Application of statistical methods in socio-economic geography and spatial management based on selected scientific journals listed in the Web of Sciences database

Sławomir Dorocki[1], Mariusz Cembruch-Nowakowski[2]

Abstract

The aim of the paper is to present an analysis of the use of statistical methods and tools in scientific articles related to socio-economic geography and spatial management published in the years 2012–2021. In order to evaluate the intensity and diversity of this phenomenon, a query was carried out using the Web of Sciences electronic academic information database. A preliminary literature search led to the decision to focus on papers published in three selected journals relating to social geography (Geoforum), economic geography (Applied Geography) and spatial management (Landscape and Urban Planning). The paper analyses the variety of the statistical tools used in the studies presented in the aforementioned journals. The frequency and type of the applied statistical methods, computer software and computing tools is correlated with the specificity of the research area.

Key words: socio-economic geography, statistical methods, computer software, electronic academic information database.

1. Introduction

Statistical methods are frequently used by researchers as auxiliary tools for data analysis in various scientific disciplines including: exact, natural, medical, and social sciences as well as humanities. However, they are particularly useful for data collection and analysis in some sub-disciplines, such as e.g. geography including socio-economic geography and spatial management. In demography and population geography, mathematics and statistics were introduced already in the 18th century (Fleszar, 1962). However, the term "quantitative geography" appeared in the literature at the beginning of the 20th century in papers published by Wallis (1912) and Huntington (1927).

---

[1] Pedagogical University of Cracow, Kraków, Poland. ORCID: https://orcid.org/00000-0001-6083-0346.

[2] Pedagogical University of Cracow, Kraków &The Central Statistical Office of Poland, Culture Statistics Center, Kraków, Poland. E-mail: mariusz.cembruch-nowakowski@up.krakow.pl. ORCID: https://orcid.org/0000-0001-8443-9915.

The real revolution, which occurred at the turn of the 50s and 60s of the 20th century, resulted in a transition from descriptive geography (idiographic) to the empirical law-making (nomothetic) geography. As a consequence, more advanced statistical methods (including multidimensional ones) were used by the researchers involved in these studies (R. P Haining, R. Haining, (2003); Runge, 2006; A. Agresti, (2009); J. E. Burt et al. (2009); A. Hanushek, J. E. Jackson, (2013); R. Kitchin, (2013); O. Schabenberger, C. A. Gotway, (2017); Czyż, Chojnicki, 2019; Hauke, 2021).

Socio-economic geography and spatial management emerged as a new discipline of social sciences in 2018. The current development of social and economic geography is tightly connected with the development of statistical tools and geostatistical methods. The contemporary computing potential and availability of statistical databases (big data) supported by IT statistical tools allow the spatial interpretations of various spatial phenomena. It should be pointed out that such a research methodology usually requires the involvement of interdisciplinary research teams. It can be observed that recently such teams consisting of geographers, statisticians, and computer scientists have been created to carry out joint research in a wide area of geography.

2. Data Collection

To analyze whether and to what extent the statistical tools are currently used in socio-economic geography and spatial management the query of papers dealing with that research subject and published in the international scientific journals in the period of time 2012–2022 was performed.

Web of Science (WoS), one of the largest electronic academic information databases, was used as the source to retrieve the related publications. The publications in which the statistical methods and IT programs were used for data analysis were selected and used for further discussion.

The sample for the analysis was selected in two step procedure. In the first step, the content analysis was applied for the search in Web of Science electronic academic information database using the following key words: socio-economic geography, spatial management, statistical methods, computer software, IT programs. The search allowed to identify 30 journals (160 512 papers) matching the subject of interest. From that set of journals three of them published by Elsevier publishing house were selected: Geoforum, Applied Geography and Landscape and Urban Planning with the content of 5507 papers dealing with the problems under investigation. These journals were selected based on their scopes and reputation in scientific community (see below for details).

In the second step, the query of the papers presenting research results in the area of socio-economic geography and spatial management published in these journals in the period of time 2012-2021 was performed. That resulted in the selection of 592 papers meeting the imposed requirements. That sample was subjected to detailed analysis assuming the confidence level α = 0.95, fraction = 0.5, maximal error 0.05%. Taking into account these parameters the sample for analysis reduced to 383 papers. However, to ensure the representativeness of the content it was decided to analyse all of the 592 papers. That resulted in increase of the maximal error to about 4%.

3. Data Selection and Analysis

The preliminary analysis indicated that the scopes of three scientific journals: Goforum, Applied Geography, and Landscape and Urban Planning are compatible with the research subjects of socio-economic geography and spatial management. In the period of time considered in the analysis (2012–2021), these journals published 5507 papers dealing with these subjects: Geoforum (1948) Landscape and Urban Planning (1832), and Applied Geography (1727). It should be pointed out that the scopes of these journals cover the representative and the most important aspects of research in geography and spatial management. Geoforum is mainly interested in papers presenting the research results of studies in social geography, including the geography of settlement and political geography. „*Geoforum is a leading international, interdisciplinary journal publishing innovative research and commentary in human geography and related fields. It is global in outlook and integrative in approach. The broad focus of Geoforum is the organisation of economic, political, social, and environmental systems through space and over time. Areas of study range from the analysis of the global political economy, through political ecology, national systems of regulation and governance, to urban and regional development, feminist, economic and urban geographies and environmental justice and resources management*"[3].

Geoforum has Cite Score 5.9, Impact Factor = 3.93, supports open access and is abstracted/indexed in 7 data bases (Current Contents, Academic Journal Guide (Chartered Association of Business Schools), Elsevier BIOBASE, Engineering Village – GEOBASE, Current Contents – Social & Behavioral Sciences, Social Sciences Citation Index and Scopus).

Applied Geography is mainly presenting papers on economic geography. "*Applied Geography is a journal devoted to the publication of research which utilizes geographic approaches (human, physical, nature-society and GIScience) to resolve human problems*

---

[3] Retrieved from: https://www.sciencedirect.com/journal/geoforum. 22.06.2022

that have a spatial dimension. These problems may be related to the assessment, management and allocation of the world's physical and/or human resources. The underlying rationale of the journal is that only through a clear understanding of the relevant societal, physical, and coupled natural-humans systems can we resolve such problems"[4].

Applied Geography has Cite Score equal to 8.3, Impact Factor 4.73, and supports open access. It is abstracted/indexed in 16 main data bases (Sage Public Administration Abstracts, Geography, Ecological Abstracts, Envirofiche, GeoRef, Oceanographic Literature Review, Elsevier BIOBASE, International Development Abstracts, Sage Urban Studies Abstracts, Social Sciences Citation Index, Current Contents, Current Geographical Publications, Environmental Abstracts, Environmental Periodicals Bibliography, Geographical Abstracts, Scopus).

Landscape and Urban Planning is dealing with special management. "Landscape and Urban Planning is an international journal aimed at advancing conceptual, scientific, and applied understandings of landscape in order to promote sustainable solutions for landscape change. Landscapes are visible and integrative social-ecological systems with variable spatial and temporal dimensions. Landscapes are increasingly urban in nature and ecologically and culturally sensitive to changes at local through global scales. Multiple disciplines and perspectives are required to understand landscapes and align social and ecological values to ensure the sustainability of landscapes. The journal is based on the premise that landscape science linked to planning and design can provide mutually supportive outcomes for people and nature"[5].

Landscape and Urban Planning, has Cite Score equal to 12.7, IF = 8.12, supports open access. It is abstracted/indexed in 13 main data bases: Science Citation Index, Elsevier BIOBASE, LandSearch, Engineering Village – GEOBASE, Applied Ecology Abstracts, BIOSIS Citation Index, Current Contents, Environmental Periodicals Bibliography, Geographical Abstracts, Scopus, Cambridge Scientific Abstracts, Environmental Abstracts, Urban Studies Abstracts. It can be summarized that these journals are representative of the subject under investigation. The statistical sample of 592 papers published in these journals in the period of time 2012–2021 was carefully reviewed to search for the application of statistical methods in data analysis (see Table 1). The number of papers selected from various journals is proportional to the total number of papers published in each of them. It should be noticed that Applied Geography is a quarterly journal while Geoforum is a monthly journal, which translates to the total number of papers published in each of them.

---

[4] Retrieved from: https://www.sciencedirect.com/science/article/abs/pii/S014362282100028X. 22.06.2022

[5] Retrieved from: https://www.sciencedirect.com/journal/landscape-and-urban-planning. 22.06.2022

Table 1: Number of analyzed papers published in the selected journals in the period of time 2012–2021

Journal/year	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Sum	IF
Applied Geography	10	12	7	8	9	9	7	11	23	17	113	4.60
Geoforum	42	25	27	11	24	10	33	20	27	32	251	3.77
Landscape and Urban Planning	28	28	18	26	17	24	20	17	24	26	228	6.85
Sum	80	65	52	45	50	43	60	48	74	75	592	

Source: own studies.

The authors of 63% papers analyzed used statistical methods as research tools (Table 2). Most often, these methods were present in the works on the special management, papers published in Landscape and Urban Planning and in economic geography, papers published in Applied Geography. The share of papers in which statistical methods were used was almost ten times smaller in the case of publication in Geoforum than for the first two magazines. This is mainly due to the fact that papers published in Geoforum were of more theoretical nature or they presented research based on the survey results, which were often purely descriptive. It is noticeable that in the articles under investigation several (more than one) statistical methods were usually used; on average, about 3.5 methods were used per article. In this analytical approach, the publications in Geoforum did not stand out considerably from the average methods used in the other two inspected journals.

In the conducted study, not only the methods used were analyzed but also the way of presenting the results in the form of charts and diagrams was considered. However, not all charts, but only those used as the basis for the statistical inference were taken into consideration. These graphical methods prevailed in Applied Geography (in 69% of papers) while in Geoforum they were rarely present (in 1.6 % of papers). It should be observed that in Landscape and Urban Planning the choropleth maps and diagrams accompanying the geostatistical methods were identified. However, they were not considered in the current analysis because there are basic substantive and methodological difficulties in the expectance of their applicability in the social and political geography where they are rarely, if ever, used.

Table 2: Fraction of papers in which the statistical methods were used in the selected journals in the period of time 2012–2021

Journal	Statistical methods used (%)	The average number of methods used per one paper	Charts used (%)
Landscape and Urban Planning	88.60	3.77	59.65
Applied Geography	88.50	3.16	69.03
Geoforum	9.96	2.28	1.59
Sum	63.34	3.47	36.82

Source: own studies.

As a result of the query, the following statistical methods were distinguished: average measures (mean), measures of variation - including standard deviations (SD), positional measures, e.g. median, dominant, etc. (position), correlation-regression (R2), statistical model (model), standardization or normalization (normal). In addition, the most commonly used advanced statistical methods and programs used to carry out calculations were reviewed. It should be noticed that in some cases there was no information on the type of statistical tools used although the results presented clearly indicated that they were applied.

Table 3: Fraction of papers in which the selected statistical methods were used in selected journals in the period of time 2012–2021

Journal	r2	mean	model	SD	position	normal
Applied Geography	58.41	46.02	46.90	39.82	43.36	45.13
Geoforum	3.19	7.57	3.19	4.78	1.59	2.39
Landscape and Urban Planning	70.61	65.35	58.33	59.65	53.51	26.32
Sum	39.70	37.16	32.77	32.60	29.56	19.76

Source: own studies.

Among these statistical methods, the correlation-regression (R2) was the most frequently used (in 39.7% of analyzed papers). That method was mostly applied to analyze the economic and social phenomena in studies carried out in the area of spatial management (published in Landscape and Urban Planning) and economic geography. Interestingly, although the analyzed phenomena had spatial dimension (geographic coordinate system) the autocorrelation of variables was neglected in their analyzes (Dorocki, Jenner 2016). Thus, these methods were applied for selection as well as for observation of the relationship between variables.

The arithmetic average was in the second place amongst the statistical methods used by the authors of the papers published in analyzed journals. That simple method

was used most often in the studies in the area of social geography (papers published in Geoforum). Two others identified by our methods such as measures of differentiation our statistical models were used to a similar extent.

Surprisingly, it was observed that the frequency of using statistical methods in the papers published in the time frame 2012-2021 was not dependent on the year of publication. It was expected that in the most recent papers there will be a higher probability of using statistical tools but that assumption was not confirmed. The use of statistical methods was shown to be dependent mainly on the research subject presented but also on the quality and quantity of data collected. In the case of publications from the area of social geography (Geoforum) usually a survey and scientific observations served as a base for analysis. The analysis of this kind of data was often limited to the very description of the relative values and the survey responses obtained. The most common methods, in this case, were average and positional measures and the Cronbach Alpha test. In papers from the area of economic geography the most frequently analyzed are secondary statistical data derived from databases shared by public institutions. In such cases, the study of dependencies between statistical variables, confidence testing as well as creation and verification of statistical models were most often undertaken. In the case of papers from the area of spatial management, the studies involved mainly spatial data. Thus, the researchers applied mostly the geostatistical methods (GIS) and analyses related to environmental protection such as risks analysis and matrix methods. Also, the spatial or mixed models were considered.

Out of all advanced statistical methods used in the literature analyzed the following, most often repeated methods are listed in alphabetical order: (ANOVA), Amsterdam mode, Cronbach Alpha, Chi-2, exploratory factor analysis (EFA), Generalized Linear Mixed Models (GLMM), generalized linear models (GLM), Kolmogorov–Smirnov test, Gaussian distribution, Kruskal–Wallis test, mixed methods, Monte Carlo method, Mann Whitney U test, generalized cross-validation (GCV), Manova (Wilk's lambda), multiple linear regression (MLR), Moran I, OLS, PCA, Pearson, Spearman, multidimensional regression tree (MRT), Shannon diversity index, index Kappa. It is worth mentioning that the use of many statistical methods is also related to the accessibility of computer statistical programs. Out of the statistical software used the most popular was R Project for Statistical Computing (R), which served as an analytical tool in 53 papers, of which 43 were published in Landscape and Urban Planning. The popularity of that program comes not only from its computing capabilities but also from its high availability; it is free statistical software working in the UNIX, Windows, and in the MacOS environments. Additionally, various packages extending the possibilities of the program are delivered.

In the second place among the statistical software used was Statistical Package for the Social Sciences (SPSS). It was originally addressed mainly to the social sciences but currently it is widely used also in other scientific disciplines. One of the most popular geostatistical tool is ArcGIS software. It allows the creation and processing/modification of the existing maps, analysis of spatial data, their visualization, and data management in geodatabases. That program was used mainly by the authors publishing papers in Landscape and Urban Planning. Most often the NVivo package was used for the analysis of qualitative data obtained from surveys, e.g. text - interview content, etc. Also, the very popular MS Excel package and built-in statistical functions are utilized. As with the selection of analytical methods for a given type of data, also in the case of statistical software, there is a diversification of the programs used (Table 4).

Table 4: Share of statistical software used by the authors of papers published in selected journals in the time frame 2012–2021

Program	Applied Geography	Geoforum	Landscape and Urban Planning	sum
R	9	1	43	53
SPSS	14	5	14	33
ArcGIS	1		17	18
NVivo		5	2	7
MS excel	3	2	2	7
GeoDa	2	1	3	6
PC-ORD			6	6
ArcMap			5	5
Matlab	1		1	2
MAXQDA			2	2
MINITAB			2	2
Qgis			2	2
SigmaPlot			2	2

Source: own studies.

4. Discussion

In the current paper the application of statistical methods in studies in socio-economic geography and spatial management was presented based on the analysis of the papers published in three deliberately selected prestigious scientific journals. These journals are published by the Elsevier publication house and have high citation sores and impact factors. These can serve as evidence that they are of interest to a wide population of the researchers. As indicated above, the results of research on the subjects related to the socio-economic geography and spatial management are also presented in several other scientific journals, including some of Polish journals. Some of them are

not recognized in WoS while others, e.g. Bulletin of Geography-Socio-Economic Series, European Spatial Research and Policy, Geographia Polonica, Miscellanea Geographica, Studies of the Industrial Geography Commission of the Polish Geographical Society, Quaestiones Geographicae are indexed only recently. It could be recommended to consider, in the extension of the analysis presented within this paper, to include the contents of some of these journals in the near future. This would allow to consider the interest of many other researchers in application of statistical methods in analysis of data relevant to the socio-economic geography. It should be observed that Polish researchers, e.g. T. Czyż, W. Ratajczak, P. Czapliński, Z. Chojnicki, D. Jędrzejczyk, T. Stryjakiewicz, I. Jażdżewska, J. Hauke, (2021) belong to that group.

The sample of 592 papers published in the journals indicated below in the period of time 2012–2021 was subjected to the detailed analysis. Interestingly, the fraction of papers in which the statistical methods reached almost 89% for publications in Landscape and Urban Planning and Applied Geography while only about 10% for publications in Geoforum. The correlation-regression (R^2) was the most frequently used method. It was applied to analyze data in almost 40% of papers reviewed. The simple, arithmetic average was the second most popular statistical method used by the authors of the papers published in analyzed journals. This was followed by differentiation our statistical models, which were used to a similar extent. Unexpectedly, that order of popularity of the statistical methods did not changed during the decade under consideration (2012–2021). Various types of statistical software were used as the analytical tools. R Project for Statistical Computing (R) was the most popular one. It was used as an analytical tool in the research resented in 53 papers. Interestingly, 43 of them were published in Landscape and Urban Planning. Such high frequency of using that software can be explained considering its high computing capabilities combining with availability, as it is free of charge statistical software working in UNIX, Windows, and in the MacOS environments. The second place most popular statistical software was Statistical Package for the Social Sciences (SPSS), currently widely used also in various scientific disciplines. For the spatial analysis the ArcGIS software is applied. The NVivo package is used for the analysis of qualitative data obtained from surveys while MS Excel package is helpful in the quantitative analysis.

5. Conclusions

The literature query allowed drawing a picture of differentiation and intensity of the statistical methods and statistical software application in studies carried out in the area of socio-economic geography and spatial management. Differences observed between the sub-disciplines reflect the specificity of the research practices characteristic

for them. For example, the papers dealing with the subject related to spatial management usually present the conclusions reached within the applied research projects which are carried out by large multidisciplinary, often international research teams. In that case, advanced analyzes based on statistical models and methods (often multidimensional methods and triangulation) are used. On the other hand, studies in the area of social geography are less costly and usually carried out by individual researchers. Thus, the application of less advanced and simpler analytical methods prevail, producing more general sets of results (often only as percentage values). The papers in the area of economic geography present mainly the results of secondary data analysis based on the tests and dependency research carried out with the use of statistical methods and models.

There is no doubt that conducting research with the support of statistical methods increases the credibility and reliability of their results as well as ensures the correctness of inference. This is particularly important for the analysis of spatial phenomena which is becoming more and more complex.

It can be concluded that geographers should constantly develop and improve their competencies in the area of applied statistics. Also, the involvement of professional statisticians in the development of statistical tools better suited for studies of various geographic phenomena would be highly beneficial for the discipline.

Limitations and future studies.

The conclusions presented above are based on the analysis of the representative but relatively small sample of the literature resources available. This can be considered as the limitation of the certainty of the results presented. The analysis of a larger sample and observation of possible changes which have occurred in recent years will be performed to draw a wider and more precise picture of the phenomena under observation.

References

Agresti, A., (2009). *Statistical methods for the social sciences*, Vol. 207. Upper Saddle River, NJ: Pearson Prentice Hall.

Burt, J. E., Barber, G. M. and Rigby, D. L., (2009). Elementary statistics for geographers. *Guilford Press*.

Czapliński, P., (2008). Problematyka badawcza przemysłu w geografii na tle nauk ekonomicznych, *Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego*, 11, pp. 46–52.

Czyż T., Ratajczak W., (1986). Mathematical methods in economic geography. *Concepts and Methods in Geography*, 1, pp. 99–126.

Czyż, T., (2016). Metoda wskaźnikowa w geografii społeczno-ekonomicznej. *Rozwój Regionalny i Polityka Regionalna*, (34), pp. 9–19.

Czyż, T., Chojnicki, Z., (2019). Rola poznańskiego ośrodka geograficznego w implementacji metod i modeli matematycznych w geografii społeczno-ekonomicznej. *Rozwój Regionalny i Polityka Regionalna*, (45), pp. 9–21.

Dorocki, S., Jenner, B., (2016). Recepta na nienormalność rozkładu i współzależność obserwacji z wykorzystaniem testów randomizacyjnych i testu Mantela – na przykładzie rozmieszczenia zasobów ludzkich w regionach Francji. *Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego*, 30(2), pp. 186–197. https://doi.org/10.24917/20801653.302.12.

Fleszar, M., (1962). Zarys historii geografii ekonomicznej w Polsce do 1939 r. Warszawa. *Wydawnictwa Geologiczne*.

Haining, R. P., Haining, R., (2003). Spatial data analysis: theory and practice. Cambridge university press.

Hanushek, E. A., Jackson, J. E., (2013). Statistical methods for social scientists. *Academic Press*.

Hauke J., (2021). Metody statystyczne w geografii społeczno-ekonomicznej: szkic historyczny oraz ograniczenia i korzyści stosowania w dobie cyfryzacji. *Czasopismo Geograficzne*, 92(1), 73–93. https://doi.org/10.12657/czageo-92-04.

Huntington, E., (1927). The quantitative phases of human geography. *The Scientific Monthly*, 25(4), pp. 289–305.

Jażdżewska, I., (2021). Od nauk geograficznych w kierunku nauki o geoinformacji. *Wydawnictwo Uniwersytetu Łódzkiego*.

Jędrzejczyk, D., (2000). Mathematical Method of Analysing Forms of the Rural Settlements of Benon Janowski. Miscellanea Geographica, *Regional Studies on Development*, 9(1), pp. 159–164.

Kitchin, R., (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, 3(3), pp. 262–267.

Runge, J., (2006). Metody badań w geografii społeczno-ekonomicznej–element metodologii, wybrane narzędzia badawcze, *Wyd. Uniwersytetu Śląskiego, Katowice*.

Schabenberger, O., Gotway, C. A., (2017). Statistical methods for spatial data analysis. *CRC press.*

Stryjakiewicz T., (2003). Nowe interpretacje starych wskaźników w geografii społeczno-ekonomicznej. [W:] H. Rogacki (red.), Problemy interpretacji wyników metod badawczych stosowanych w geografii społeczno-ekonomicznej i gospodarce przestrzennej. *Bogucki Wydawnictwo Naukowe*, Poznań, pp. 135–142.

Wallis, B. C., (1912). The Use of Formulæ in Quantitative Geography. *Geographical Journal*, pp. 175–178.