal fills the space when zoomed out. Based on the indicator D , the predictability index P can be calculated [41] as follows:

$$P = 2|D - 1.5|$$

and value P increases when value D less or more than 1.5. If $D < 1.5$, then there is an antipersistence in the structure of the time series R . If $D > 1.5$, then the time series R is persistent. If value P is close to 0, then it can be argued that the time series R is random or close to random. Determining the value of the predictability index in combination with the assessment of the Hurst exponent enables the selection of optimal parameter values for forecasting models of emission time series.

Next, the dynamic changes in the Hurst exponent can be evaluated. A family of fixed-length time series h can be built, constructed sequentially from time series values R , and the DFA method can be applied for each of them. That is, for time series

$$R^h = (r(t_1), r(t_2), \dots, r(t_h))$$

$$R^{h+1} = (r(t_2), r(t_3), \dots, r(t_{h+1})), \dots$$

we plot a time series of indicators $\alpha_h, \alpha_{h+1}, \dots, \alpha_n$, $h < n$. Each value α_q corresponds to the predictability index value P_q , $q = \bar{h}, n$. Creating this time series can help monitor air pollution from harmful substances. In this case, the following interpretation of the results can be given:

1. If $\frac{1}{2} < \alpha_q \leq 1$, $0 < P_q \leq 1$, $q \in [h, n]$, then this may indicate the presence of long-term memory or persistence of the time series. In the context of analyzing the time series of air pollution indicators, such Hurst exponents may mean that the main trend of changes in the values of air pollution indicators persists. In other words, the emissions are stable and predictable. For such a time series, traditional forecasting methods based on long-term trends can be used. This is important for the operation of air pollution monitoring systems, because the case of persistence of a time series of air pollution indicators is a sign of the stability, foresight, and manageability of emissions.
2. If $\alpha_q \simeq \frac{1}{2}$, $P_q \simeq 0$, $q \in [h, n]$, this means that the time series of the outliers is random or close to random. In the context of analyzing the time series of air pollution indicators, this may mean that the level of pollution changes erratically, and the emissions of pollutants cannot be regulated or predicted. This may be due, for example, to a violation of the processes of standards and requirements in the operation of coal-fired power stations, leading to the exceeding the permissible emission levels, especially if persistence was recorded in the time series before the identification of randomness, i.e., $\frac{1}{2} < \alpha_q \leq 1$. Violating the standards in the operation of coal-fired power stations may relate to non-compliance with the temperature regime of coal burning. If coal is burned at too low a temperature, it can lead to incomplete combustion, increasing carbon monoxide (CO) emissions, volatile organic compounds, and solid particles. The use of low-quality coal may also be a violation, leading to increased emissions of sulfur oxides and heavy metals. Additionally, there may be violations in the operation of flue gas purification systems, non-compliance with the operating standards of boiler equipment, and inadequate maintenance.
3. If $0 \leq \alpha_q < \frac{1}{2}$, $0 < P_q \leq 1$, $q \in [h, n]$, then the time series is ergodic; that is, the process is antipersistent. In analyzing the time series indicators of air emissions, pollution tends to temporarily decrease after high values or, conversely, increase after low values. The trend of such a time series changes faster than does a random time series. Another important finding is that a drop α_q below the persistence threshold indicates possible improper emission control. There may be falsification or a lack of monitoring of pollutant emissions. This does not allow us to adequately assess the compliance of the pollution level with the standards.

The value of the Hurst exponent for the time series of air pollution in regards to dynamics allows for the effective monitoring of the level of air pollution, identification of critical situations, and quick response to exceedances of permissible emissions levels. If the time series of outliers is persistent, traditional models that consider the trend component can be chosen to predict it.

3.2. Area of Study

This section provides a description of the region where the study was conducted. We analyzed data for pollutant emissions recorded by sensors at the Ekibastuz coal-fired power station. The station is located in the Ekibastuz coal-mining center, which is a cluster of large power stations and one of the largest energy generating areas in Kazakhstan. The district is located in the northeast of the Republic of Kazakhstan in the city of Ekibastuz, with a population of 152,509 people as of 2023 [13]. The region is known for its large coal reserves.

The Ekibastuz coal basin is home to some of the world's largest open-pit coal mines. There are two power stations in the area (GRES-1, GRES-2). One station (GRES-1) has a capacity of 4000 MW and uses low-grade brown coal extracted from nearby mines. Another nearby station (GRES-2) has a capacity of about 1000 MW and boasts the world's tallest chimney (419.7 m). The electricity generated by the stations is used for domestic consumption and is exported to neighboring countries. Figure 1 shows the location of the study area.



Figure 1. Location of the study area—Ekibastuz city, Ekibastuz coal-mining center, and coal-fired power stations (GRES-1, GRES-2), Republic of Kazakhstan.

3.3. Experimental Methods and Instruments

This section justifies the use of the DFA method and describes the specific measurement instruments that were used in the study. Industrial-grade, non-reference instruments were used to measure emissions. For monitoring gas components such as O₂, CO, NO, and SO₂, the ULTRAMAT 23 sensor (manufactured by Siemens, Germany) was employed. This device operates on the principle of infrared radiation with a double- or triple-layer detector, which ensures high selectivity and measurement accuracy. Thanks to automatic air calibration, the sensor maintains the long-term stability of the measured values without the need for expensive reference gases. Verification with certified reference gases is conducted at least once per year, while air-purge calibration is performed every four hours. To measure particulate matter and dust, the SOLER DUST 200 dust analyzer (produced by Promanalit, Republic of Kazakhstan) was used. This device is designed for the automatic measurement of the mass concentration of suspended particles (dust) in flue gases and chimney ducts. It operates based on the scattered light reflection method: a red laser diode light illuminates dust particles in the measurement volume of the gas duct, and the light scattered backward by these particles is detected and analyzed. The measurement process is non-contact, continuous, and does not require sampling from the flow. According to the technical specifications, the relative measurement error for mass dust concentration using the SOLER DUST 200 is within $\pm 15\%$. Both devices undergo annual verification using certified calibration gases. Measurements of gas and dust concentrations are conducted as part of the mandatory emission monitoring system. Device reliability is verified once a year to ensure the validity of the collected data.

This paper implemented a DFA of air pollution indicators collected at the Ekibastuz coal-fired power station to achieve the research goal. The method allows us to detect long-term correlations in nonlinear signals. The time series of air pollution indicators

collected are characterized by the presence of significant consecutive gaps. Therefore, to determine the Hurst exponent for establishing long-term memory, we implemented a marked fluctuation analysis, which enabled us to calculate fractal characteristics for relatively small time series dimensions. Similar studies were performed to establish the phased characteristics of air pollution indicators in Astana and indicators of mine water releases from metal ore mining facilities in the Northern Balkhash region and the Sayak ore district (Republic of Kazakhstan) [30,37]. However, the R/S analysis method was used in these studies. This was possible because areas of significant volume without start-ups were identified in the time series of outliers and releases, which enabled the accurate calculation of the Hurst exponent's values and an examination of the Hurst exponent's behavior regarding dynamics.

4. Results

4.1. Collection of Data

The emissions of pollutants at the Ekibastuz coal-fired power station are measured using industrial sensors installed directly on stationary emission sources. To monitor the concentration of suspended particles at the emission source, dust meters are installed that measure the optical density of the smoke flow, allowing for the determination of the level of dust contamination. Gas analyzers are also used and are installed near the source of release. Sampling employs a sampling probe that draws flue gas from the pipe. Before supplying gas to the gas analyzer, the sample is subjected to a preparation stage that includes filtration and cooling to remove condensate and solid impurities that may affect measurements accuracy. After preparation, the sample enters the measuring chamber of the gas analyzer, where the concentration of the corresponding substances is determined.

Data for the period from 1 March 2023 to 31 December 2024 were collected and analyzed. Indicators were recorded every 10 min. During this period, 67,527 observations of indicators were completed. The values of 14 indicators were recorded (Table 1), among which 2 pollution indicators were identified and used in the following studies (the data are specified in the Supplementary Materials).

Table 1. Description of indicators that were monitored at Ekibastuz coal-fired power station.

№	Indicator Code	Description	Units of Measurement
1	N-2-HC-CALC-WTR_CO.UNIT2@NET0	Mass concentration of carbon monoxide (CO) in wet gas	mg/m ³
2	N-2-HC-CALC-WTR_NO.UNIT2@NET0	Mass concentration of nitric oxide (NO) in wet gas	mg/m ³
3	N-2-HC-CALC-WTR_NO2.UNIT2@NET0	Mass concentration of nitrogen dioxide (NO ₂) in wet gas	mg/m ³
4	N-2-HC-CALC-WTR_O2.UNIT2@NET0	Mass concentration of oxygen (O ₂) in wet gas	mg/m ³
5	N-2-HC-CALC-WTR_SO2.UNIT2@NET0	Mass concentration of sulfur dioxide (SO ₂) in wet gas	mg/m ³
6	N-2-HC-MIP-CURR_DUST2.5.UNIT2@NET0	Mass concentration of PM _{2.5} in emissions	mg/m ³
7	N-2-HC-MIP-CURR_DUST10.UNIT2@NET0	Mass concentration of PM ₁₀ in emissions	mg/m ³
8	N-2-HC-CALC-WTR-CURR_CO.UNIT2@NET0	Current mass consumption of CO	g/s
9	N-2-HC-CALC-WTR-CURR_NO.UNIT2@NET0	Current mass consumption of NO	g/s
10	N-2-HC-CALC-WTR-CURR_NO2.UNIT2@NET0	Current mass consumption of NO ₂	g/s
11	N-2-HC-CALC-WTR-CURR_O2.UNIT2@NET0	Current mass consumption of O ₂	g/s
12	N-2-HC-CALC-WTR-CURR_SO2.UNIT2@NET0	Current mass consumption of SO ₂	g/s
13	N-2-HC-CALC-MIP-CURR_DU.UNIT2@NET0	Current mass consumption of dust	g/s
14	N-2-HC-CALC-WTR_NOX.UNIT2@NET0	Mass concentration of external nitrogen oxides (NO + NO ₂)	mg/m ³

The first pollution indicator group consists of the mass concentrations of CO, NO, NO₂, O₂, SO₂, PM_{2.5}, and PM₁₀. Indicators of this group are measured in mg/m³. The

second group includes the current mass consumption of CO, NO, NO₂, O₂, CO₂, and dust (PM_{2.5} and PM₁₀ in total). Indicators of this group of pollutants are measured in grams per second. A dataset of indicators in CSV format is available in the Supplementary Materials. Table 2 shows some descriptive statistics.

Table 2. Descriptive statistics of time series of air pollution indicators at Ekibastuz coal-fired power station.

Indicator	CO	NO	Mass Concentration				CO	Current Mass Consumption				
			NO ₂	SO ₂	PM _{2.5}	PM ₁₀		NO	NO ₂	SO ₂	NO _x	DUST
Number of observations	47,772	50,558	33,538	63,027	65,920	65,920	47,134	48,683	32,361	61,920	64,860	64,668
Mean	7.77	74.18	100.47	688.68	379.15	379.15	2.05	19.30	24.80	173.87	109.770	96.04
Std	8.72	119.28	44.20	463.71	246.38	246.38	1.87	30.47	11.46	119.87	158.31	62.41
Min	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max	624.00	499.00	200.00	3339.00	1000.00	1000.00	140.00	119.0	51.00	710.00	697.00	371.00

4.2. Application of Detrended Fluctuation Analysis to Calculate the Hurst Exponent

This section presents the results of the experiment assessing the long-term memory and the predictability index for the time series of emissions of air pollutants. Based on the DFA analysis for each time series of pollution indicators, the Hurst exponent, predictability index values were calculated, and the presence of long-term memory in the time series for the established period from 1 March 2023 to 31 December 2024 was investigated. Figure 2 shows the change in indicators for the entire specified time period. For these, the values of the Hurst exponent and the predictability index were calculated based on the number of points shown in Table 2. A different number of points was recorded in each time series. This is because the time series of the pollution indicators exhibited gaps that were not considered. Fractal analysis indicators were calculated for time series sections that did not contain gaps.

To validate the choice of sliding window length in the DFA method, a sensitivity analysis of the Hurst exponent α was conducted across a wide range of pollutants. The estimation was performed for multiple window sizes (50, 100, 300, 500, 750, and 1000 data points) in order to evaluate the stability of the α value under varying window lengths.

All analyzed time series can be conditionally divided into two groups. The first group includes indicators for which the Hurst exponent values remain close to 0.5. This applies in particular to the concentrations of CO and nitrogen oxides (NO). For these series, the α values show extremely low sensitivity to window length: the difference in estimates between the smallest and largest window sizes is less than 0.003. Such stability indicates the absence of pronounced long-term memory in the data structure, which is characteristic of nearly random or noise-like processes. In this case, selecting a window length in the range of 300–500 data points is considered safe and does not significantly affect the conclusions. The second group includes time series characterized by Hurst exponent values in the range of 0.7–0.8. This group comprises concentrations of NO₂, SO₂, PM_{2.5}, PM₁₀, and dust, as well as aggregated NO_x indicators. These series exhibit noticeable sensitivity to the choice of window length. For instance, when using short windows (50–100 data points), α values can reach as high as 0.81, whereas increasing the window length to 1000 data points results in a decrease to approximately 0.68. This variability, ranging from 0.1 to 0.13, suggests that short windows may overestimate the level of persistence, while longer windows tend to smooth out local fluctuations, diminishing part of the correlation structure.

Thus, selecting a window length of 500 data points represents a reasonable compromise for both types of time series. For data with low α values, this choice has little impact on the results, while for series exhibiting long-term dependencies, it provides stable estimates

without excessively suppressing information about underlying trends. This approach ensures methodological robustness and preserves the informativeness of the results for further analysis or forecasting.

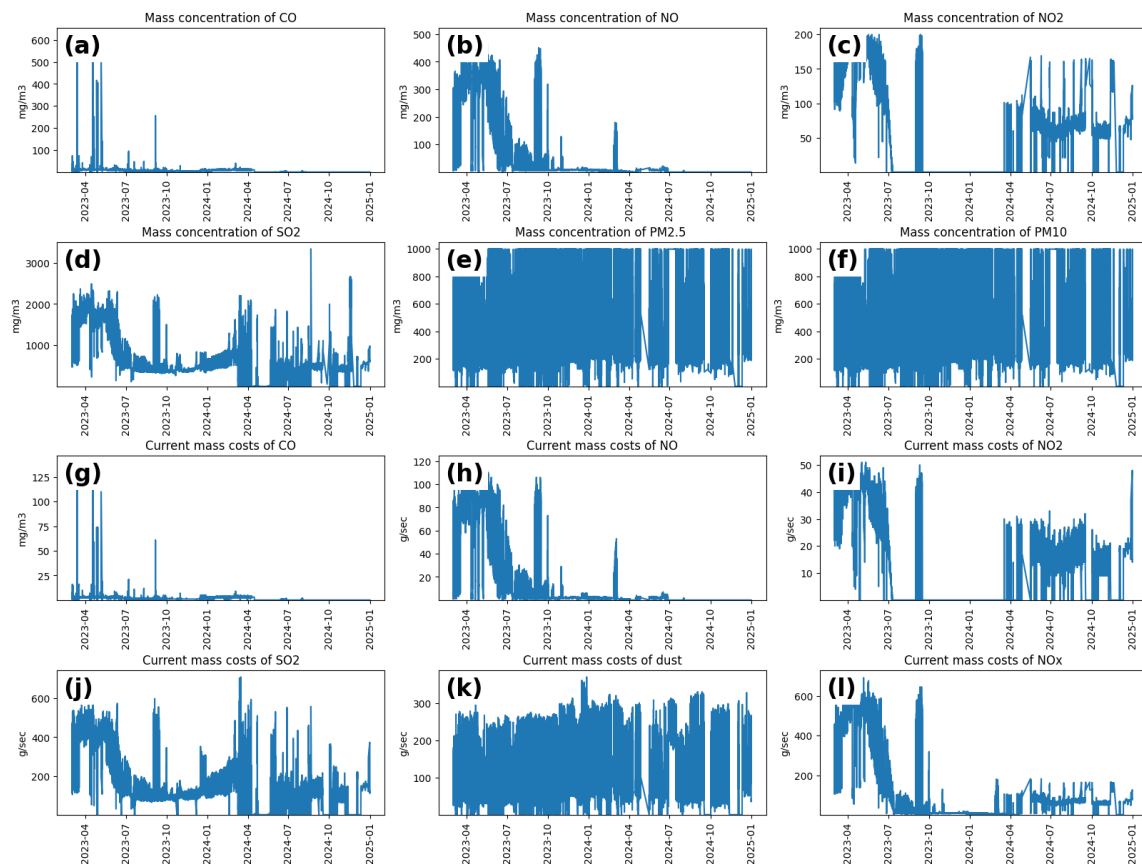


Figure 2. Time series of air pollution values recorded at Ekibastuz coal-fired power station for the period from 1 March 2023 to 31 August 2024 (smoothed using a 24 h moving average window): (a) mass concentration of CO, (b) mass concentration of NO, (c) mass concentration of NO₂, (d) mass concentration of SO₂, (e) mass concentration of PM_{2.5}, (f) mass concentration of PM₁₀, (g) current mass consumption of CO, (h) current mass consumption of NO, (i) current mass consumption of NO₂, (j) current mass consumption of SO₂, (k) current mass consumption of dust, (l) current mass consumption of NO_x.

During the data preprocessing stage, an assessment of missing data patterns was conducted across all time series. It was found that the total number of gaps was relatively small (ranging from 2 to 4 segments per series), but the duration of individual gaps was substantial, varying from 124 to 1560 data points. As a result, the use of interpolation methods to fill such long gaps was deemed inappropriate, as it could artificially distort the internal structure of the time series.

Meanwhile, in cases where short gaps were present (up to several dozen data points), the use of specialized interpolation techniques (such as spline interpolation) was indeed considered at other stages of the modeling process. However, in our study, all detected gaps were too long for such methods to be applied effectively. As a result, the corresponding segments were entirely excluded from the analysis. This ensured the integrity of the processed subsequences and maintained compliance with the methodological requirements of the chosen approaches. In addition, the dataset was examined for the presence of statistical outliers. A classical approach was applied, based on identifying deviations exceeding three standard deviations (3σ) from the mean. As a result, 158 observations exceeding 3σ were detected for SO₂, and 346 for NO_x.

None of these observations exceeded the threshold of 5σ , which allows them to be classified as anomalously high, but not extreme, values. These observations accounted for only 0.254% of the total dataset. Given their small number and limited deviation from the mean, it was decided not to exclude them from further analysis, as their impact on the overall structure of the time series is minimal. The moving window method with a fixed length of 500 points was used to process time series of air pollution indicators. Calculations were performed only for segments of the series that did not contain missing values, which corresponds to standard practices for ensuring statistical correctness in fractal analysis. Due to gaps in the original data, the number of processed points differed for different time series. These omissions were automatically excluded from the analysis, and the corresponding areas without results are marked on the graphs (Figures 2 and 3), allowing for the visual identification of areas with insufficient data quality for inclusion in the study. We visually analyzed the time series for each pollutant. However, due to high variability and significant overlap between the curves, displaying the raw trajectories in full resulted in interpretation difficulties. To address this, we applied data smoothing using a 24 h moving average window. This approach reduced the influence of short-term fluctuations, highlighted long-term trends, and improved the clarity of the plots. Figure 3 presents the smoothed trajectories of the Hurst exponent, allowing for a clearer comparison of long-term memory dynamics across different pollutants.

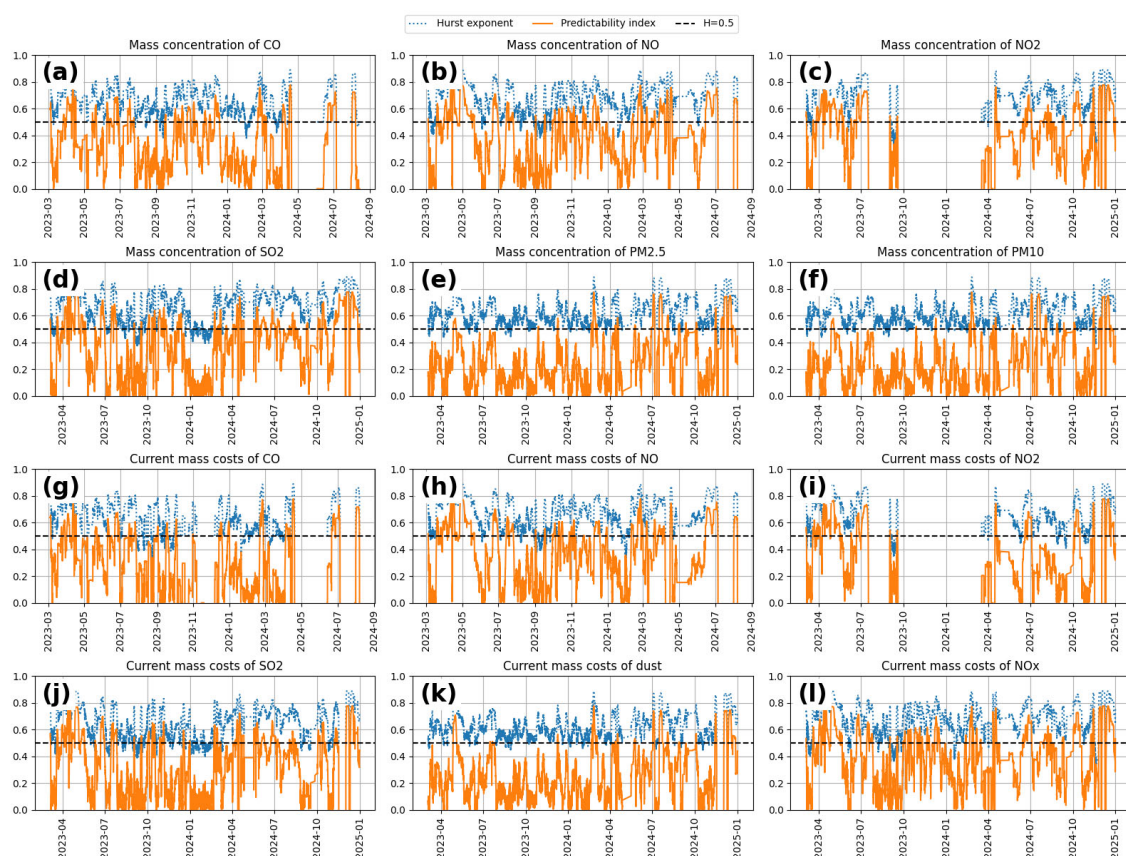


Figure 3. Time series of α_q values for air pollution indicators that were recorded at the Ekibastuz coal-fired power station for the period from 1 March 2023 to 31 August 2024: (a) mass concentration of CO, (b) mass concentration of NO, (c) mass concentration of NO₂, (d) mass concentration of SO₂, (e) mass concentration of PM_{2.5}, (f) mass concentration of PM₁₀, (g) current mass consumption of CO, (h) current mass consumption of NO, (i) current mass consumption of NO₂, (j) current mass consumption of SO₂, (k) current mass consumption of dust, (l) current mass consumption of NO_x.

Next, the Hurst exponent and predictability index were calculated, and the results are presented in Table 3. After analyzing the results obtained, we can distinguish three main groups of indicators. For the mass concentration of CO and NO, as well as for the current mass consumption of CO and NO, the predictability index value is close to zero, indicating the random nature of the pollution series. For the mass concentration indicators of dust and the current mass flow rate of NO₂, SO₂, and dust, the predictability index value is close to 0.45, and for the mass concentration indicators of NO₂ and SO₂ and the current mass flow rate of NO_x, the value is close to 0.55. For both groups of indicators, we can conclude that there is a long-term memory or persistence of the time series of the pollutants.

Table 3. Value of the Hurst exponent and predictability index.

Value		Parameters	
		Hurst Exponent	Predictability Index
Mass concentration	CO	0.4998	0.0003
	NO	0.4998	0.0003
	NO ₂	0.7658	0.5316
	SO ₂	0.7678	0.5356
	PM _{2.5}	0.7229	0.4459
	PM ₁₀	0.7229	0.4459
Current mass consumption	CO	0.4998	0.0003
	NO	0.4998	0.0003
	NO ₂	0.7176	0.4352
	SO ₂	0.7196	0.4392
	DUST	0.7205	0.4411
	NO _x	0.7658	0.5316

Changes in the values of the Hurst exponent and the predictability index were calculated using the flow window method to identify the causes of such different behaviors of the pollutants over the specified period. The graph of the changes in the indicators is shown in Figure 3. An analysis of the graphs shows that the predictability index value varies from 0 to 0.8 for all pollution indicators. This index is characterized by sharp variability over time, indicating instability in the predictability of pollution levels. Such fluctuations are probably caused by abnormal emissions that lack a clear periodicity and occur due to unpredictable human-caused or weather factors. Short-term peaks in the emissions of pollutants recorded by individual sensors significantly affect the structure of the time series and reduce the stability of the dynamics, as reflected in the decrease in predictability index values during the corresponding periods. This suggests that the time series of mass concentrations of CO and NO, as well as their current mass consumption, are not fundamentally different from the time series of other pollutants. It should also be noted that predictability index values close to zero can sometimes be caused by missing data at the end of the time series. Such gaps violate the integrity of the signal fluctuation structure and reduce the accuracy of estimating the predictability index. This problem can be partially eliminated by selecting an alternative point for calculation, applying averaging methods, or increasing the analysis window length, which allows for smoother results and reduces the impact of local breaks.

Therefore, based on the analysis results, it can be concluded that for all the time series of air pollution indicators from the Ekibastuz coal-fired power station, the characteristic periods of growth of the Hurst exponent and the predictability index also correspond to periods of decrease in these indicators. The Hurst exponent drops below 0.5, and the predictability index drops to 0. Predictability is index-sensitive to the insignificant influence of random factors on the time series. That is, in the event of a change in the emissions of air pollution that does not correspond to the current trend, this indicator is directed to 0. In general, the results indicate the stability and predictability of pollutant emissions, as most of

the time series of pollution indicators are persistent. Thus, using the example of data from the Ekibastuz coal-fired power station, it can be argued that the described tools are effective and can be used to monitor the state of air pollution. When calculating the Hurst exponent using a 500-point window of the air pollution time series, we identified two distinct clusters: periods of persistence ($\alpha \in [\alpha_i^*, 0.82]$) and periods of randomness ($\alpha \in [0.40, \alpha_i^*]$). Thus, the operational system of the coal-fired power station can be characterized as alternating between two distinct states, each corresponding to one of the identified clusters. Threshold value α_i^* is defined as the expected value of the Hurst exponent for a standard normal process, calculated using the following formula [47]:

$$\alpha_i^* = \sqrt{\frac{2}{\pi(i-1)}} \sum_{j=1}^{i-1} \sqrt{\frac{i-j}{j}} \quad (1)$$

Kristoufek found that for a window size of 512 data points, the average value of α_i^* is approximately 0.5316 [48]. Thus, in our case, it was established that the obtained Hurst exponent values were analyzed alongside the log files of the coal-fired power station's operation. It was found that the system's state depends on the boiler load. Specifically, under stable operation with a boiler load of 90–100%, the system tends to remain in a state corresponding to the persistence cluster, with the Hurst exponent values for the emission time series falling within this range. During transitional periods, the boiler load fluctuates between 30% and 70%, and the system enters a state of randomness, with the Hurst exponent values for the emission time series reflecting this random behavior. During these periods, the emissions exhibit unstable fluctuations, indicating a shift in the technological operating mode. This helps explain when and under what conditions the system transitions between states corresponding to persistent and random phases in the emission time series. Proper planning of the coal-fired power station's operation and timely maintenance allow the system to remain in the persistent cluster state for as long as possible. This is crucial for forecasting and maintaining stability in air pollutant emissions. In particular, calculating the Hurst exponent enabled the development and testing of deep sparse transformer networks for predicting pollutant time series. The implementation of this approach and the verification of results based on data from the Ekibastuz coal-fired power station are detailed in study [49].

5. Discussion

5.1. Findings, Implications, and Recommendations for Urban Monitoring

The operation of the Ekibastuz coal-mining center is of significant economic importance for the energy sector of Kazakhstan and Central Asia as a whole. Its work plays a key role in providing electricity to industry and maintaining the stability of energy security in this region. However, the operation of such powerful power stations has a negative environmental impact, producing air pollution and greenhouse gas emissions, which significantly harm health and affect the lives of the local population. The population of the Pavlodar region, where the Ekibastuz coal-mining center is located, is approximately 750,000 people [13]. Studies [19,20] indicate a significant prevalence of chronic lung diseases in this region. Currently, projects are being implemented to modernize and improve the efficiency of these stations to reduce harmful emissions. Therefore, monitoring air pollution indicators at the power station is an important element in ensuring the effectiveness of these projects.

This study collected and analyzed data for the period from 1 March 2023 to 31 December 2024 at the Ekibastuz coal-fired power station. During the specified period, 14 indicators (67,527 observations) were collected at 10 min intervals. A DFA was applied,

a predictability index was constructed, and changes in the Hurst exponent regarding dynamics were investigated. In general, the analysis of the Hurst exponent showed that all the time series are persistent, and there is a long-term memory in their structure. Such series can be predicted using traditional statistical forecasting models or machine learning models. The results also identified moments when the Hurst exponent decreased to values close to 0.5, which indicates an increase in the influence of random factors on the time series. In this case, the predictability index value is reduced to 0. Such areas may indicate unstable changes in pollutant emissions. A change in the emissions of pollutants with a change in the main trend of the time series is a sign of a sharp change in the current level of emissions, which may indicate emission overages. This information can serve as the basis for air pollution monitoring systems, especially in such a challenging region (from an ecological perspective) as the Ekibastuz coal-mining center, near which tens of thousands of people reside.

To further support the validity of using the DFA method, we conducted additional modeling based on classical time series analysis approaches, particularly ARIMA. These models were applied to all investigated indicators, with a primary focus on evaluating short-term autocorrelation, residual structure, and information criteria. As a result of modeling using the baseline ARIMA(2,1,1) model [50], short-term memory was detected, evidenced by statistically significant coefficients of the autoregressive components. However, the residuals of this model exhibited autocorrelation (Ljung–Box, $p < 0.001$), high variance ($\sigma^2 \approx 3775$), and deviation from normality (Jarque–Bera, $p < 0.001$), indicating limited explanatory power. To improve the quality of approximation, an ARIMA(20,1,1) model was applied, incorporating a longer lag structure. This adjustment significantly reduced the information criteria (AIC dropped by more than 13,000 points), decreased the residual variance (to ≈ 3102), and completely eliminated residual autocorrelation (Ljung–Box, $p = 0.80$), formally indicating model adequacy.

However, even under these conditions, the residuals exhibited significant heteroscedasticity ($\alpha = 0.21$, $p < 0.001$), along with pronounced skewness and excess kurtosis (kurtosis > 13), indicating a complex internal dynamic in the series that was not fully captured by the ARIMA model. Thus, while classical linear models proved suitable for describing short-term memory, they failed to provide reliable estimates of long-term dependencies and large-scale fluctuations—factors that are critical for environmental indicators with potentially fractal structures. In this context, the DFA method was chosen as more suitable for detecting long-term memory properties in non-stationary time series. Unlike traditional approaches, DFA does not require prior stationarization, is robust to trends and structural shifts, and enables direct assessment of large-scale autocorrelation characteristics. While we acknowledge the potential value of wavelet analysis as a multi-scale approach, the primary focus of this study was to evaluate monofractal properties, which are effectively characterized using DFA.

Since the emissions of pollutants were recorded directly at the coal-fired power station, it is important to determine whether these emissions impact air quality in the city of Ekibastuz. The Ekibastuz coal-fired power station is located on the northern shore of Lake Zhengeldy, 16 km north of the city of Ekibastuz [51]. This relatively short distance indicates that emissions from the station do reach the city, as supported by previous studies on pollutant dispersion distances around coal-fired power plants [26–28]. Analyzing and forecasting changes in emissions near the source and how these are influenced by the operational conditions of the power station makes it possible to predict their impact on urban areas through the application of dispersion models. Based on the calculation of the Hurst exponent using a 500-point window in the air pollution time series, we identified two distinct clusters: periods of persistence and periods of randomness. The operational

state of the coal-fired power station alternates between these two modes. It was established that the system's state is directly influenced by boiler load. When the system is in a state corresponding to the randomness cluster, emissions exhibit unstable fluctuations, which reflect changes in the technological operating mode. Such changes may be caused by equipment failures or maintenance work, leading to reduced boiler load and increased pollutant emissions. In contrast, under stable operation without technological disturbances, the boiler operates at full capacity, and emissions remain controlled. These findings help explain when and under what conditions the system transitions between persistent and random phases in the emission time series. Furthermore, the calculated values of the predictability index were used to fine-tune the parameters of a neural network for forecasting emission time series, offering an effective solution for air pollution monitoring. This research is discussed in greater detail in study [49].

5.2. Limitations and Future Research Lines

Based on the results of the study, some restrictions can be formulated. The first limitation is the small dataset that has been collected and is available for analysis. The period is one and one-half years (from March 2023 to August 2024). The data were collected every 10 min, ensuring a sufficient amount of data for conducting the study and applying the DFA method. However, it would be interesting to study possible changes in the structure of the time series of air pollution indicators that were recorded in previous years. To maintain methodological rigor, long gaps in the dataset (up to 1560 missing points) were excluded from the analysis. While this preserves statistical integrity, it may omit important events such as shutdowns, malfunctions, or extreme emissions episodes, which are critical from a policy perspective.

Another limitation is that other facilities in this region can release pollutants into the air. In general, this study makes a statement about the complex environmental problems of the region where the Ekibastuz coal-fired power station is located. Accordingly, these environmental problems can be harmful to health and worsen the quality of life for the local population. Despite this, data was selected for analysis from the pipe, as well as directly from the source of emissions. Therefore, it is safe to say that the emissions of pollutants was not affected by other objects or weather conditions. However, the direction of wind propagation and the geographical features of the region in which the settlements are located can affect the concentration of pollutants at a distance from the source of emissions. The stations are located north of Ekibastuz. The prevailing southerly wind can impact the air quality in the city. Therefore, for a comprehensive analysis, it is also necessary to analyze the air directly in the city of Ekibastuz, especially when the trend of the Hurst exponent time series changes from the level of persistence to the level of randomness and in the event of a drop in the predictability index. In the future, it will be important to gain access to a larger dataset over a longer period of time, as well as from other coal-fired power stations and coal mining and consumption facilities in the region. This is the only way to conduct a comprehensive analysis of environmental problems and form recommendations for improving the environmental situation in this region.

The results obtained are important for environmental monitoring services in cities located near regions with intensive coal production and consumption, specifically, coal-fired power stations. Prompt response to violations of regulatory requirements for changes in air pollutant concentrations is key to maintaining health and a high quality of life for city residents near emission sources.

As an additional example of operational application, the predictability index can be used as a real-time signal for automated monitoring and anomaly detection in emission control systems at thermal power plants. In particular, this study proposes establishing a threshold

value for the index (e.g., $p < 0.1$), justified through empirical analysis of the historical time series. When the predictability index drops below this threshold, it signals a transition in the time series from stable, trend-driven behavior to a noise-like, random regime characterized by the dominance of short-term fluctuations and the loss of long-term memory.

In practice, such a change may correspond to technological disturbances in boiler operation, such as transitions to low or unstable load levels, deviations from optimal combustion temperatures, or failures in flue gas cleaning systems. These events often lead to unpredictable and potentially excessive pollutant emissions that cannot be adequately captured by static forecasting models.

To implement this approach, the predictability index can be continuously computed within a moving time window using real-time sensor data streams. Such processing allows for near-instant detection of threshold crossings. Upon detection, the monitoring system can initiate a multi-level response protocol, as follows:

1. Automatic flagging and logging of the event in the environmental management system for regulatory reporting and audit purposes.
2. Sending notifications (e.g., SMS, email, or integration with SCADA) to plant operators or environmental services for immediate investigation.
3. Automatic adjustment of the forecasting model parameters (for example, switching to models with shorter memory or retraining on new data) to maintain prediction accuracy under changing dynamics.
4. Initiating diagnostic procedures to check technological parameters such as fuel quality, combustion temperature, and the effectiveness of flue gas cleaning equipment.

By defining such threshold values for the predictability index, plant operators can create emission monitoring systems that do not merely passively record data but actively detect regime changes in emission behavior. This enables preventive maintenance and supports compliance with environmental standards, reducing the risk of unnoticed exceedances of permitted emission levels.

Moreover, thanks to the high sensitivity of the predictability index to subtle changes in the statistical structure of emission data—even before large-scale exceedances occur—it can serve as an early warning system, providing time to mitigate the problem before it escalates.

Integration of this approach with existing SCADA systems or environmental decision-support platforms enhances real-time monitoring capabilities. Such a system can continuously assess the statistical predictability of emissions, dynamically adapt forecasting strategies, and generally improve the effectiveness of ambient air quality management. A promising direction for future research is refining the selection of threshold values based on historical data analysis, as well as extending the approach to multi-component scenarios using combined predictability indices or multivariate models with long-term memory.

6. Conclusions

Long-term memory was studied in the time series of air pollution indicators that were recorded at the Ekibastuz coal-fired power station (Republic of Kazakhstan) for the period from 1 March 2023 to 31 August 2024, including mass concentrations of CO, NO, NO₂, SO₂, PM_{2.5}, PM₁₀ and current mass consumption of CO, NO, NO₂, SO₂, dust, and NO_x. For the analysis, a marked fluctuation analysis was used. Since time series exhibit consecutive spaces, and areas without gaps are of small length, this method is best suited for determining the presence of long-term memory in the time series structure. The values of the Hurst exponent and predictability index were calculated. Changes in the Hurst exponent regarding dynamics were also calculated and visualized. The dynamics of the indicator indicate a generally persistent nature of all indicators. On average, the Hurst exponent for all the time series is kept at 0.7. There are also areas of decrease in the

Hurst exponent to values close to 0.5 and the predictability index to values close to 0. This means that significant emissions of pollutants occurred during the recorded periods, i.e., the influence of random factors on the time series increased. At the same time, the sensitivity of the predictability index is quite high, since a slight drop in the Hurst exponent provoked a sharp drop in the predictability index values. In general, time series are trend-resistant and can be predicted using traditional time series forecasting methods or machine learning models.

Identifying the time series plots for which the Hurst exponent is falling (analysis of the indicator of dynamics) and the predictability index is a sign of an increase in the influence of random factors on the time series. This is a sign of changes in the dynamics of pollutant release emissions and may indicate possible excess emissions that need to be controlled. Thus, the information obtained can help build air monitoring systems at coal mining and consumption facilities and metal ore processing facilities, generally ensuring effective management of facilities that may harm environmental safety and the health and quality of life of the population living near these facilities.

The dynamic analysis of the Hurst exponent for the emission time series made it possible to identify two clusters corresponding to periods of persistence and randomness in the operation of the coal-fired power station. It was found that the operational state of the system is determined by the boiler load. When the system enters a state of randomness, this typically indicates a reduction in boiler load and changes in the technological operating modes. As a result, emissions exhibit unstable fluctuations. Conversely, when the boiler operates at maximum load, emissions remain stable and are characterized as persistent, according to the DFA method. The study also demonstrated that calculating the predictability index enables the fine-tuning of forecasting model parameters for emission time series, which is crucial for developing reliable air quality monitoring systems.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/urbansci9070273/s1>, Table S1: Data for air pollution indicators recorded at the Ekibastuz coal-fired power station from 1 March 2023, to 31 August 2024.

Author Contributions: Conceptualization, O.K. and Y.A.; methodology, O.K.; software, Y.A.; formal analysis, O.K. and Y.A.; investigation, A.B. and S.B. (Svitlana Biloshchytska); data curation, A.N. and S.B. (Sergiy Bronin); writing—original draft preparation, O.K.; writing—review and editing, Y.A., S.B. (Svitlana Biloshchytska), S.B. (Sergiy Bronin), and A.B.; visualization, Y.A.; supervision, A.N.; project administration, A.B. All authors have read and agreed to the published version of the manuscript.

Funding: 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”.

Data Availability Statement: All data are available in this publication.

Acknowledgments: The authors thank the reviewers and editors for their generous and constructive comments that have improved this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Energy Institute. Statistical Review of World Energy. 2024. Available online: <https://www.energyinst.org/statistical-review> (accessed on 6 April 2025).
2. Finkelman, R.B.; Wolfe, A.; Hendryx, M.S. The future environmental and health impacts of coal. *Energy Geosci.* **2021**, *2*, 99–112. [CrossRef]
3. Amster, E. Public health impact of coal-fired power plants: A critical systematic review of the epidemiological literature. *Int. J. Environ. Health Res.* **2019**, *31*, 558–580. [CrossRef]

4. Mahlangeni, N.; Kapwata, T.; Webster, C.; Howlett-Downing, C.; Wright, C.Y. Exposure to air pollution from coal-fired power plants and impacts on human health: A scoping review. *Rev. Environ. Health* **2025**, *2025*, 1–22. [CrossRef]
5. American Lung Association. Toxic Air: The Case for Cleaning up Coal-Fired Power Plants. 2011. Available online: <https://www.lung.org/getmedia/c3b2b744-7c7e-4941-b0cd-5a5e468515d1/toxic-air-report.pdf> (accessed on 6 April 2025).
6. International Agency for Research on Cancer. *Indoor Emissions from Household Combustion of Coal*, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans; World Health Organization: Lyon, France, 2012. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK304396/> (accessed on 6 April 2025).
7. World Health Organization. WHO Global Ambient Air Quality Database Update. 2018. Available online: <https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database/2018> (accessed on 6 April 2025).
8. Henneman, L.; Choirat, C.; Dedoussi, I.; Dominici, F.; Roberts, J.; Zigler, C. Mortality risk from United States coal electricity generation. *Science* **2023**, *382*, 941–946. [CrossRef]
9. Furlow, B. Coal power plant pollution linked to deaths in the USA. *Lancet Respir. Med.* **2023**, *12*, 193. [CrossRef] [PubMed]
10. Wesz, J.G.B.; Miron, L.I.G.; Delsante, I.; Tzortzopoulos, P. Urban Quality of Life: A Systematic Literature Review. *Urban Sci.* **2023**, *7*, 56. [CrossRef]
11. ISO 37120:2018; Sustainable Cities and Communities—Indicators for City Services and Quality of Life. International Organization for Standardization (ISO): Geneva, Switzerland, 2018. Available online: <https://www.iso.org/standard/68498.html> (accessed on 6 April 2025).
12. Ekibastuz, Kazakhstan, Britannica. 2025. Available online: <https://www.britannica.com/place/Kazakhstan/People> (accessed on 6 April 2025).
13. The Population of Kazakhstan by Individual Ethnic Groups at the Beginning of 2021. 2021. Available online: <https://stat.gov.kz/> (accessed on 6 April 2025).
14. Kuznetsova, E.; Vaillancourt, K.; Energy Transition in Monocities. Coal Phase-out Roadmap and Just Transition Action Plan for Ekibastuz (Kazakhstan). *2030 Roadmap and Action Plan*. 2023. Available online: <https://www.un-page.org/static/dc10131616a8f9f1411d844e623463c9/esmia-undp-kazakstan-ekibastuz-2023-06-29-final.pdf> (accessed on 6 April 2025).
15. Carbon Border Adjustment Mechanism. 2025. Available online: https://taxation-customs.ec.europa.eu/carbon-border-adjustment-mechanism_en (accessed on 6 April 2025).
16. UN World Water Development Report. 2021. Available online: <https://www.unwater.org/publications/un-world-water-development-report-2021> (accessed on 6 April 2025).
17. Assanov, D.; Zapasnyi, V.; Kerimray, A. Air Quality and Industrial Emissions in the Cities of Kazakhstan. *Atmosphere* **2021**, *12*, 314. [CrossRef]
18. IQAir, World’s Most Polluted Countries & Regions. 2024. Available online: <https://www.iqair.com/world-most-polluted-countries> (accessed on 6 April 2025).
19. Nugmanova, D.; Feshchenko, Y.; Iashyna, L.; Gyrina, O.; Malynovska, K.; Mammadbayov, E.; Akhundova, I.; Nurkina, N.; Tariq, L.; Makarova, J.; et al. 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 Pulm. Med.* **2018**, *18*, 26. [CrossRef]
20. Nugmanova, D.; Sokolova, L.; Feshchenko, Y.; Iashyna, L.; Gyrina, O.; Malynovska, K.; Mustafayev, I.; Aliyeva, G.; Makarova, J.; Vasylyev, A.; et al. The Prevalence, Burden and Risk Factors Associated with Bronchial Asthma in Commonwealth of Independent States Countries (Ukraine, Kazakhstan and Azerbaijan): Res. of the CORE Study. *BMC Pulm. Med.* **2018**, *18*, 110. [CrossRef]
21. Kenessary, D.; Kenessary, A.; Adilgireiuly, Z.; Akzholova, N.; Erzhanova, A.; Dosmukhametov, A.; Syzdykov, D.; Masoud, A.-R.; Saliev, T. Air Pollution in Kazakhstan and Its Health Risk Assessment. *Ann. Glob. Health* **2019**, *85*, 133. [CrossRef]
22. Semenova, Y.; Zhunussov, Y.; Pivina, L.; Abisheva, A.; Tinkov, A.; Belikhina, T.; Skalny, A.; Zhanaspayev, M.; Bulegenov, T.; Glushkova, N.; et al. Trace Element Biomonitoring in Hair and Blood of Occupationally Unexposed Population Residing in Polluted Areas of East Kazakhstan and Pavlodar Regions. *J. Trace Elem. Med. Biol.* **2019**, *56*, 31–37. [CrossRef]
23. Kerimray, A.; Assanov, D.; Kenessov, B.; Karaca, F. Trends and Health Impacts of Major Urban Air Pollutants in Kazakhstan. *J. Air Waste Manag. Assoc.* **2020**, *70*, 1148–1164. [CrossRef]
24. Biloshchytskyi, A.; Kuchanskyi, O.; Andrashko, Y.; Yedilkhan, D.; Neftissov, A.; Biloshchytska, S.; Amirgaliyev, B.; Vatskel, V. Reducing Outdoor Air Pollutants through a Moss-Based Biotechnological Purification Filter in Kazakhstan. *Urban Sci.* **2023**, *7*, 104. [CrossRef]
25. Biloshchytskyi, A.; Kuchansky, A.; Andrashko, Y.; Neftissov, A.; Vatskel, V.; Yedilkhan, D.; Herych, M. Building a model for choosing a strategy for reducing air pollution based on data predictive analysis. *East.-Eur. J. Enterp. Technol.* **2022**, *117*, 23–30. [CrossRef]
26. Issakhov, A.; Alimbek, A.; Issakhov, A. A numerical study for the assessment of air pollutant dispersion with chemical reactions from a thermal power plant. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 1035–1061. [CrossRef]

27. Levy, J.; Spengler, J.D.; Hlinka, D.; Sullivan, D.; Moon, D. Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: Model sensitivity and implications. *Atmos. Environ.* **2002**, *36*, 1063–1075. [\[CrossRef\]](#)
28. Goyan, N.; Vyas, A.D. Dispersion Modelling of Air Pollutants of Coal Based Thermal Power Station. In Proceedings of the 5th International Conference on Advances in Chemical, Biological & Environmental Engineering (ACBEE-17), New Delhi, India, 14–15 September 2013; pp. 233–235.
29. Datt, G.; Maitra, P.; Menon, N.; Ray, R. Coal plants, air pollution and anaemia: Evidence from India. *J. Dev. Stud.* **2023**, *59*, 533–551. [\[CrossRef\]](#)
30. Biloshchytskyi, A.; Neftissov, A.; Kuchanskyi, O.; Andrashko, Y.; Biloshchytska, S.; Mukhatayev, A.; Kazambayev, I. Fractal Analysis of Air Pollution Time Series in Urban Areas in Astana, Republic of Kazakhstan. *Urban Sci.* **2024**, *8*, 131. [\[CrossRef\]](#)
31. Wang, Q.; Zhao, T.; Wang, R.; Zhang, L. Backward Trajectory and Multifractal Analysis of Air Pollution in Zhengzhou Region of China. *Math. Probl. Eng.* **2022**, 2022, 2226565. [\[CrossRef\]](#)
32. Liu, X.; Hadiatullah, H.; Tai, P.; Xu, Y.; Zhang, X.; Schnelle-Kreis, J.; Schlöter-Hai, B.; Zimmermann, R. Air pollution in Germany: Spatio-temporal variations and their driving factors based on continuous data from 2008 to 2018. *Environ. Pollut.* **2021**, *276*, 116732. [\[CrossRef\]](#)
33. Gil-Alana, L.A.; Yaya, O.S.; Awolaja, O.G.; Cristofaro, L. Long Memory and Time Trends in Particulate Matter Pollution (PM_{2.5} and PM₁₀) in the 50 U.S. States. *J. Appl. Meteorol. Climatol.* **2020**, *59*, 1351–1367. [\[CrossRef\]](#)
34. Meraz, M.; Rodriguez, E.; Femat, R.; Echeverria, J.C.; Alvarez-Ramirez, J. Statistical persistence of air pollutants (O₃, SO₂, NO₂ and PM₁₀) in Mexico City. *Phys. A Stat. Mech. Its Appl.* **2015**, *427*, 202–217. [\[CrossRef\]](#)
35. Nikolopoulos, D.; Moustiris, K.; Petraki, E.; Koulougliotis, D.; Cantzos, D. Fractal and Long-Memory Traces in PM₁₀ Time Series in Athens, Greece. *Environments* **2019**, *6*, 29. [\[CrossRef\]](#)
36. Varotsos, C.; Ondov, J.; Efstathiou, M. Scaling properties of air pollution in Athens, Greece and Baltimore, Maryland. *Atmos. Environ.* **2005**, *39*, 4041–4047. [\[CrossRef\]](#)
37. Biloshchytskyi, A.; Kuchanskyi, O.; Neftissov, A.; Andrashko, Y.; Biloshchytska, S.; Kazambayev, I. Fractal Analysis of Mining Wastewater Time Series Parameters: Balkhash Urban Region and Sayak Ore District. *Urban Sci.* **2024**, *8*, 200. [\[CrossRef\]](#)
38. Xue, Y.; Pan, W.; Lu, W.Z.; He, H.D. Multifractal nature of particulate matters (PMs) in Hong Kong urban air. *Sci. Total Environ.* **2015**, *532*, 744–751. [\[CrossRef\]](#)
39. Thompson, J.R.; Wilson, J.R. Multifractal detrended fluctuation analysis: Practical applications to financial time series. *Comput. Simul.* **2016**, *126*, 63–88. [\[CrossRef\]](#)
40. Kantelhardt, J.W.; Zschiegner, S.A.; Koscielny-Bunde, E.; Havlin, S.; Bunde, A.; Stanley, H.E. Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A Stat. Mech. Its Appl.* **2002**, *316*, 87–114. [\[CrossRef\]](#)
41. Parmar, K.S.; Bhardwaj, R. Water quality index and fractal dimension analysis of water parameters. *Int. J. Environ. Sci. Technol.* **2013**, *10*, 151–164. [\[CrossRef\]](#)
42. Kantelhardt, J.W.; Koscielny-Bunde, E.; Rego, H.H.A.; Havlin, S.; Bunde, A. Detecting long-range correlations with detrended fluctuation analysis. *Physica A* **2001**, *295*, 441–454. [\[CrossRef\]](#)
43. Bryce, R.M.; Sprague, K.B. Revisiting detrended fluctuation analysis. *Sci. Rep.* **2012**, *2*, 315. [\[CrossRef\]](#)
44. Hardstone, R.; Poil, S.; Schiavone, G.; Jansen, R.; Nikulin, V.; Mansvelder, H.; Linkenkaer-Hansen, K. Detrended Fluctuation Analysis: A Scale-Free View on Neuronal Oscillations. *Front. Physiol.* **2012**, *3*, 450. [\[CrossRef\]](#)
45. Rangarajan, G.; Sant, D.A. Fractal dimensional analysis of Indian climatic dynamics. *Chaos Solitons Fractals* **2004**, *19*, 285–291. [\[CrossRef\]](#)
46. Peters, E.E. *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1994; p. 336.
47. Annis, A.; Lloyd, E.H. The expected value of the adjusted rescaled Hurst range of independent normal summands. *Biometrika* **1976**, *63*, 111–116. [\[CrossRef\]](#)
48. Krištoufek, L. *Classical and Modified Rescaled Range Analysis: Sampling Properties Under Heavy Tails*; IES Working Paper; Charles University in Prague, Institute of Economic Studies (IES): Prague, Czech Republic, 2009; p. 26.
49. Andrashko, Y.; Kuchanskyi, O.; Biloshchytskyi, A.; Neftissov, A.; Biloshchytska, S. Forecasting Air Pollutant Emissions Using Deep Sparse Transformer Networks: A Case Study of the Ekibastuz Coal-Fired Power Plant. *Sustainability* **2025**, *17*, 5115. [\[CrossRef\]](#)

-
50. Smith, T.G.; Pmdarima. Tips to Using Auto_Arima. Available online: https://alkaline-ml.com/pmdarima/tips_and_tricks.html (accessed on 23 May 2025).
 51. Limited Liability Partnership Ekibastuz GRES-1. 2023. Available online: https://gres1.kz/en/page.php?slug=contact-us&utm_source (accessed on 10 June 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.