**DOI: 10.37943/19IKWT5637**

**Andrii Biloshchytskyi**
Doctor of Technical Sciences, Professor, Vice-Rector for Science and Innovation
a.b@astanait.edu.kz, orcid.org/0000-0001-9548-1959
Astana IT University, Kazakhstan
Professor of Information Technologies Department, Kyiv National University of Construction and Architecture, Ukraine

**Oleksandr Kuchanskyi**
Doctor of Technical Sciences, Professor of the Computing and Data Science Department
kuczanski@gmail.com, orcid.org/0000-0003-1277-8031
Astana IT University, Kazakhstan
Department of Information Control Systems and Technologies, Uzhhorod National University, Ukraine

**Alexandr Neftissov**
PhD, Associate Professor, Director of the Science and Innovation Center "Industry 4.0"
alexandr.neftissov@astanait.edu.kz, orcid.org/0000-0003-4079-2025
Astana IT University, Kazakhstan

**Svitlana Biloshchytska**
Doctor of Technical Sciences, Professor of the Computing and Data Science Department
bsvetlana2007@ukr.net, orcid.org/0000-0002-0856-5474
Astana IT University, Kazakhstan
Professor of Information Technologies Department, Kyiv National University of Construction and Architecture, Ukraine.

**Arailym Medetbek**
Master student of Digital Public Administration,
arailym.medetbek@astanait.edu.kz, orcid.org/0009-0001-7192-0360
Astana IT University, Kazakhstan

# INTEGRATED MODEL FOR FORECASTING TIME SERIES OF ENVIRONMENTAL POLLUTION PARAMETERS

**Abstract:** The quality of life in large urban areas is considerably diminished by air pollution, with major contributors being motor vehicles, industrial activities, and fossil fuel combustion. A major contributor to air pollution is coal-fired and thermal power plants, which are commonly found in emerging markets. In Astana, Kazakhstan, a rapidly expanding city's significant reliance on coal for heating and considerable building exacerbate air pollution. This research is essential for improving urban development practices that support sustainable growth in rapidly expanding cities. Using time series data from four monitoring stations in Astana using fractal R/S analysis, the study looks at long-term patterns in air pollutant levels, especially PM10 and PM2.5. The stations' Hurst exponents were determined to be 0.723, 0.548, 0.442, and 0.462. Additionally, the flow window method was used to study the Hurst exponent's dynamic behavior. The findings showed that one station's pollution levels had long-term memory, which suggests that the time series is persistent. While anti-persistence was noted in the third and fourth sites, data from the second station indicated nearly random behavior. The Hurst exponent values explain the October 2021 spike in pollution levels, which is probably


Received: 11.09.2024 Accepted: 24.09.2024 Published: 30.09.2024

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

caused by thermal power plants close to the city. The fractal analysis of time series could serve as an indicator of environmental conditions in a given region, with persistent pollution trends potentially aiding in predicting critical pollution events. Anti-persistence or temporary pollution spikes may be influenced by the observation station's proximity to pollution sources. Overall, the findings suggest that fractal time series analysis can act as a valuable tool for monitoring environmental health in urban areas.

**Keywords**: urban air pollution; R/S analysis; time series analysis; Hurst exponent; PM10; PM2.5

**Introduction**

Chemical, physical, or biological pollutants that alter the composition of the atmosphere are referred to as indoor or outdoor air pollutants by the World Health Organization (WHO) [1]. Common sources of pollution include the usage of cars, industrial processes, and the combustion of fossil fuels. Landfills, mining, waste management, and agricultural activities are other causes. The extensive usage of thermal and coal-fired power plants is the primary cause of pollution in emerging nations. In order to create mechanical energy that powers electric generators, thermal power plants transform fuel's chemical energy into thermal energy. In a similar vein, coal-fired power plants use coal, a major environmental pollutant, to provide both heat and energy, particularly during the colder months. Greenhouse gases that contribute to climate change are also released when fossil fuels are burned. However, carbon dioxide emissions in the global energy system have only dropped by 1% in more than three decades since the UN Framework Convention on Climate Change was adopted. Global power generation is still dominated by fossil fuels, with renewable energy sources making up only 8.2% of the total. In 2021, energy-related emissions reached historic highs due to a 59% rise in the world's energy consumption [2].

Air pollution has detrimental effects on public health, including higher mortality, neurological problems, respiratory and cardiovascular illnesses, and pregnancy issues [2], [3]. Additionally, there is evidence that greater air pollution levels are associated with higher type II diabetes death rates [4]. As the consequences of climate change and air pollution from fuel combustion intensify over time, the detrimental effects are particularly apparent for pregnant women and children [5].

The public's health is seriously threatened by pollutants such as sulfur dioxide, nitrogen dioxide, carbon monoxide, and particulate matter. PM2.5 particle pollution killed 4.2 million people in 2020, the same number as in 2015. There was a 5% drop in the average mortality rate per 100,000 inhabitants. Anthropogenic emissions were responsible for 80% of these fatalities, whereas fuel combustion was directly responsible for 35%. Notably, the number of coal-burning-related fatalities dropped from 687,000 in 2015 to 561,000 in 2020, an 18% decrease. Stricter air pollution regulations in China and less coal use in Europe is mostly to blame for this decline. The quality of life is greatly impacted by the rise in air pollution, particularly in places with high population densities. Approximately 99 percent of people on the planet breathe air that has at least one pollutant level beyond tolerable limits, according to WHO estimates [1]. This issue disproportionately affects low and middle-income nations. Seven indicators spanning urban services, economics, culture and leisure, urban transportation, social connections, safety, and environmental conditions have been established based on the ISO 37120:2018 standard to evaluate the quality of life in cities [6], [7], [8]. Air and noise pollution, climatic comfort, cleanliness, and wastewater treatment are examples of environmental indicators. The combined impacts of several pollution sources, including industry, coal-fired power plants, and extensive transportation infrastructure, deteriorate air quality in major urban agglomerations. The purpose of this study is to examine the features of urban air

DOI: 10.37943/19IKWT5637



pollution and provide methodical approaches for tracking air quality and purification levels, which are essential for urban growth and planning. If these initiatives are successful, the urban population's quality of life may improve, and morbidity may decline.

This research aims to investigate the long-term memory of time series data on hazardous material concentrations gathered from four sites in Astana. The following tasks were established in order to accomplish this goal:

1. To better understand the dynamic behavior of the Hearst index, compute the Hearst index using fractal analysis of time series of air pollution data and find long-term memory models.

2. Examine variations in Astana's pollution levels with dangerous compounds, including variables like the weather and the city's closeness to possible pollution sources, particularly emissions from burning coal during the heating season.

**Materials and Methods**

Urban air pollution is a complicated issue brought on by a number of variables. Every city has unique features that impact pollution levels, including climate, topography, industrial activity, and transportation hubs. These elements must be examined independently in order to create efficient urban planning techniques that lower pollution levels and enhance inhabitants' health and standard of living. As previously stated, a rise in the concentration of dangerous chemicals in the air is linked to a number of illnesses, which further strains the healthcare system. Furthermore, the rise in consumption brought on by population expansion in big cities makes pollution much worse. Thus, it is essential to take action to enhance urban environmental conditions. Planning the location of possible pollution sources away from residential areas and increasing the usage of renewable energy sources are advised for rapidly expanding metropolitan regions.

Large urban agglomerations and rapidly expanding cities like Kazakhstan's capital, Astana, are particularly affected by this problem. The Bureau of National Statistics estimates that there were 1.458 million people living in Astana as of May 1, 2024. The city's yearly population growth rate surpasses 4%, indicating a healthy balance between migration and natural population increase. Along with extensive building, this population boom has increased the demand for water and power, which in turn has increased the emissions of toxic compounds. The city's heavy reliance on coal for heating systems is another significant element influencing the degree of environmental pollution.

Astana relies so largely on coal, during the heating season, the air's concentration of particulate matter (PM2.5) rises sharply. Astana is the second coldest capital in the world, with an average annual temperature of 3.5°C. Its location in a climatically distinct continental zone guarantees a long warm season. Despite the fact that the city's high average wind speeds aid in the natural dispersion of dangerous materials, the winter pollution levels nevertheless surpass WHO guidelines. Although Astana's average yearly wind speed is between 5 and 7 m/s [10], the city nonetheless has a significant pollution concentration, particularly during the lengthy warm season. The city's extensive usage of coal for heating systems and electricity generation accounts for a major portion of this. Central heating season of Astana often lasts for over seven months, increasing the amount of solid particles and combustion products in the air.

The city's total pollution level is influenced by a number of variables, including the fast population expansion of over 4% annually, in addition to the pollution caused by heating. The majority of Astana's pollution sources are stationary, in line with the research [11]. According to WHO guidelines [13], Astana's average yearly air quality index (AQI) is around 53 points, with summertime averages of 35 points [12]. However, because coal burning is so common in the winter, pollution levels rise sharply.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

75% of Kazakhstan's heat supply, including central heating systems and coal-based electricity generation, comes from coal combustion, according to the World Bank [14]. The nation's substantial reliance on coal has resulted in persistently high levels of air pollution; the average PM2.5 concentration is around 22 micrograms per cubic meter, far higher than the WHO-recommended threshold of 5 micrograms per cubic meter.

Kazakhstan's air is polluted by about 3.5 thousand industrial companies. They operate in the oil and gas complex, non-ferrous and ferrous metallurgy, and thermal power engineering. They are spread throughout 80 cities. As to the 2015 statistics, Karaganda (596.4 thousand tons), Pavlodar (552.9 thousand tons), Aktobe (134.3 thousand tons), East Kazakhstan (127.2 thousand tons), and Atyrau (110.6 thousand tons) are the five biggest cities in terms of air pollution (Figure 1). Approximately 2 million Kazakhs now reside in regions with high air pollution levels. The Karaganda region's industrial cities of Temirtau and Balkhash are on the list of cities with the worst air, as reported by Kazhydromet.

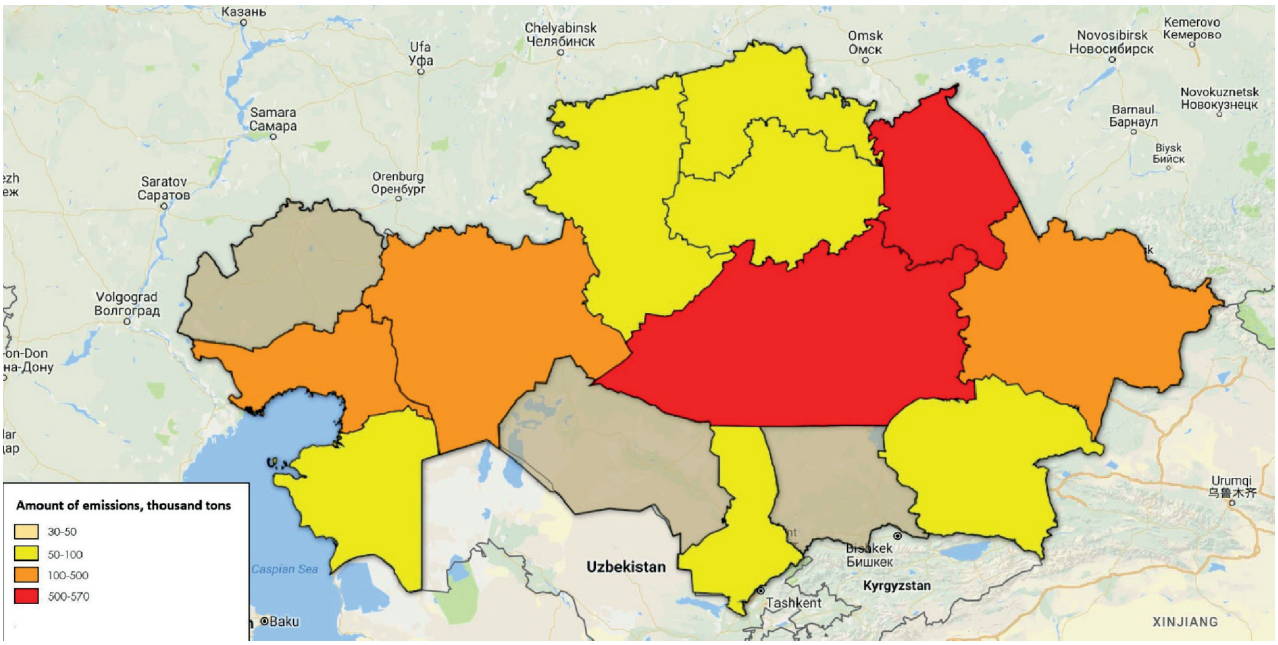


Figure 1. The amount of emissions in the Republic of Kazakhstan.

Introducing biotechnological filtration systems, which employ certain moss species to filter the air by releasing oxygen, is one potential way to enhance Astana's air quality. The study's proposal calls for the development of a network of moss-based biological filters that may be integrated into the Smart City system [15]. The Astana University of Information Technology is currently installing and testing the system. To find the best place for these filters, a model created [16] in a different study resolves the discrete optimization issue. This methodology enables filters to be strategically placed to enhance monitoring and air quality by forecasting air quality indicators in many places [17]. However, because of the nonlinear dispersion of pollution and the features of Astana, it is crucial to examine the structure of time series and the existence of long-term memory prior to utilizing techniques for predicting pollution levels. Fractal analysis techniques like range scaling analysis (R/S analysis) and trend deviation oscillation analysis (DMI) can be applied for this.

The quality of long-term memory has been applied to air pollution time series by several writers. For instance, the dynamic link between atmospheric carbon dioxide emissions and the Organization for Economic Cooperation and Development's (OECD) Industrial Production Index (IPI) across both short and long time periods was examined using multifractal approach-

DOI: 10.37943/19IKWT5637



es in [18]. In order to ascertain if the time series of changes in air pollution contains several distinguishing features, the MF-DFA approach was employed in [19] since the expansion of industry and urbanization upsets the natural equilibrium and raises air pollution. The city of Zhengzhou in China was the focus of this inquiry. Singularity spectrum, extended Hurst index, Renyi index, and logistic graphs of fluctuation functions were used to show the multifractality of Zhengzhou air quality time series. Another study [20] examined long-term data on variations in meteorological conditions, gaseous pollutants (CO, NO2, NOX, SO2, and O3), and particulate matter (PM2.5, PM10) concentrations at 412 stationary monitoring stations in Germany between January 2008 and December 2018.The Hearst index research indicates that Germany's air pollution levels exhibited a consistent trend during the course of the study. Other nations have carried out comparable research. Reference [21] examined the dynamics of time series of air quality indicators in 50 US states after examining the statistical properties of particulate matter (PM10 and PM2.5). The study made use of a long-term memory system that was partially integrated [21]. The study's overall findings differed by state: western regions showed greater resilience, whereas eastern states frequently saw a fall in the Hearst index. In [22], the long-term reliance of Mexico City on air pollution was assessed using the R/S methodology. However, this trait has evolved throughout time. R/S analysis was employed in [23] to demonstrate the erratic behavior and long-term memory of PM10 particle emissions gathered in Athens, Greece. Furthermore, a number of studies have estimated the Hurst parameter using nonparametric techniques, such as deviant variation analysis (DFA) [24] and multifractal deviant oscillation analysis (MF-DFA) [25]. Specifically, the DFA approach was used to estimate shorter time series of air pollution [26], [27]. Considering the pertinent empirical features, these techniques are all appropriate for examining the time series of air pollution as they are all founded on the idea of long-term memory and are connected. Fractal analysis of air pollution time series is therefore crucial for assessing the persistence of hazardous material emissions into the atmosphere and determining long-term memory, particularly in metropolitan agglomerations.

**Experimental Methods and Area of Study**

More than 1.4 million people live in the metropolitan agglomeration of Astana [9]. Situated on a steppe plain, the city occupies an area of 797.33 km$^2$. Its climate is distinctly continental, with long, severe winters marked by strong winds and dry summers. Astana's geographic position has a beneficial impact on air pollution levels since the wind direction helps disperse dangerous compounds, bringing their concentration down below WHO guidelines. However, the operation of coal-fired power plants for district heating and the usage of coal for heating residential buildings cause pollution levels to rise dramatically during the lengthy cold season, which lasts for about seven months.

As part of a collaborative government project, private organizations including Astana IT University (Astana, Republic of Kazakhstan), S. Toraighyrov University (Pavlodar, Republic of Kazakhstan), and Promanalit LLP (Pavlodar, Republic of Kazakhstan) and Prometeo Chain System KZ Ltd (Astana, Republic of Kazakhstan) evaluated the amount of air pollution in Astana after examining the PM10 and PM2.5 particulate matter content (refer to the financing section). Pollution was measured at four facilities in Astana (see Figure 2). Four times a day, at 13:00, 19:00, 13:00, and 19:00, data is gathered. using analyzers that are automatically employed for this purpose. Each automatic analyzer is equipped with microprocessors that generate time series data on pollution levels. All collected values are transmitted to a central station for processing and storage. The data analyzed in this study spans the period from June 1, 2021, to July 30, 2022.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

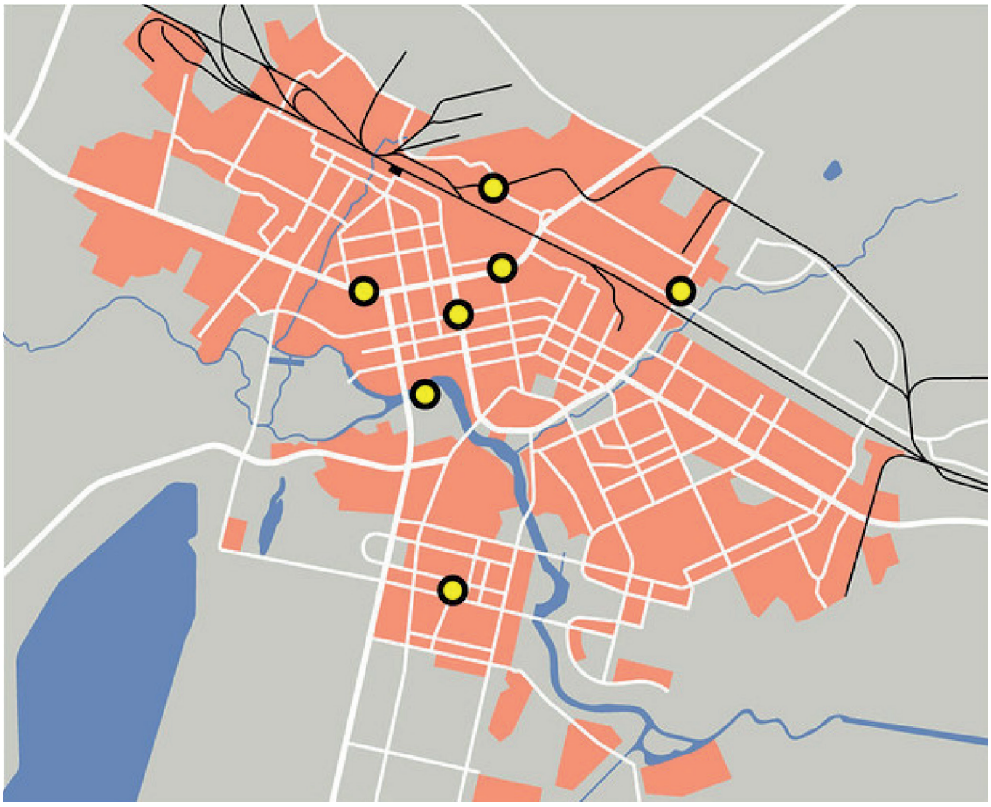

Figure 2. Location of air pollution measurement stations within the city of Astana.

Integrated model for forecasting time series of environmental pollution parameters

The complete model for predicting a time series of environmental pollution parameters is discussed in this paragraph. It combines several forecasting models created by predictive statistical analysis of pollution indicators. One of the key characteristics of this model is its capacity to incorporate popular forecasting methods while permitting parameters to be changed in response to the outcomes of fractal analysis-based prediction estimations.

The task is to use a time series

$$Q = \left(q(t_1), q(t_2), \ldots, q(t_n)\right)$$

to determine estimates of the values in order to generate the most accurate prediction, that is, to establish the behavior of the time series of pollutant parameters for a specific number of points ahead.

$$\hat{q}(t_{n+1}), \hat{q}(t_{n+2}), \ldots, \hat{q}(t_{n+k}),$$

where $\hat{q}(t_{n+i})$ – time series forecast $Q = \left(q(t_1), q(t_2), \ldots, q(t_n)\right)$, $i = \overline{1,k}$.

The challenge must be solved by identifying a functional dependency that, given the time series' known values, would roughly correspond to the anticipated value needed.

$$\hat{q}(t_{n+1}) = \Phi\left(q(t_c), q(t_{c+1}), \ldots, q(t_n)\right),\ c < n,$$

$$\hat{q}(t_{n+2}) = \Phi\left(q(t_{c+1}), q(t_{c+2}), \ldots, q(t_n), \hat{q}(t_{n+1})\right),$$

$$\ldots$$

$$\hat{q}(t_{n+k}) = \Phi\left(q(t_{c+k-1}), q(t_{c+k}), \ldots, q(t_n), \hat{q}(t_{n+1}), \ldots, \hat{q}(t_{n+k-1})\right).$$

Several metrics, including standard deviation, average absolute error, and average relative error, can be used to measure the forecast's accuracy.

DOI: 10.37943/19IKWT5637



Let a set of models be given $\Phi_1, \Phi_2, \ldots, \Phi_M$, which, based on the time series Q allow to make the most accurate forecast, i.e. find the value:

$$\hat{q}_j(t_{n+1}) = \Phi_j\left(\alpha_1^j, \alpha_2^j \ldots \alpha_m^j, q(t_c), q(t_{c+1}), \ldots, q(t_n)\right),\ c < n, \tag{6}$$

$$\hat{q}_j(t_{n+2}) = \Phi_j\left(\alpha_1^j, \alpha_2^j \ldots \alpha_m^j, q(t_{c+1}), q(t_{c+2}), \ldots, q(t_n), \hat{q}(t_{n+1})\right), \tag{7}$$

$$\ldots$$

$$\hat{q}_j(t_{n+k}) = \Phi_j\left(\alpha_1^j, \alpha_2^j \ldots \alpha_m^j, q(t_{c+k-1}), q(t_{c+k}), \ldots, q(t_n), \hat{q}(t_{n+1}), \ldots, \hat{q}(t_{n+k-1})\right),$$

$$j = \overline{1, M},$$

where $\hat{q}_j(t_{n+i})$ – time series forecast Q, $i = \overline{1,k}$ based on the model $j = \overline{1,M}$, $\alpha_d^j$ – parameters of the forecasting model j, $d = \overline{1,m}$, m – number of parameters of the forecasting model j.

Let the model $\Phi_1$ is represented by the usual exponential model of order $p \geq 0$. Exponential order model $p \geq 0$ is determined by the formula:

$$x_n^{[p]} = \alpha \cdot x_n^{[p-1]} + (1-\alpha) x_{n-1}^{[p]},$$

where $x_n^{[0]} = q(t_1)$, $x_0, x_0^{[2]}, \ldots$ – beginning conditions for the relevant order's exponential averages [28], $\hat{q}_j(t_{n+1}) = x_n^{[p]}$, $\alpha \in [0,1]$. That is, this model will be defined as $\Phi_1\left(\alpha, q(t_1), q(t_2), \ldots, q(t_n)\right)$.

Let the model $\Phi_2$ is represented by the Holt exponential smoothing model, which is used to model time series with a pronounced trend component [29]:

$$\hat{q}_2(t_{n+1}) = x_n + y_n,$$

$$x_n = \alpha_1 q(t_n) + (1-\alpha_1)(x_{n-1} + y_{n-1}),$$

$$y_n = \alpha_2 (x_n - x_{n-1}) + (1-\alpha_2) y_{n-1},$$

$\hat{q}_2(t_{n+1})$ – a prediction made using the Holt time series model one point in advance Q, $\alpha_1, \alpha_2 \in [0,1]$, accordingly, the model has the form $\Phi_2\left(\alpha_1, \alpha_2, q(t_1), q(t_2), \ldots, q(t_n)\right)$.

The Winters model $\Phi_3$ It is applied to systems that show multiplicative seasonality and an additive trend component. In this model, the time series value is analyzed separately by smoothing the trend, the seasonal component, and the non-seasonal component.

$$\hat{q}_3(t_{n+1}) = (x_n + y_n) s_{n-P+1},$$

$$x_n = \alpha_1 \frac{q(t_n)}{s_{n-P}} + (1-\alpha_1)(x_{n-1} + y_{n-1}),$$

$$y_n = \alpha_2 (x_n - x_{n-1}) + (1-\alpha_2) y_{n-1},$$

$$s_n = \alpha_3 \frac{q(t_n)}{s_n} + (1-\alpha_3) s_{n-P},$$

where $\alpha_1, \alpha_2, \alpha_3 \in [0,1]$ – smoothing parameters, P – seasonal cycle period, $s_n$ – assessment of the seasonal component of the model. This model is denoted by $\Phi_3\left(\alpha_1, \alpha_2, \alpha_3, q(t_1), q(t_2), \ldots, q(t_n)\right)$.

Autoregressive model $\Phi_4$ is determined by the formula:

$$\hat{q}(t_{n+1}) = \gamma_0 + \gamma_1 q(t_{n-1}) + \gamma_2 q(t_{n-2}) + \ldots + \gamma_c q(t_{n-c}),$$

Undefined parameters are computed using the criterion to minimize the sum of RMS errors. Since the memory capacity determines how many points the autoregressive model should employ, we can construct the model as $\Phi_4\left(q(t_{n-c+1}), q(t_{n-c+2}), \ldots, q(t_n)\right)$.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

The number of points utilized to compute the prediction value is determined by a parameter that defines the moving average model. The larger the memory in the time series, the larger this parameter should be. That is, the model $\Phi_5\left(q\left(t_{n-c+1}\right),q\left(t_{n-c+2}\right),\ldots,q\left(t_n\right)\right)$ is defined as follows:

$$\hat{q}\left(t_{n+1}\right)=\frac{1}{\alpha}\sum_{j=0}^{c-1}q\left(t_{n-j}\right),\ \alpha>0.$$

Weighted moving average with a set of normalized weights $\left\{\omega_1,\omega_2,\ldots,\omega_c\right\}$, $\sum_{j=1}^{c}\omega_j=1$, is determined by the formula:

$$\hat{q}\left(t_{n+1}\right)=\frac{1}{c}\sum_{j=0}^{c-1}\omega_{j+1}q\left(t_{n-j}\right),\ c>0.$$

This model is denoted by $\Phi_6\left(q\left(t_{n-c+1}\right),q\left(t_{n-c+2}\right),\ldots,q\left(t_n\right)\right)$. The number of models can be increased. However, it should be noted that the initial parameters should be chosen considering the results of the pre-prognostic analysis. By using this method, it will be feasible to provide a more accurate prediction of environmental contamination parameters and minimize forecasting error for time series. For the development of an efficient environmental monitoring system, this precision is essential.

A comprehensive forecasting model that incorporates all the previously mentioned models should be developed, along with the findings from the pre-prognostic fractal analysis of the time series. The ability to dynamically modify your settings in response to changes in the environment is how this approach varies from others. Each of the described models produces a certain level of error when implemented. The accuracy of the model is determined by the minimum forecasting error. Let the time series plot $Q^o=\left(q\left(t_z\right),q\left(t_{z+1}\right),\ldots,q\left(t_n\right)\right)$ the accuracy of the forecasting models is monitored, then the error can be calculated for each model $G_n^i\left(Q^o,Q\right)$ by the formula:

$$G_n^i\left(Q^o,Q\right)=\sqrt{\frac{1}{n-z+1}\sum_{j=z}^{n}\left(\hat{q}_i\left(t_j\right)-q_i\left(t_j\right)\right)^2},\ i=\overline{1,6},$$

$G_n^i\left(Q^o,Q\right)$ – prediction error of point $q(t_n)$, based on the model $\Phi_1$.

Using the following formula, we determine the criterion for choosing a model to execute the forecast for each model:

$$\tilde{G}_n^i=\eta G_n^i(Q^o,Q)+(1-\eta)\tilde{G}_{n-1}^i,\ \eta\in[0,1],$$

where $G_n^i$ – exponentially smoothed point prediction error $q(t_n)$, based on the model $\Phi_1$.

If the error argument is the model on which it was calculated, i.e. $\tilde{G}_n^i=\tilde{G}_n(\Phi_i)$. Then, the model for which the condition of minimum forecasting error is met is selected to perform the forecast:

$$\Phi_i^*=arg\min_i\left(\tilde{G}_n^i\right).$$

For the selected list of models, we will use their connection with the results of the pre-forecast analysis. That is, with the calculated value of $H(Q^n)$.

The smoothing parameter for the first model indicates how much the series' past values affect the predicted outcome. If this parameter approaches one, the forecast produced by this model will closely resemble a naive forecast. Conversely, a forecast may still appear naive if the series being predicted exhibits strong persistence. Therefore, a rational formula can be established to determine the optimal smoothing parameter within the model.

DOI: 10.37943/19IKWT5637



$$\Phi_1\left(\alpha, q\left(t_1\right), q\left(t_2\right), \ldots, q\left(t_n\right)\right):$$

$$\alpha = \begin{cases} H\left(Q^n\right), H\left(Q^n\right) \geq 0{,}5 \\ 0.5, H\left(Q^n\right) < 0{,}5 \end{cases}.$$

In the second model $\Phi_2\left(\alpha_1, \alpha_2, q\left(t_1\right), q\left(t_2\right), \ldots, q\left(t_n\right)\right)$ first smoothing parameter $\alpha_1$ – determines the trend, and the second $\alpha_2$ – the random component. Accordingly, these parameters can be determined by the formula:

$$\alpha_1 = \begin{cases} H\left(Q^n\right), H\left(Q^n\right) \geq 0{,}5 \\ 0.5, H\left(Q^n\right) < 0{,}5 \end{cases},$$

$$\alpha_2 = \begin{cases} \max\left\{1, 0.5 + 10\left|H_h^T - H\left(Q^n\right)\right|\right\}, 0.5 \leq H\left(Q^n\right) \leq H_h^T \\ 0.5, H\left(Q^n\right) > H_h^T \text{ or } H\left(Q^n\right) < 0.5 \end{cases}.$$

In the third model $\Phi_3\left(\alpha_1, \alpha_2, \alpha_3, q\left(t_1\right), q\left(t_2\right), \ldots, q\left(t_n\right)\right)$ first smoothing parameter $\alpha_1$ – determines the trend, and the second $\alpha_2$ – random component, the third – $\alpha_3$ – seasonality. The following formula may be used to compute these parameters if the trend changes points of the statistics curve V match the seasonality of P:

$$\alpha_1, \alpha_3 = \begin{cases} H\left(Q^n\right), H\left(Q^n\right) \geq 0{,}5 \\ 0.5, H\left(Q^n\right) < 0{,}5 \end{cases},$$

$$\alpha_2 = \begin{cases} \max\left\{1, 0.5 + 10\left|H_h^T - H\left(Q^n\right)\right|\right\}, 0.5 \leq H\left(Q^n\right) \leq H_h^T \\ 0.5, H\left(Q^n\right) > H_h^T \text{ or } H\left(Q^n\right) < 0.5 \end{cases}.$$

In the models $\Phi_4\left(q\left(t_{n-c+1}\right), q\left(t_{n-c+2}\right), \ldots, q\left(t_n\right)\right)$, $\Phi_5\left(q\left(t_{n-c+1}\right), q\left(t_{n-c+2}\right), \ldots, q\left(t_n\right)\right)$ and $\Phi_6\left(q\left(t_{n-c+1}\right), q\left(t_{n-c+2}\right), \ldots, q\left(t_n\right)\right)$ the value of the parameter is determined by the presence of long-term memory in the time series. It can be determined by visual inspection of the V statistics curve. If the point that shows a sharp change in the upward and downward trend of this curve is equal to P, $P \in \mathrm{N}$, then the parameter c can be set to P, i.e., models of this form will be built:

$$\Phi_t\left(q\left(t_{n-P+1}\right), q\left(t_{n-P+2}\right), \ldots, q\left(t_n\right)\right),\ t = \overline{4,6}.$$

This model is unique in that it is adaptable in addition to taking into account the findings of the predictive fractal analysis. In other words, it could adjust the model's parameters in response to environmental changes, which is particularly crucial when utilizing these models in environmental monitoring systems.

If a time series fluctuates above a certain threshold, it is likely to indicate a stable pollution situation, where either a decrease or increase in pollution levels occurs, but the trend remains predictable (see Fig. 3). To forecast pollution levels in this scenario, methods that account for a strong trend component should be employed, which is crucial for effectively building the system for monitoring the environment. Conversely, if such a time series' values fall under the established threshold, it can signal an emergency or unstable changes in pollution levels (see Fig. 4). In this case, models that incorporate a significant random component should be utilized for predicting pollution levels.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

Additionally, it may indicate a greater impact of random events on the process in question if the time series values were originally above the threshold value before beginning to abruptly decrease and fall below it. In this situation, you must examine the V statistics' value to confirm that the time series' long-term memory was indeed lost. This, in turn, means possible emission emergencies that need to be addressed promptly. Figure 5 demonstratews such a case when the value of the Hurst index decreases sharply, crossing the threshold level. All other cases that may occur in the case of time series analysis

$$H(Q)=\left(H(Q^{1,w}),H(Q^{2,w+1}),\ldots,H(Q^{n-w+1,n})\right)$$

should be analyzed separately.

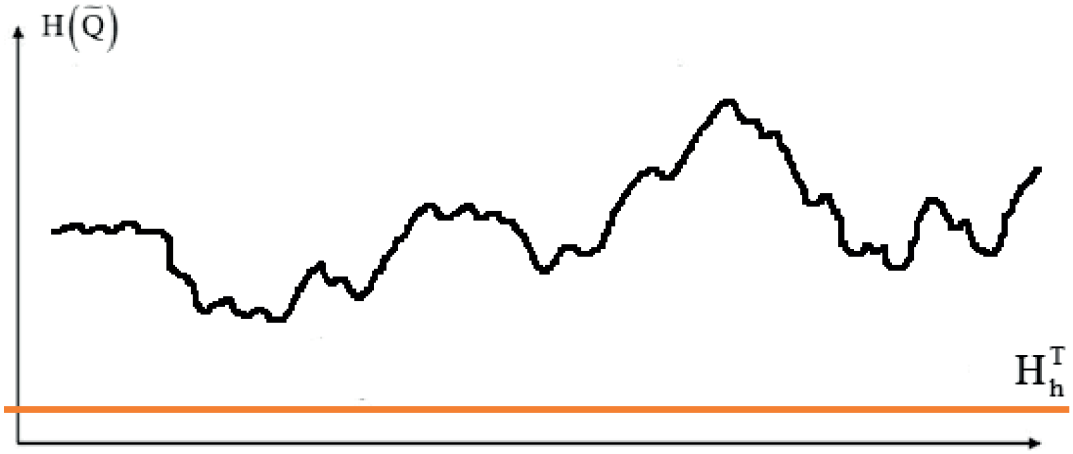


Figure 3. The position of the time series $H(\tilde{Q})$ is above the threshold value $H_h^T$.

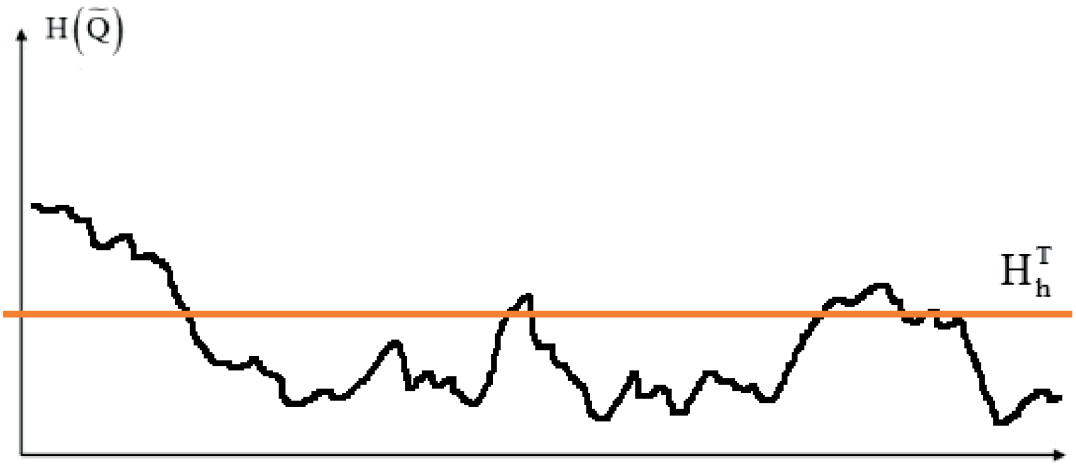


Figure 4. The position of the time series $H(\tilde{Q})$ is below the threshold value $H_h^T$.

DOI: 10.37943/19IKWT5637



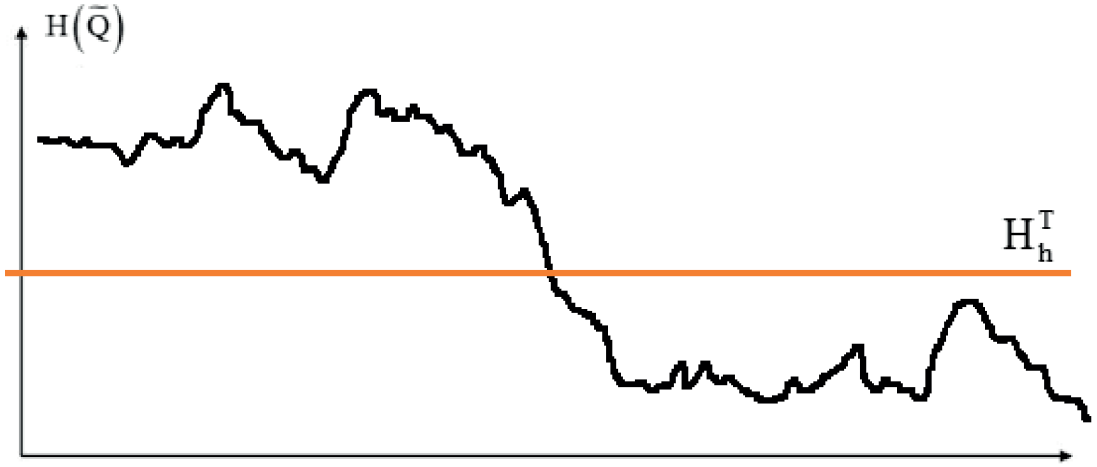


Figure 5. The position of the time series $H(\tilde{Q})$ decreases sharply from above the threshold $H_h^T$ to below the threshold $H_h^T$.

Critical statistical estimates that may be used to develop forecasting models and an efficient environmental monitoring system are provided by the fractal analysis of the time series of environmental pollution parameters.

**Results**

In the period from June 1, 2021, to July 30, 2022, air pollution levels were monitored in four settlements of Astana, Republic of Kazakhstan, to assess air quality and study long-term memory. The data was collected four times a day at 13:00, 19:00, 19:00 and 19:00. The location of the stations is shown in Figure 2. Three stations are located on the right bank, and one on the left bank of the Ishim River. At each station, the content of significant pollutants was monitored every six hours, including particulate matter (PM10, PM2.5), sulfur dioxide, carbon monoxide, nitrogen dioxide, hydrogen fluoride, benzopyrene, sulfates, nitric oxide, benzene, ethylbenzene, chlorobenzene, paraxylene, methaxylene, cumene, orthoxylene, cadmium (Cd), copper (Cu), zinc (Zn), chromium VI (Cr), arsenic (As) and other significant pollutants. Since most of these indicators had values in time series close to zero, the study focused on the cumulative time series of suspended particles PM2.5 and PM10.

The dataset is accessible in CSV format in other publications. Descriptive statistics for time series data are shown in Table 1.

Table 1. Descriptive statistics of pollution time series.

| Indicator | Station | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Number of observations | 1384 | 1038 | 1038 | 1038 |
| Mean | 0.219487 | 0.225809 | 0.222071 | 0.225241 |
| Std | 0.354448 | 0.421328 | 0.366821 | 0.417980 |
| Min | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| Max | 3.400000 | 3.900000 | 3.400000 | 3.900000 |

The research was conducted using a sliding window of one calendar month with 120 data points. The rolling() function from the Pandas package was used to do a fractal analysis while examining time data. Descriptive statistics for the smoothed series data are shown in Table 2.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

Table 2. Descriptive statistics of time series of pollution after smoothing.

| Indicator | Station | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Number of observations | 1280 | 1038 | 1038 | 1038 |
| Mean | 0.227144 | 0.222413 | 0.226039 | 0.221515 |
| Std | 0.216896 | 0.233680 | 0.234858 | 0.219416 |
| Min | 0.090833 | 0.082500 | 0.085000 | 0.095833 |
| Max | 1.084167 | 1.229167 | 1.141667 | 1.145833 |

The issue of missing data was resolved by smoothing the data, which successfully decreased the time series' variance and range.

To investigate the acute emissions of PM10 and PM2.5 particulate matter, data on the locations and operations of thermal power plants in Astana were gathered (see Fig. 3). Stations 2, 3, and 4 are situated closer to thermal power plants, it was observed. Therefore, the primary cause of the consistent rise in pollution levels seen in the time series is station 1's considerable distance from these sources of pollution. Table 3 displays the maximum capacity of thermal power plants as well as the separation between stations and air quality monitoring stations.

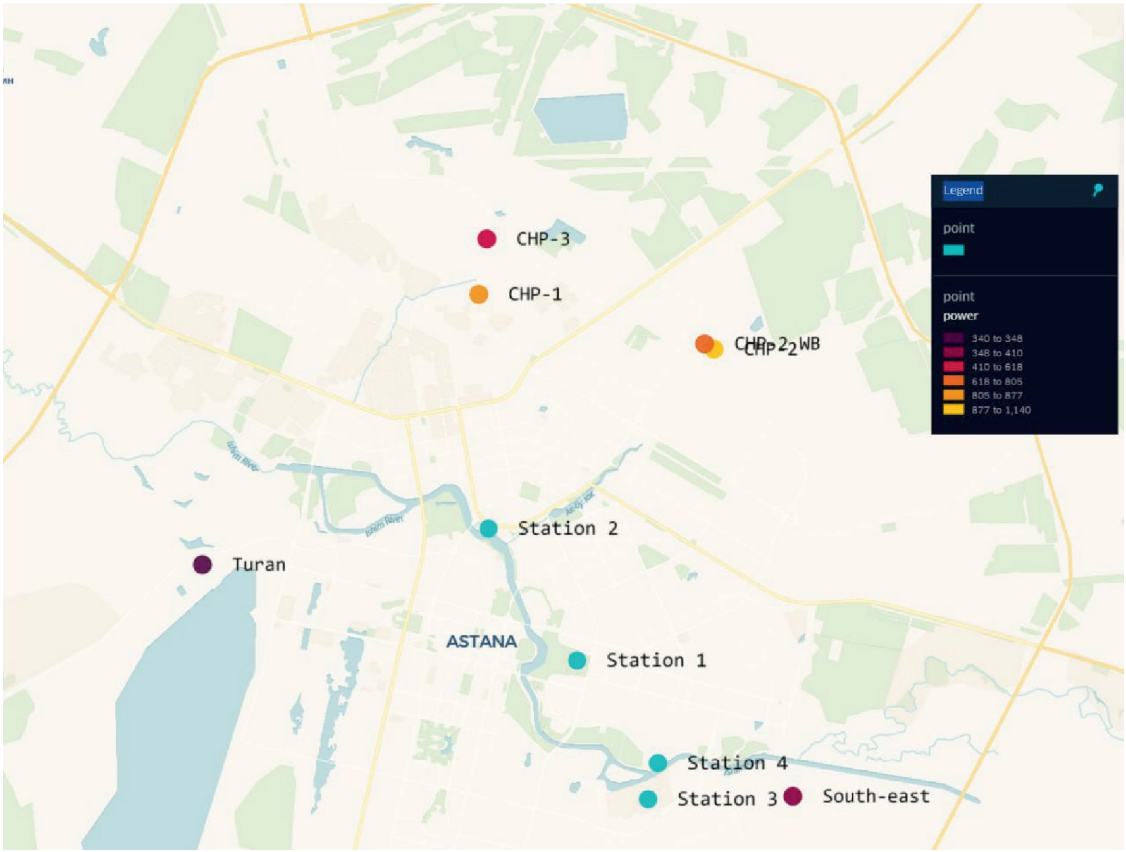


Figure 3. Placement of air pollution monitoring stations and thermal power plants.

DOI: 10.37943/19IKWT5637



Table 3. Distances between observation stations and thermal power plants.

| Thermal Power Plant | Maximum Capacity, Gcal/Year | Distance to the Observation Station, km | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| CHP 1 | 824 | 8.93 | 5.52 | 12.55 | 11.83 |
| CHP 2 | 1140 | 8.01 | 6.8 | 10.71 | 9.84 |
| CHP 2 WB | 795 | 8.04 | 6.71 | 10.81 | 9.94 |
| CHP 3 | 440 | 10.15 | 6.81 | 13.73 | 12.99 |
| South-west | 350 | 6.02 | 9.57 | 3.42 | 3.28 |
| Turan | 340 | 9.14 | 6.82 | 11.9 | 11.75 |
| Min distance | | 6.02 | 5.52 | 3.42 | 3.28 |

Long-term memory research was made easier by calculating the Hurst index for every time series using R/S analysis. Additionally, a map of thermal networks was examined and data on the locations of thermal power plants was gathered in light of the unique features of the city's pollution.

This method may be utilized as an indicator of the local environment, according to the findings of dynamic computations of the Hearst index and the identification of potential pollution sources caused by coal burning during Astana's heating season. When the Hearst indicator shows that a time series of pollution is persistent, the source of the pollution is most likely situated distant from the monitoring station. In these situations, a gradual rise in pollution levels may be anticipated, allowing for the early detection of potentially disastrous pollution situations.

However, if the time series shows instability or unpredictability, it indicates the critical or near-critical condition of the environment with unpredictable emissions when the monitoring station is situated close to the pollution source. The timely adoption of actions to lessen the consequences of pollution can be facilitated by this knowledge, which is essential for sensible environmental management.

**Discussion**

The findings derived from the fractal R/S analysis of air pollution time series in Astana, collected from four monitoring stations, highlight significant environmental challenges faced by the city. An examination of the air pollutant concentration time series recorded at these four locations revealed the presence of long memory. The fractal R/S analysis yielded Hurst exponents of 0.723, 0.548, 0.442, and 0.462 for the respective stations. Additionally, the dynamics of the Hurst exponent were explored using the sliding window method.

According to the research, station 1's time series demonstrates persistence and long-term memory, but station 2's time series is described as almost random. The time series from stations 3 and 4, on the other hand, show anti-persistence. Notably, these Hurst exponent values were impacted by the notable spike in pollution levels that was recorded in October 2021.

During the heating season from October to December, a significant spike in PM10 and PM2.5 particulate matter levels were recorded, reaching critical thresholds. This rise can be attributed to the inefficiencies of thermal and coal-fired power plants designed for centralized heating, as well as the extensive use of combustible fuels for residential heating. The location of monitoring stations relative to pollutant emission sources significantly impacts the structure of the pollutant concentration time series. Specifically, time series from stations near thermal power plants exhibit anti-persistent or random characteristics, complicating their analysis with standard methods. This phenomenon signifies abrupt emissions that deviate from a consistent trend in harmful substance concentrations, potentially indicating a critical environmental situation that necessitates prompt action from city environmental services.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

For the time series associated with station 1, situated in a park area and relatively distant from pollution sources, the Hurst exponent exceeds 0.7, indicating persistence. Its persistent character is further supported by the Hurst exponent's dynamic analysis for this series. But like other locations, this time series also shows a notable rise at the start of the summer and a fall in the average daily air temperature. Because of the less noticeable rise, long-term memory is confirmed, and the Hurst index may be calculated more easily. These results align with similar research, such as the fractal analysis of PM10 concentration time series in Athens [23].

The air pollution data for Astana, prepared for analysis, were collected at four monitoring stations with a six-hour interval. While the dataset is ample for analysis, it spans a limited timeframe of one year (from June 2021 to July 2022). This limitation restricts the ability to evaluate how pollution levels in Astana changed during the heating seasons in previous years, which could provide valuable insights into annual trends in pollutant emissions during the autumn and winter months.

Also, the factor that may impact conclusions about air quality in Astana is the geographical distribution of the monitoring stations. Only one of the four stations is on the left side of the Ishim River; the other three are on the right bank. The right bank represents an older part of the city, developed several decades ago, and is characterized by aging thermal power plants and infrastructure that require modernization. Furthermore, this area contains many private households that typically rely on burning fossil fuels for heating. As a result, a significant escalation in air pollution was likely observed during the start of the heating season, from October to December.

The Ishim River's left bank, on the other hand, is a relatively recent location, with most of the building occurring within the previous 20 years. This region has experienced improvements in infrastructure and greater access to natural gas for heating. If monitoring stations were positioned in this part of the city, it is likely that critical levels of pollution would not have been detected.

The findings from this study are significant for the city's environmental monitoring services, as prompt action in response to such pollutant emissions is crucial for safeguarding the health of residents and ensuring environmental safety in the region.

**Conclusion**

Thermal power plants in Astana, which are situated closer to monitoring stations 2, 3, and 4, are responsible for the noticeable rise in PM10 and PM2.5 particulate matter levels. In contrast, the more gradual rise in pollution levels recorded at station 1 is attributed to its greater distance from these pollution sources. Furthermore, weather conditions also play a role in influencing the intensity of coal combustion and the resulting emissions.

The state's strategic goal is to facilitate a transition to heating systems that utilize natural gas, alongside the development of the necessary infrastructure. Future research could focus on examining the effects of this transition on Astana's environmental conditions, particularly concerning air pollution levels of PM10 and PM2.5. Moreover, investigating the impact of the city's rapid population growth, which exceeds 4% annually, could provide insights into the pressures placed on the energy system. One promising approach for mitigating air pollution is the application of modified biochar, especially for significant emitters like thermal power plants [30].

The findings presented in this study are crucial for urban development, particularly regarding the sustainable growth of rapidly expanding urban agglomerations. The results suggest that fractal time series analysis can serve as a valuable tool for assessing the ecological state of a region. This technique can be used to forecast possible critical pollution levels if a pollution source is situated distant from a monitoring station and the Hurst exponent shows a con-

sistent time series. On the other hand, the station's closeness to pollution sources may cause anti-persistence or unpredictability in the pollution time series.

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

## References

[1] WHO. Air Pollution. 2024. Available online: https://www.who.int/health-topics/air-pollution#tab=tab_1 (accessed on 2 July 2024).
[2] Romanello, M.; Di Napoli, C.; Drummond, P.; Green, C.; Kennard, H.; Lampard, P.; Scamman, D.; Arnell, N.; Ayeb-Karlsson, S.; Ford, L.B.; et al. The 2022 report of the Lancet Countdown on health and climate change: Health at the mercy of fossil fuels. *Lancet* 2022, *400*, 1619–1654. https://doi.org/10.1016/S0140-6736(22)01540-9 .
[3] Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and health impacts of air pollution: A review. *Front. Public Health* 2020, *8*, 14.
[4] GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019, *393*, 1958–1972.
[5] Pereira, F.; Nadeau, K. Climate Change, Fossil-Fuel Pollution, and Children's Health. *N. Z. J. Med.* 2022, *386*, 24, 2303–2314. https://doi.org/10.1056/NEJMra2117706 .
[6] *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 2 July 2024).
[7] 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.
[8] To, T.; Zhu, J.; Terebessy, E.; Zhang, K.; Fong, I.; Pinault, L.; Jerrett, M.; Robichaud, A.; Menard, R.; van Donkelaar, A.; et al. Does exposure to air pollution increase the risk of acute care in young children with asthma? An Ontario, Canada study. *Environ. Res.* 2021, *199*, 111302.
[9] Bureau of National Statistics of Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Astana City. Summary of the Socio-Economic Development of the Region. 2018. Available online: https://stat.gov.kz/en/region/astana/ (accessed on 2 July 2024).
[10] Global Wind Atlas, 2018. Available online: https://globalwindatlas.info/en (accessed on 2 July 2024).
[11] Kerimray, A.; Bakdolotov, A.; Sarbassov, Y.; Inglezakis, V.; Poulopoulos, S. Air pollution in Astana: Analysis of Recent Trends and Air Quality Monitoring System. *Mater. Today Proc.* 2018, *5*, 22749–22758.
[12] IQAir. Air Quality in Astana. Available online: https://www.iqair.com/us/kazakhstan/astana (accessed on 2 July 2024).
[13] IQAir. New WHO Air Quality Guidelines Will Save Lives. Available online: https://www.iqair.com/us/newsroom/2021-WHOair-quality-guidelines (accessed on 2 July 2024).
[14] The World Bank. Strengthening Public Finance for Inclusive and Resilient Growth. Public Finance Review. 2023. Available online: https://thedocs.worldbank.org/en/doc/90dbef81d187b403bbb2a9acc2f460d8-0080062024/original/Kazakhstan-PFR-full-Report-January-2024-en.pdf (accessed on 2 July 2024).
[15] 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. https://doi.org/10.3390/urbansci7040104 .

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

[16] 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.
[17] He, Y.; Kuchansky, A.; Paliy, S.; Shabala, Y. Problems in Air Quality Monitoring and Assessment. In Proceedings of the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 28–30 April 2021; pp. 28–30. https://doi.org/10.1109/SIST50301.2021.9465915 .
[18] Bozkus, S.K.; Kahyaoglu, H.; Mahamane Lawali, A.M. Multifractal analysis of atmospheric carbon emissions and OECD industrial production index. *Int. J. Clim. Change Strateg. Manag.* 2020, *12*, 411–430. https://doi.org/10.1108/IJCCSM-08-2019-0050 .
[19] 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, *2226565*, 17. https://doi.org/10.1155/2022/2226565 .
[20] Liu, X.; Hadiatullah, H.; Tai, P.; Xu, Y.; Zhang, X.; Schnelle-Kreis, J.; Schloter-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. https://doi.org/10.1016/j.envpol.2021.116732 .
[21] Gil-Alana, L.A.; Yaya, O.S.; Awolaja, O.G.; Cristofaro, L. Long Memory and Time Trends in Particulate Matter Pollution (PM2.5 and PM10) in the 50 U.S. States. *J. Appl. Meteor. Climatol.* 2020, *59*, 1351–1367. https://doi.org/10.1175/JAMC-D-20-0040.1.
[22] 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. https://doi.org/10.1016/j.physa.2015.02.009 .
[23] Nikolopoulos, D.; Moustris, K.; Petraki, E.; Koulougliotis, D.; Cantzos, D. Fractal and Long-Memory Traces in PM10 Time Series in Athens, Greece. *Environments* 2019, *6*, 29. https://doi.org/10.3390/environments6030029 .
[24] Varotsos, C.; Ondov, J.; Efstathiou, M. Scaling properties of air pollution in Athens, Greece and Baltimore, Maryland. *Atmos. Environ.* 2005, *39*, 4041–4047. https://doi.org/10.1016/j.atmosenv.2005.03.024 .
[25] 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. https://doi.org/10.1016/j.scitotenv.2015.06.065.
[26] Thompson, J.R.; Wilson, J.R. Multifractal detrended fluctuation analysis: Practical applications to financial time series. *Comput. Simul.* 2016, *126*, 63–88. https://doi.org/10.1016/j.matcom.2016.03.003.
[27] 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. https://doi.org/10.1016/S0378-4371(02)01383-3 .
[28] Anis, A.; Lloyd, E. The expected value of the adjusted rescaled Hurst Range of independent normal summands. *Biometrika* 1976, *63*, 111–116.
[29] Peters, E.E. *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1994; p. 336.
[30] Jia, L.; Cheng, P.; Yu, Y.; Chen, S.H.; Wang, C.X.; He, L.; Nie, H.T.; Wang, J.C.; Zhang, J.C.; Fan, B.G.; et al. Regeneration mechanism of a novel high-performance biochar mercury adsorbent directionally modified by multimetal multilayer loading. *J. Environ. Manag.* 2023, *326*, 116790. https://doi.org/10.1016/j.jenvman.2022.116790