

WSPOMAGANIE DECYZJI Z WYKORZYSTANIEM NARZĘDZI GIS – RYZIKO ZWIĄZANE Z DOKŁADNOŚCIĄ DANYCH ŹRÓDŁOWYCH

GIS – A TOOL IN DECISION SUPPORT – RISK CAUSED BY DATA QUALITY EFFECT

Beata Hejmanowska

Zakład Fotogrametrii i Informatyki Teledetekcyjnej,
Akademia Górniczo-Hutnicza w Krakowie

SŁOWA KLUCZOWE: GIS, dokładność danych, wspomaganie decyzji

STRESZCZENIE: Systemy GIS oferują bogate narzędzia analityczne do przetwarzania danych, które również w Polsce stopniowo zaczynają być wykorzystywane do wspomagania decyzji. Pojawia się w związku z tym pytanie: jakie jest ryzyko związane z decyzją podjętą na podstawie analiz GIS? Na ryzyko podjęcia nieprawidłowej decyzji ma wpływ na pewno jakość danych źródłowych znajdujących się w bazie. Zwykle jednak nie tylko dane są podstawą podjęcia decyzji, ale wyniki ich przetworzeń. Dane źródłowe podlegają przekształceniom, są wykorzystywane jako dane wejściowe w różnych modelach. Wynik analizy jest obciążony zarówno niedokładnością danych źródłowych, jak również niedokładnością wykorzystywanych algorytmów i nieadekwatnością modeli do rzeczywistości. W systemach GIS, tworzonych w Polsce, rzadko gromadzi się informacje o dokładności danych źródłowych. Z kolei wykorzystywane oprogramowanie GIS tylko sporadycznie umożliwia analizę wpływu dokładności danych na wynik przeprowadzonej analizy. Problem ten jest pozostawiony do rozwiązania indywidualnemu użytkownikowi. W niektórych narzędziach GIS ma on do wyboru najbardziej, popularną obecnie, metodę Monte Carlo. Metoda ta, pomimo, że jest uniwersalna, tzn. może być wykorzystywana właściwie we wszystkich analizach GIS, ma dwie wady: analizy trzeba wykonywać setki razy, żeby oszacować dokładność wyniku analizy, a poza tym nie pozwala ona na określanie przestrzennego rozkładu dokładności analizy. Alternatywą są metody analityczne lub empiryczne, o ile znane są wykorzystywane w systemie GIS algorytmy obliczeniowe. Przedstawiona powyżej problematyka, dotycząca wpływu dokładności danych na ryzyko podejmowania decyzji z wykorzystaniem systemów GIS, została omówiona szerzej w niniejszej publikacji. Porównano w niej wyniki analizy przestrzennych wykonanej z uwzględnieniem wpływu dokładności danych źródłowych oraz bez uwzględnienia tego faktu. Problem przenoszenia się dokładności danych źródłowych na wynik analizy przestrzennej pokazano na przykładzie wyboru nowej lokalizacji pod zabudowę. W analizie uwzględniono dokładność numerycznego modelu terenu (DTM). Umożliwiło to oszacowanie ryzyka związanego z podjęciem decyzji wyboru lokalizacji i koniecznością wykonania nieplanowanych prac ziemnych w celu wyrównaniem terenu.

1. WPROWADZENIE

Niedokładność danych źródłowych w systemach GIS nie jest obecnie uwzględniana podczas wykonywania analiz przestrzennych. W zasadzie nie gromadzi się również

informacji o dokładności danych w systemach GIS tworzonych w Polsce (np. LPIS – Land Parcel Identification System, OKI – Ośrodki Koordynacyjno Informacyjne przy RZGW – Regionalne zarządy Gospodarki Wodnej, SIOT – System Informacji o Terenie LGOM).

Wydaje się jednak, że istnieje znaczące ryzyko związane z podejmowaniem decyzji bez uwzględniania przy tym dokładności danych.

2. STAN WIEDZY

W momencie, kiedy zaczynamy naprawdę wykorzystywać GIS dla wspomagania decyzji, czyli kiedy decydenci podejmują faktyczne decyzje korzystając z danych GIS, pojawia się pytanie – jakie jest ryzyko związane z tak podjętą decyzją?

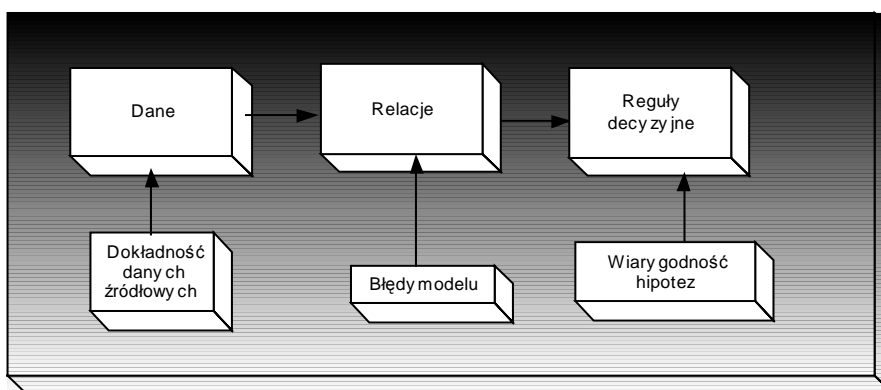
Dopóki dane GIS są rozumiane właściwie tylko jako tło kartograficzne, dopóty nie mamy pełnej świadomości ryzyka związanego z korzystaniem z tych danych. Problem zwykle w takim przypadku jest ograniczony do instrukcji geodezyjnych, zgodnie, z którymi mapy są tworzone. Pojawienie się systemów GIS, które mogą integrować dane w różnych skalach, formatach, układach współrzędnych, pozwoliło, już choćby wizualnie zweryfikować jakość zgromadzonych danych. Jak wiemy, w niektórych przypadkach oznacza to, że dokładności geodezyjne to jedno, a rzeczywistość to inna sprawa. Docelowo jednak chcielibyśmy mieć w systemie dane spójne i wolne od błędów grubych i zapewne po pewnym czasie tak będzie.

Jednakże nawet w takim przypadku dane zgromadzone w systemie nie będą bezbłędne i będzie to rzutowało na ich wykorzystanie, dokładność danych będzie wpływała na dokładność wyników dalszego ich przetwarzania. Zjawisko to do tej pory jest z reguły ignorowane podczas korzystania z danych GIS. Jedynie w sytuacjach, kiedy w grę wchodzi pieniądze lub bezpieczeństwo ludzi podejmuje się pewne próby uwzględniania dokładności danych. Dobrym przykładem może być system związany z dopłatami bezpośrednimi do rolnictwa tzw. ZSZIK lub związany z ochroną przeciw powodziową tzw. OKI (Hejmanowska, 2005; Krause *et al.*, 2003).

2.1. Teoria decyzji

W teorii decyzji istnieją generalnie dwa podejścia: deskryptywne i normatywne. W podejściu deskryptywnym, będącym przedmiotem badań psychologicznych, analizie poddawane są reakcje ludzkie związane ogólnie z podejmowaniem decyzji, nie koniecznie optymalnych (Sokołowska, Pohorille, 2000). Drugie podejście nazywane jest normatywnym, gdy celem jest podjęcie decyzji optymalnej. W tym przypadku decyzja może być podjęta w tzw. warunkach pewności, ryzyka i niepewności. Jeśli decyzja jest podejmowana w warunkach pewności, czyli każda decyzja pociąga za sobą określone, znane konsekwencje, to stanowi ona przedmiot deterministycznych metod teorii decyzji. W obu pozostałych przypadkach mamy do czynienia z niedeterministycznymi metodami teorii decyzji. Decyzja podejmowana w warunkach ryzyka to każda decyzja, która pociąga za sobą więcej niż jedną konsekwencję, znamy zbiór możliwych konsekwencji i prawdopodobieństwa ich wystąpienia. Decyzje podejmujemy w warunkach niepewności, kiedy nie znamy prawdopodobieństw wystąpienia konsekwencji danej decyzji.

Proces podejmowania decyzji z wykorzystaniem systemów GIS można przeanalizować na podstawie poniższego schematu (rys. 1). Dane zgromadzone w systemie są poddawane różnego rodzaju przetwarzaniom, wykorzystującym relacje pomiędzy danymi. Ostatecznie zostaje podjęta określona decyzja. Każdy element może, zatem być czynnikiem obciążającym ostateczną decyzję. Zarówno jakość danych źródłowych, jak i dokładność modeli i wiarygodność hipotez stanowią ryzyko w podejmowaniu decyzji z wykorzystaniem danych i narzędzi GIS.



Rys. 1. GIS jako narzędzie wspomaganie decyzji, źródła błędów

2.2. Jakość danych

Jakość danych GIS definiowana jest za pomocą szeregu standardów (ISO/TC 211, CEN TC287, DIGEST, OEEP, ISPRS, FIG, ISO/DIS 19113). Zgodnie z ostatnim z nich jakość można scharakteryzować za pomocą: kompletności, spójności logicznej, dokładności przestrzennej, czasowej i tematycznej. Jeśli założymy, że dane są kompletne, spójne logicznie i aktualne, czyli takie, jakie chcielibyśmy mieć w systemie, to jakość danych będzie rozumiana jako dokładność przestrzenna (dokładność położenia) i dokładność tematyczna, czyli dokładność atrybutów.

2.3. Analizy GIS

Dane w systemie GIS mogą być zgromadzone w modelu rastrowym i wektorowym. Warunkuje to w pewnym sensie rodzaj analiz przestrzennych, które można przeprowadzić. Generalnie w obu modelach można wykonywać zapytania do bazy danych poprzez atrybut jak i poprzez lokalizację. Również operacje logiczne (iloczyn, suma, różnica przestrzenna), buforowanie, analizy widoczności, grupowanie, generowanie spadków i ekspozycji są dostępne w obu modelach: wektorowym i rastrowym. Natomiast bardziej skomplikowane zadania (np. modelowanie hydrologiczne, generowanie odległości, miękkie analizy) są możliwe do wykonania jedynie w modelu rastrowym.

Uwzględnianie dokładności danych w analizach GIS jest możliwe z wykorzystaniem wzorów analitycznych lub empirycznych wiążących skutek z przyczyną lub też wykorzystując metody probabilistyczne, np. metodę Monte Carlo. W metodzie Monte Carlo wykonuje się analizę przestrzenną, następnie zaburza się dane wejściowe błędem i wykonuje analizę powtórnie. Należy przeprowadzić takie postępowanie wielokrotnie (setki razy). Analiza różnicowa wyników pozwala szacować dokładność analizy typu GIS.

Ważną jest świadomość przenoszenia się błędu danych źródłowych na wynik analizy przestrzennej i jego wpływu na ryzyko związane z podejmowaniem decyzji w oparciu o wynik przetworzenia danych w systemie GIS. Poniżej zaproponowano wykorzystanie formuły na ilościowe określenia ryzyka dla potrzeb szacowania ryzyka związanego z analizami przestrzennymi GIS.

2.4. Ryzyko

Ryzyko jest rozumiane jako potencjalna szkoda wywołana określonym działaniem (zjawiskiem), przy czym zwykle określa się prawdopodobieństwo wystąpienia niepożądanego zjawiska. Z pojęciem ryzyka związane są następujące terminy: prawdopodobieństwo, wartość oczekiwana, zysk, strata, poziom ryzyka, scenariusz. Scenariusz (prawdopodobny scenariusz) łączy ze sobą w jedną całość: ryzyko, stratę i zysk na pewnym poziomie prawdopodobieństwa (<http://encyclopedia.worldsearch.com>).

Spośród różnych definicji ryzyka, można podać również definicję ilościową (Kaplan, Garrick, 1981), która łączy ze sobą scenariusz (S), prawdopodobieństwo zajścia zdarzenia (P) oraz miarę skutków (C) wywołanych przez scenariusz (S).

$$R = S \cdot P \cdot C \quad (1)$$

Kluczową sprawą jest, więc prawdopodobieństwo (P) zajścia zdarzenia (S). Powstaje pytanie – w jaki sposób określić prawdopodobieństwo zajścia zdarzenia S, oraz co to jest zdarzenie S? Zdarzeniem S, w rozumieniu analiz przestrzennych GIS może być wynik analizy przestrzennej. Przykładowo, jeśli wykonujemy typową analizę GIS w celu znalezienia nowej lokalizacji, na przykład dla zabudowy, to poddajemy dane źródłowe szeregowi przekształceń i w efekcie otrzymujemy obszary, spełniające zadane na początku warunki. Jakość danych źródłowych wpływa na wiarygodność wyniku ich przetworzenia. Skoro dane nie są bezbłędne, to i wynik analizy przestrzennej też nie jest bezbłędny. Oznacza to, że moglibyśmy określić prawdopodobieństwo, że ten teren spełnia określone warunki zamiast pokazać obszary, które teoretycznie je spełniają.

Narzędzia GIS najczęściej nie dostarczają wprost informacji o dokładności zaimplementowanych funkcji. Do dyspozycji pozostają więc metody analityczne, zaproponowane w literaturze lub uzyskane w wyniku własnych analiz wybranych algorytmów (Hejmanowska, 2000, 2005), metody oparte o wzory empiryczne (Ehlschlaeger, Shortridge, 1996; Florinsky, 1998; Qihao Weng, 2002; Ritter, 1987; Skidmore, 1989) lub metoda Monte Carlo (Krause *et al.*, 2003; Wechsler, 2001). Niezależnie od sposobu obliczenia, wraz z wynikiem analizy można, znając dokładność danych źródłowych wygenerować mapę dokładności wyniku analizy, która z kolei pozwala

na określenie prawdopodobieństwa (P) zajścia zdarzenia (S) i realizację powyższego wzoru na ryzyko (1).

3. METODYKA

Opisana poniżej analiza przestrzenna miała na celu porównanie wyniku uzyskanego metodą twardą, bez uwzględniania dokładności danych źródłowych i metodą miękką, z uwzględnieniem tego faktu. Przedmiotem analizy był wybór lokalizacji pod nową zabudowę.

Przyjęto następujące warunki, które ta lokalizacja powinna spełniać:

- teren płaski, nachylenie $< 3^\circ$,
- odległość od wód > 250 m,
- odległość od dróg < 100 m,
- teren niezabudowany.

Analizie poddano obszar w atrakcyjnej części Krakowa, okolice Lasu Wolskiego. Wykorzystano dane GIS województwa małopolskiego w modelu wektorowym (wody, drogi, teren zabudowany) i w modelu rastrowym (DTM).

3.1. Analiza bez uwzględnienia dokładności danych źródłowych

Wybór lokalizacji spełniającej powyższe założenia wykonuje się z reguły w sposób twardy wykorzystując dostępne narzędzia GIS. W związku z tym, że dane są zarówno w modelu wektorowym jak i rastrowym wykorzystuje się narzędzia oprogramowana umożliwiającego przetwarzanie obu typu danych (czasem jest to związane z wykorzystywaniem zupełnie różnych programów). Z numerycznego modelu terenu (DTM) generujemy w modelu rastrowym mapę nachyleń. Następnie dokonujemy takiego przetworzenia, żeby pokazać obszary o nachyleniu mniejszym niż 3 stopnie. Wynik tego przetworzenia transformujemy do modelu wektorowego i dalsze analizy wykonujemy już w tym modelu. Generujemy obszar buforowy względem dróg i rzek oraz wyłączamy obszary już zabudowane. Ostatnim etapem jest pokazanie obszarów spełniających wszystkie cztery powyższe warunki.

3.2. Analiza z uwzględnieniem dokładności DTM

Jakość danych źródłowych (w tym ich dokładność) obciążają z pewnością wyniki opisanej powyżej analizy. W celu zademonstrowania wpływu dokładności danych na wynik analizy przestrzennej wykonano powyższą analizę uwzględniając dokładność DTM. Jeśli założymy najprostszy sposób obliczania nachylenia, czyli wykorzystania 4 pikseli otaczających piksel, dla którego jest ono liczone (czyli piksel górny, dolny, lewy i prawy) i obliczenia wypadkowego nachylenia ze składowej x i y to błąd nachylenia można obliczyć następująco (Hejmanowska, 2005):

$$m_s = \pm \frac{\sqrt{2}}{2\Delta h(1 + \tan^2 S)} m_z \quad (2)$$

gdzie:

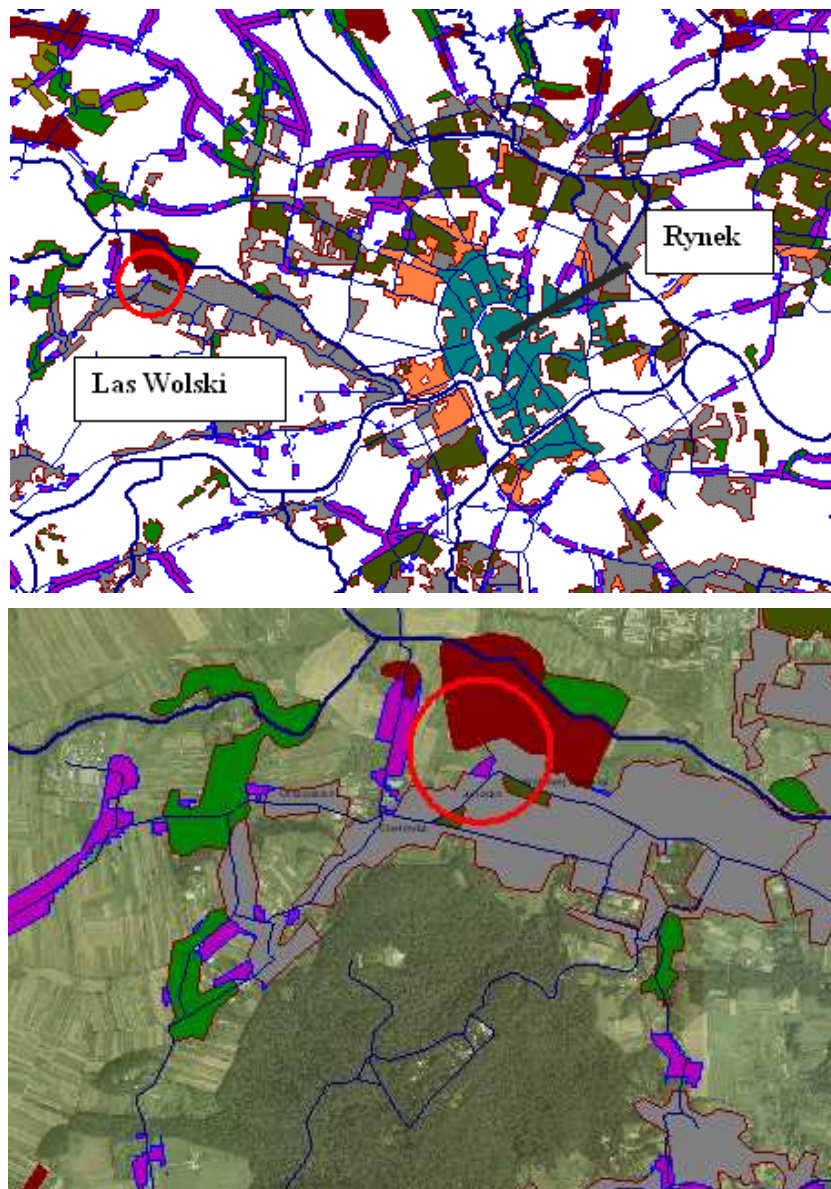
- m_s – niedokładność nachylenia,
- Δh – oczko siatki DTM,
- S – nachylenie,
- m_z – błąd DTM.

Zaproponowany wzór pozwala, równocześnie z mapą nachyleń, wygenerować mapę błędów nachyleń. Oznacza to, że zamiast pokazać obszary płaskie, czyli o nachyleniu mniejszym niż 3 stopnie można wygenerować mapę prawdopodobieństwa, że teren jest płaski. Następnie można określić obszary, dla których prawdopodobieństwo, że teren nie jest płaski jest określone, np. wynosi 30 %. Ostatecznie więc, można omawianą analizę zmodyfikować przedstawiając, jakie jest ryzyko wyboru danej lokalizacji związane w tym przypadku z dokładnością tylko jednej analizowanej warstwy: DTM.

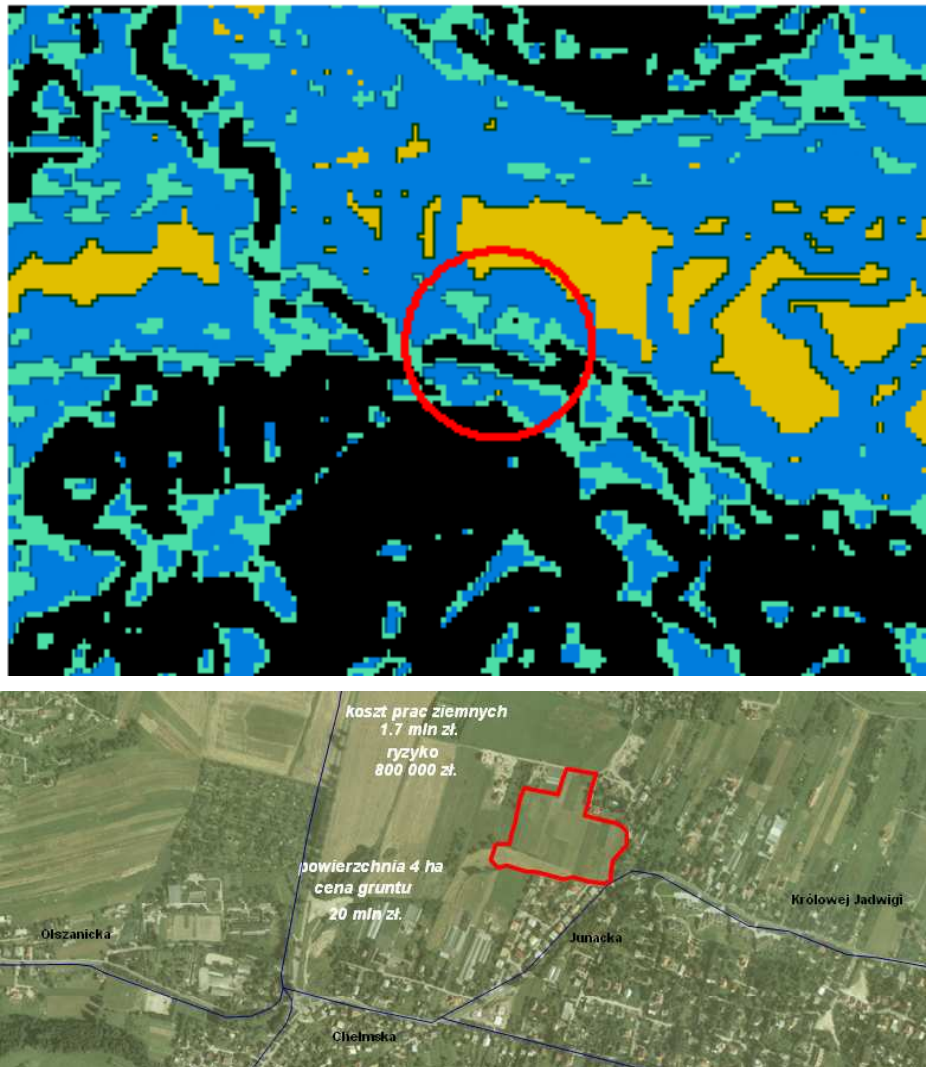
4. WYNIKI

W wyniku tradycyjnie przeprowadzonej analizy (metodą twardą) uzyskano nowe lokalizacje zaznaczone kolorem fioletowym na rysunku 2. Przeanalizujemy lokalizację u zbiegu ulic Królowej Jadwigi i Junackiej, zaznaczoną okręgiem na tym rysunku.

W dalszej kolejności przeprowadzono analizę z uwzględnieniem dokładności DTM, przyjmując jednorodny błąd DTM, $m_z = \pm 1$ m. Obliczono prawdopodobieństwo, że teren jest płaski, nachylenie poniżej 3 stopnie (rys. 3). W kolorze żółtym przedstawione zostały tereny płaskie (prawdopodobieństwo równe 1), w kolorze niebieskim tereny, dla których prawdopodobieństwo, że nachylenie jest poniżej 3 stopnie wynosi od 0.95 do 1, a obszary o prawdopodobieństwie od 0.5÷0.95 na zielono. Pozostałe obszary (nachylenie powyżej 3 stopni), zostały wyłączone z analizy i zaznaczone na czarno na rysunku. Ostatecznie przeprowadzoną analizę związane z kosztami wyrównania terenu, wzięwszy pod uwagę fakt, że teren nie byłby płaski, zgodnie z przyjętymi kryteriami. Ostatecznie, koszt prac ziemnych oszacowano na 1.7 mln zł, co daje ryzyko ok. 800 000 zł.



Rys. 2. Lokalizacje pod nową zabudowę (kolor fioletowy) na tle ortofotomapy



Rys. 3. Mapa prawdopodobieństwa, że teren jest płaski:
kolor żółty = 1, kolor niebieski - 0.95÷1, kolor zielony - 0.5÷0.95 (u góry);
ryzyko związane z wyborem zaznaczonej lokalizacji - ca 800 000 zł
(u dołu)

5. WNIOSKI

Wyniki analizy przestrzennej dotyczącej wyboru lokalizacji nowej zabudowy przeprowadzonej z uwzględnieniem niedokładności DTM pozwalają stwierdzić, że obszary dla których prawdopodobieństwo niespełnienia warunku, że teren jest płaski nie są zanedbywane małe i w niektórych przypadkach mogą powodować określone ryzyko związane ewentualnie z koniecznością wykonania nieplanowanych znacznych prac ziemnych. Oznacza to, że chcąc racjonalnie zarządzać ryzykiem związanym z podejmowaniem decyzji w oparciu o analizy GIS należy uwzględniać jakość wykorzystywanych danych, w tym przypadku ich dokładność.

6. LITERATURA

Ehlschlaeger C.R., Shortridge A., 1996. Modeling Elevation Uncertainty in Geographical Analyses. Proceeding of the International Symposium on Spatial Data Handling, Delft, Netherlands. 9B.15-9B.25.

Florinsky I.V., 1998. Accuracy of local topographic variables derived from digital elevation models. *Int. J. Geographical Information Science*, Vol. 12, no 1, s. 47-61.

Hejmanowska B., 2000. Wpływ niedokładności Cyfrowego Modelu Terenu na błąd nachyleń i ekspozycji. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, Vol. 10. s. 41/1-41/14.

Hejmanowska B., 2005. Wpływ jakości danych na ryzyko procesów decyzyjnych wspieranych analizami GIS. *AGH Uczelniane Wydawnictwa Naukowo-Dydaktyczne*, Kraków 2005, ISDN 0867-6631.

Kaplan S., Garrick B.J., 1981. On the quantitative definition of risk. *Risk Analysis*, Vol. 1, s. 11-27.

Krause J., Uhrich S., Bormann H., Dieckrüger B., 2003. Uncertainty Analysis of a Floodrisk Mapping Procedure Applied in Urban Areas”, Hydrology Division. Geographical Institute University of Bonn, <http://www.giub.uni-bonn.de/hrg/Poster%20and%20Presentations/nizza02.pdf>

Qihao Weng, 2002. Quantifying Uncertainty of Digital Elevation, Models Derived from Topographic Maps. *Advances in Spatial Data Handling*, Edited by D. Richardson and P. van Oosterom, Springer-Verlag, New York, s. 403-418.

Ritter P., 1987. A vector-based slope and aspect generation algorithm. *Photogrammetric Engineering and Remote Sensing*, 53(8), s. 1109-1111.

Skidmore A. K., 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *International Journal of Geographical Information Systems*, 3(4), s. 211-232.

Sokołowska J., Pohorille A., 2000. Model of risk and choice: challenge or danger. *Acta Psychologica*, 104 (2000), s. 339-369.

Wechsler P. S., 2001. Digital Elevation Model (DEM) Uncertainty: Evaluation and Effect on Topographic Parameters. http://www.csulb.edu/~wechsler/Dissertation/P262/_Author_Information

Prezentowane wyniki badań prowadzono w ramach tematu AGH: 11.11.150.459.

**GIS – A TOOL IN DECISION SUPPORT – RISK CAUSED
BY DATA QUALITY EFFECT**

KEY WORDS: GIS, data accuracy, decision support

Summary

There are many GIS tools for data analysis, also applied in Poland to decision support. The following question can appear: what is the risk of a decision based on the GIS analysis? GIS data quality certainly affects on the decision risk. Usually not only raw data are analyzed, but also the results of data transformations. Data are often transformed in GIS systems and used in different models as input data. Results of transformation are influenced by the original data quality (especially accuracy) and model/algorithm reliability.

In GIS databases in Poland, information on data quality is very rare gathered. On the other hand, GIS software only occasionally delivers tools for algorithm and function accuracy analysis. The problem is left individually to the user. In some software, the popular Monte Carlo method is implemented. The method is very universal, it means that can be applied in almost all GIS analyses. Monte Carlo has also disadvantages: analyses must be carried out hundreds times, and the spatial distribution of the results error is difficult to obtain. Analytical or empirical methods are an alternative if the algorithms implemented in GIS software are known.

In the paper, the risk connected to the GIS analyses based on the data with known accuracy is presented. Spatial analysis is performed by hard methods, without data quality consideration and assuming some data uncertainty. Error propagation is presented in the paper based on a new urban area example and the DTM accuracy was considered in the spatial analysis. The risk of ground leveling caused by DTM inaccuracy was evaluated.

Dr hab. inż. Beata Hejmanowska
e-mail: galia@agh.edu.pl
tel. +12 6172288