

Geostatistical methods in the assessment of the spatial variability of the quality of river water

Małgorzata Krasowska^{1, *}, and *Piotr Banaszuk*¹

¹Białystok Technical University, Faculty of Civil and Environmental Engineering, Department of Environmental Protection and Management, ul. Wiejska 45E, 15-351 Białystok, Poland

Abstract. The research was conducted in the agricultural catchment in north-eastern Poland. The aim of this study was to check how geostatistical analysis can be useful for the detection zones and forms of supply stream by water from different sources. The work was included the implementation of hydrochemical profiles. These profiles were made by measuring the electrical conductivity (EC) values and temperature along the river. On the basis of these results, the authors calculated the coefficient of Moran I and performed semivariogram and found that the EC values are correlated on a stretch of about 140 m. This means that the spatial correlation between samples of water in the stream is readable over a distance of about 140 meters. Therefore it is believed that the degree of water mineralization on this section is shaped by water entering the river channel migration in different ways: through tributaries, leachate drainage and surface runoff. In the case of the analyzed catchment, the potential sources of pollution were drainage systems. Therefore, the spatial analysis allowed the identification pollution sources in a catchment, especially in drained agricultural catchments.

1 Introduction

In recent years many new methods, as well as conceptual and mathematical models were developed to quantify the transfer of non-source pollutions from the catchment to surface waters [1–4]. Among the methods, which allows understanding better phenomena occurring in the geographical space is the geostatistical analysis. The characteristic feature of most natural phenomena is an autocorrelation or spatial dependence a situation when phenomenon being researched within one spatial unit causes an increase or decrease of that phenomenon presence in the adjacent units [5, 6]. This dependence is linked to so-called “first law of geography” by Tobler who in 1970 stated that “everything is related to everything else, but near things are more related than distant things” [7].

Methods based on statistics can be useful to detect the correlation between variables and to identify factors that are not available in direct observation [8]. Therefore, currently, geostatistical methods are used for modeling and forecasting of the pollution of different compartments of the environment and identification of hidden structures of chosen selected measurement series of integrated environmental monitoring [9–11]. Many authors confirm

* Corresponding author: m.krasowska@pb.edu.pl

that those methods are an efficient tool which lets to define the temporal and spatial dimension of water quality fluctuation, both surface-, and groundwater [12–14]. Yan et al. [15] stated that geostatistical analysis helped to identify the polluted risky regions and could be useful and valuable to pollution control strategies, as well as plan and management on the watershed. Besides, they were also helpful to further research on water quality simulation and validate the simulation accuracy in watershed space. The study by Einax and Soldt [16] and Eneji et al. [17] confirmed the usefulness of geostatistical analyses to assess the water quality and identify sources of pollution. Geostatistical methods can be used to describe the spatial distribution and variability of the quality of riverine ecosystems [18]. Therefore, in this paper, we attempt to use selected geostatistics to verify their usefulness in the detection of the spatial structure of the chemical composition of the stream during the period of high waterlogging. Simultaneously we checked how semivariance analysis and Moran's I could be applied for identifying zones and forms which supply solutes the stream in the agricultural catchment.

2 Study area

The study was performed within a small catchment (187 ha) in the north-east Poland. That stream is a left-bank tributary of Horodnianka River. The catchment is rural with 75% of arable lands, 16% – grasslands, 3.5% – woodlands and 5.5% built-up areas and wastelands. Approximately 60% of the agricultural land is artificially drained, mainly by underground pipes. The analyzed stream has no natural outflow and originates from an underground drainage system. Its length between the origin and the gauging profile was about 1550 m. At this section, the stream has three tributaries (Fig. 1).

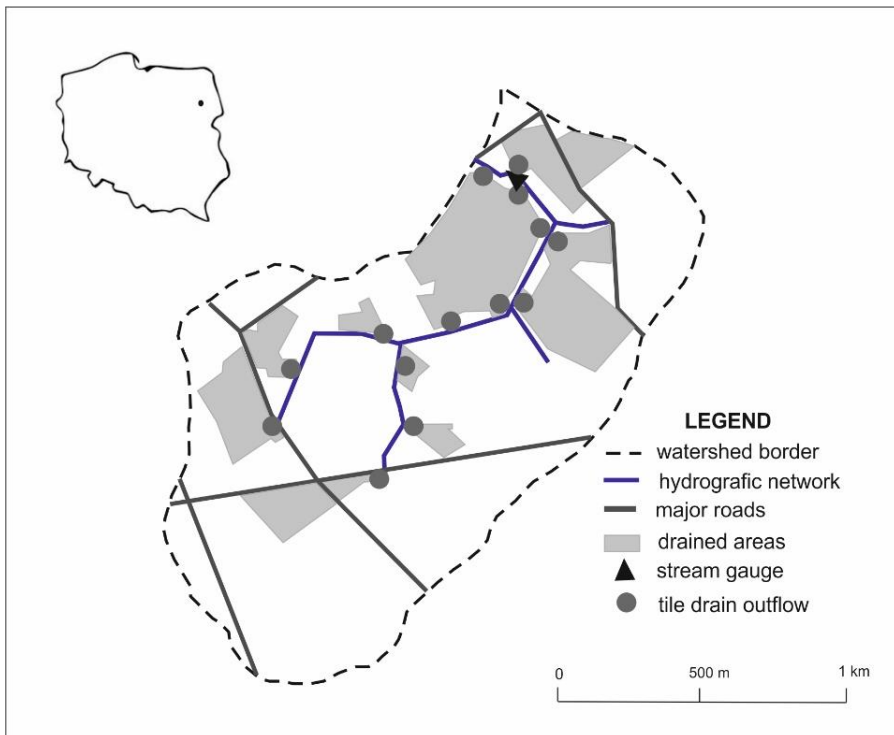


Fig. 1. Location and features of the studied catchment.

3 Methods

For statistical analyses, we used data collected during hydrochemical profiling of stream in March 2010, during a high waterlogging of catchment and high stream flow. They included measurements of electrolytic conductivity, water temperature and the concentration of NO_3^- , Cl^- , $\text{Si}_2\text{O}_3^{2-}$, performed along the creek course at regular intervals of 10 m. The measurement of electrolytic conductivity and temperature of the water was determined by using Hach Lange HQ40D, while the concentration of ions was determined by spectrophotometry using Slandi LF300.

It was assumed that the conductivity could be employed as a proxy for the content of solutes in water [19]. Furthermore, it was assumed that changes in water chemistry along the stream resulted from a supply of chemicals and were not related to the biogeochemical reactions that occur in the riverbed [20]. This assumption seems to be fulfilled because the measurements were performed in the winter season when biota was dormant phase. To detect the spatial structures underlying the observed variability of the water quality we used Moran's I statistic calculated according to the formula (1):

$$I = \frac{n \sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i \sum_j w_{ij} \sum_i (z_i - \bar{z})^2} \quad (1)$$

where:

n – number of observations,

w_{ij} – element of a matrix of spatial weights,

z_i, z_j – variable value at a particular location i or j ,

\bar{z} – the mean of the variable [21].

Moran's I statistics [22] belongs to the one of the oldest and most commonly used in studies of spatial autocorrelation. It is interpreted as the coefficient a correlation occurs between the values of the variable in the spatial unit. The value of Moran's I generally varies between 1 and -1. Positive autocorrelation in the data translates into positive values of Moran's I; negative autocorrelation produces negative values. No autocorrelation results in a value close to 0. Moran's coefficient can be used in the assessment of the pattern of dispersion, randomness, and concentration. Moran's I values may range from -1 (dispersed) to +1 (clustered) [6, 23, 24]. The result of the Moran analysis is correlogram. From the intersection of correlogram line with the x-axis, we can evaluate the type of spatial distribution [25]. The SAM 4 software was used to calculate the spatial correlation [26].

The semivariance is defined as half the average squared difference between values (2):

$$\gamma = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u_\alpha) - z(u_\alpha + h)]^2 \quad (2)$$

where:

$N(h)$ – number of pairs of sampling locations located at distance h from one another

Z – the value of the variable u_α at the sampling location α [27].

Empirical semivariogram was made to characterize the spatial continuity which exists in the analyzed data set. Semivariogram model was characterized by three parameters:

- range, which represents the large-scale correlation,
- nugget effect, which represents uncorrelated small-scale variability,
- sill, which is the upper bound of γ [25, 28].

In practice analysis of semivariograms allows determining the decrease of self-similarity with increasing distance between measurements. Semivariograms models were made by using Surfer 9.

4 The results

4.1 Hydrochemical profiles

During the hydrochemical mapping of the stream measurements of electrolytic conductivity (EC) and temperature (T) along its were made (Fig. 2).

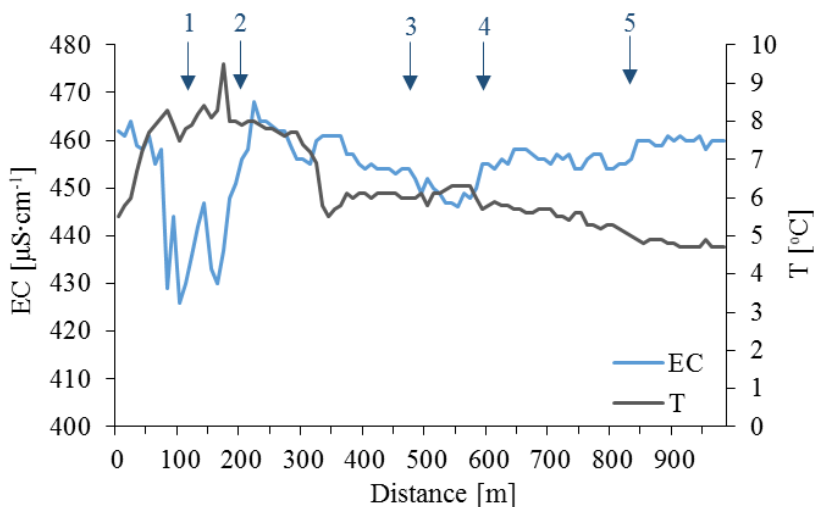


Fig. 2. Variation in the electrolytic conductivity (EC) and temperature (T) of water down the stream (potential sources of pollution: 1, 2, 4 – tributaries; 3, 5 – drain tile).

We observed a sharp decrease in EC value at 100–150 meter stream (from the source), with the fast recovery to initial values tens meters further (with one peak at 250 m), where EC was aligned at about $450 \mu\text{S}\cdot\text{cm}^{-1}$ and did not show any clear trend. The water temperature showed the highest value up to 150 m after that there was a marked fall. Besides, along with the water course, an evident variability in physio-chemical properties of water was observed (Tab. 1). NO_3^- changed from $14 \text{ mg}\cdot\text{dm}^{-3}$ at the 150 meter of the water course to $30 \text{ mg}\cdot\text{dm}^{-3}$ at 650 meter. Its variability was similar to changes in $\text{Si}_2\text{O}_3^{2-}$ (Fig. 3). Based on hydrochemical mapping it is possible to detect zones of the stream supply with solutes.

Table 1. Statistics of the chosen physio-chemical properties measured along the water course.

	EC [$\mu\text{S}\cdot\text{cm}^{-1}$]	Temp. [$^{\circ}\text{C}$]	NO_3^-	$\text{Si}_2\text{O}_3^{2-}$	Cl ⁻
			[$\text{mg}\cdot\text{dm}^{-3}$]		
M	457	5.8	25.5	1.9	12.7
\bar{X}	457 (58)	6.0 (5.4)	24.9 (31.6)	1.9 (1.5)	12.8 (4.4)
SD	10.8	1.3	6.4	0.4	1.1
CV	2.4	21.7	25.9	21.7	8.5

M – median, \bar{X} – arithmetic mean, *SD* – standard deviation, *CV* – coefficient of variation, (in brackets the value of the range)

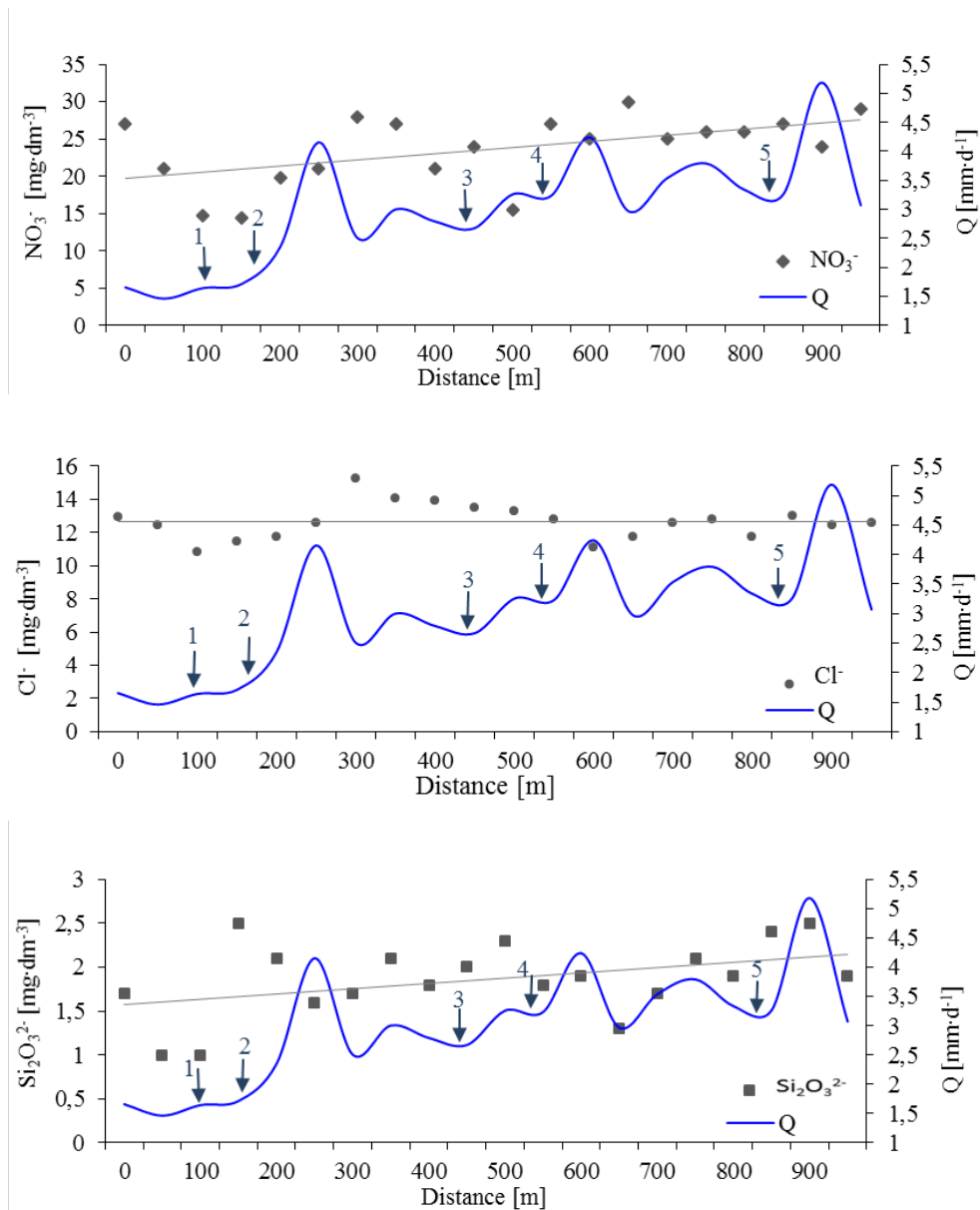


Fig. 3. Dynamics of the stream discharge (Q) and concentration of dissolved components down the stream (potential sources of pollution: 1, 2, 4 – tributaries; 3, 5 – drain tile).

4.2 Semivariance analysis

The semivariogram, which was made for the measurement of electrolytic conductivity showed the presence of autocorrelation at a distance of 100–120 m (Fig. 4). The best-matched model was the spherical model, which the overall formula is (3, 4):

$$\gamma(h) = C_0 + C \left(1,5 \left(\frac{h}{A_0} \right) - 0,5 \left(\frac{h}{A_0} \right)^3 \right) \text{ for } h \leq A_0 \quad (3)$$

$$\gamma(h) = C_0 + C \text{ for } h \geq A_0 \tag{4}$$

where:

C_0 – nugget effect,

C_0+C – sill

A_0 – range,

h – (distance) between samples.

The value of variance identified the size of the measurement error, which was relatively small and amounted to about $2,5 \mu\text{S}\cdot\text{cm}^{-1}$. Whereas sill, which is the upper bound of semivariance was about 82. It was found that the spatial correlation (A_0) between the samples of water in the stream is readable on the stretch to about 140 m. At this point, an increase in the value of semivariance has been "detained."

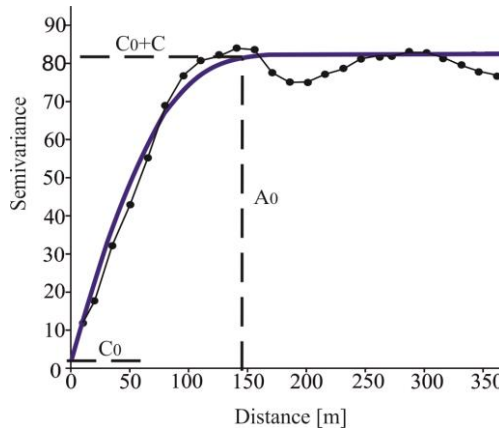


Fig. 4. Semivariogram with their model for analyzed electrolytic conductivity.

4.3 Moran's analysis

There are similar conclusions about Moran's statistics. Correlogram made for EC measurements indicated that the autocorrelation coefficient is positive on the stretch to about 140 m. This is the point where the intersection of a line with the X-axis occurred (Fig. 5). The correlation between the variables decreases from +0.2 to -0.2 at a distance of about 240 m, and then rise again and reaches a low positive value at a distance of 735 m. Such a correlogram results indicate gradual structure creation. Moreover, the cyclical fluctuation was discovered at every 140–200 m.

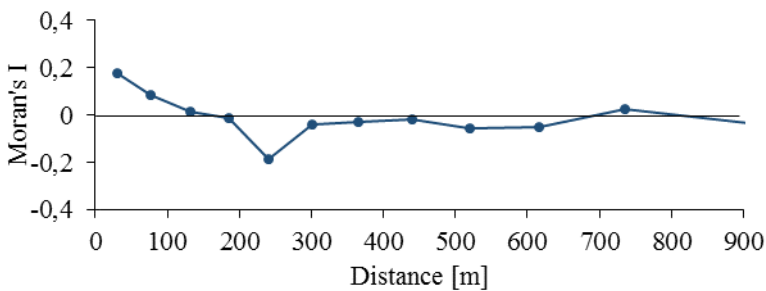


Fig. 5. Correlogram showing the variation of spatial autocorrelation in the time function for the measurement of electrolytic conductivity.

5 Summary

The study was conducted in the winter half-year, under the conditions of high catchment wetness when intensive movement of solutes to the stream occurs [19, 20, 29, 30]. Contaminants can migrate in different ways: with surface runoff, shallow subsurface, and deeper groundwater flow. In agricultural catchments, which in the north-eastern Polish are largely drained, drainage systems play an important role in the movement of solutes from the catchment to stream [30]. The agricultural catchments water from drains are usually are contaminated with nutrients [31].

The research shows that the sources and routes of migration in the basin, which are rich in components dissolved, had an impact on the mineralization of the streamwater. By hydrochemical mapping, it has been found that the variability of electrolytic conductivity in the stream water was driven mainly by water inflow from surface tributaries and tile drains. On the basis of hydrochemical mapping we can only approximately zones which supply waters from the catchment. In this paper was performed geostatistical analysis for the precise identification of these forms. The geostatistical methods used in this study fully confirmed observations made in the field. The analysis of the semivariance and Moran's spatial correlation showed that on the section length of 140 m, there was a spatial correlation, which reflected the real influence of solutes supplied by tributaries and tile drains regularly placed every 140 m. This is the maximum range of the autocorrelation, which is a real similarity between samples [20]. The impact of the sources of dissolved substances was fading over this distance. The geostatistical methods, have shown the occurrence of relatively frequent and regular source of solute, which are situated along the water course. Thus they can be used for quite correct identification of the hidden spatial structures that affect the water quality, as it was stated by Stach [20], Liu et al. [13], Eneji and Onuche [17], Basu et al. [14], Yan et al. [15].

The acknowledgments: research was funded by Ministry of Science and Higher Education as a part of the project S/WBiŚ/01/17.

References

1. B. Namysłowska-Wilczyńska, A. Wilczyński, *Ochr. Sr.* **2**, 65 (1997)
2. S.P. Inamdar, M.J. Mitchell, *Water Resour. Res.* **42** (2006)
3. M.J. Gromiec, (IMGW Warszawa, 2008)
4. H.S. Wheater, N. McIntyre, T. Wagener, (Cambridge University Press, 2008)
5. R. Bivand, PWN Poznań, 23–28 (1980)
6. K. Janc, *Dokumentacja Geograficzna* **33**, 76–83 (2006)
7. H.J. Miller, *Ann. Assoc. Am. Geogr.* **94**, 284–289 (2004)
8. J.B. Racine, H. Reymond, (PWN, Warszawa, 1977)
9. P. Goovaerts, (Oxford University Press, 1997)
10. B. Nas, *Polish J. of Environ. Stud.* **18**, 6, 1073–1082 (2009)
11. A.S. Gharbia, S.S. Gharbia, T. Abushbak, H. Wafi, A. Aish, M. Zelenakova, F. Pilla, *J. Geosci. Environ. Protect.* **4**, 89–103 (2016)
12. H. Bu, X. Tan, S. Li, O. Zhang, *Ecotoxicol. Environ. Saf.* **73**, 5 (2010)
13. W.C. Liu, H.L. Yu, C.N. Chung, *Int. J. Environ. Res. Public Health* **8**, 4, 1126–1140 (2011)

14. S. Basu, K.S. Lokesh, *Int. J. Water Res. Environ. Eng.* **5**, 10, 591–596 (2013)
15. C.A. Yan, W. Zhang, Z. Zhang, Y. Liu, C. Deng, N. Nie, *PLoS ONE* **10**, 3 (2015)
16. J.W. Einax, U. Soldt, *Elsevier* **46**, 1, 79–91 (1999)
17. I.S. Eneji, A.P. Onuche, R.S. Ato, *J. of Environ. Protect.* **3**, 915–921 (2012)
18. B.G. Kotak, A. Selinger, *Manitoba Model Forest Report 05-2-63*, (2006)
19. M. Stolarska, *Acta Geographica Lodzienia*, **94** (2008)
20. A. Stach, *Przegląd Naukowy Inż. i Kszt. Środ. Wyd. SGGW Warszawa* **11**, 2 (25), 130–143 (2003)
21. P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, (Wyd. PWN, Warszawa, 2006)
22. P.A.P. Moran, *Biometrika* **37**, 17–23 (1950)
23. K.P. Overmars, G.H.J. de Koning, A. Veldkamp, *Ecol. Model.* **164**, 2, 257–270 (2003)
24. X.N. Huo, H. Li, D.F. Sun, L.D. Zhou, B.G. Li, *Int. J. Environ. Res. Public Health* **9**, 3, 995–1017 (2012)
25. M.J. Fortin, M.R.T. Dale, (Cambridge University Press, 2005)
26. T.F. Rangel, J.A.F. Diniz-Filho, L.M. Bini, *Ecography* **33**, 1, 46–50 (2010)
27. E. Gringarten, C.V. Deutsch, *Springer Berlin* **33**, 34, 507–534 (2001)
28. N.Z. Sun, A.Y. Sun, (Springer-Verlag, New York, 2015)
29. M. Klein, *Earth Surf. Process. Landforms* **6**, 173–178 (1981)
30. P. Banaszuk, M. Krasowska, A. Kamocki, *Springer-Verlag Berlin*, 63–77 (2011)
31. C.S. Tan, C.F. Drury, M. Soultani, I.J. van Wesenbeeck, H.Y.F. Ng, J.D. Gaynor, T.W. Welacky, *Wat. Sci. Tech.* **38**, 4–5, 130–110 (1998)