UDC 004.31
DOI: 10.15587/2312-8372.2019.172157

Putrenko V.,
Pashynska N.

DATA MINING OF SUSTAINABLE DEVELOPMENT PROCESS WITH USING NIGHTLIGHT INDICATORS

Об'єктом дослідження є процес сталого розвитку територіальних одиниць на прикладі регіонів України. Концепція сталого розвитку стала провідною стратегією розвитку для більшості країн світу. Однією із найбільших проблем залишається отримання повних та верифікованих даних для моделей оцінювання розвитку. В роботі було використано методику розрахунку індексу сталого розвитку, яка розроблена в Світовому центрі даних з геоінформатики та сталого розвитку Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського». Дана методика ґрунтується на розрахунку метрики індексу сталого розвитку на основі вимірів якості життя населення та безпекової компоненти проживання для окремих країн та регіонів. Для застосування методики на регіональному рівні було запропоновано використання інформації про нічне освітлення території, яка отримується засобами дистанційного зондування Землі із супутникових систем. Досліджено характер та тісноту зв'язку між яскравістю нічного освітлення та індикаторами сталого розвитку. З'ясовано, що найбільш значущий зв'язок наявний між показниками індексу економічного розвитку, індексом впливу на зміну клімату та нічним освітленням території регіонів України. На основі геоінформаційного аналізу програмного забезпечення ArcGIS компанії ESRI було застосовано інструментарій статистичного зонування, який надає можливості для статистичної обробки даних супутникових знімків у межах виокремлених за адміністративним принципом регіонів або інших полігональних областей.

На основі математичного апарату інтелектуального аналізу даних було здійснено глобальний та локальний регресійний аналіз зв'язку між виявленими показниками. Розгляд тісноти цього зв'язку в територіальному розрізі дозволив виявити зони з найбільшою та низькою тіснотою впливу, що пояснюється особливостями соціально-економічного розвитку даних територій. Завдяки цьому забезпечується можливість отримання розрахунку аналогічних показників на більш детальних територіальних рівнях, які відповідають окремим районам або містам обласного значення.

Ключові слова: *інтелектуальний аналіз, сталий розвиток, нічне освітлення, просторова регресія, геопросторовий аналіз.*

Received date: 27.11.2018
Accepted date: 14.12.2018
Published date: 30.06.2019

1. Introduction

In 2015, Ukraine joined the global efforts to achieve the goals of sustainable development (CSR) within the United Nations (UN) Summit for the adoption of the Development Agenda for the post-2015 period. This integrated action plan is based on the harmonization of economic, environmental and socio-institutional development dimensions, in particular, to overcome poverty in all its forms and strengthen the fight against climate change. The CSR Global Assessment System covers around 230 indicators [1], receiving high-quality data that is often a challenge for national statistical authorities.

Adequate data are definitely important for making informed decisions and tracking progress towards achieving the CSR. This requires the collection, processing, analysis and dissemination of unprecedented levels of data and statistical information at all levels, including data from new and innovative sources, as outlined in the CSR 2016 report [2]. The report also highlights the importance of researching new data sources, new data collection technologies, integrating different data sources, in particular, the combination of geospatial information and statistics. The global reports on sustainable development [3, 4] have repeatedly highlighted the potential of using Big Data approach in measuring progress on sustainable development, using satellite remote sensing and communication technology.

Therefore, it is important to study the possibilities of data verification for models of sustainable development assessment using remote sensing of the Earth.

2. The object of research and its technological audit

The object of the study is the process of sustainable development of territorial units by the example of the regions of Ukraine. The UN Millennium Declaration, accepted in 2000 by the 189 countries at the UN Millennium Summit, identified 17 Millennium Development Goals (MDGs), that is, a comprehensive framework of values, principles and key development drivers by 2015. After the development of universal MDGs in many countries of the world, the MDGs were adapted and targets were set, taking into account the specifics of the national situation. Ukraine has joined the UN Millennium Declaration. The assessment of the sustainable development process is based on the use of statistical data, sensory data, data from remote sensing of the Earth, content analysis in social networks and the Internet. This provides an opportunity to assess the economic, social and environmental components of the sustainable development process. In many cases, the necessary statistical indicators are missing or not going to, which complicates the adoption of managerial decisions. In this case, the calculation of estimates is based on the use of related factors, which requires the definition of the nature, level and closeness of the links between factors that influence on the process of sustainable development.

One of the most problematic places is the definition of the distribution rule and the strength of the links between components of the sustainable development process and the data of remote sensing.

3. The aim and objectives of research

The aim of research is determination of the availability and strength of statistical dependencies between light exposure at night on the level of regions of Ukraine and the main indicators characterizing the process of sustainable development.

The objectives of research are:

1. To determine the input data used to estimate nighttime illumination at the regional level.
2. To make preliminary analysis of geospatial data in order to obtain zonal statistics by region.
3. To determine the type of links between light data and components of sustainable development.
4. To establish global and local correlations and regression relationships between these groups of data.

4. Research of existing solutions of the problem

A striking example of the implementation of this approach can be NASA's ARSET program [5], which provides training on the implementation of satellite data in decision-making. For example, one of the ARSET webinars in 2017 was devoted to the use of satellite data on suspended particle contents of less than 2.5 microns for air quality assessment, based on the indicators of CSR 3 «Healthcare» and CSR 11 «Sustainable City» [6]. But these studies do not provide an opportunity to assess the impact of individual factors in the structure of the sustainable development model. The papers [7, 8] deals with research of using of Remote Sensing Data for modeling air quality in the Ukraine's cities. However, they do not provide a quantification assessment. One more – from the soil and vegetation cover to analyze progress towards achieving the CSR 15 «Ensuring Life on Earth» [9]. However, the described parameters do not have a developed system of criteria for evaluation.

Determining the economic status of a country or region – the issue is rather difficult. One way of indirect estimation of economic activity is based on the analysis of the change in illumination in the area under observation at night, based on satellite images [10]. But for each type of region economy this assessment will be different. The paper [11] estimates the economies of sub-Saharan Africa. However, this region is very different from Ukraine. The paper [12] attempted to assess the economic situation in the East of Ukraine by studying the change in lighting during the period of hostilities. And in [13, 14] – based on data on electricity consumption in these territories received from satellite images. But these works do not provide an actual assessment. The intensity of illumination is closely linked to an important indicator of economic activity – electricity consumption. The authors of the relevant method [12] have found that in countries with low and middle income per capita, the change in night illumination by 1 % is approximately equal to a change in income by 1 %. However, this estimate is correct only for a well-developed economy.

Thus, the results of the analysis allow to conclude that the relevant indicators of the sustainable development process for the territory of Ukraine were not mathematically formalized, which requires an expanded study.

5. Methods of research

In this work it was used data from Suomi NPP satellites that were orbited in 2011. The satellite has 5 onboard sensors:

1) Advanced Technology Microwave Sounder (ATMS);
2) Cross-track Infrared Sounder (CrIS);
3) Ozone Mapping and Profiler Suite (OMPS);
4) Visible Infrared Imaging Radiometer Suite (VIIRS);
5) Clouds and the Earth's Radiant Energy System (CERES).

For thematic processing of the filming data, the Day/Night Band (DNB) product developed on the basis of images from the VIIRS device was used – the monthly composite nightlight coverage of the night shots.

Using the ArcGIS 10.5 program, a statistical analysis of the illumination of the area was made to compare with the integral index of the economic dimension used in the assessment of sustainable development.

The formation of geostatistical data takes place using zonal statistics operations implemented in the ArcGIS Spatial Analyst module. The tool «Zonal statistics in a table» is suitable, which allows you to summarize the values of the raster within the zones of another set of data, is used for process data of a qualitative distribution of classification features (Table 1).

The calculation of zonal statistics occurs with the use of all cells that fall into the boundary of the input zones. As zones, spatial objects with the same values can be used as well as a group of objects that fall into one class. The size and configuration of the zones can be defined as a thematic raster containing qualitative signs, as well as vector features containing the appropriate attributes for zoning.

The zonal statistics can be calculated for both geometric parameters of features that fall into the zone, and for their attributive or quantitative attributes. Also, the tool for calculating zonal statistics is used to fill the zones with minimum values at the boundary of the zone (Table 1).

Table 1

Tools for working with zones are divided into two categories in terms of application

Category of tools for working with zones	Tools
Processing of zone shape statistics	Zonal Geometry Zonal Geometry As Table
Processing attribute statistics for zones	Zonal Statistics Zonal Statistics as Table
Tools that determine the proportion of classes within the zone	Tabulate Area
Tools that determine the frequency distribution of the values of one input raster in zones defined by another raster	Zonal Histogram
Tools to fill the values of the zone	Zonal Fill

ISSN 2226-3780

The administrative-territorial division of Ukraine by regions was selected as the basic unit of zonal statistics. As a result, statistics were obtained on night lighting for each region. These data were compared with the data of the model of sustainable development assessment. As a result, a correlation matrix was derived between indicators of threats and the mean value of illumination for the region.

The methodology of evaluation and analysis of sustainable development, as it is partially outlined in [14, 15] and used in this study. This methodology includes a model of sustainable development, which is an interdisciplinary generalization of the models known in the scientific, economic and social fields of science. The technique of methodology is applying formal statistical methods and expert evaluation methods for the analysis of sustainable development processes.

According to the methodology sustainable development assessment, the process of sustainable development will be characterized by two main components: security (C_{sl}) and quality (C_{ql}) of people life, and generalized measure of sustainable development will be defined by quaterion:

$$\{Q\} = j\,w_{sl}C_{sl} + w_{ql}\overrightarrow{C_{ql}}\left(I_{ec}, I_e, I_s\right). \quad (1)$$

Quaterion $\{Q\}$ contains an imaginary weighted scalar part $j\,w_{sl}C_{sl}$ which describes the security of people life and actual weighted part of a vector which describes the quality of life in the space of three dimensions:

1) economic (I_e);
2) environmental (I_{ec});
3) socio-institutional (I_s).

Weight coefficients w_{sl} and w_{ql} in formula (2) are used in order to align the scale of security and quality of life components. For the purpose of quantitative assessment of sustainable development we use the principles of constructing a hierarchical system of indicators and indices that are defined as L_1-norms:

$$I_i = \sum_{j=1}^{n} w_j x_{i,j},\ i = \overline{1,m},\ \sum_{j=1}^{n} w_j = 1, \quad (2)$$

in the space of indicators $X^1 \times X^2 \times \ldots \times X^m$ characterizing the economic, ecological and socio-institutional development of each region. Weights w_j in formula (3) are determined expertly.

The value of security of life components C_{sl} for j-th region is defined as the rate of Minkowski vector $\vec{S}_j = (s_i^j)$, $s_i^j = 1 - t_i^j$, $i = \overline{1,n}$:

$$C_{sl} = \left\|\vec{S}_j\right\| = \left(\sum_{i=1}^{n}\left(s_i^j\right)^p\right)^{\frac{1}{p}}, \quad (3)$$

with a parameter $p = 3$.

In order to determine the strength of the link between indicators of sustainable development and illumination of the regions, key indicators that characterize the sustainable development index were selected [14].

6. Research results

As a result, a correlation matrix was obtained for checking the connectivity between the statistical indicators of region illumination and their reflection in the main parameters of sustainable development (Table 2).

Table 2

Value of correlation for indicators of illumination and components of the index of sustainable development by regions of Ukraine

Values of illumination / Components of Index of sustainable development	Max	Min	Mean
Component of Quality of Life	–0.3062362	0.6958459	**0.775869**
Component of Security of Life	–0.0870352	0.209811661	**0.5144**
Economic prosperity	–0.3388719	–0.031903556	0.053557
Index of environmental dimension	–0.4851694	0.030830184	–0.12258
Index of economic dimension	0.16050311	0.66635917	**0.837093**
Human Development Index	–0.0076972	0.086519007	–0.0017
Competitiveness Index	0.19815784	0.573736492	**0.766881**
Quality of life index	–0.2961539	0.113519109	0.140841
Index of social dimension	–0.3372712	0.29152067	0.317593
Sustainable Development Index	–0.2552884	0.590410857	**0.757561**
Ecological security	–0.3308463	–0.044158404	0.168294
Economic security	0.12583802	0.345941902	**0.637313**
Security (other)	0.33190022	0.061210228	–0.09038
Impact on climate change	–0.0620561	0.77810648	**0.997755**
Environmental stress and danger	–0.4012843	–0.093752585	–0.31661
The state of the environment	–0.3183306	0.149893923	0.144667

The analysis of this matrix allows establishing the tightness of the link between the regions' illumination and the components of the assessment of sustainable development. Verification of statistical emissions of maximum and minimum values indicates a lack of close dependencies. The calculations of the mean values, on the contrary, indicate the presence of significant dependencies between the indicators. Among the economic indicators, the greatest tightness of dependencies is indicated by the index of economic measurement 0.837093, which indicates the correctness of the mathematical model of calculation. The largest correlation index has an impact on climate change, which is equal to almost 1 (0.997755). The threat «Impact on Climate Change» has been assessed as an inescapable contribution to climate change in each region of Ukraine by using the Carbon Dioxide Density Index:

$$x_{CCH} = \frac{x_{CCH1}}{x_{ACH1}}, \quad (4)$$

where x_{CCH} – the regional contribution index for climate change, tons CO_2-eq/km^2; x_{CCH1} – carbon dioxide emissions into the atmosphere, ths. tons. Such a high correlation between the indicators shows a direct relationship between CO_2 emissions and illumination of the regions (Fig. 1).

The highest value is Kyiv. Despite the hostilities, the Donetsk region continued to have highlight. The following positions occupy Kharkiv, Dnipropetrovsk, Kyiv, Lviv and Volyn regions. The smallest values of lighting are Kherson, Kirovograd and Zhytomyr regions.

Spatial correlation and spatial regression indicators are used to establish relationship.

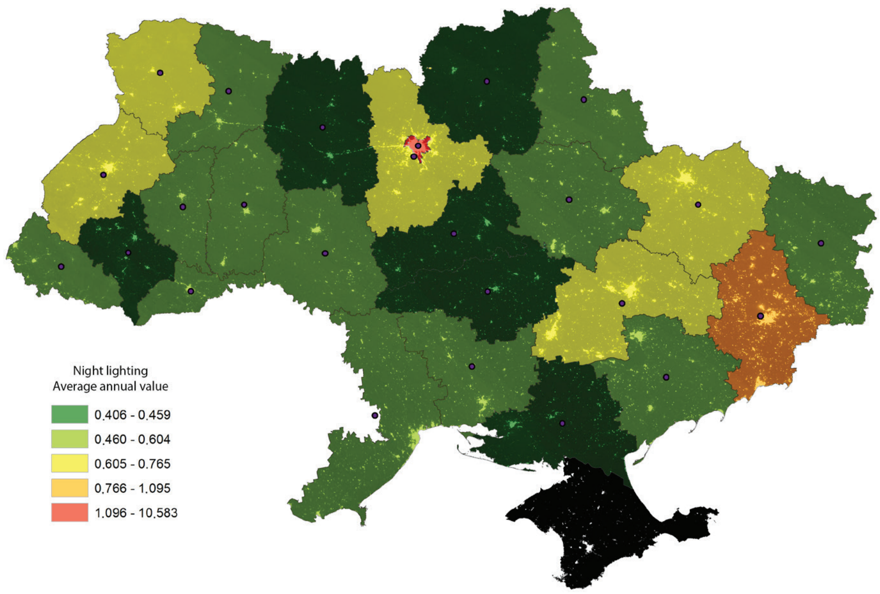

Fig. 1. Average distribution of light indicators in 2016 by regions of Ukraine

Geographic weighted local regression is defined as:

$$u_i(x_k,y_k)=\beta_{0i}(x_k,y_k)+\beta_{1i}(x_k,y_k)\cdot v_{1i}+$$
$$+\beta_{2i}(x_k,y_k)\cdot v_{2i}+...+\beta_{pi}(x_k,y_k)\cdot v_{pi}+\varepsilon_i, \quad (5)$$

for observation i, where $u_i(x_k,y_k)$ is the dependent variable evaluated in place i; v_{pi} – explanatory variable; $\beta_p(x_k,y_k)$ – local coefficients of regression; p – the number of variables, and the remainder is estimated at the location i. Each local regression equation is solved with different weights of observations based on the decay function with distance, centered on observation i.

Geographically weighted local Pearson correlation coefficient allows to establish a tight relation between spatially distributed data. To do this, it is calculated for each pair of values of two phenomena in the places where they are placed according to the formula:

$$r_i(x_k,y_k)=\frac{\sum_{j=1}^{n}\varpi_j\cdot(u_j-\bar{u})\cdot(v_j-\bar{v})}{(n-1)\cdot S_u\cdot S_v}=\beta_1(x_k,y_k)\frac{S_u}{S_v}, \quad (6)$$

where $r_i(x_{ki},y_k)$ – the correlation coefficient; (x_k,y_k) – the place of observation i; u_i and v_i – a separate observation; $\bar{u}$ and $\bar{v}$ – the mean values of two variables; ϖ_j – the weight that each measurement has, depending on the function of decay with distance, centered on observation i; n – the size of the sample; S_u and S_v – standard deviations of two variables; $\beta_1(x_k,y_k)$ – calculated parameter for two-dimensional local regression of observation i.

To use the geographically weighted regression method and correlation, the transformation of attribute data from average illumination and the index of economic dimension are used. The organization of data in such a model allows to apply geo-weighted correlation algorithms in the language of Python (Fig. 2).

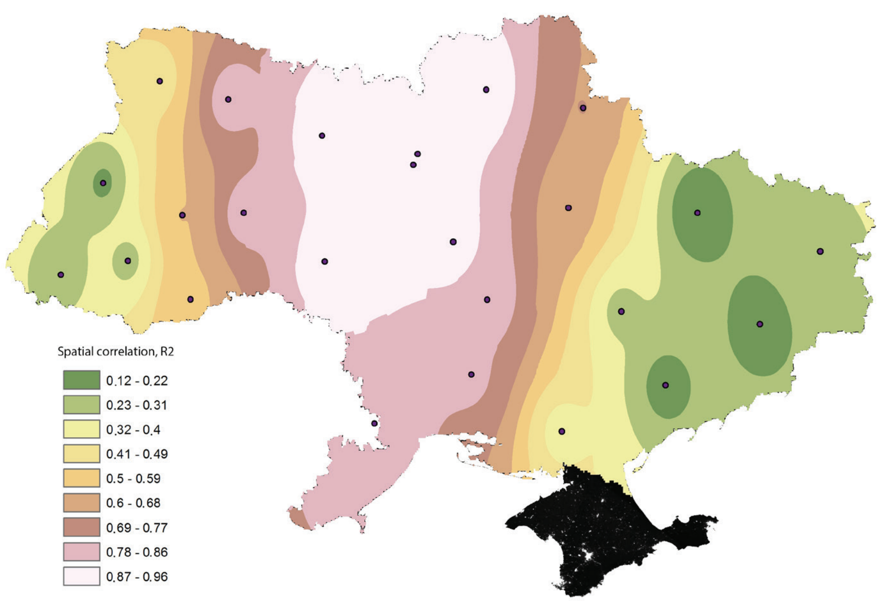

Fig. 2. The value of the spatial correlation of regions of Ukraine in terms of average illumination and the index of economic dimension

ISSN 2226-3780

The estimation of the spatial correlation between the data shows the existence of the closest dependence between the indicators for the central part of Ukraine and its deviation for the eastern and western parts of the country. In the east, the decline in the tightness of dependence may be due to military actions and the occupation of the territory. In the west the decline in the tightness of dependence may be due to ignoring part of economic income by means of official statistics.

7. SWOT analysis of research results

Strengths. The positive effect of the use of such external factors of assessment like the brightness of nightlights is the possibility of a more detailed assessment of the economic component of the process of sustainable development in the context of subregional units and individual settlements. This allows you to reduce the cost of collecting field statistics, increase the speed of data processing and reduce the time it takes to calculate model data.

Weaknesses. The negative effect are the difficulty of obtaining statistically accurate estimates of the process due to the fact that the cohesiveness of relate between individual indicators of sustainable development and nightlights varies in space and time.

Opportunities. Further studies are connected with the use of machine learning for the possibility of evaluating sustainable development components at a more detailed level of administrative regions, cities and communities.

Implementation of the methods may allow to increase the efficiency of the development of regional policy in Ukraine under the conditions of the decentralization process. This study may be conducted for other countries and regions of the world, taking into account local specificities.

Threats. Instability of the military-political and economic situation in Ukraine negatively affects the assessment of components of sustainable development, taking into account the weakening of the security component.

8. Conclusion

1. Satellite composite images from the Suomi NPP, which provide the highest resolution and are represented by composite images that accumulate information for the year, have been selected as incoming illumination data. It is determined that the use of composite images gives the best result in the mathematical model of relations with economic indicators of the process of sustainable development.

2. Instrumentation of zonal statistics for the automated calculation of the relevant statistical indicators of the brightness of illumination in a regional section has been applied.

3. The mathematical type of connection between the indicators of illumination and indicators of sustainable development is determined. The calculation of the index of economic measurement has a fairly high correlation value with illumination (0.83), which confirms the correctness of the calculations. The highest correlation value has indicators of regional illumination with the threat «Impact on climate change». Thus, illumination statistics can also be used to estimate carbon emissions.

4. The indicators of global and local regression equations for different regions of Ukraine are established. It is determined that the closest connection between the indicators is characteristic of the central regions of Ukraine.

References

1. *Revised list of global Sustainable Development Goal indicators*. Available at: https://unstats.un.org/sdgs/indicators/Official%20Revised%20List%20of%20global%20SDG%20indicators.pdf Last accessed: 26.06.2019
2. The Sustainable Development Goals Report 2016 (2016). *UN*. Available at: http://ggim.un.org/documents/The%20Sustainable%20Development%20Goals%20Report%202016.pdf
3. Prototype Global Sustainable Development Report – Executive Summary (2013). *UN-DESA*. Available at: https://sustainabledevelopment.un.org/content/documents/975GSDR%20Executive%20Summary.pdf
4. Global Sustainable Development Report – 2016 (2016). *UN*. Available at: https://sustainabledevelopment.un.org/content/documents/2328Global%20Sustainable%20development%20report%202016%20(final).pdf
5. The Applied Remote Sensing Training (ARSET) program (2017). *NASA*. Available at: https://arset.gsfc.nasa.gov/
6. Satellite Derived Annual PM2.5 Data Sets in Support of United Nations Sustainable Development Goals (2017). *NASA*. Available at: https://arset.gsfc.nasa.gov/airquality/webinars/AQ-SDG-17
7. Putrenko, V. V., Pashynska, N. M. (2017). The use of remote sensing data for modeling air quality in the cities. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-5/W1*, 57–62. doi: http://doi.org/10.5194/isprs-annals-iv-5-w1-57-2017
8. Putrenko, V. (2017). Data mining of relationship in crowdsourcing projects and social activities of citizens. *Proceedings IEEE 1st Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 1060–1065. doi: http://doi.org/10.1109/ukrcon.2017.8100413
9. Remote Sensing of Land Indicators for Sustainable Development Goal 15. *NASA*. Available at: https://arset.gsfc.nasa.gov/land/webinars/sdg15
10. Lieskovsky, A. J. (2014). *Electricity Use as an Indicator of U.S. Economic Activity*. Washington: Energy Information Administration, 87.
11. Henderson, J. V., Storeygard, A., Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review, 102 (2)*, 994–1028. doi: http://doi.org/10.1257/aer.102.2.994
12. Coupe, T., Myck, M., Najsztub, M. (2016). And the Lights Went Out — Measuring the Economic Situation in Eastern Ukraine. *VoxUkraine*. Available at: https://voxukraine.org/en/and-the-lights-went-out-measuring-the-economic-situation-in-eastern-ukraine-en/
13. Lyalko, V. I., Sakhatsky, O. I., Elistratova, L. O., Apostolov, A. A. (2017). Use of space images of NPP / VIIRS at night to assess the economic crisis in the East of Ukraine (Donetsk and Luhansk regions). *Bulletin of the National Academy of Sciences of Ukraine, 2*, 48–53. doi: http://doi.org/10.15407/visn2017.02.048
14. Zghurovskyi, M. Z., Akimova, O. A., Boldak, A. O., Vavulin, P. L., Dzhyhyrei, I. M., Yefremov, K. V., Putrenko, V. V. (2014). *Sustainable Development Analysis: Global and Regional Contexts. Part 2. Ukraine in the Indicators of Sustainable Development (2013)*. Kyiv: NTUU «KPI», 172.
15. Zgurovsky, M. Z., Boldak, A. A., Yefremov, K. V. (2013). Intelligent analysis and the systemic adjustment of scientific data in interdisciplinary research. *Cybernetics and Systems Analysis, 49 (4)*, 541–552. doi: http://doi.org/10.1007/s10559-013-9539-7

***Putrenko Viktor,** PhD, Senior Researcher, Education and Scientific Complex «World Data Center for Geoinformatics and Sustainable Development», National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, e-mail: putrenko@wdc.org.ua, ORCID: http://orcid.org/0000-0002-0239-9241*

***Pashynska Nataliia,** PhD, Senior Researcher, Department of Intellectual and Information Systems, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: n_pashynska@ukr.net, ORCID: http://orcid.org/0000-0002-0133-688X*