

## Spatial distribution and statistic analysis of the anthropogenic line forms on the different basic fields

Bartłomiej Szypuła

*Department of Geomorphology, Faculty of Earth Sciences, University of Silesia, Będzińska Str. 60, 41-200 Sosnowiec, Poland  
E-mail address: bartlomiej.szypula@us.edu.pl*

### ABSTRACT

Despite the rapid development of geoinformation technology and GIS - a classic cartogram is still widely used method for presenting geographic features and phenomena, especially with regard to the relative values connected with the basic fields. The aim of this article was to investigate how the size and shape of the different basic fields influence the results of the phenomenon presentation (in this case anthropogenic line forms). In the experiment were used fields in the shape of: square, hexagon, circle and triangle with different sizes: 1 km, 2 km, 4 km, 8 km and 10 km. Different field areas with the same height, but of a different shape affected to varied quantitative characteristics within them. However, different field heights have caused an obvious increase or decrease the detail of the results. To take a look at the image of the spatial distribution of line forms compared cartograms with another, independent method – kernel density analysis. After setting kernel density image with cartograms one turned out that basic fields shape did not change the image of spatial relations significantly and well-characterized them in general. For this study area the best results obtained after the application fields with heights of 2 km and 4 km in the shape of squares and hexagons. It appears that the hexagons better than squares reflect the spatial image of the forms (hexagons allow better representation of the directions and shapes of the studied phenomena), however, they are less common in a geostatistical researches, and that's why they are more difficult to use, especially for comparative analysis.

KEY WORDS: spatial statistics, analysis field, cartogram, kernel density, ArcGIS, Silesian Upland

### 1. Introduction

Cartogram as a quantitative method, is ideal for showing relationships in the spatial distribution of geographical phenomena. Generally accepted rule is to present relative values (i.e. relating to a specific area unit) by this method. Therefore making cartogram requires careful selection of the basic field size, which implies a detail of the result report. Another important element of the constructing of cartogram is dividing into the class values. Because of the many different methods of determining the class breaks one can use a traditional approach, based on the analysis of the histogram, and determine the class to strive for a similar size within them (see PASŁAWSKI, 2006) or try to group the ranges with similar values (see SALISZCZEW, 1998). While the nature of the data (number, distribution, statistics, etc.) is constant,

the basic field is defined by the researcher arbitrarily. The main goal of this article is to test how the size and shape of the basic field of analysis influence assessment of the spatial distribution of anthropogenic line forms in the southern part of the Silesian Upland. Is the size of the analytic field cause the scale generalization only or can disturb the spatial relationship of data values? and does the field shape has a significant impact on this image?

Reflections on this subject above are important, because the cartogram method is still often and readily utilized in the earth sciences, particularly with regard to the line objects. Examples would be the density analysis of the river network of the Vistula and the Bug interfluvium (SZALKIEWICZ, 1947), methodological issues relating to the structure of the density maps of the water network in relation to the Lublin area (SIRKO, 1971,

1978). In addition, a comparative study of short satellite lineaments and radar networks with the structure and morphotectonics of the marginal zone of Carpathian region (JAROSZEWSKI & PIĄTKOWSKA, 1988) and basic field analysis to estimate of valleys network density of episodic outflow (KĘSIK, 1961), gullies network density and intensity of gullies erosion (BURACZYŃSKI, 1964, 1977, 1984, 1989/1990; MARUSZCZAK, 1973; JÓZEFACIUK & JÓZEFACIUK, 1992; GAWRYSIK & HARASIMIUK, 2009).

## 2. Study area

The study area is located in the southern part of the Silesian Upland between 50°05' - 50°26'N and 18°33' - 19°13'E coordinates (Fig. 1). According

to the Polish regional and geomorphic divisions (GILEWSKA, 1986; KONDRACKI, 2001) this study site belongs to the Silesian Upland and partially to the Silesian Lowland (in the west) and the Oświęcim Basin (in the south). The study area occupies about 1.920 km<sup>2</sup> (48 km x 40 km). True heights are more than 200 m (elevation range: 198 to 400 m a.s.l., mean: 268 m a.s.l.), so the local relief is diversified. The highest elevations are built of resistance Carboniferous rocks (sandstones, shales) and Triassic (limestones, dolomites and marls). Subsidences are filled up with Pleistocene deposits (boulder clays) and Holocene (sands and valley gravels). The whole area is cut through by the biggest rivers in the NW-SE direction (among other things Kłodnica, Przemsza, Brynica).

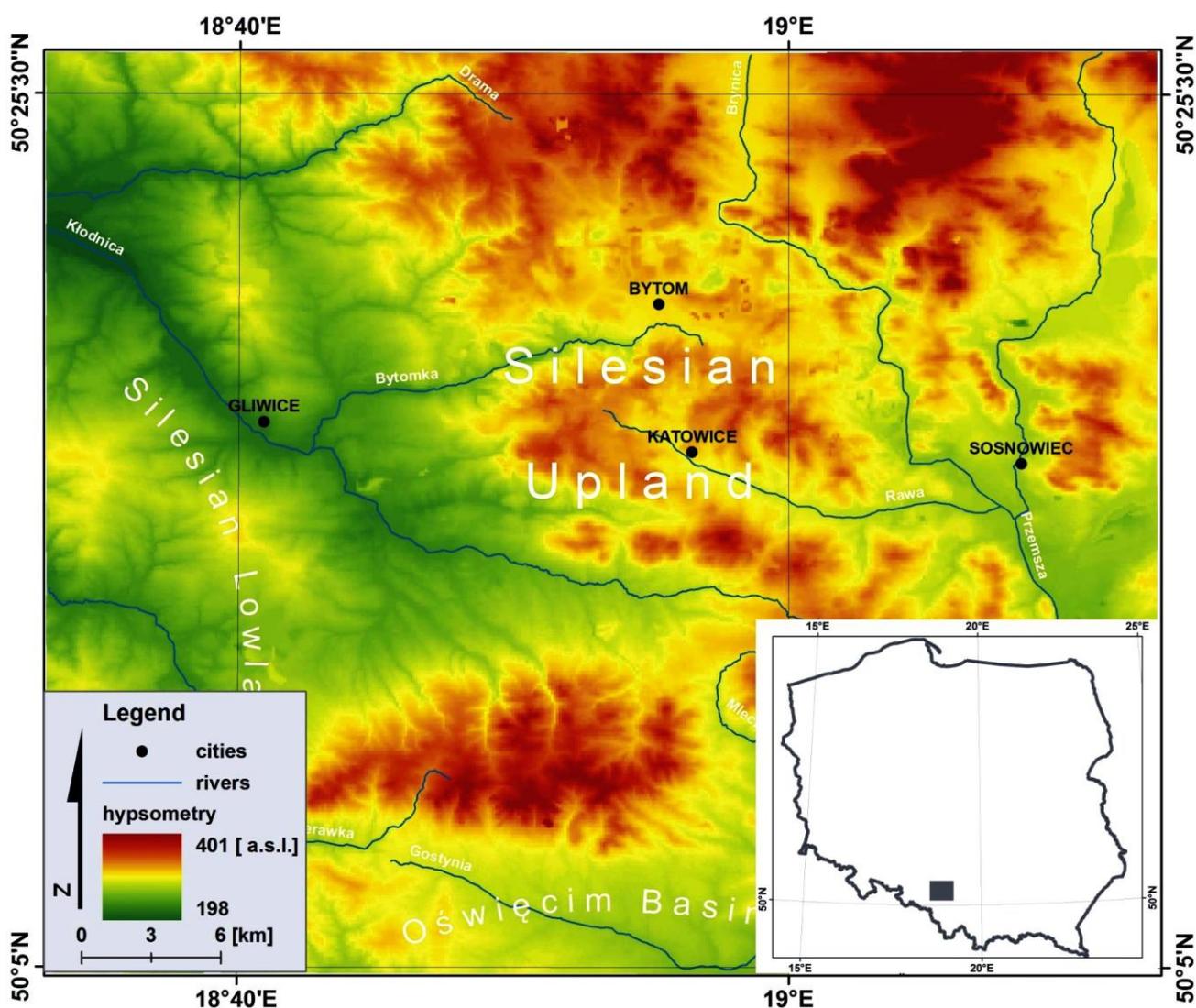


Fig. 1. Hypsometry and location of study area

Due to the nature of the research this part of the Silesian Upland was chosen deliberately - the surface is very heavily transformed by human activity. The coal mining, zinc and lead ores, filling

sand and the related industrial development and urbanization caused intensive development of various anthropogenic forms in this area (see DULIAS, 2013).

### 3. Source and research methods

The basic materials used for this work included the selected sheets of Topographic Map 1:50 000 (1995). The scanned maps were subject to rectification to the current reference polish system (ETRS 1989 Poland CS92). The next step was vectorizing of interesting convex line anthropoforms - railway and road embankments (quantitative data used from SZYPUŁA, 2013) (Fig. 2). The aim of the study was to capture the spatial and quantitative relationships among these anthropogenic forms, and then analyse and answer the question: which basic fields are the best to present the analysed phenomenon?

The presentation of the experiment results one used cartogram method, which (thanks to the simplicity of the design and results interpretation) is very well suited for benchmarking the relative values of a given phenomenon in relation to the established real spaces.

When you create a cartogram there are two key issues: the values division of the class and the

basic field. When it comes to the division of the data into classes - one has to arbitrarily assumed grouping into classes using the "natural breaks" as the best method which reflects the spatial continuous data. The method is based on a natural grouping that identifies break points by picking the class breaks that best group similar values and maximize the differences between classes. The features are divided into classes whose boundaries are set where there are relatively big jumps in the data values (ARCGIS., 1999-2008). In order to examine the influence of basic fields for spatial image of the line forms the fields of different sizes and shapes was used. It was decided to basic field in the shape of: square, hexagon, circle and triangle with heights: 1 km, 2 km, 4 km, 8 km and 10 km (Fig. 3). Further line data preparation relied on intersecting them to the basic fields of every shape. The intersection was intended to allow statistical calculations for each basic field separately (Fig. 4). In effect 20 sets of line data were made, which served for making the cartograms.

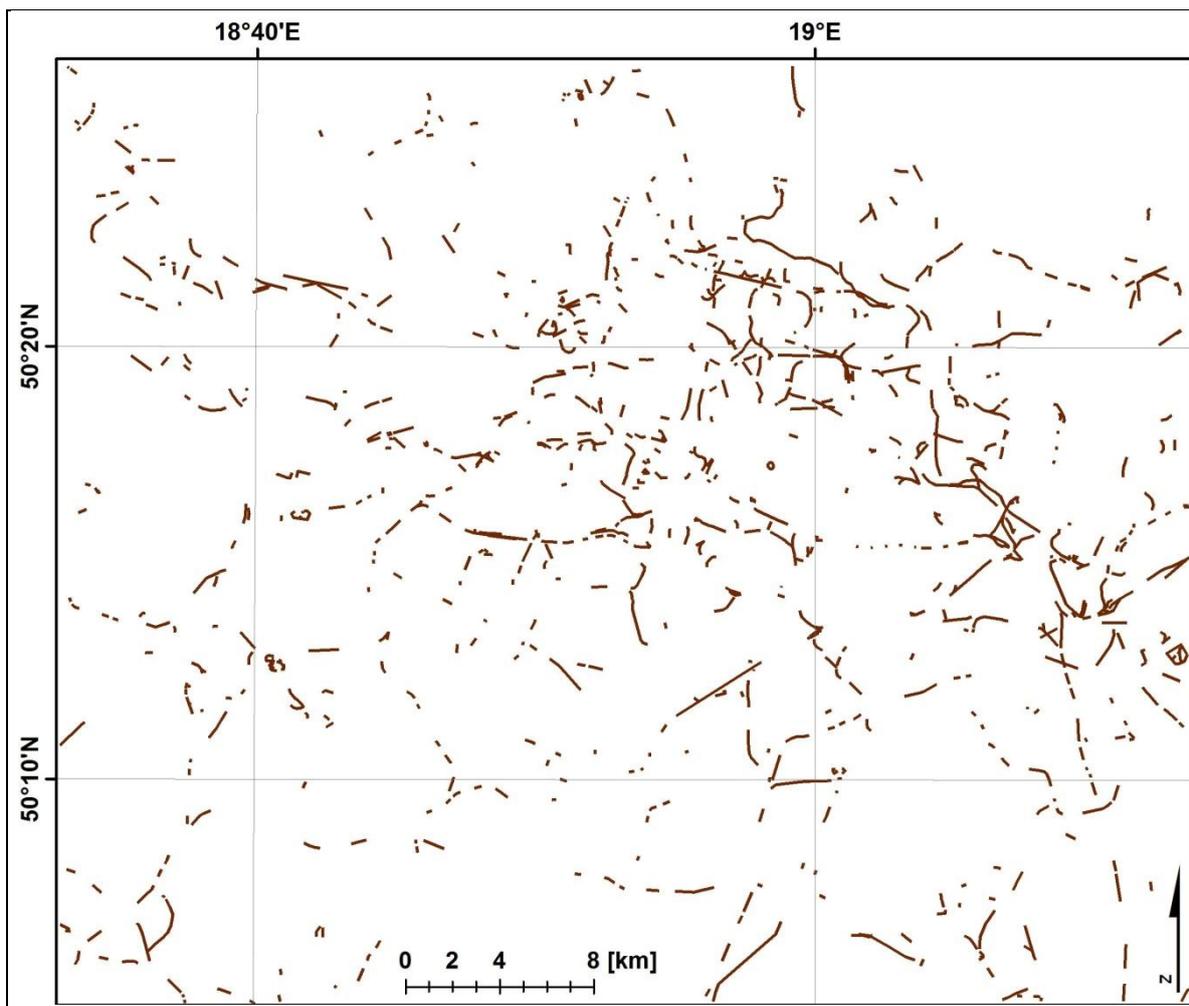


Fig. 2. Spatial distribution of analysed line forms

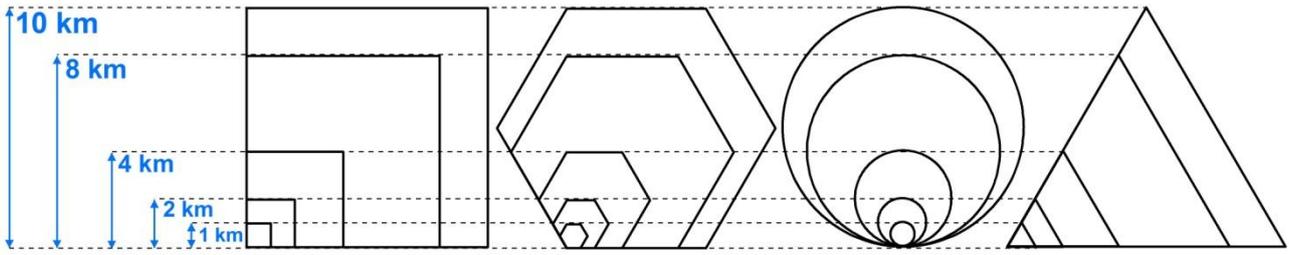


Fig. 3. Sizes and shapes of used basic fields

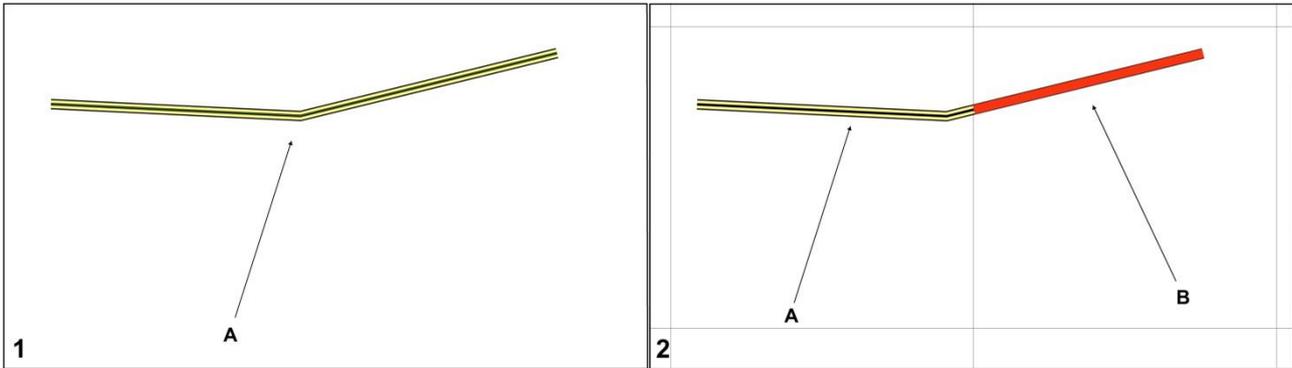


Fig. 4. Line data before (1) and after intersection (2) by the square basic fields

#### 4. Results

(1) Statistical structure analysis of raw data (Tab. 1) showed, that data are characterized by strong positive skewness of distribution, which is clearly visible in the histogram (Fig. 5). Typically in the normal distribution about 68% of the observations is within one standard deviation from the mean (GREGORY, 1970). In this case nearly 89% of all values are within one standard deviation from the average. The range data (5.56 km) and

a standard deviation (0.58 km) suggest the data are relatively diverse. However, the analysis of the histogram (Fig. 5) and calculating a variation coefficient (95%) convinces us that the data are characterized by very high variability (e.g. isolated high values above 3). It can be said that on the one hand, there is a huge predominance of small values (up to 1-1.5 km), which imply a low average value (0.60 km), on the other, with a very high variability of the occurrence of a few, but high values (from 2 to 5.7 km).

Table 1. Statistical data of line forms length

Parameter	Value
minimum	0.06 km
maximum	5.62 km
range	5.56 km
mean	0.60 km
standard deviation	0.58 km
median	0.58 km
mode	0.30 km
number of pieces	797
variation coefficient	95%

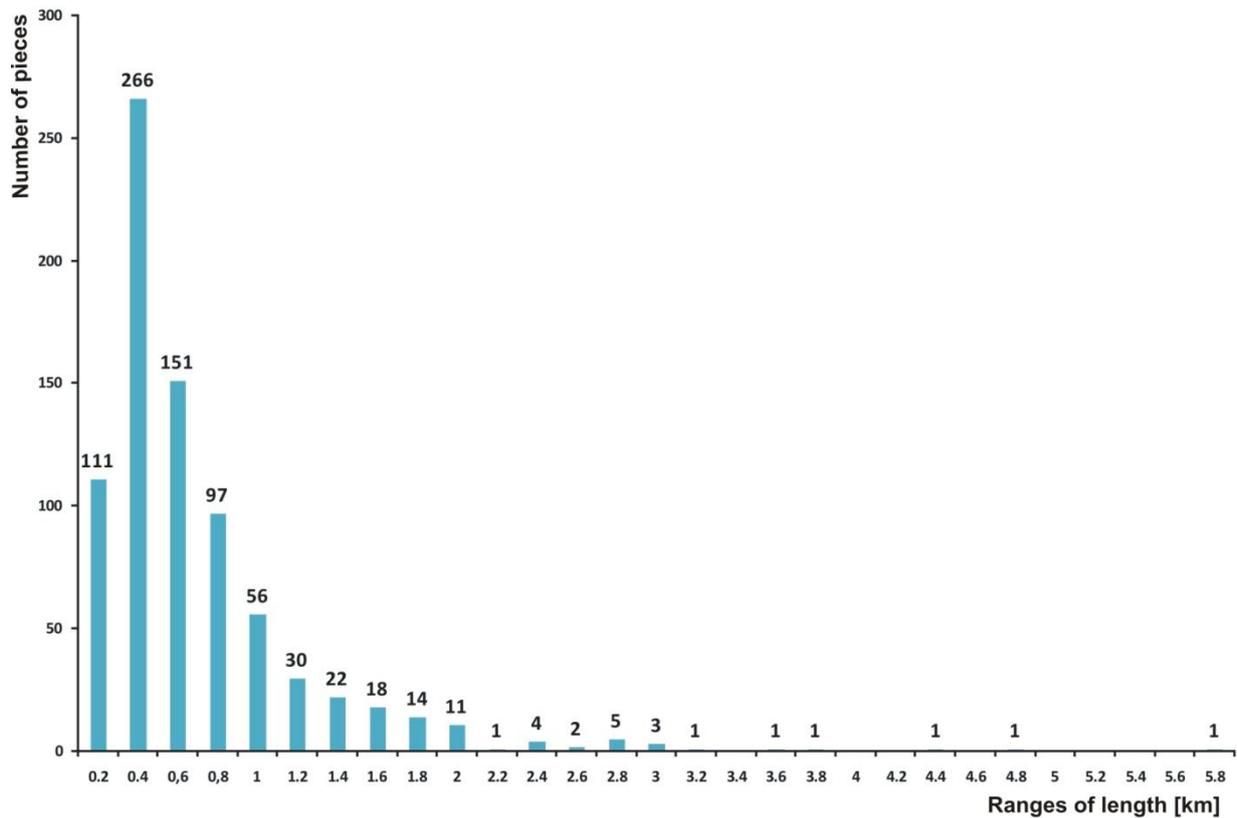


Fig. 5. Histogram of line forms length data

(2) The same height basic fields but in different shapes have different areas. Square fields occupy the largest area (Fig. 6). Fields of other shapes are smaller than the square of: 14%

(hexagon), 22% (circle) and 43% (triangle). This fact was not without influence on the statistical and spatial image results.

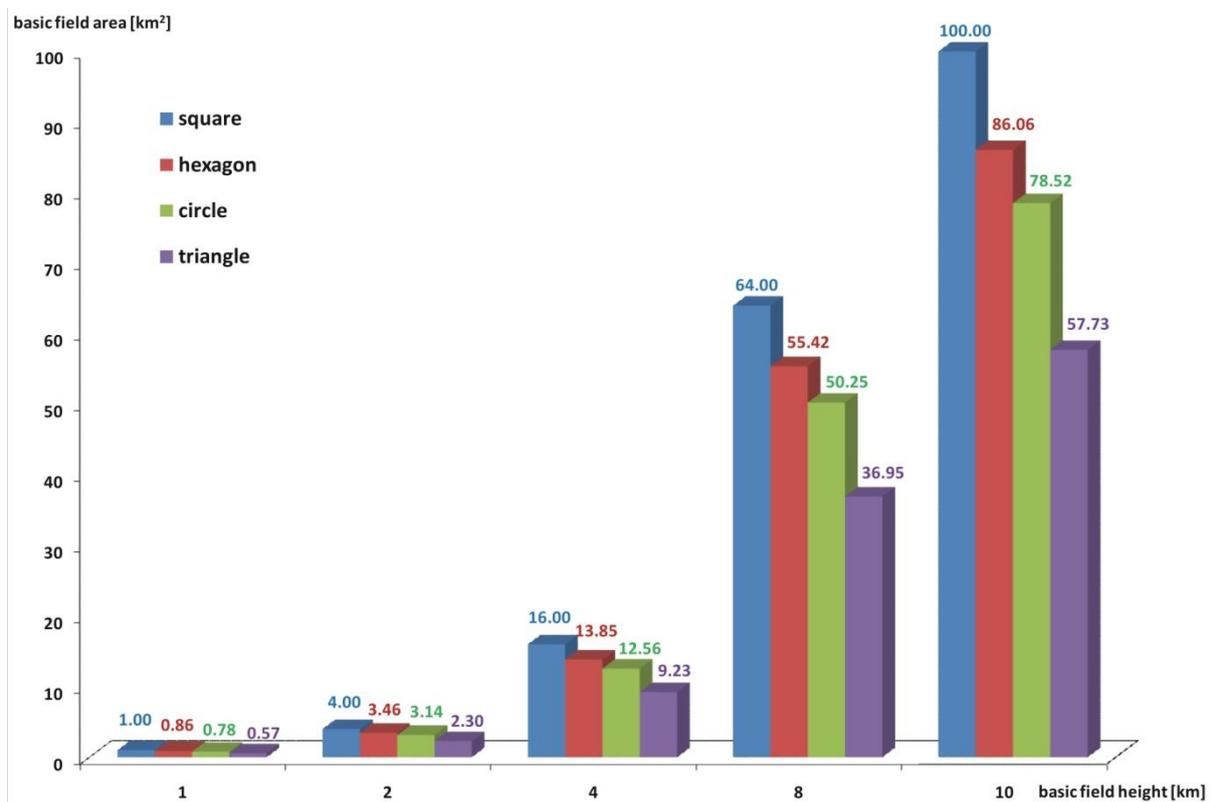


Fig. 6. Basic fields sizes and areas

(3) The results of the statistical characteristics related to fields of different shapes were grouped in size shown in the following Tables 2-6 (these results relate to processed data - intersected by basic fields). The range of values is the greatest for square (1, 2, 4 km) and hexagonal fields (8, 10 km),

and the lowest for circular (1, 4, 8 km) and triangular fields (2, 10 km). The maximum difference (48%) in the values range occurred in the 10 km size fields between hexagons (68 km) and triangles (35 km). This is probably related to occupied space (hexagons are 1.5 times larger than triangles).

Table 2. Statistical data of basic fields (1 km)

Parameter	Field shape			
	square	hexagon	circle	triangle
minimum	0.00 km	0.00 km	0.00 km	0.00 km
maximum	3.03 km	3.19 km	2.17 km	2.41 km
range	3.03 km	3.19 km	2.17 km	2.41 km
mean	0.25 km	0.21 km	0.19 km	0.14 km
standard deviation	0.46 km	0.42 km	0.38 km	0.30 km
variation coefficient	184.0%	200.0%	200.0%	214.28%
amount	1920	2200	1920	3320

Table 3. Statistical data of basic fields (2 km)

Parameter	Field shape			
	square	hexagon	circle	triangle
minimum	0.00 km	0.00 km	0.00 km	0.00 km
maximum	6.87 km	6.70 km	6.12 km	5.58 km
range	6.87 km	6.70 km	6.12 km	5.58 km
mean	1.00 km	0.85 km	0.77 km	0.58 km
standard deviation	1.29 km	1.13 km	1.07 km	0.84 km
variation coefficient	129.0%	132.94%	138.96%	144.82%
amount	480	560	480	820

Table 4. Statistical data of basic fields (4 km)

Parameter	Field shape			
	square	hexagon	circle	triangle
minimum	0.00 km	0.00 km	0.00 km	0.00 km
maximum	16.68 km	17.84 km	12.47 km	13.16 km
range	16.68 km	17.84 km	12.47 km	13.16 km
mean	4.00 km	3.40 km	3.05 km	2.27 km
standard deviation	3.76 km	3.38 km	2.90 km	2.42 km
variation coefficient	94.0%	99.41%	95.08%	106.6%
amount	120	140	120	210

Table 5. Statistical data of basic fields (8 km)

Parameter	Field shape			
	square	hexagon	circle	triangle
minimum	0.54 km	0.54 km	0.54 km	0.29 km
maximum	43.35 km	49.04 km	34.71 km	38.94 km
range	42.81 km	48.50 km	34.17 km	38.65 km
mean	16.01 km	13.52 km	12.38 km	9.26 km
standard deviation	11.54 km	10.60 km	9.40 km	7.99 km
variation coefficient	72.07%	78.4%	75.92%	86.28%
amount	30	35	30	50

Table 6. Statistical data of basic fields (10 km)

Parameter	Field shape			
	square	hexagon	circle	triangle
minimum	7.90 km	0.00 km	6.38 km	0.00 km
maximum	61.73 km	68.23 km	49.25 km	35.33 km
range	53.83 km	68.23 km	42.87 km	35.33 km
mean	24.02 km	18.71 km	17.60 km	13.07 km
standard deviation	15.61 km	15.91 km	12.60 km	10.13 km
variation coefficient	64.98%	85.03%	71.59%	77.50%
amount	20	25	20	36

Similarly, maximum average values occurred in square fields, and minimum in triangles. The differences in average values reached up to 45% – which was due to the large differences in surface area. Along with the arithmetic mean values followed the standard deviation (except for 10 km fields, which standard deviation was higher for the hexagonal fields – see Tab. 6).

The variation coefficient is a consequence of the mean and standard deviation value. It tells us about how much data within a set (in this case – the basic field) are varied; it shows the percentage of the data set variability (see GREGORY, 1970; ŁOMNICKI, 2003). In general, it can be observed very strong differentiation of the data: the greatest variation coefficient value was for the triangular fields (the only exceptions are basic fields of 10 km height, where the greatest variability value hexagons reached), and the lowest values for the square fields. This is an obvious relationship with the occupied surface area (see Fig. 6).

(4) It was decided to calculate the kernel density to compare the spatial data distribution presented in cartograms of the different basic fields. Kernel Density calculates the density of features in a neighbourhood around those features, ie, in a space of defined radius from the point or line by using the kernel function (see ARCGIS..., 1999–2008). The use of the kernel function for lines was adapted from a quadratic kernel function for point density described in SILVERMAN (1986) and well suited for working with spatial data. In order to be possible comparison – the kernel density calculations were made in such way that the search radius was such the same as the comparing basic field height (Fig. 7–11).

The resulting image was too detailed and thus fragmented and diffuse for basic fields of height 1 km (Fig. 7).

Cartograms with squares and hexagons reached the most similar effects of the spatial distribution to the kernel density values. Hexagon shape better reflects the direction geometry than the square. The best results of line forms spatial character were observed for cartograms with 2 km and 4 km height (Fig. 8 and 9). From the initial colour chaos emerged area with the dominant values where the phenomenon concentration turned out to be great. Peripheral zones with the minimum values in the south-western and northern parts of the area marked quite clearly.

Cartograms with 8 km and 10 km field height were characterized by too much generalization (Fig. 10 and 11). An excessive generalization level highlighted the peak phenomenon position in the central-eastern part of the area, but lost a lot of local maxima (see Fig. 8).

(5) Analysis of the value ranges associated with various basic field shapes could be observed that:

a) ranges for squares and hexagons developed similarly, but an interesting fact was the hexagonal values were slightly higher (than squares), although the hexagonal fields were smaller than the square fields (Fig. 6)

b) values in the ranges of circular and triangular fields were significantly lower. For triangular fields, this was bound up with almost half the size of the surface (relative to the squares) and for circular fields – with the lack of full coverage of the same type of basic fields in the test area (i.e. the occurrence of gaps).

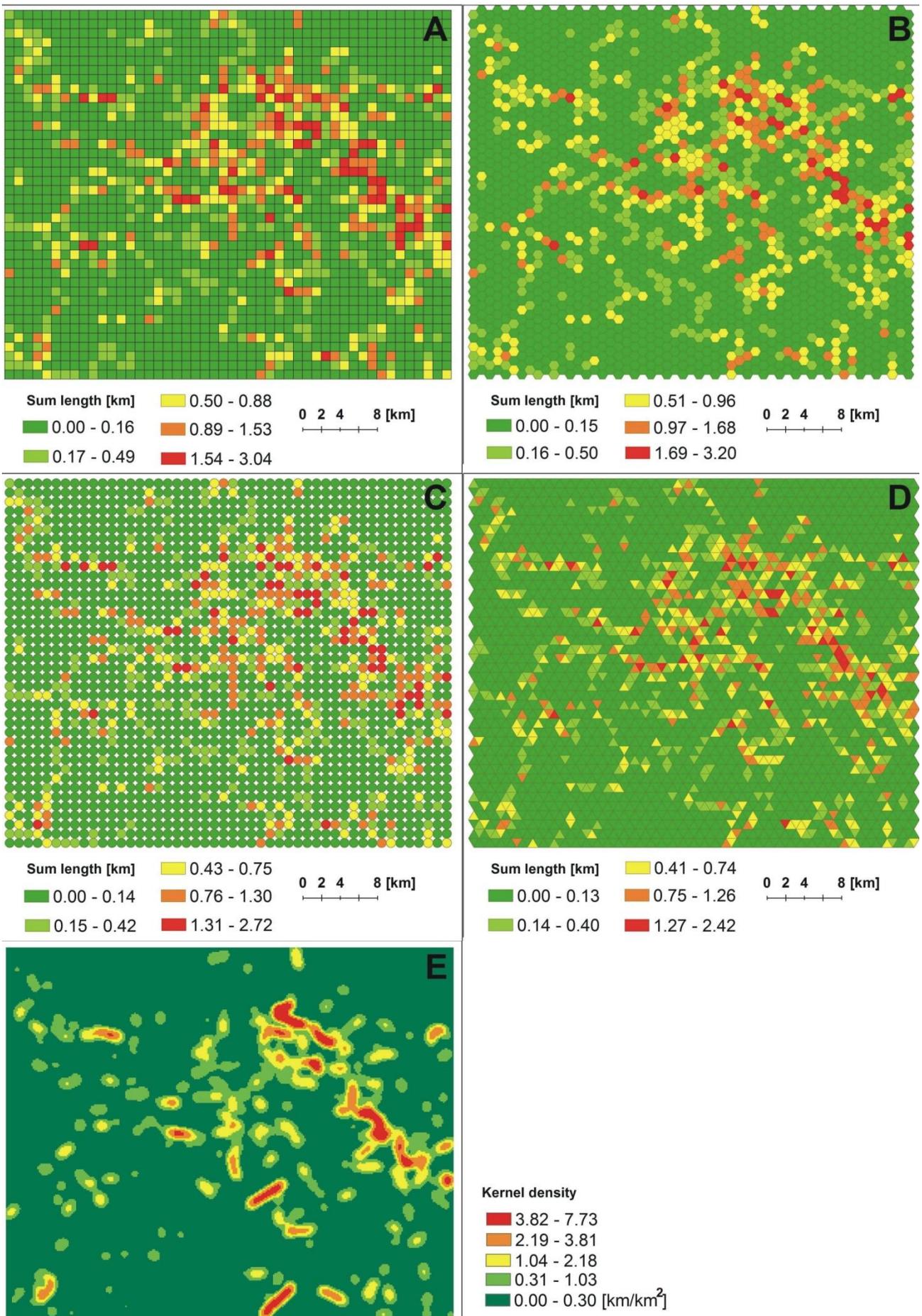


Fig. 7. Cartograms of line forms density in basic fields (1 km) A - squares, B - hexagons, C - circles, D - triangles and E - kernel density

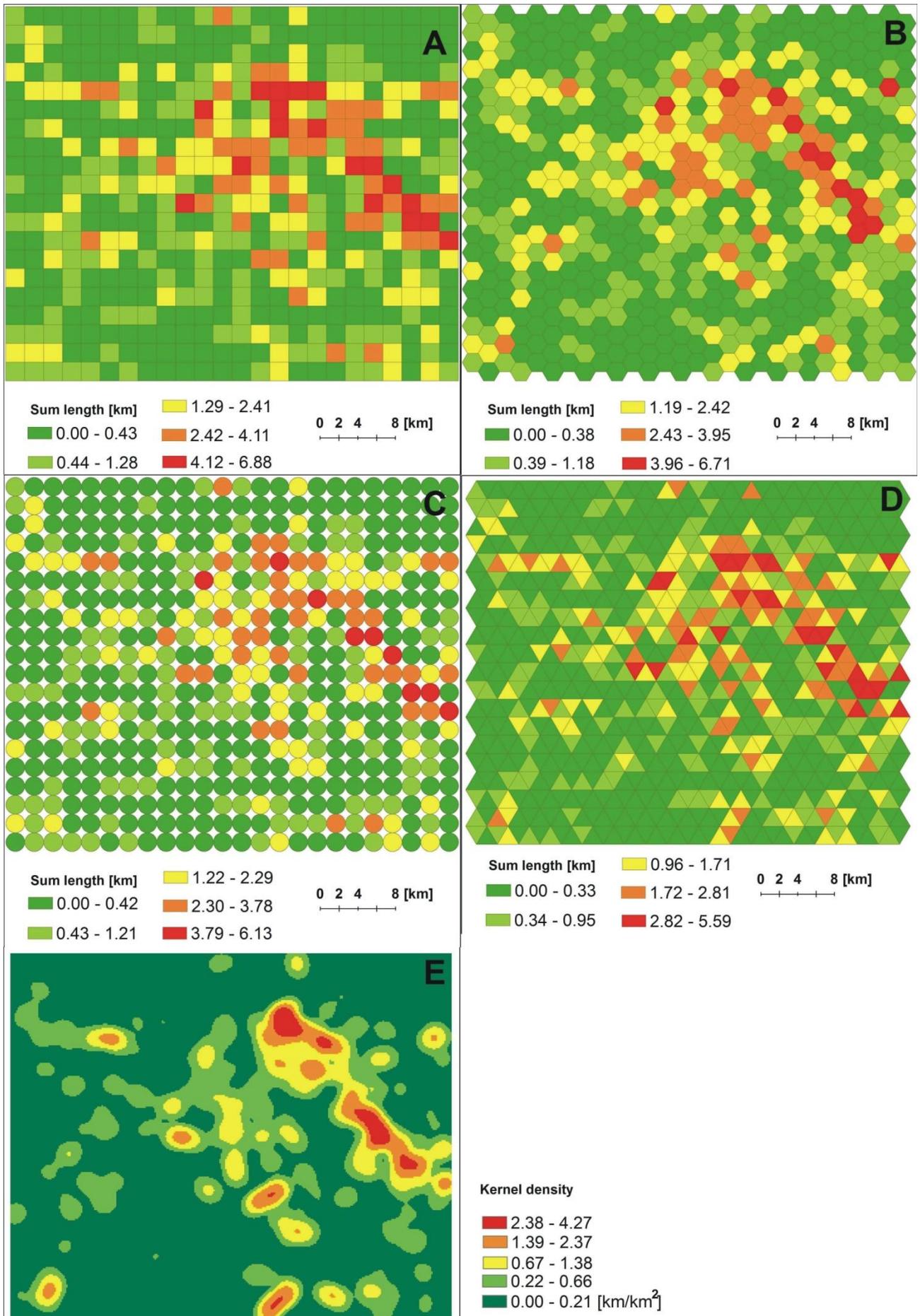


Fig. 8. Cartograms of line forms density in basic fields (2 km) A - squares, B - hexagons, C - circles, D - triangles and E - kernel density

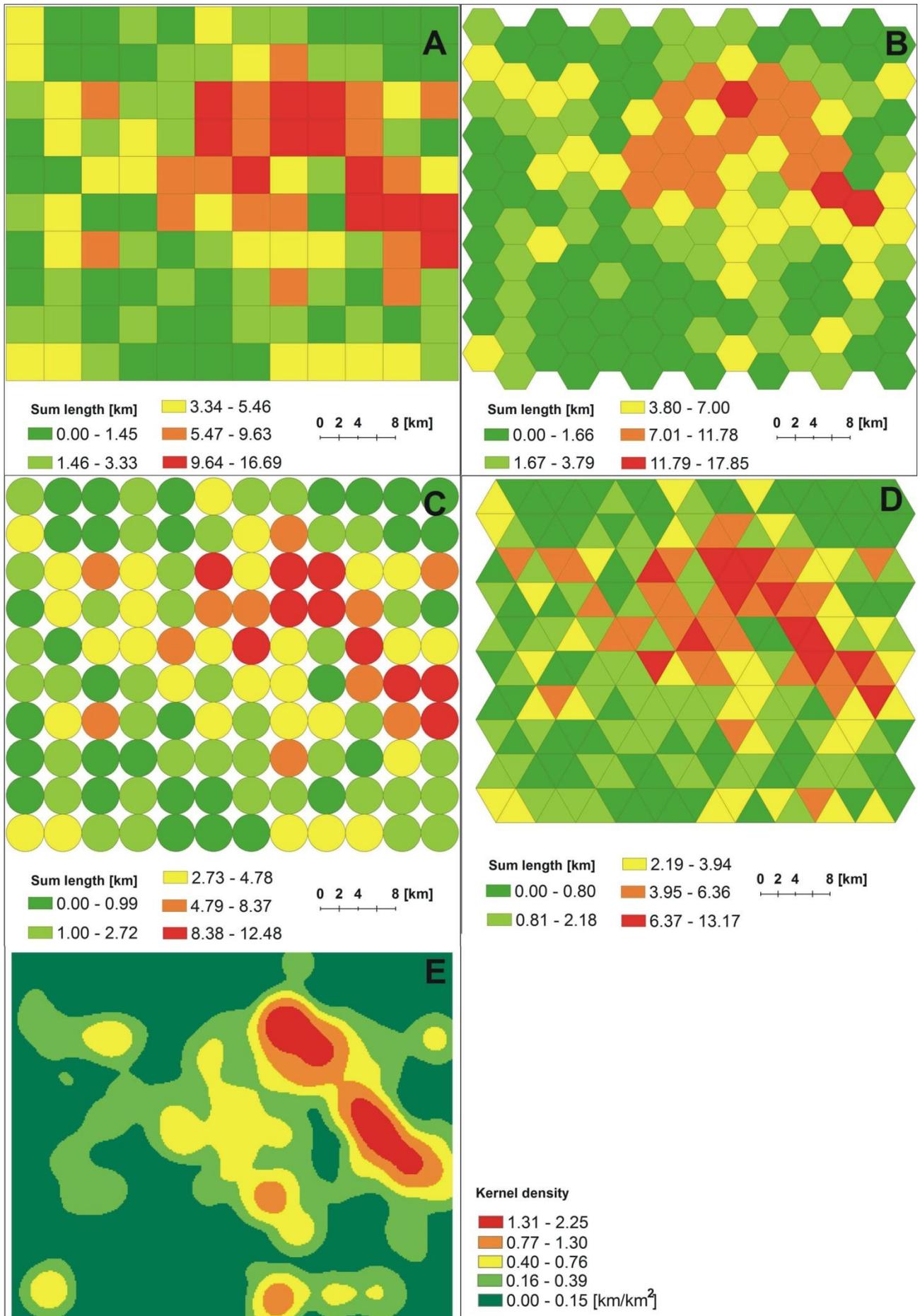


Fig. 9. Cartograms of line forms density in basic fields (4 km) A - squares, B - hexagons, C - circles, D - triangles and E - kernel density

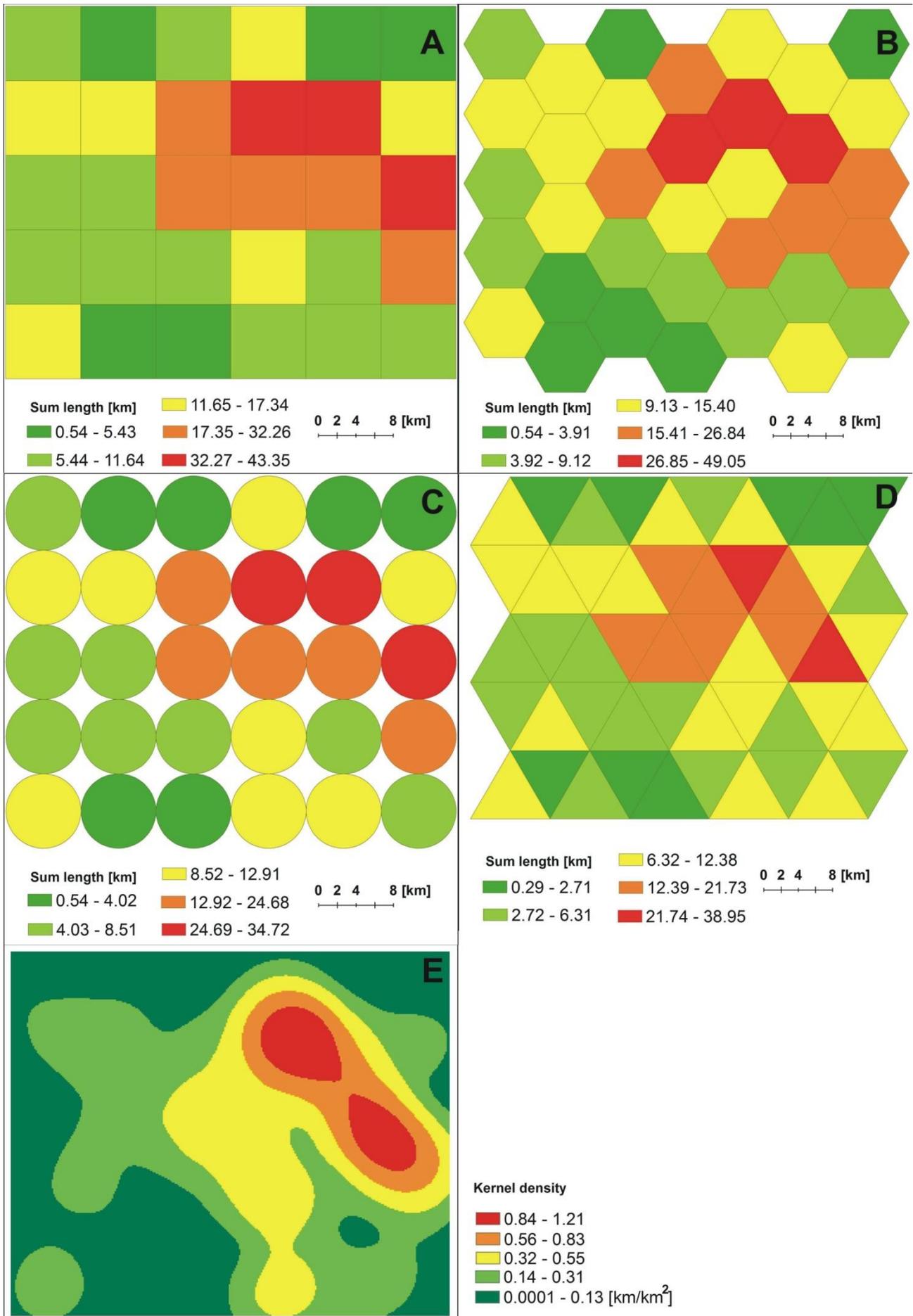


Fig. 10. Cartograms of line forms density in basic fields (8 km) A - squares, B - hexagons, C - circles, D - triangles and E - kernel density

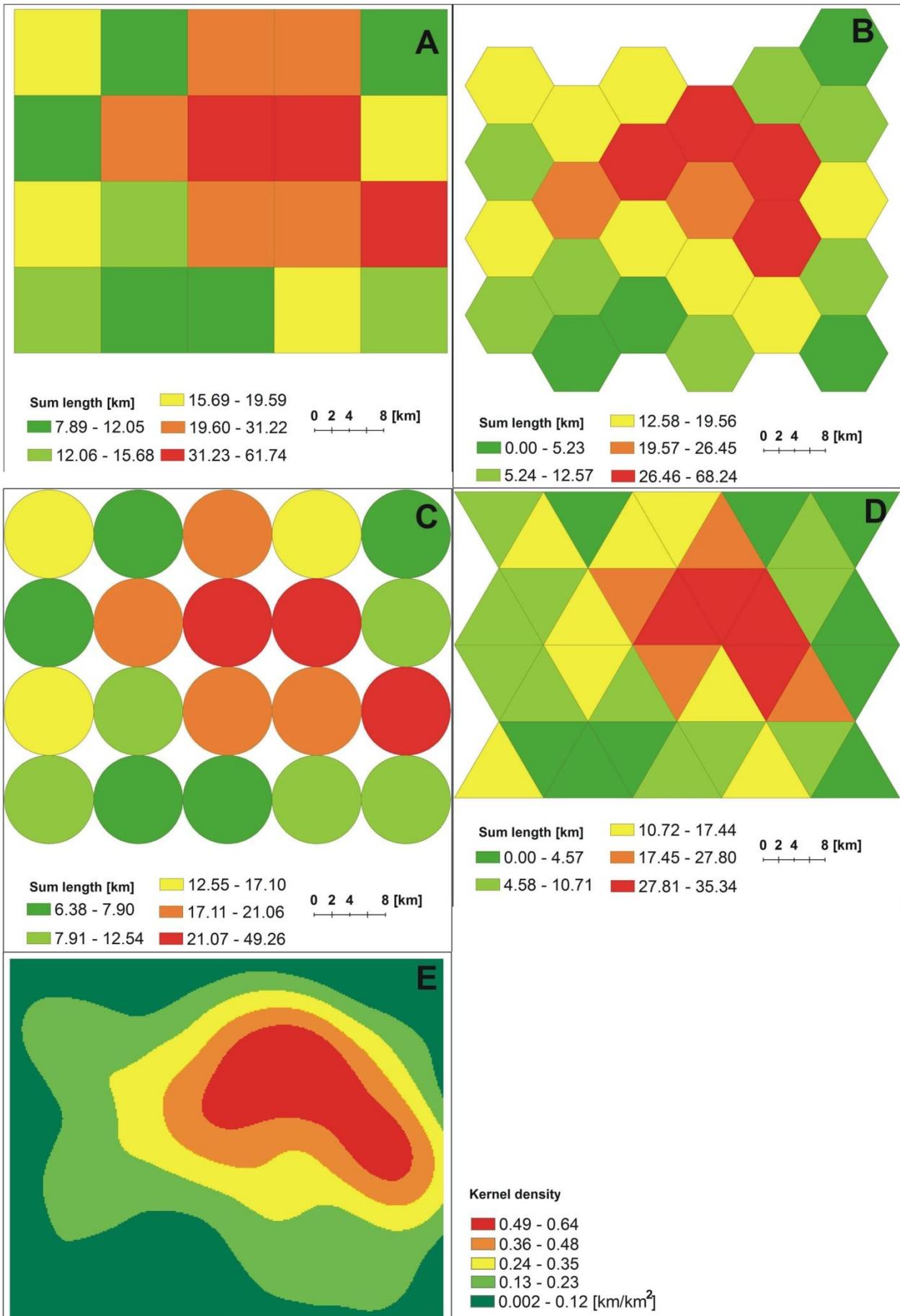


Fig. 11. Cartograms of line forms density in basic fields (10 km) A - squares, B - hexagons, C - circles, D - triangles and E - kernel density

## 5. Conclusions

(1) The spatial distribution image of the linear anthropogenic forms was similarly represented by all the basic fields (in general terms). However, different sizes and different shapes of the analysis fields resulted in a diverse quality of the presented effects.

(2) The best sizes of basic fields to present the spatial distribution of anthropogenic line forms were fields in the size of 4 km<sup>2</sup> and 16 km<sup>2</sup>. On the one hand - these fields have shown the variation and large spatial variability very well, on the other - characteristic places of the maximum and minimum values. The smallest fields (1km<sup>2</sup>) proved to be too detailed, resulting in data blurring. On the other hand, the largest fields (64 km<sup>2</sup> and 100 km<sup>2</sup>) over-simplified the image (the scale generalization), and blurred the differences, leaving one dominant value concentration.

(3) Different shapes of the basic fields of the analysis did not affect on the spatial image of line forms significantly, because every cartogram preserved the basic nature of the diversity of values. However the field shape, turned out to be significant for quantitative characteristics. Triangular fields, by its smallest size, resulted in the greatest fragmentation of data representation and the worst geometric representation of spatial relations also.

(4) Circular fields did not change the spatial image of line data in a significant way (although the lowered areas), so they are suitable for the presentation of spatial relationships of phenomena. However they could not be taken into account to the quantitative characteristics, because they did not cover the whole coverage of the study area, only 78% of its surface.

(5) Square and hexagonal fields were the best for these spatial analyses. The square is the easiest geometrical shape to analyse, especially for comparisons. Besides the square is basic shape of digital terrain models (GRID) and natural shape of raster data (the smallest element of the digital image is a square pixel) - what it makes the most popular and the most commonly used analysis mask. However, in the light of the analysed data and experiments in different sizes and shapes of basic fields - hexagon turned out to be the most suitable. This is probably due on the one hand with its shape - closer to the circle (than the square), but in contrast to the circle - filling the entire space, leaving no voids between adjacent fields.

(6) Another important factor may be the question about the distances between the centres of adjacent cells: for hexagonal fields they are always the same for all adjacent fields (distances between square centres contacted by sides are smaller than contacted by corners). The third indicator may be a greater number of sides (than in the square), which geometrical fitting of irregularly arranged data in space is improved, there are preserved directions and shapes more faithfully (confirmation of this fact is comparison of hexagonal field results with kernel density analysis, which gave the most similar images of spatial distribution of analyzed line forms).

(7) Maybe it will be an interesting experiment to create GRID DEM with hexagonal fields? [LI ET AL.](#) (2005) mentioned, that GRID model may be build of the basic fields of different regular shapes (equilateral triangles, squares, hexagons, parallelograms) or irregular. However, as I know, there is still not implemented in GIS programmes the tools using hexagons as basic fields to create GRID DEM.

(8) Of course, a method of space division by hexagonal fields has its limitations and disadvantages also. Undoubtedly, it is difficult to precisely locate the individual hexagons in the study space while it is making comparisons of processed data from different sources. Besides, there is trouble on the edges of the study area - there are fields incompletely filled with data. [KOZŁOWSKI](#) (1971) also leveled charges about the artificiality and excessive regularity of such reference fields.

(9) To sum up, the investigation suggests that the most adequate basic fields to present and analysis of the spatial relationships of the environmental elements (in this case the anthropogenic line forms) are hexagons. This fact confirms the assumption put forward long ago by [CZEKAŁSKI](#) (1933), which concluded that the most appropriate basic field model is the circle. However, due to the circle limitations associated with the incomplete coverage area - the most similar figure turned out to be just a hexagon. The same assumptions received [ŁAWNICZAK](#) (2008) who analysed in his study the morphometric characteristics of NW Poland relief and [ŚLESZYŃSKI](#) (2012), who also divided the polish area for hexagons.

### **Acknowledgements**

*I would like to thank Renia Dulias for encouragement and drive to action, and for many valuable comments so that article became better.*

## References

- ArcGIS Desktop Helps 9.3. 1999-2008. ESRI.
- Buraczyński J. 1964. Les entailles d'erosion recentes (ravins) du Roztocze Occidental. *Biul. Lubel. Tow. Nauk.*, D, 3-4: 23-26.
- Buraczyński J. 1977. Natężenie erozji wąwozowej i erozji gleb na Roztoczu Gorajskim. *Zesz. Probl. Post. Nauk Roln.*, 193: 91-99.
- Buraczyński J. 1984. Wpływ tektoniki na rozwój dolin strefy krawędziowej Roztocza. *Rocz. Pol. Tow. Geol.*, 54, 1/2: 209-225.
- Buraczyński J. 1989/1990. Rozwój wąwozów lessowych na Roztoczu Gorajskim w ostatnim tysiącleciu. *Ann. UMCS*, sec., B, 44/45: 95-104.
- Czekalski J. 1933. Mapa izarytmiczna a obraz rzeczywisty. Próba analizy metody. *Wiad. Służby Geogr.*, 3: 202-234.
- Dulias R. 2013. *Denudacja antropogeniczna na obszarach górniczych na przykładzie Górnośląskiego Zagłębia Węglowego*. Wyd. Uniw. Śl., Katowice.
- Gawrysiak L., Harasimiuk M. 2009. Analysis of the gorges distribution in Lublin Upland and Roztocze. [in:] Zgłobicki W., Dotterweich M., Rodzik J. (eds.), *Gully erosion in Lublin Region*. Lublin.
- Gilewska S. 1986. Geomorfologiczny podział Polski. *Przegl. Geogr.*, LVIII, 1-2: 15-40.
- Gregory S. 1970. *Metody statystyki w geografii*. PWN, Warszawa.
- Jaroszewski W., Piątkowska A. 1988. O naturze niektórych lineamentów (na przykładzie Roztocza). *Ann. Soc. Geol. Pol.*, vol. 58: 423-443.
- Józefaciuk C., Józefaciuk A. 1992. Gęstość sieci wąwozowej w fizjograficznych krainach Polski. *Pamiętniki Puławskie*, 101: 51-66.
- Kęsik A. 1961. Valles des terrains loessiques de la partie l'Ouest du Plateau de Nałęczów. *Ann. UMCS*, sec., B, 15: 123-153.
- Kondracki J. 2001. *Geografia regionalna Polski*. PWN, Warszawa.
- Kozłowski J. 1971. Pola jednakowej wielkości jako podstawa ilościowych opracowań kartograficznych. *Mat. Ogólnopol. Konf. Kartogr., "Problemy Kartografii Tematycznej"*. Lublin-Warszawa.
- Li Z., Zhu Q., Gold C. 2005. *Digital terrain modeling: principles and methodology*. CRC Press, Washington.
- Ławniczak R. 2008. *Morfometryczne cechy rzeźby a geneza wybranych form Polski północno-zachodniej*. Ser. Geografia, 80. UAM. Poznań.
- Łomnicki A. 2003. *Wprowadzenie do statystyki dla przyrodników*. PWN, Warszawa.
- Mapa Topograficzna 1:50 000. 1995 Arkusze: M-34-50-C, M-34-50-D, M-34-51-C, M-34-62-A, M-34-62-B, M-34-62-C, M-34-62-D, M-34-63-A, M-34-63-C.
- Maruszczak H. 1973. Erozja wąwozowa we wschodniej części pasa Wyżyn południowopolskich. *Zesz. Probl. Post. Nauk Roln.*, 151: 15-30.
- Paślowski J. (red.). 2006. *Wprowadzenie do kartografii i topografii*. Wyd. Nowa Era, Wrocław.
- Saliszczew K.A. 1998. *Kartografia ogólna*. PWN, Warszawa.
- Silverman B.W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.
- Sirko M. 1971. Metoda „profilu” jako podstawa konstrukcji map gęstości sieci rzecznej. *Folia Soc. Sc. Lubl.*, D, 11: 121-127.
- Sirko M. 1978. Z metodyki konstrukcji map gęstości sieci rzecznej. *Pol. Przegl. Kartogr.*, 10, 4: 145-153.
- Szalkiewicz B. 1947. Gęstość sieci rzecznej międzyrzecza Wisły i Bugu. *Ann. UMCS*, sec., B, 2, 6: 151-162.
- Szypuła B. 2013. Quantitative changes of anthropogenic relief over the last 100 years in the Silesian Upland (south Poland). *Zeit. für Geomorphologie*. DOI: 10.1127/0372-8854/2013/0111.
- Śleszyński P. 2012. A geomorphometric analysis of Poland based on the SRTM-3 data. *Geogr. Pol.*, 85, 4: 45-59.