The importance of principal component analysis for environmental biodiagnostics of Donbass

Vitalii Nespirnyi[1,2], and *Andrei* Safonov[1*]

[1]Donetsk State University, 283001, Universitetskaya St., 24, Donetsk, Russia
[2]Economic Research Institute, 283048, Universitetskaya St., 77, Donetsk, Russia

Abstract. The ecological situation in the Central Donbas remains tense today and requires timely diagnosis of the state of ecosystems. Major causes of the tense environmental situation in Donbass are high level of industrialization (mining, metallurgical, mining and chemical industries), urbanization, agricultural technologies, landscape transformation and warfare from 2014 to the present. With all the variety of biological and chemical methods in obtaining large numerical information, the importance of mathematical approaches is of particular significance. Using the example of principal component analysis, a method for visualizing data in assessing transformed ecotopes of Donbass has been tested. An attempt has been made to reconstruct some missing data from the list of numerical characteristics. The values of probability and reliability of the data have been established. This allows to have more accurate information in monitoring and assessing the environment in the region. Principal component analysis complements the available cartographic materials, highlights the most significant processes considering general degradation of the state of Donbass ecosystems (by indicator plants *Bryum argenteum* Hedw., *Ceratodon purpureus* (Hedw.) Brid, *Amblystegium subtile* (Hedw.) Schimp., *Centaurea diffusa* Lam., *Cichorium intybus* L., *Tripleurospermum inodorum* (L.) Sch. Bip. et al.). The cause and effect relationships in the peculiarities of landscape transformation are grouped in order to further restore the integrity and functionality of the historical and geographical environment.

1 Introduction

In regions with a high level of anthropogenic transformation [1, 2], the task of implementing innovative technologies with high-quality and timely assessment of polluted landscapes is urgent [3, 4]. According to the severity of environmental problems at the present stage, the territory of Donbass is distinguished [5, 6], for which attempts are being made to mathematically and geophysically comprehend the processes taking place on the basis of the data on the state of plants [7, 8].

Methods of mathematical data processing are relevant in geoecological [2, 9, 10] studies: establishment of statistical significance, least squares method, functional analysis,

* Corresponding author: andrey_safonov@mail.ru

correlation analysis, geographic information systems and others [3, 11, 12]. However, each method has its own tasks and goals to use [13, 14].

This study was also conducted because of the lack of up-to-date information on the environmental problems of Donbass, since the hostilities impede and limit the work of departmental environmental laboratories. Historically, over the last 70 years, this region has had a high level of urbanization in combination with a developed system of mining, metallurgical, coke, and pharmaceutical industries. Over the last 120 years, open and underground mining methods and the intensification of the agricultural sector have radically transformed the natural geochemical cycles in natural systems. Up to 96-98% of the surface of such territories are considered to be anthropogenically disturbed. The laboratories for phytoindication monitoring of Donetsk State University [5, 7] have accumulated a large amount of field material on the condition of indicator plants in the Northern Azov region (1996-2024). Large statistical arrays of data for 185 plant species and more than 200 collection sites need analytical processing to understand the current environmental situation throughout the territory of Donbass.

The purpose of the research is to test principal component analysis for the research of the ecological imbalance in the natural systems of Donbass on the basis of phytoindication monitoring data.

2 Materials and Methods

The approach associated with the determining the principal components in multidimensional data is a widespread technique in the quantification of environmental elements [8, 15, 16], therefore, when analyzing the information obtained during a series of experiments, this approach was used, taking into account the recommendations of other Russian scientists [17, 18]. The initial data for the analysis were the results of a phytoindication study conducted on the territory of the Donetsk region [19, 20] with a multi-element analysis of the accumulation of chemical elements in the bodies of mosses as indicators of air pollution in the observation network [8, 21]. These papers contain tabular values and the results of the analysis of the principal components for some subsets of chemical elements.

The field experiment was implemented using the methods of active and passive environmental phytomonitoring [20, 21]. For ingredient analysis, plant samples were collected, which had initially been grown under control conditions and transplanted to special places in the localization nodes of the monitoring network throughout the territory of Central Donbass. The phytoindication method is based on the structural plasticity [19] of species with a wide ecological amplitude.

Ingredient monitoring in this series of experiments [7, 22] is based on data of the content of Na, Mg, Al, Si, P, S, Cl, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Br, Rb, Sr, Zr, Mo, Cd, Sb, I, Cs, Ba, La, Ce, Nd, Sm, Eu, Tb, Dy, Yb, Hf, Ta, W, Hg, Pb, Th, U in the gametophyte of mosses (*Amblystegium subtile* (Hedw.) Schimp., *Brachythecium campestre* (Müll.Hal.) Bruch et al., *Bryum argenteum* Hedw., *Bryum caespiticium* Hedw., *Bryum capillare* Hedw., *Ceratodon purpureus* (Hedw.) Brid, *Pylaisia polyantha* (Hedw.) Schimp.) and tissues of flowering indicator plants (*Berteroa incana* (L.) DC., *Capsella bursa-pastoris* (L.) Medik., *Centaurea diffusa* Lam., *Cichorium intybus* L., *Diplotaxis muralis* (L.) DC., *Echium vulgare* L., *Plantago lanceolata* L., *Plantago major* L., *Reseda lutea* L., *Senecio vulgaris* L., *Tanacetum vulgare* L., *Tragopogon major* Jacq., *Tripleurospermum inodorum* (L.) Sch. Bip.) [8, 20].

When processing field research data, methods of statistical processing and R-modeling were used – in systems of paired correlation between data sets, determining the closeness of the relationship between the indicators of the ingredient composition of indicator plants and the territorial proximity of these values. To identify clusters of elements whose concentrations have a high degree of correlation, a graphical image was constructed [8, 20]. The corresponding calculations were performed in the free R software environment for statistical computing.

The use of mosses is a widely applicable methodology for assessing the level of air pollution [23, 24]. For the analysis, data were used for both 24 and 113-component monitoring networks in Central Donbas. The problem is that, as a rule, 3-5% of the data from the total array fall out (are missing) from the analysis for various technical reasons, therefore it is difficult to conduct a correct correlation analysis in such cases. The mathematical apparatus used is also in demand as a tool in making a forecast about the state of natural environments [25, 26]. Such manipulations can be carried out when detecting and calculating steadier trends [27] in ecosystem change [28], if the trend is confirmed over 4-5 years of observation.

3 Results

If we use all the actual material on the accumulation of 46 chemical elements in mosses for a sparse monitoring network, then the weight coefficients of the contribution of the concentration of each chemical element to the first two components can be represented as follows (Fig. 1).

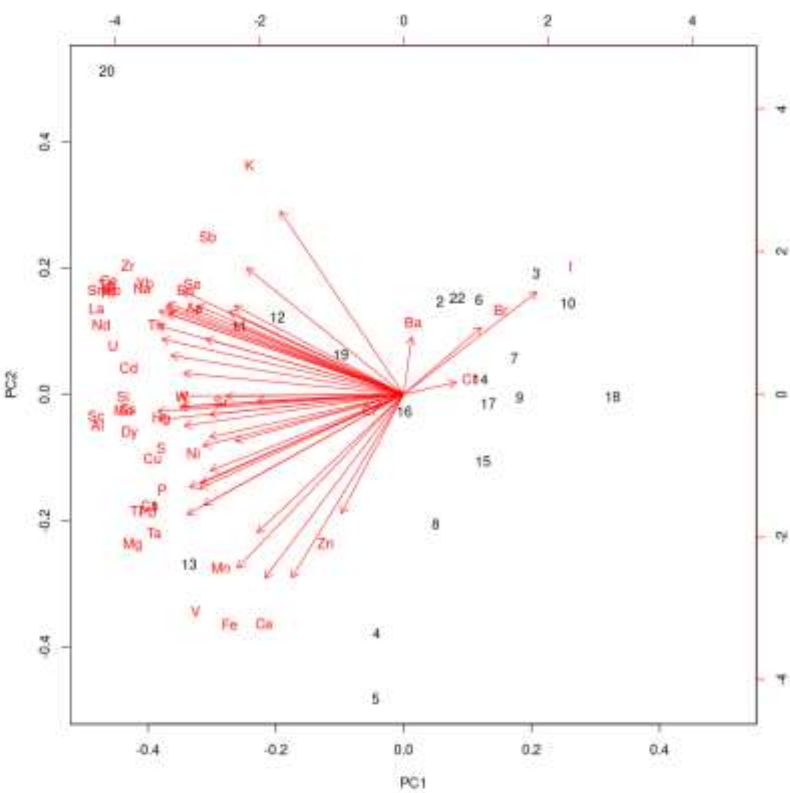

Fig. 1. Conjugate processes of elemental accumulation in mosses-indicators of the fragmented monitoring network of Donbass in projection on the first two principal components.

The established high concentrations of Mn, Ni, Cu and Pb when analysing natural and territorial complexes are explained by the proximity of the local production facilities for the extraction and processing of minerals, places of metallurgical and chemical industry.

It should be noted that the initial data were scaled so that the standard deviations for each element were the same (equal to 1). As can be seen from the presented figure, almost all elements make a negative contribution to the first principal component, and therefore it can be associated with the overall level of pollution and (or) disturbances in ecosystems. The elements Sc, La, Al, Sm and Nd make the largest absolute contribution to this component. The concentrations of these elements are recommended to be used as an indicator of the degree of disturbance of the ecosystems of Donbass in the period 2018-2023. When analyzing the available material according to the data [8, 20, 21], it was found out that complete information is available only for 21 points from the 24-component monitoring network. Accordingly, only 21 concentration vectors were used to construct the main directions. The coordinates of these vectors projected onto the plane of the first two components are shown in Fig. 2.

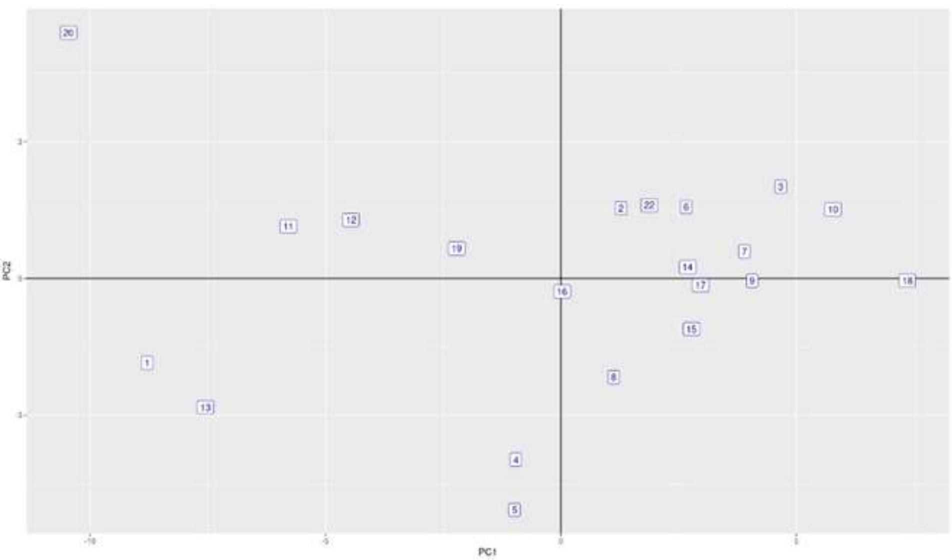

Fig. 2. A system of monitoring points with complete information on the elemental composition, grouped by principal component analysis.

Taking into account the interpretation of the first component described above, it can be concluded that the highest degree of pollution is observed at points numbered 20, 1, 13, and the lowest at points 18, 10. The numbers of monitoring sampling sites are shown on the map in previous publications [8, 21].

Since there is no concentration of the Nd element in the data for the points of the monitoring network 21, 23, 24, and it makes a significant contribution to the first principle component, it is not possible to accurately determine the projections of the vectors of element concentrations at these points in the main directions.

However, it is possible to find points with concentrations as close as possible to those measured during the experiment on a manifold stretched over the first few components. Then the coordinates corresponding to the concentration of Nd found points can be used as estimates for the missing values. Considering that the total proportion of variance explained by the principal components PC1-PC6 is 0.825180, it can be expected that when restoring unknown concentrations for the first six components, the relative measurement error will not exceed 18%.

4 Discussion

For the implementation of such an experiment, the following plants were identified as important: *Berteroa incana* (L.) DC., *Centaurea diffusa* Lam., *Cichorium intybus* L., *Plantago lanceolata* L., *Reseda lutea* L., *Senecio vulgaris* L., *Tragopogon major* Jacq., *Tripleurospermum inodorum* (L.) Sch. Bip., and in further testing of the mathematical apparatus, other species can be used: *Glaucium corniculatum* (L.) Rudolph, *Portulaca oleracea* L., *Dianthus campestris* M. Bieb, *Gipsophila paniculata* L., *Moehringia trinervia* (L.) Clairv., *Stellaria subulata* Boeber ex Schlecht., *Atriplex mircantha* C.A.Mey., *Campyliadelphus chrysophyllus* (Brid.) R.S.Chopra, *Dicranella cerviculata* (Hedw.) Schimp., *Orthotrichum diaphanum* Brid., *Brachythecium velutinum* (Hedw.) Schimp., *Pleuridium acuminatum* Lindb.

In the implemented monitoring experiment [29, 30], it is necessary to select species of native flora and those with a wide ecological amplitude. In such cases, the informativeness of the method increases and its use makes it possible to draw reliable conclusions [31, 32]. It is also favourable that the entire territory of Central Donbass is a flat area with a predominance of steppe landscapes – open ecological systems with tiers in phytocenoses mainly in the underground part. This fact is due to the limited reserves of water resources in the summer.

Attempts have been made to use other methods of mathematical interpretation of data (algorithmization based on geographic information systems, methods of multidimensional statistics, design of information scales, decoding of vegetation, automated biomonitoring, quotas for anthropogenic load on landscapes and local natural complexes, urban diagnostics), however, principal component analysis has revealed the greatest demonstrativeness and is in demand in the practice of environmental services in the region.

5 Conclusions

The expediency of introducing principal component analysis for the indication of Donbass systems has been proved, which confirms the need to use the apparatus of mathematical statistics in generalizing and understanding environmental processes.

The method of ecosystem assessment is based on the use of indicator plants (*Centaurea diffusa* Lam., *Cichorium intybus* L., *Tripleurospermum inodorum* (L.) Sch. Bip. et al.) and is associated with some structural changes in them according to the following indicators: pollen structure, terate structure of the flower and inflorescences, deformation of trichomes on the leaf blade, changes in conformational tissues, dystopia in integumentary tissues, shape and size of seeds and plant embryos, shoot architectonics, changes in leaf venation, terates in the structure of fruits.

Due to the mathematical methods of analyzing large amounts of information on botany and chemistry, it is possible to present the entire set of numerous data in a single model. This is widely used to implement a decision-making program for the restoration and development of an anthropogenically transformed region in Eastern Europe.

Completed within the framework of the work of the Azov-Black Sea Mathematical Center (Agreement No. 075-02-2024-1446 dated February 29, 2024).

The task has been completed within the framework of the youth laboratory "Diagnostics and mechanisms of adaptation of natural and anthropogenically transformed ecosystems of Donbass" (№ 1023110700153-4-1.6.19;1.6.11;1.6.12).

References

1. S. Yeprintsev, S. Kurolap, O. Klepikov, P. Vinogradov, E3S Web of Conferences **389**, 03030 (2023)

2. S. V. Shekoyan, S. A. Yeprintsev, P. M. Vinogradov, L. A. Lepeshkina, A. A Voronin, IOP Conference Series: Earth and Environmental Science, 543 (1), 012025 (2020)
3. V. V. Ermakov, S. F. Tyutikov, Geochemistry International, **58**, 3 (2020)
4. V. O. Korniyenko, V. N. Kalaev, Contemporary Problems of Ecology, **15**, 7 (2022)
5. T. V. Avraimova, A. I. Safonov, Scientific and Technical Libraries, **14**, 3 (2023)
6. O. Frunze, E3S WoC, **462**, 02004 (2023)
7. A. Safonov, A. Glukhov, BIO WoC, **31** (2021)
8. I. Zinicovscaia, A. Safonov, A. Kravtsova, O. Chaligava, E. Germonova, Physics of Particles and Nuclei Letters, **21**, 2 (2024)
9. S. Yeprintsev, S. Kurolap, O. Klepikov, S. Shekoyan, E3S WoC, **215**, 03009 (2020)
10. E. I. Mirnenko, Ecosystem Transformation, **5**, 2 (2022)
11. T. I. Kutyavina, T. Ya. Ashikhmina, Theoretical and Applied Ecology, **2** (2022)
12. S. I. Kolesnikov, N. A. Vernigorova, A. A. Kuzina, K. Sh. Kazeev, J. of Experimental Biology and Agricultural Sciences, **5**, 5 (2017)
13. S. Yeprintsev, S. Kurolap, O. Klepikov, S. Shekoyan, AIP Conference Proceedings **2647**, 070030 (2022)
14. Y. A. Semenishchenkov, Russian Journal of Biological Invasions, **8**, 3 (2017)
15. S. Vyas, L. Kumaranayake, Health Policy and Planning, **21**, (2006)
16. S. Chen, L. Yu, C. Zhang, Y. Wu, T. Li, J. of Env. Manag., **339** (2023)
17. I. Ivanov, A. Baranov, S. Mironov, S. Akbari, IEEE Sensors Letters, **6**, 5 (2022)
18. A. A. Cherednikova, A. G. Kuzmin, M. M. Guzenko, St. Petersburg State Polytechnical University Journal. Physics and Mathematics ,**16**, S1.2. (2023)
19. A. Safonov, BIO WoC, **43**, 03002 (2022)
20. A. I. Safonov, A. S. Alemasova, I. Zinicovscaia, K. N. Vergel, N. S. Yushin, A. V. Kravtsova, O. Chaligava, Geochemistry International, **61**, 10 (2023)
21. I. Zinicovscaia, K. Vergel, A. I. Safonov, Ecosystem Transformation, **6**, 3 (2023)
22. A. Safonov, E3S WoC, **431**, 01031 (2023)
23. I. Zinicovscaia, C. Hramco, O. Chaligava, N. Yushin, K. Vergel, Plants **10**, 3 (2021)
24. P. Świsłowski, K. Vergel, I. Zinicovscaia, Ecological Indicators, **138** (2022)
25. N. N. Kharchenko, V. N. Kalaev, V. O. Kornienko, IOP Conf. Series: Earth and Environmental Science, **875**(1), 012049 (2021).
26. V. V. Ermakov, Geochemistry International, **53**, 3 (2015)
27. I. T. Bayouli, H. T. Bayouli, A. Dell'Oca, Ecological Indicators, **125**, 107508 (2021)
28. X. Luo, Ch. Wu, Y. Lin, W. Li, J. of Env. Sc. **125** (2023)
29. S. Zheng, C. Xu, G. Lv, H. Shuai, Q. Zhang, Environmental Pollution, **333** (2023)
30. M.-C. Lafrenière, J.-F. Lapierre, D. E. Ponton, F. Guillemette, M. Amyot, Geochimica et Cosmochimica Acta, **353** (2023)
31. M. A. Outram, M. Figueroa, Current Opinion in Plant Biology, **67** (2022)
32. J. Kattge, G. Bonisch, S. Diaz, S. Lavorel, Global Change Biol., **26**, 1 (2020)