Vol. 6, No. 3, 2021

METHODS OF COMPREHENSIVE STATISTICAL ANALYSIS OF AGGLOMERATION ATMOSPHERIC AIR POLLUTION

Inesa Loyeva[1], Salem Rabeea Bazar[2], Oleksii Burhaz[1], Olena Vladymyrova[1]

[1]Odessa State Environmental University
15, Lvivska Str., Odesa, 65016, Ukraine
[2]Environmental Sciences & Marine Biology College, Hadhramout University
PO Box 50512, Al Mukalla, Hadramaut, Yemen
alexburgaz84@gmail.com

https://doi.org/10.23939/ep2021.03.130

Received: 27.05.2021

Abstract. The article presents an algorithm of a complicated statistical analysis of atmospheric air pollution over certain urban territories, based on methods of multivariate statistical analysis and a statistical decisions theory. Statistical analysis of the concentrations obtained from the observation posts network provides the data on the atmospheric air pollution background characteristics. This analysis enables us to find the spatial correlations between the concentrations of the ingredient under various synoptic conditions. The proposed set of statistical methods is advisable to be applied in the development of a state monitoring program for atmospheric air protection over specific zones or territories (urban agglomeration) as part of the implementation of a new management procedure for state monitoring of atmospheric air protection in Ukraine.

Keywords: air pollution; statistical methods; level of pollution assessment; monitoring.

1. Introduction

Air pollution destroys human health, affects food security, hinders economic development, directly contributes to climate change and degrades the environment that sustains our livelihoods.

The importance of air protection issue is especially acute in urban areas, where industrial complexes, traffic activities and populated areas are predominant. Consequently, there is an actual need for a comprehensive analysis of physical-geographical, technogenic, socio-economic processes occurring in urbanized areas.

The intensity of deposition or dispersion of air pollutants and the conditions for their transboundary transfer depend on the state of the atmospheric boundary layer. (Berljand, 1975). Therefore, when studying the processes of atmospheric pollution, it is necessary to study the meteorological conditions.

Without solving the problem of organizing a consistent monitoring system in atmospheric air protection that meets the requirements of optimality concerning space and time to objectively assess the impact of air pollution on the environment, health and livelihood of the population it is impossible to ensure control of atmospheric air quality, assess and predict its changes and the degree of danger, develop scientifically grounded recommendations for making managerial decisions.

2. Results and Discussion

The previous study of the state of atmospheric air is directly related to the solution of this problem by using statistical analysis based on modern methods.

The set of statistical methods, which, in our opinion, can solve this problem more comprehensively, is shown in a block diagram, Fig. 1 (Loyeva et al., 2010).

The block diagram includes three groups of statistical methods that allow us to better analyze the air quality conditions, which, in its turn, will enable the development of recommendations regarding convenient decisions in the management process of air quality protection.

For citation: Loyeva I., Rabeea Bazar S., Burhaz O., Vladymyrova O., 2021. Methods of comprehensive statistical analysis of agglomeration atmospheric air pollution. Journal Environmental Problems. Vol. 6, No. 3. p. 130–134. DOI: https://doi.org/10.23939/ep2021.03.130

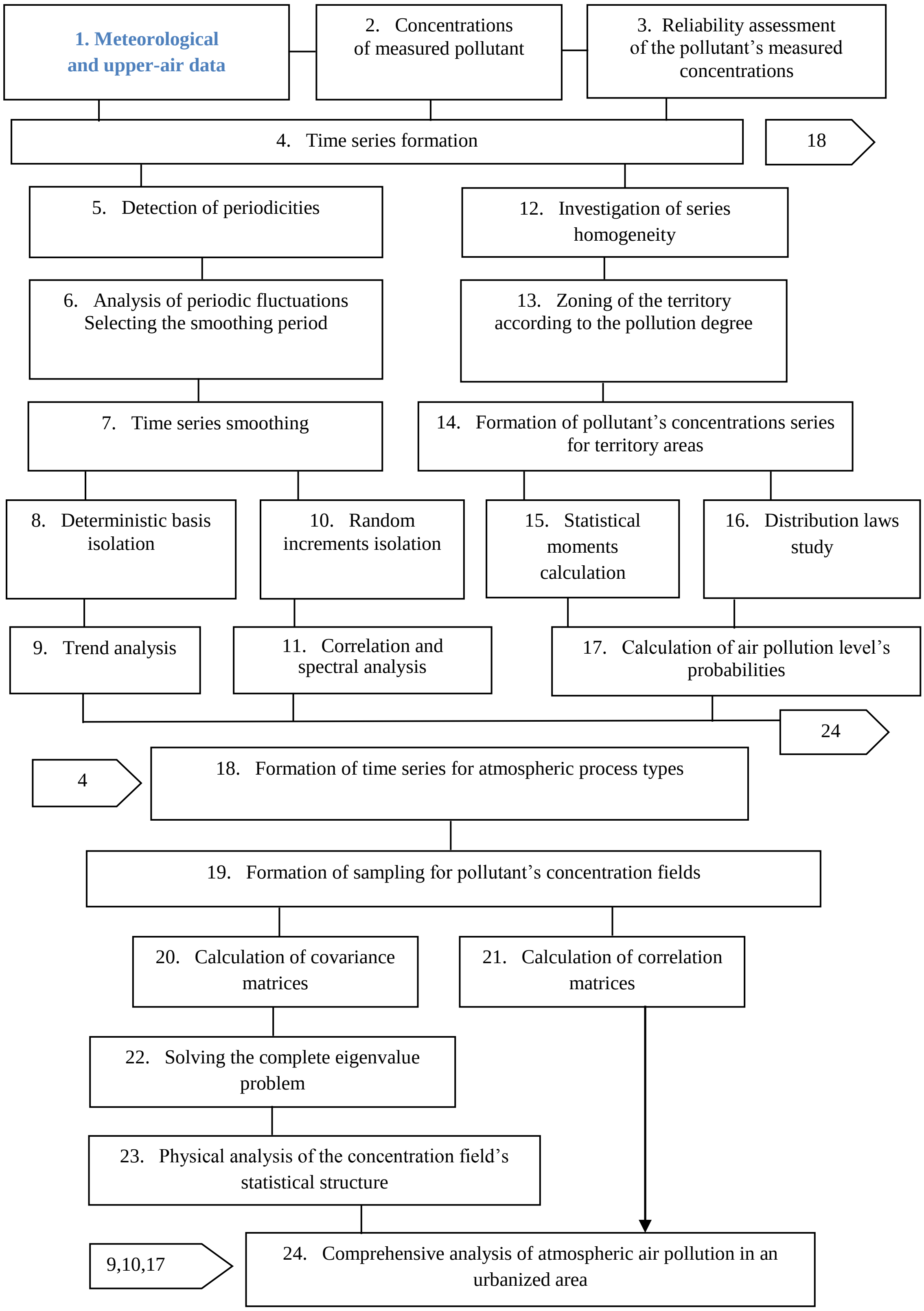

Fig. 1. Block diagram of a comprehensive statistical analysis of atmospheric air pollution in the agglomeration

The measured data must meet specific requirements. These include reliability, systematicity and sufficient volume, which will make it possible to form representative statistical series.

The statistical pollutant's one-time concentration for a temporary sample is generated to assess the homogeneity of the statistical series terms. For this purpose, a hypothesis is formulated to state that the series term's extreme values belong to the same general entity as well as all its other terms. The hypothesis is tested at a given significant level using the Student's t criterion. The check is repeated at a given significant level till the numerical series members that make up the sample are homogeneous (Shkol'nyy et al., 1999; Honcharova et al., 2007).

The first group of methods (blocks 4-11). As known, the process of accumulation of impurities in the atmosphere is a non-stationary process. Different states of the atmospheric boundary layer create various conditions that lead to the deposition or the dispersion of these pollutants, which also determine the tendency (trend) in the change of pollutants concentration. Periodic fluctuations are superimposed on the trends due to the daily, seasonal or annual variation of meteorological values. In addition, due to the atmospheric turbulence and the non-stationarity of the pollutant's emission from the sources, a random component arises. Thus, a change in the pollutant's concentration at a certain point of the territory is a non-stationary random process that includes a deterministic component (including trends and long-period fluctuations) and a random component. The process of non-stationarity makes it difficult to study the statistical structure of time series. The problem of transforming a non-stationary time series into a quasi-stationary one is solved by subtracting its deterministic basis from the original series. The selection is carried out by smoothing using the method of moving average, while the deterministic component is distinguished in the best way if the smoothing period is equal to the greatest frequency (Kendal et al., 1976).

Investigation of time series statistical structure of the ingredient's concentration comprises two stages. The first stage is revealing the latent periodicity in the pollutant's concentration series and the second one is smoothing the time series.

To identify the periodicities characteristic of the process in this study, we propose a method based on the Fourier transform. It allows us, with no additional research, to determine the frequencies, amplitudes and initial phases of the periodic components that are contained in the time series (Serebrjannikov et al., 1965).

As a result of the transformation, the obtained amplitude-frequency indicators of the identified periodicities allow us to build graphs. For statistically significant values (with a probability of 68 %), we take the periodicity that corresponds to an amplitude that goes beyond the upper limit of the confidence interval. The maximum frequency, as a rule, corresponds to the lowest frequency of significant amplitude (Loyeva et al., 2010).

The deterministic basis of the process is extracted by filtering or smoothing the original time series. Typically, filtration is achieved by using exponential smoothing. To isolate the deterministic component from the initial series, it is necessary, depending on the nature of the process under the study, to choose the length of the smoothing segment that would correspond to the period of harmonic oscillations which is inherent in this process. The selection of frequency from those available in the research process is determined by the researcher depending on which frequency of fluctuations needs to be filtered (Loyeva et al., 2010).

The analysis of the deterministic component makes it possible to assess trends in the atmospheric air quality state, which makes it possible to assess the effectiveness of the state's environmental policy.

The random component of the process is also important for the assessment of the air quality process, as its characteristics indicate the time scales of fluctuations and the magnitudes of the variability of the air pollutant's concentrations at one point or other points of the given territory. With the help of correlation and spectral analysis, you can determine these characteristics (Loyeva et al., 2010).

The second group of methods (blocks: 4; 12-17). According to the level of atmospheric air pollution, the zoning of the territory is carried out by testing the statistical hypothesis of the concentration-time series homogeneity using the nonparametric Wilcoxon criterion (Gmurman, 1977).

Differences like pollution with harmful impurities in different regions are established by calculating inversions for all possible pairs of the pollutants concentration time series. For each ingredient, both matrices of inversion and upper and lower boundaries of the critical region are calculated. A joint analysis of these matrices makes it possible to identify regions that are statistically homogeneous in terms of atmospheric pollution nature with one or another harmful impurity.

To determine the main statistical properties of the time series of ingredients concentrations in the regions, the calculation and assessment of statistical moments are achieved. The analysis of the obtained statistical characteristics of the ingredient concentration's distribution, in addition to studying of distribution laws make it possible to assess the atmo To determine the main statistical properties of the time series of ingredient concentrations in the regions, they carry the calculation and assessment of statistical moments out. The analysis of the obtained statistical characteristics of the ingredient concentration distribution, in addition to studying distribution laws, make it possible to assess the atmospheric air pollution probability in identified areas of urbanized territory (Shkol'nyy et al, 1999).

The third group of methods (blocks: 4; 18-23). The formation of the pollutant concentration field significantly depends on the state of the atmospheric boundary layer and is a random field. Multivariate statistical analysis, as a set of probable statistical methods complex, is widely used in studying the statistical structure of many fields as development of statistical forecasting models. By the pollutant concentration field, we mean the sum of their values at a set of territory points that are related to a certain moment in time. With this formulation, the individual properties of the field are not considered, but the general features of the statistical ensemble of random fields are established, which are usually called their statistical structure. The mathematical expectation and the covariance matrix characterized the statistical structure of random fields, the statistical estimates of which are found by averaging over the entire set of the fields of possible realizations (Kendal et al, 1976).

As known, a certain sequence of pollutant concentrations fields contains information on the spatial and temporal characteristics of the pollutant. With the help of component analysis, it is possible to separate the information related to the features of the spatial structure of the field, which is concentrated in the covariance matrix eigenvectors, and the information on the features of their temporal structure. The latter is contained in the random components of the fields that are their transformation in the basis of eigenvectors or the expansion in terms of natural orthogonal functions (n.o.f.). Eigenvectors can be identified with functions since each of their coordinates refers to a specific measurement point, that is, is a coordinate function (Bagrov, 1959).

Decomposition of concentration fields in n.o.f. has the advantage that allows us to isolate from the investigated random fields the components the variances of which exhaust most of the total variance of the fields. These components are commonly referred to as main components. Their variances are eigenvalues. Thus, comparing the successive sums of the first few eigenvalues with the total variance of the field which is equal to the trace of the covariance matrix (the sum of its elements is located on the main diagonal or, which is all the same, the sum of all eigenvalues) and, given a certain threshold, we can determine n.o.f. which quite fully characterize the main (Shkol'nyy et al., 1999).

The dispersion of the first principal component has the highest value, then it characterizes the largest-scale processes that influence the formation of the random variables field. Obviously, the first eigenvector also possesses the same property. Subsequent principal components reflect the impact of smaller-scale processes. Thus, component analysis allows, first, to compress the information about the investigated field, and second, to identify the processes of various scales responsible for the formation of investigated fields. If the random vector is not a field but a predictor vector, then component analysis makes it possible to reduce the number of predictors in the predictive regression equation or discriminant function, as well as to carry out their orthogonalization (Shkol'nyy et al., 1999).

3. Conclusions

The created algorithm of complex analysis of atmospheric air pollution over certain selected regions results from research achieved by the authors' team over a long time period.

Statistical analysis of available concentrations from the observational network enables obtaining the background characteristics of atmospheric air pollution. It will help to investigate the spatial correlations between the concentrations of the ingredient under various synoptic conditions. Correlations are the basis that is used in developing the statistical models for prediction of the atmospheric air pollution level of the given targeted areas as a whole for the models of regulating the rate of emissions from pollution sources and to achieve specified concentration fields of ingredients under unfavourable meteorological conditions, as well as to assess the contribution of emission sources to the formation of pollutant concentration fields in atmospheric air over the considered area. The proposed set of statistical methods can be successfully applied in the development of a state monitoring program related to atmospheric air protection over such considered zones as a part of the new procedure in the implementation plan of the state monitoring program of atmospheric air protection (Deyaki pytannya…, 2019). The development of the program requires an analysis of atmospheric air quality and the convenient selection of observation modes based on a preliminary assessment of the spatial distribution of pollutant concentrations, and the establishment of an assessment mode for the targeted zone (urban agglomeration). Based on the results of the analysis, there is a possibility to design an observation and assessment network: location and number of the observation station; and give the recommendation to develop an optimized network of observation stations.

References

Berljand, M. (1975). *Sovremennye problemy atmosfernoj diffuzii zagrjaznenie atmosfery*. Leningrad: Gidrometeoizdat. [in Russian]

Bagrov. N. (1959). Analiticheskoe predstavlenie posledovatel'nosti meteorologicheskih polej posredstvom estestvennyh ortogonal'nyh sostavljajushhih. *Trudy Central'nogo instituta prognozov*, (74), 3–24. [in Russian]

Deyaki pytannya zdiysnennya derzhavnoho monitorynhu u v haluzi okhorony atmosfernoho povitrya. Kabinet ministriv Ukrayiny (2019). Retrieved from https://zakon.rada.gov.ua/laws/show/827-2019-%D0%BF#Text [in Ukrainian]

Gmurman, & V. (1977). *Teorija verojatnostej i matematicheskaja statistika.* Moskva: Vysshaja shkola. [in Russian]

Honcharova, L., & Shkol'nyy, Ye. (2007). *Metody obrobky ta analizu hidrometeorolohichnoyi informatsiyi (zbirnyk zadach i vprav): Navchal'nyy posibnyk.* Odesa: Ekolohiya. [in Ukrainian]

Kendal, M.Dzh., & St'juart, L. (1976). *Mnogomernyj statisticheskij analiz i vremennye rjady*. Moskva: Nauka. [in Russian]

Loyeva, I., Vladymyrova, O., & Verlan, V. (2010). *Ocinka stanu zabrudnennja atmosfernoho povitrja velykoho mista: metody analizu, prohnozu, rehuljuvannja: Monohrafija.* Odesa: Ekolohija. [in Ukrainian]

Shkol'nyy. Ye., Loyeva. I., & Honcharova, L. (1999). *Obrobka ta analiz hidrometeorolohichnoyi informatsiyi*. Odesa. [in Ukrainian]

Serebrjannikov, I., & Pervozvanskij, A. (1965). *Vyjavlenie skrytyh periodichnostej.* Kiev: Nauka. [in Russian]