

Simplification of Lines Containing Fixed Points Using the Douglas-Peucker Algorithm

Nada VUČETIĆ*, **Svetozar PETROVIĆ****, **Ante STRUNJE***

* *University of Zagreb, Faculty of Geodesy, Kačićeva 26, 10000 Zagreb, Croatia, nvucetic@geof.hr, astrunje@net.hr*

** *Berlin, Germany, sveto117@yahoo.de*

74

Abstract: *The paper presents the results of simplification by Douglas-Peucker algorithm of lines containing fixed points (points which are not allowed to be removed). The investigations were performed both with lines which were first split in segments between fixed points with consecutive simplification of all individual segments as well as on the integral (i.e. without a splitting) lines with fixed points.*

Key words: *line simplification, Douglas-Peucker algorithm, fixed point, visualization, logical consistency, positional accuracy*

1 Introduction

Generalization is a very complex process composed of various procedures. One of them, used in a digital environment, is the line simplification consisting of reduction of number of points of the original line retaining at the same time its essential characteristics (McMaster and Shea 1992).

In the literature different algorithms for line simplification can be found (see e.g. Ivanov 1965, Ramer 1972, Douglas and Peucker 1973, Cromley 1991, Visvalingam and Whyatt 1993, Wang and Müller 1998, Guo et al. 2001). The most well-known among them, which is implemented in various commercial software packages like ArcInfo, MGE Map Generalizer or Autodesk Map, is the Douglas-Peucker algorithm (Douglas and Peucker 1973).

This algorithm tries to select the points which are considered as characteristic points of the line in the manual generalization. Investigating reasons why a cartographer selects specific points, some authors came to the idea that the distance of the point on the curved line from a straight line segment could probably justify such a choice. Therefore, they introduce an arbitrarily chosen maximum distance (tolerance or threshold). If the distance of all points of the original line from the straight line segment is less than the tolerance, the straight line segment will suffice to represent the whole line. If this condition is not satisfied, the furthest point is selected as characteristic, since it is considered to probably be the apex of the sharpest angle, and the same procedure is repeated with new segments.

Discussing the generalization in general, as well as its particular procedures like line simplification, frequently only their influence on the quality of visualization is considered, and their influence on the quality of spatial data (positional accuracy, attribute accuracy, logical consistency and completeness) is left out of consideration.

Influence of line simplification on the quality of spatial data is subject of investigation of various authors (see e.g. McMaster 1986, 1987, João 1998, Veregin 1999, 2000, Skopeliti and Tsoulos 2000, Shahriari and Tao 2002, Cheung and Shi 2004).

This contribution presents results of the investigation of the influence of simplification of lines containing fixed points using Douglas-Peucker algorithm on the positional accuracy. Under fixed points we mean here the points of original line which are not allowed to be removed in the course of simplification, like intersections of routes, mouths of river tributaries, junctions of three borders, etc.

Pojednostavljenje linija s fiksnim točkama Douglas-Peuckerovim algoritmom

Nada VUČETIĆ*, Svetozar PETROVIĆ**, Ante STRUNJE*

* Sveučilište u Zagrebu, Geodetski fakultet, Kačićeva 26, 10000 Zagreb
nvucetic@geof.hr, astrunje@net.hr

** Berlin, sveto117@yahoo.de

75

Sažetak: U radu se daju rezultati pojednostavljenja Douglas-Peuckerovim algoritmom linija s fiksnim točkama (točkama koje ne smiju biti uklonjene). Ispitivanja su provedena kako na linijama koje su prvo razlomljene na segmente od jedne do druge fiksne točke pa potom svaki taj segment posebno pojednostavljen, tako i na cijelim (nerazlomljenim) linijama s fiksnim točkama.

Ključne riječi: pojednostavljenje linije, Douglas-Peuckerov algoritam, fiksna točka, vizualizacija, logička konzistentnost, položajna točnost

1. Uvod

Generalizacija je vrlo složen proces koji se sastoji od niza postupaka. Jedan je od njih, u digitalnom okruženju, pojednostavljenje linija koji se sastoji od redukcije broja točaka izvorne linije zadržavajući bitne njezine karakteristike (McMaster i Shea 1992).

Za pojednostavljenje linija u literaturi se mogu pronaći različiti algoritmi (vidi npr. Ivanov 1965, Ramer 1972, Douglas i Peucker 1973, Cromley 1991, Visvalingam i Whyatt 1993, Wang i Müller 1998, Guo i dr. 2001). Najpoznatiji je među njima Douglas-Peuckerov algoritam (Douglas i Peucker 1973), koji je ugrađen u mnoge komercijalne softverske pakete kao što su ArcInfo, MGE Map Generalizer ili Autodesk Map.

Taj algoritam nastoji izabrati točke koje se smatraju karakterističnim točkama linije pri ručnoj generalizaciji. Ispitujući razloge zbog kojih kartograf izabire neke točke, autori dolaze na ideju da bi najvjerojatnije udaljenost točke zakrivljene linije od pravocrtnog segmenta mogla opravdati takav izbor. Zbog toga uvode proizvoljno zadanu maksimalnu udaljenost (toleranciju). Ako je udaljenost svih točaka izvorne linije od pravocrtnog segmenta manja od zadane tolerancije, tada će pravocrtni segment biti dovoljan za prikaz cijele linije. Ako taj uvjet nije ispunjen, tada se izabire najudaljenija točka jer se smatra da je ona najvjerojatnije vrh najoštrijeg kuta, te se isto ponovi s novim segmentima.

Pri razmatranju generalizacije općenito, pa tako i pojedinih njezinih postupaka kao što je pojednostavljenje linija, često se promatra njezin utjecaj samo na kvalitetu vizualizacije, a zaboravlja se njezin utjecaj na kvalitetu prostornih podataka (položajnu točnost, točnost atributa, logičku konzistentnost i potpunost).

Utjecaj pojednostavljenja linija na kvalitetu prostornih podataka predmet je istraživanja mnogih istraživača (vidi npr. McMaster 1986, 1987, João 1998, Veregin 1999, 2000, Skopeliti i Tsoulos 2000, Shahriari i Tao 2002, Cheung i Shi 2004).

U ovome radu daju se rezultati istraživanja utjecaja pojednostavljenja linija s fiksnim točkama Douglas-Peuckerovim algoritmom na položajnu točnost. Pod fiksnim točkama podrazumijevaju se ovdje točke izvorne linije koje se pojednostavljenjem ne smiju ukloniti, npr. križanja prometnica, ušća pritoka vodotoka, tromeđa granica područja, itd.

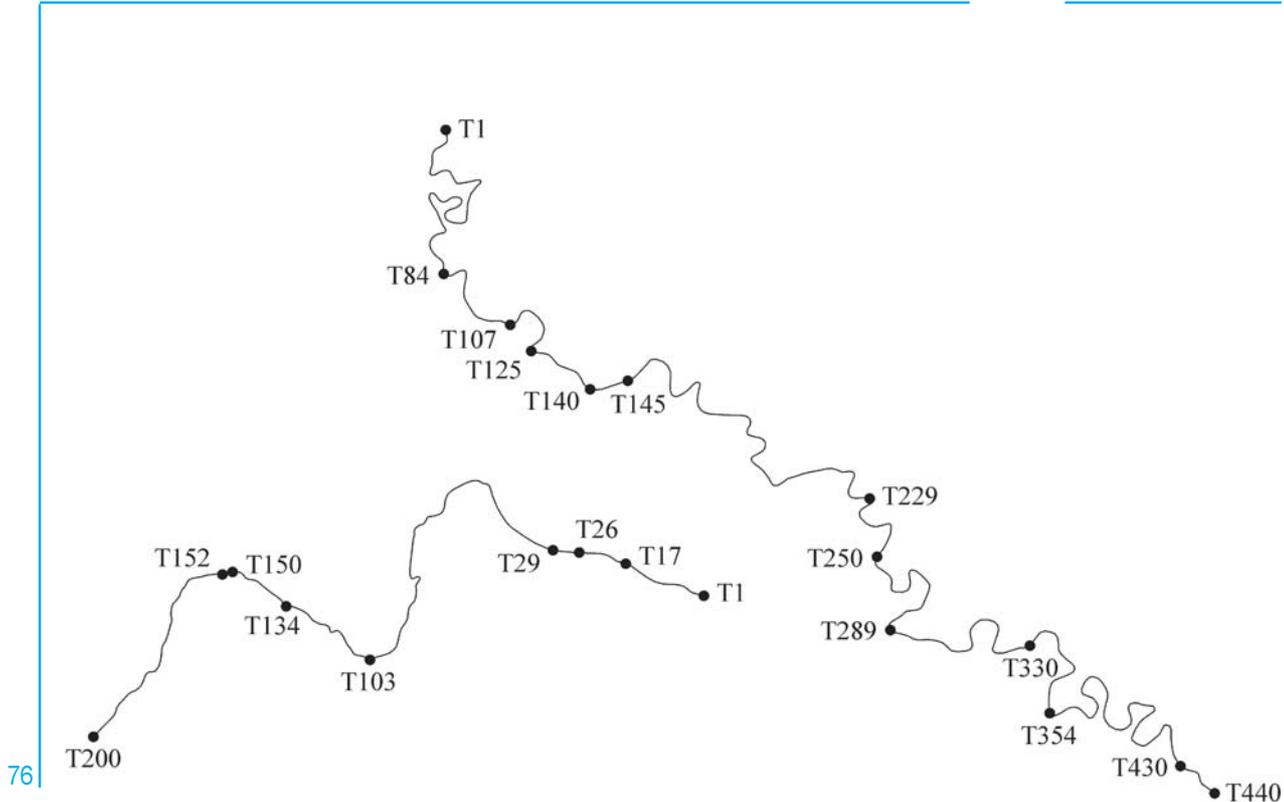


Fig. 1. River Sunja (left) and a part of Sava (right) with marked fixed points, scale 1:500 000

Slika 1. Prikaz rijeke Sunje (lijevo) i dijela Save (desno) s označenim fiksnim točkama u mjerilu 1:500 000

2 Logical Consistency and Positional Accuracy of Simplified Lines

Each algorithm which is based on reduction of number of points of a line, and the Douglas-Peucker algorithm is one of these, surely does not remove the first and the last point of the original line. Which points lying between these two are going to be removed, depends on the way of performing the reduction. Hence, it might happen that some points are removed which should not be removed, which leads to logical inconsistencies, e.g. that two roads do not cross after the reduction, or that a tributary stream does not flow into a river any more.

In order to retain logical consistency, prior to simplification the lines are split into segments between fixed points. Next, each of these line segments is simplified separately. Due to this intentional splitting the logical consistency will for sure not be disturbed; however, the question of positional accuracy of the simplified line arises.

Is it better to mark the fixed points of the original line and to simplify the line as a whole, without any splitting? In order to give an answer to this question, we modified the Douglas-Peucker algorithm in such a way, that, if some fixed point should be removed, because its distance from the considered straight line segment is less than the chosen tolerance, it is nevertheless retained.

3 Simplification of the Parts and of the Whole Original Line Without Splitting

Investigation of the simplification of the split and whole original line have been done using an imaginary line, complete streams of river Sunja and Vrbas, as well as on the selected parts of rivers Sava and Česma. First investigations of line simplification in individual segments and taking into account fixed points have been done in the frame of the diploma thesis of Ante Strunje (Strunje 2006), supervised by Assist. Prof. Nada Vučetić.

The source for the used water streams is a copy of a part of the blue color reproduction original for the region of Posavina and the western part of Slavonija in scale 1:500 000, produced in 1994 in the Institute for Cartography (presently Institute for Cartography and Photogrammetry) of the Faculty of Geodesy, University of Zagreb. This original was scanned by a drum scanner CalComp SCANPlus III 1800T in resolution of 800 x 800 dpi using the software Adobe Photoshop. Manual vectorization of rivers was done using a software package Autodesk Map. During the vectorization it was necessary to include fixed points, i.e. mouths of all tributaries of the considered rivers. In Fig. 1 vectorized streams of rivers Sava and Sunja with marked fixed points are given. The number of points of the river Sunja is 200, and the length amounts to 123,4 mm (in the map scale, corresponding to 61,7 km in nature). The number of points of the considered

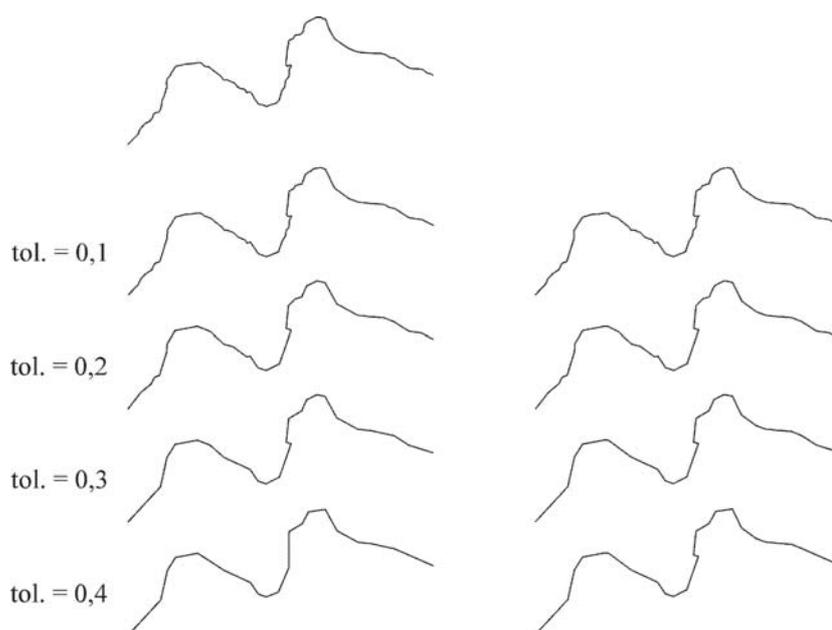


Fig. 2. River Sunja: source reduced to scale 1:1 000 000 (top), the result of simplifying individual segments using tolerances 0, 1–0,4 mm (left), the result of simplifying integral line with fixed points using tolerances 0, 1–0,4 mm (right)

Slika 2. Rijeka Sunja: izvornik smanjen u mjerilo 1:1 000 000 (gore), zajednički prikaz pojednostavljenih segmenata linije tolerancijama 0, 1–0,4 mm (lijevo), prikaz pojednostavljene nerazlomljene linije s fiksnim točkama tolerancijama 0, 1–0,4 mm (desno)

2. Logička konzistentnost i položajna točnost pojednostavljenih linija

Svaki algoritam koji se temelji na redukciji broja točaka linije, pa tako i Douglas-Peuckerov algoritam, jedino sigurno ne uklanja prvu i zadnju točku izvorne linije. Koje će točke između njih biti uklonjene ovisi o samom načinu provođenja redukcije. Tako se može dogoditi da budu uklonjene i točke koje se ne bi smjele ukloniti, što dovodi do logičke nekonzistentnosti, npr. da se dvije prometnice ne križaju ili da se prtok nekog vodotoka ne ulijeva u njega.

Kako se ne bi narušila logička konzistentnost, linije se prije provođenja postupka pojednostavljenja lome na segmente od jedne do druge fiksne točke. Zatim se svaki takav pojedini linijski segment posebno pojednostavljuje. Tim namjernim lomovima zasigurno se neće narušiti logička konzistentnost, ali se postavlja pitanje položajne točnosti pojednostavljene linije.

Je li možda bolje u skupu točaka izvorne linije označiti fiksne točke pa liniju pojednostavniti cijelu odjednom, tj. nerazlomljenu? Da bismo mogli odgovoriti na to pitanje modificirali smo Douglas-Peuckerov algoritam: ako bi se neka fiksna točka trebala ukloniti, jer je njezina udaljenost od promatranog pravocrtnog segmenta manja od zadane tolerancije, ona se ipak ne uklanja.

3. Pojednostavljenje dijelova razlomljene i cijele nerazlomljene izvorne linije

Ispitivanja pojednostavljenja razlomljene i nerazlomljene izvorne linije provedena su na izmišljenoj liniji, na cijelim tokovima rijeka Sunje i Vrbasa, te na odabranim dijelovima tokova rijeka Save i Česme. Prva ispitivanja pojednostavljenja linija po segmentima i s uzimanjem u obzir fiksnih točaka provedena su u okviru diplomskog rada Ante Strunje (Strunje 2006), mentorica je bila doc. dr. sc. Nada Vučetić.

Izvornik za upotrijebljene tokove rijeka kopija je dijela reprodukcijskog originala za plavu boju područja Posavine i zapadnog dijela Slavonije u mjerilu 1:500 000, izrađenog u Zavodu za kartografiju (danas Zavod za kartografiju i fotogrametriju) Geodetskog fakulteta Sveučilišta u Zagrebu 1994. godine. Taj je izvornik skeniran valjkastim skenerom CalComp SCANPlus III 1800T u rezoluciji 800 x 800 dpi pomoću programa Adobe Photoshop. Vektorizacija skeniranog izvornika provedena je programskim paketom Autodesk Map. Pri vektorizaciji trebalo je obavezno uključiti fiksne točke, tj. ušća svih pritoka navedenih rijeka. Na slici 1 dan je prikaz vektoriziranih tokova rijeka Save i Sunje s označenim fiksnim točkama. Broj točaka toka rijeke Sunje je 200, a duljina iznosi 123,4 mm (u mjerilu karte, što odgovara 61,7 km u prirodi). Broj točaka dijela toka

part of the river Sava is 440, the length being 275,2 mm (in the map scale, corresponding to 137,6 km in nature). It should be stressed that more line points have been vectorized than necessary in order to get as smooth lines as possible, i.e., without visible sharp angles.

The lines were simplified for scales two and five times smaller than the original applying tolerances of 0,1–0,4 mm in the resulting scale. Each line was first divided into segments and each of them simplified separately. Further, in the data file of the original line fixed points were marked and the whole line simplified at once (without splitting).

Here we present the results only for two rivers, Sunja and the part of Sava, for the scale two times smaller than the original. Other results are very similar. Figure 2 and 3 presents simplified lines. In these figures it is seen that the simplified whole line is visually more similar to the original line than the line obtained by splitting it first and simplifying consecutively.

For the estimation of positional accuracy areal and vector displacements of the simplified to the original line were determined. We define areal displacement as the sum of areas of polygons formed by the original and the simplified lines divided by the length of the original line,

78

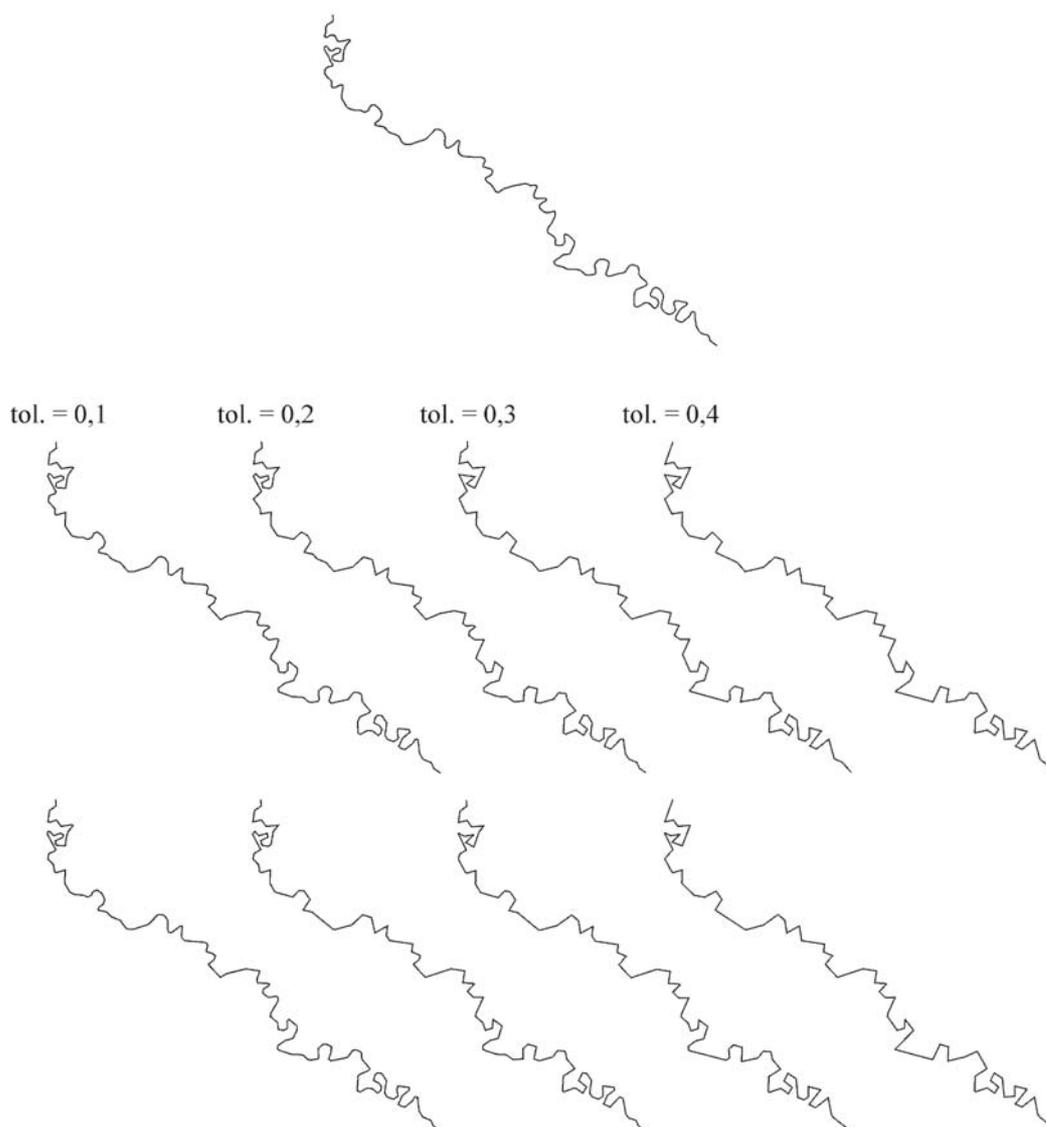


Fig. 3. Part of river Sava: source reduced to scale 1:1 000 000 (top), the result of simplifying individual line segments using tolerances 0,1–0,4 mm (center), the result of simplifying integral line with fixed points using tolerances 0,1–0,4 mm (bottom)

Slika 3. Dio rijeke Save: izvornik smanjen u mjerilo 1:1 000 000 (gore), zajednički prikaz pojednostavljenih segmenata linije tolerancijama 0,1–0,4 mm (sredina), prikaz pojednostavljene nerazlomljene linije s fiksnim točkama tolerancijama 0,1–0,4 mm (dolje)

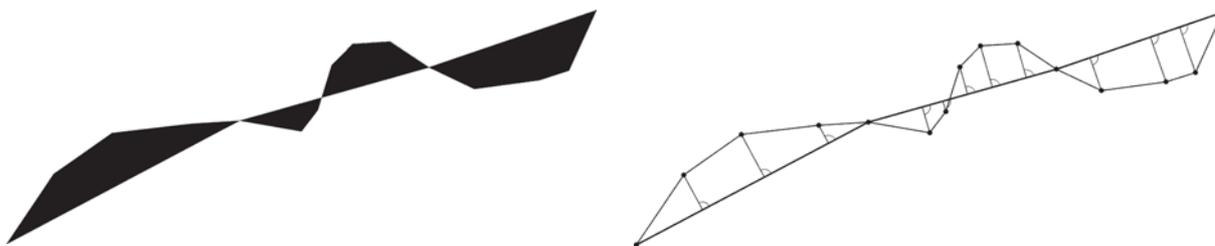


Fig. 4. The areas of polygons formed by the original and the simplified line (left) and the euclidean distances of points of the original from the simplified line (right)

Slika 4. Površine poligona koje formiraju izvorna i pojednostavljena linija (lijevo) i euklidske udaljenosti točaka izvorne od pojednostavljene linije (desno)

rijeke Save je 440, a duljina je 275,2 mm (u mjerilu karte, što odgovara 137,6 km u prirodi). Treba napomenuti da je vektoriziran veći broj točaka linija od potrebnoga kako bi linije bile što je moguće glatkije, tj. bez vidljivih lomova.

Izabrane linije pojednostavljene su za dva i pet puta sitnije mjerilo od izvornog i to s tolerancijama 0,1–0,4 mm u izvedenom mjerilu. Svaka linija prvo je razlomljena na segmente i svaki od tih segmenata je potom posebno pojednostavljen. Zatim su u skupu točaka izvorne linije označene fiksne točke i pojednostavljena je cijela (nerazlomljena) linija.

Ovdje ćemo dati rezultate tih ispitivanja samo za dva vodotoka, Sunju i dio Save, i to za dva puta sitnije mjerilo od izvornog jer su ostali rezultati vrlo slični. Na slikama 2 i 3 dani su prikazi pojednostavljenih linija.

Na tim je slikama vidljivo da je pojednostavljena nerazlomljena linija uglavnom vizualno sličnija izvornoj liniji, nego razlomljena pa pojednostavljena izvorna linija.

Za određivanje položajne točnosti računano je površinsko i vektorsko odstupanje pojednostavljene od izvorne linije. Površinsko odstupanje definiramo kao zbroj površina poligona koje formiraju izvorna i pojednostavljena linija podijeljen s duljinom izvorne linije, a vektorsko kao zbroj euklidskih udaljenosti točaka izvorne od pojednostavljene linije podijeljen s brojem točaka izvorne linije (sl. 4). Dakle, oba odstupanja su prosječne mjere.

Rezultati dobiveni pojednostavljenjem linija po segmentima i linija razmatranih kao cjeline s označenim fiksnim točkama u datoteci dani su u tablicama 1 i 2. Iz podataka u tablicama vidljivo je da su vrijednosti odstupanja pojednostavljene nerazlomljene linije od izvorne linije u većini slučajeva manje od vrijednosti odstupanja razlomljene linije. To isto vrijedi i kada su duljine pojednostavljenih linija jednake. Kada bi broj točaka izvorne linije bio manji nego ovdje upotrijebljen, zasigurno bi ta razlika u odstupanjima bila više izražena.

79

Table 1. River Sunja – computed values of areal and vector deviations of the simplified from original line, as well as the number of points and the lengths of simplified lines (all values in resulting scale).

Tablica 1. Rijeka Sunja – izračunane vrijednosti površinskih i vektorskih otklona pojednostavljenih linija od izvorne linije, te broj točaka i duljine pojednostavljenih linija (sve su navedene vrijednosti u izvedenom mjerilu)

Tol.	Simplified split original line				Simplified integral original line			
	Pojednostavljena razlomljena izvorna linija				Pojednostavljena nerazlomljena izvorna linija			
Tol.	Areal displacement	Vector displacement	No. of points	Length of line	Areal displacement	Vector displacement	No. of points	Length of line
[mm]	[mm ² /dm]	[μm]	Broj točaka	Duljina linije [mm]	Površinski otklon [mm ² /dm]	Vektorski otklon [μm]	Broj točaka	Duljina linije [mm]
0,1	2,7	27	53	61,1	2,6	26	56	61,1
0,2	5,0	51	34	60,4	4,4	47	37	60,4
0,3	7,0	74	25	59,9	6,4	71	27	60,0
0,4	10,2	104	21	59,2	7,7	81	25	59,8
Length of the original line = 61,7 mm; No. of points of the original line = 200								
Duljina izvorne linije = 61,7 mm; Broj točaka izvorne linije = 200								

Table 2. Part of the river Sava – computed values of areal and vector deviations of the simplified from original line, as well as the number of points and the lengths of simplified lines (all values in resulting scale).

Tablica 2. Dio rijeke Save – izračunane vrijednosti površinskih i vektorskih otklona pojednostavljenih linija od izvorne linije, te broj točaka i duljine pojednostavljenih linija (sve su navedene vrijednosti u izvedenom mjerilu).

Tol.	Simplified split original line				Simplified integral original line			
	Pojednostavljena razlomljena izvorna linija				Pojednostavljena nerazlomljena izvorna linija			
Tol.	Areal displacement	Vector displacement	No. of points	Length of line	Areal displacement	Vector displacement	No. of points	Length of line
[mm]	Površinski otklon [mm ² /dm]	Vektorski otklon [μm]	Broj točaka	Duljina linije [mm]	Površinski otklon [mm ² /dm]	Vektorski otklon [μm]	Broj točaka	Duljina linije [mm]
0,1	2,8	30	157	135,5	2,8	29	163	135,5
0,2	5,0	53	109	133,4	5,6	57	109	133,3
0,3	8,4	89	80	130,9	7,1	76	93	131,9
0,4	10,2	110	70	129,6	10,8	116	76	129,6
Length of the original line = 137,6 mm; No. of points of the original line = 440								
Duljina izvorne linije = 137,6 mm; Broj točaka izvorne linije = 440								

80

and the vector displacement as the sum of euclidean distances of points of the original line from the simplified line divided by the number of points of the original line (Fig. 4). Hence, both displacements are average measures.

Results obtained by simplifying individual line segments and integral lines with marked fixed points are given in Table 1 and 2. From these it is visible that the deviations of the simplified integral line from the original line are in most cases less than the deviations obtained for the split line. The same holds also when the lengths of simplified lines are equal. Should the number of points of the original line be less than the one used here, the difference in deviations would surely be more expressed.

4 Conclusion

Simplifying divided and integral (without splitting) original line with fixed points by Douglas-Peucker algorithm yields different results. First splitting and then simplifying parts of the original line suffices to guarantee logical consistency of lines; however, in most cases the positional accuracy is less than positional accuracy of the simplified integral line with marked fixed points. In the automatic line simplification every cartographer should bear this in mind, since the simplification of lines influences not only the quality of visualization, but the quality of spatial data as well.

4. Zaključak

Pojednostavljenjem razlomljene i cijele (nerazlomljene) izvorne linije s fiksnim točkama Douglas-Peuckerovim algoritmom dobivaju se različiti rezultati. Lomljenjem pa pojednostavljenjem dijelova izvorne linije osigurana

je logička konzistentnost linija, ali je u većini slučajeva položajna točnost manja od položajne točnosti pojednostavljene nerazlomljene linije s označenim fiksnim točkama. Pri automatskom pojednostavljenju linija svaki bi kartograf trebao to imati na umu, jer pojednostavljenje linija ne utječe samo na kvalitetu vizualizacije, nego i na kvalitetu prostornih podataka.

References / Literatura

- Cheung, C. K., Shi, W. (2004): Estimation of the Positional Uncertainty in Line Simplification in GIS. *The Cartographic Journal*, Vol. 41, No. 1, 37–45.
- Cromley, R. G. (1991): Hierarchical Methods of Line Simplification. *Cartography and Geographic Information Systems*, Vol. 18, No. 2, 125–131.
- Douglas, D. H., Peucker, T. K. (1973): Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *The Canadian Cartographer*, Vol. 10, No. 2, 112–122.
- Guo, Q. S., Brandenberger, C., Humi, L. (2001): A Progressive Line Simplification Algorithm. Working Paper of the 4th Workshop on Progress in Automated Map Generalization, Beijing, China, August 2–4, http://www.geo.unizh.ch/ICA/docs/beijing2001/papers/guo_v1.pdf
- Ivanov, V. V. (1965): O nekim mogućnostima automatizacije topografskih kart (On Some Possibilities of Automation of Topographic Maps, in Russian). *Geodezija i kartografija*, 1, 62–66.
- João, E. M. (1998): Causes and consequences of map generalisation. *Research Monographs in Geographical Information Systems*. Taylor & Francis, London.
- McMaster, R. B. (1986): A Statistical Analysis of Mathematical Measures for Linear Simplification. *The American Cartographer*, Vol. 13, No. 2, 103–116.
- McMaster, R. B. (1987): Automated line generalization. *Cartographica*, Vol. 24, No. 2, 74–111.
- McMaster, R. B., Shea, K. S. (1992): *Generalization in Digital Cartography*. Association of American Geographers, Washington.
- Ramer, U. (1972): An Iterative Procedure for the Polygonal Approximation of Plane Curves. *Computer Graphics and Image Processing*, 1, 244–256.
- Shahriari, N., Tao, V. (2002): Minimising Positional Errors in Line Simplification Using Adaptive Tolerance Values. Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, 9–12 July, <http://www.isprs.org/commission4/proceedings/pdfpapers/063.pdf>
- Skopeliti, A., Tsoulos, L. (2000): Estimating Positional Accuracy of Linear Features. *Spatial Information Management – Experiences and Visions for the 21st Century, Workshop of FIG–Commission 3*, Athens, Greece, 4–7 October, http://www.fig.net/com_3_athens/pdf/S55.%20A.Skopeliti,%20L.Tsoulos.pdf
- Strunje, A. (2006): Pojednostavljenje toka linije. Diplomski rad. (Simplification of line flow, Diploma thesis, in Croatian.) Geodetski fakultet, Zagreb, mentor: Vučetić, Nada.
- Veregin, H. (1999): Line Simplification, Geometric Distortion, and Positional Error. *Cartographica*, Vol. 36, No. 1, 25–39.
- Veregin, H. (2000): Quantifying positional error induced by line simplification. *International Journal of Geographical Information Science*, Vol. 14, No. 2, 113–130.
- Visvalingam, M., Whyatt, J. D. (1993): Line Generalization by Repeated Elimination of Points. *The Cartographic Journal*, Vol. 30, No. 1, 46–51.
- Wang, Z., Müller, J.-C. (1998): Line Generalization Based on Analysis of Shape Characteristics. *Cartography and Geographic Information Systems*, Vol. 25, No. 1, 3–15.