

Adam DOSKOCZ¹
Wojciech REJCHEL²

PROPOZYCJA AUTOMATYZACJI ANALIZY DOKŁADNOŚCI BAZ DANYCH MAP WIELKOSKALOWYCH

Streszczenie

W pracy przedstawiono propozycję automatyzacji oceny dokładności opracowania sytuacyjnego wielkoskalowych map cyfrowych. Badania wykonano w zakresie czterech różnych metod pozyskania danych do bazy mapy: nowy pomiar tachimetrem, przeliczenie wcześniejszych pomiarów bezpośrednich zrealizowanych poprzez pomiary ortogonalne i biegunowe, ręczna wektoryzacja rastrowego obrazu ortofotomapy i graficzno-numeryczne przetwarzanie map analogowych. Uzyskane rezultaty oceny dokładności z zastosowaniem zaproponowanej analizy statystycznej są zgodne z wnioskami sformułowanymi na podstawie klasycznej oceny dokładności (zgodnie z wymogami standardów technicznych). Potwierdzono wysoką dokładność mapy wykonanej z pomiarów tachimetrem elektronicznym. Dokładność mapy utworzonej na podstawie wcześniejszych pomiarów także jest satysfakcjonująca. Badania wykazały wysoką dokładność (w zakresie szczegółów I grupy dokładnościowej, pokryw studzienek uzbrojenia podziemnego) ortofotomapy cyfrowej. Najniższą dokładność stwierdzono w przypadku mapy wykonanej poprzez przetworzenie graficzno-numeryczne map analogowych (poprzez wektoryzację map rastrowych).

Zgodnie z opinią autorów, pożądanym jest ustanowienie procedur automatyzacji oceny dokładności map cyfrowych i ich zastosowania w odpowiednim systemie eksperckim. Zapewni to sprawną weryfikację dokładności zgromadzonych baz danych map wielkoskalowych zgodną z oczekiwaniami ich użytkowników, dostarczając rozszerzonego zbioru metadanych.

Słowa kluczowe: baza danych, mapa wielkoskalowa, analiza dokładności

1. Wprowadzenie

Obecne działania i najbliższe zamierzenia Głównego Urzędu Geodezji i Kartografii w Polsce, uwzględniające stan prawny Unii Europejskiej, skierowane są na integrację danych gromadzonych, aktualizowanych i udostępnianych przez Państwową Służbę Geodezyjną i Kartograficzną [1]. Szczególnie implementacja zapisów ustawy o INSPIRE (Infrastructure for Spatial Information in the European Community, [2]) w znacznej mierze przyspiesza porządkowanie krajowych zbiorów danych przestrzennych oraz budowanie interoperacyjnych baz danych [3]. Jak wiadomo, podstawowym celem tworzenia infrastruktur informacji przestrzennej jest ułatwienie dostępu do danych przestrzennych, gromadzonych przez administrację rządową i samorządową oraz w różnych sektorach gospodarczych, wszystkim zainteresowanym podmiotom [4]. W przypadku państwowego zasobu geodezyjnego i kartograficznego opracowywana infrastruktura danych przestrzennych ma zawierać m.in. jego podstawowe części składowe tj.: kataster, mapę zasadniczą, geodezyjną ewidencję sieci uzbrojenia terenu, osnowy geodezyjne [5].

¹dr inż., Katedra Geodezji Szczegółowej, Wydział Geodezji i Gospodarki Przestrzennej, Uniwersytet Warmińsko-Mazurski w Olsztynie

²dr, Zakład Statystyki Matematycznej i Analizy Danych, Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu

Ustawa dotycząca INSPIRE wprowadza nowelizację ustawy z dnia 17 maja 1989 r. Prawo geodezyjne i kartograficzne (Dz. U. z 2005 r. nr 240, poz. 2027, z późn. zm.), m.in. w zakresie obowiązku założenia i prowadzenia baz danych obiektów topograficznych o stopniu szczegółowości zapewniającym realizację standardowych opracowań kartograficznych w skalach 1:500-1:5000. Nakłada także na samorządy terytorialne obowiązek prowadzenia bazy danych mapy zasadniczej, a tym samym przekształcenia tradycyjnych map analogowych na ich postać cyfrową do 31 grudnia 2013 roku. Wraz z postępem technologicznym zmianom ulegają także standardy techniczne. W Głównym Urzędzie Geodezji i Kartografii prowadzone są prace w zakresie opracowania nowych standardów technicznych w formie stosownych rozporządzeń ministerialnych. Ukazało się już m.in. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 9 listopada 2011 r. w sprawie standardów technicznych wykonywania geodezyjnych pomiarów sytuacyjnych i wysokościowych oraz opracowywania i przekazywania wyników tych pomiarów do państwowego zasobu geodezyjnego i kartograficznego. Zgodnie z art. 35 ustawy z dnia 4 marca 2010 roku o infrastrukturze informacji przestrzennej oraz stanowiskiem GUGiK, dotychczasowe standardy techniczne (instrukcje i wytyczne) formalnie przestają obowiązywać po dniu 8 czerwca 2012 roku. Jednakże nowe wymagania techniczne, zawarte w projektowanych rozporządzeniach, bazują na dotychczasowych instrukcjach oraz wytycznych technicznych (w tym m.in. na: O-1, O-2, O-3, G-3, G-4 i G-7).

Wielkoskalowe mapy cyfrowe wykonywane są wieloma technologiami, o specyficznych cechach i różnej jakości produktu finalnego [6]. Dlatego też autorzy opracowań cyfrowych oraz krąg ich użytkowników powinni wiedzieć jakiej dokładności opracowaniami dysponują, a w związku z tym do realizacji jakich zadań mogą być one zastosowane. Jest to także niezwykle istotne w zakresie modernizacji i utrzymania odpowiedniej jakości państwowego zasobu geodezyjnego i kartograficznego, w związku z aktualnie zachodzącymi przeobrażeniami w zakresie zasobu map wielkoskalowych i ich konwersji do państwowego układu współrzędnych geodezyjnych PUGW „2000”. Sprawne zarządzanie przestrzenią i efektywne wykonywanie opracowań planistycznych także wymaga zastosowania systemów informacji przestrzennej i najnowszych produktów geoinformatycznych [7, 8, 9].

Celem niniejszej pracy jest wskazanie możliwości automatyzacji analizy i oceny dokładności wielkoskalowych map cyfrowych poprzez konstruowanie przedziałów ufności oraz weryfikację parametrycznych i nieparametrycznych testów statystycznych (dotyczących błędów metod wyznaczenia położenia szczegółów I grupy dokładnościowej zgromadzonych w bazach danych).

2. Badania dotyczące oceny dokładności map wielkoskalowych

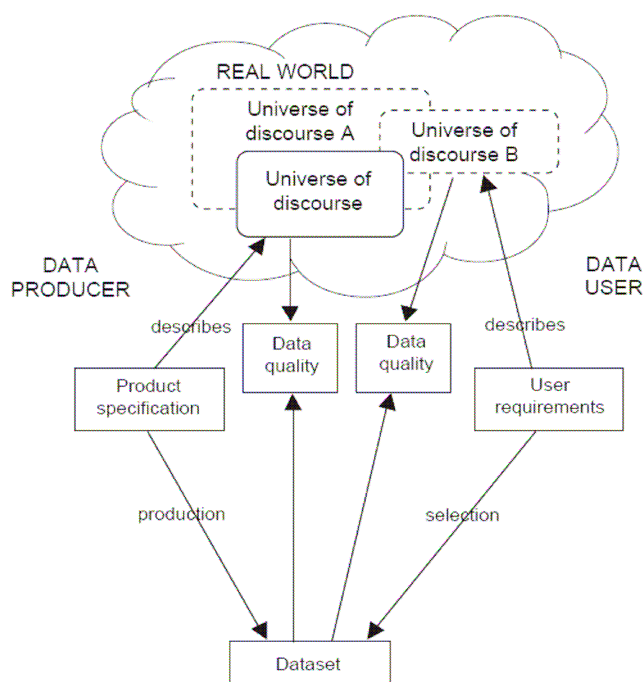
Odpowiednia jakość baz danych cyfrowych oznacza spełnienie obecnych i przyszłych oczekiwań użytkowników danych [10]. Podstawowymi cechami opisującymi jakość danych cyfrowych są genealogia, dokładność, kompletność, zgodność i aktualność [11]. Powszechnie znana jest opinia, że zgromadzone dane powinny być kompletne, zgodne i aktualne w stopniu maksymalnym z możliwych. Natomiast znaczenie dokładności baz danych nie zawsze jest dostrzegane. Tymczasem niezwykle istotna jest także dokładność bazy danych mapy cyfrowej, gdyż jest kluczowym atrybutem w procesie integracji danych geograficznych i ich interoperacyjności w procesie budowy infrastruktury danych przestrzennych [12].

Ocena dokładności map cyfrowych nie jest problemem rozpatrywanym tylko w Polsce. Badania w tym zakresie prowadzone są przez wielu uczonych na świecie, co przedstawili w swojej pracy López i Gordo [13]. W omawianych badaniach stosowane są analizy statystyczne [14, 15] oraz inne współczesne metody badawcze [16, 17, 18, 19]. Badania prowadzone w ośrodkach zagranicznych koncentrują się na mapach katastralnych oraz rolniczych opracowywanych w skalach średnich i małych. Natomiast w Polsce, oprócz mapy ewidencji gruntów, opracowywane są także inne mapy wielkoskalowe w tym mapy zasadnicze i sytuacyjno-wysokościowe do celów projektowych, które w związku z ich wykorzystywaniem do opracowań gospodarczych stanowią ważny materiał badawczy.

Z uwagi na globalne znaczenie wymiany danych geograficznych, prace z zakresu geodezji i kartografii powinny być prowadzone w oparciu o międzynarodowy dorobek normalizacyjny [20]. Międzynarodowe normy serii ISO sukcesywnie są przyjmowane (uznawane) przez Europejski Komitet Normalizacyjny CEN jako EN - normy europejskie oraz przez Polski Komitet Normalizacyjny jako PN - Polskie Normy [21]. Spośród norm opracowanych przez Komitet

Techniczny PKN/KT 297 ds. Informacji geograficznej, dla tematyki poruszanej w niniejszej pracy, szczególnie interesująca jest PN-EN ISO 19113:2005 zawierająca podstawy opisu jakości danych geograficznych oraz PN-EN ISO 19114:2005 określająca procedury oceny jakości danych geograficznych [22].

W opisie jakości danych geograficznych zgodnym z normą PN-EN ISO 19113:2005 stosowana jest zmienna logiczna (ang. "boolean variable"), która określa czy prawdziwe jest stwierdzenie, że jakość określonej cechy atrybutowej danych jest odpowiednia. Norma PN-EN ISO 19113:2005 zawiera także podstawową terminologię oraz ogólną koncepcję charakterystyki jakości danych geograficznych (rys. 1).



Rys. 1. Ramowa koncepcja zapewnienia jakości danych [23]

W procedurach oceny jakości danych geograficznych zawartych w normie PN-EN ISO 19114:2005 ocena dokładności sytuacyjnej danych (ang. "positional accuracy", w Polsce poprawnym i pełnym sformułowaniem jest: dokładność opracowania sytuacyjnego mapy lub bazy danych) realizowana jest poprzez obliczenie średniego błędu położenia punktu (ang. root mean square error, RMSE) wyznaczonego na podstawie różnicy współrzędnych punktów zapisanych w bazie danych i współrzędnych rzeczywistych (kontrolnie) wyznaczonych:

$$RMSE = \sqrt{\Sigma(Z^2)/n} \quad (1)$$

gdzie: $Z = \sqrt{(X_{II} - X_I)^2 + (Y_{II} - Y_I)^2}$ - długość wektora przesunięcia punktu kontrolnego,

X_{II}, Y_{II} - współrzędne punktu kontrolnego pozyskane z bazy danych ocenianej mapy cyfrowej,

X_I, Y_I - wzorcowe współrzędne punktu kontrolnego,

n - liczba punktów kontrolnych.

Zgodnie z Aneks D, wyjaśniającym pomiar jakości danych w normie międzynarodowej PN-EN ISO 19114:2005, należy podać liczbę punktów, których błąd położenia przekracza limit specyfikacji dokładności. Warto też uzyskaną ocenę dokładności zbioru danych wyrazić wielkością błędu względnego wyrażonego w procentach (Annex D, [24]).

Zaproponowaną w niniejszej pracy propozycję automatyzacji analizy dokładności opracowania sytuacyjnego wielkoskalowych map cyfrowych sformułowano w nawiązaniu do wymogu Instrukcji technicznej O-2, zgodnie z którym błąd położenia punktu sytuacyjnego I grupy dokładnościowej nie powinien przekroczyć ± 0.3 mm w skali mapy względem najbliższych punktów poziomej osnowy

geodezyjnej [25]. Oznacza to, że analizowane bazy danych map wielkoskalowych obszarów miejskich, powinny spełniać standard mapy o stopniu szczegółowości w skali 1:500 i dokładność ich opracowania sytuacyjnego nie powinna być mniejsza niż 15 centymetrów ($0.0003 \text{ m} \times 500 = 0.15 \text{ m}$) względem punktów poziomej osnowy geodezyjnej.

Prezentowane badania zrealizowano w zakresie wielkoskalowych map cyfrowych obejmujących tereny zurbanizowane. W badaniach rozpatrywane są bazy (zbiory) danych map utworzonych w oparciu o cztery metody wyznaczenia położenia sytuacyjnego szczegółów: Metodą A – w której, współrzędne szczegółów sytuacyjnych wyznaczono z pomiarów tachimetrem elektronicznym. Analizę dokładności wykonano na podstawie zbioru empirycznego A o liczności 484 punktów kontrolnych.

Metodą B – w której, współrzędne szczegółów I grupy pozyskano z istniejących wyników pomiarów sytuacyjnych zrealizowanych w latach 1974-99 głównie metodą domiarów prostokątnych, a w ostatnim okresie metodą biegunową z zastosowaniem tachimetru elektronicznego. Rozpatrywany zbiór empiryczny B liczy 1636 punktów kontrolnych.

Metodą C – w której, współrzędne określające sytuacyjne położenie punktów kontrolnych na rastrowym obrazie ortofotomapy, opracowanej w skali bazowej 1:2000, wyznaczono w wyniku manualnej wektoryzacji tych punktów. Analizowany zbiór empiryczny C zawiera 311 punktów kontrolnych.

Metodą D – w której, współrzędne szczegółów I grupy pozyskano metodą przetwarzania graficzno-numerycznego (w technologii wektoryzacji poprzedzonej skanowaniem map analogowych w skali 1:500 wraz z nakładkami uzbrojenia w skali 1:1000). Analizę dokładności wykonano na podstawie zbioru empirycznego D o liczności 2287 punktów kontrolnych.

Współrzędne analizowanych szczegółów sytuacyjnych wyznaczonych metodą A, B i D uzyskano w formie wykazów tekstowych bądź raportów z bazy danych ocenianych map, a w przypadku metody C także sporządzono wykaz współrzędnych poprzez zapis do pliku wyników manualnej wektoryzacji rastrowego obrazu ortofotomapy (wyników tzw. monoplottingu) zgromadzonych na nowo utworzonej wektorowej warstwie mapy. Natomiast wzorcowe współrzędne punktów kontrolnych wyznaczono poprzez nowe pomiary bezpośrednie, zawierające obserwacje nadliczbowe, wykonane tachimetrem elektronicznym. W oparciu o zrealizowane wyrównanie ściśle wyników pomiarów i ocenę ich dokładności stwierdzono, że referencyjne położenie punktów kontrolnych wyznaczono z dokładnością $m_p < 0.03 \text{ m}$.

Zbiory empiryczne stanowią obiekty kontrolne obejmujące szczegóły sytuacyjne I grupy dokładnościowej wyznaczone w ramach określonej metody pozyskania danych. Lokalizację obiektów kontrolnych na obszarze poszczególnych map ustalono losowo, z uwzględnieniem uwag kierownictwa administracji geodezyjnej poszczególnych miast. Analizowane zbiory punktów kontrolnych stanowią pierwotne (nieprzetworzone) zbiory empiryczne, zawierające także punkty o błędach odstających.

3. Propozycja automatyzacji analizy dokładności baz danych map wielkoskalowych

W przeprowadzonej analizie dokładności baz danych map wielkoskalowych skonstruowano przedziały ufności i przeprowadzono testy parametryczne i nieparametryczne dotyczące błędów metod wyznaczenia położenia sytuacyjnego szczegółów (metod A-D opisanych w 2. rozdziale pracy). Za pomocą przedziałów ufności z wysokim prawdopodobieństwem oszacowano, właściwy każdej metodzie, składnik systematyczny jej błędu. Natomiast wspierając się odpowiednimi testami statystycznymi w zakresie rozpatrywanych par metod, każdorazowo starano się wskazać lepszą z nich. Pierwszą część analizy oparto na własnościach średniej z próby jako estymatora wartości oczekiwanej badanej zmiennej. W części drugiej zastosowano metody rangowe, zaliczane do metod statystyki nieparametrycznej, w których próba jest na wstępie rangowana - oznacza to, że każda wartość badanej cechy jest zastępowana jej pozycją (rangą) na uporządkowanej rosnąco liście wszystkich wartości tej cechy. W przypadku zbiorów zaburzonych obserwacjami odstającymi metody nieparametryczne dają lepsze rezultaty [26, 27]. Wcześniej wykonane badania [28, 29] potwierdziły takie zaburzenia, w analizowanych zbiorach punktów kontrolnych występują punkty wykazujące błędy grube i rozkład populacji empirycznych wykazuje brak zgodności z rozkładem normalnym. Prace obliczeniowe w ramach prezentowanej analizy dokładności wykonano przy pomocy programu „R” [30].

3.1. Przedziały ufności i weryfikacja hipotez dotyczących średnich błędów metod

W analizie ustalono, że $z_1^A, z_2^A, \dots, z_n^A$ ($n=484$) oznaczają zaobserwowane błędy poszczególnych punktów kontrolnych metody A. Ponadto μ_A oznacza nieznaną średnią błąd metody A (jego wartość oczekiwaną). Analogiczne oznaczenia przyjęto dla metod B, C i D. Główną rolę w zastosowanym rozwiązaniu odegrało centralne twierdzenie graniczne, które stanowi, iż średnia z próby jest asymptotycznie normalnym estymatorem wartości oczekiwanej rozważanej zmiennej [27]. Fakt ten pozwolił zbudować asymptotyczne (przybliżone) przedziały ufności dla nieznanego dokładności (μ_A, μ_B, μ_C oraz μ_D) opisywanych metod wyznaczenia położenia sytuacyjnego szczegółów. Przedział ufności dla średniego błędu wyznaczenia położenia punktu sytuacyjnego na poziomie ufności 0.999 wynosi:

$$\mu_A \in \left[\bar{z}^A - \frac{u s^A}{\sqrt{n}}, \bar{z}^A + \frac{u s^A}{\sqrt{n}} \right],$$

gdzie $\bar{z}^A = \frac{1}{n} \sum_{i=1}^n z_i^A$ jest wartością oczekiwaną (przeciętną wielkością błędu) wyrażoną jej

estymatorem - średnią arytmetyczną oraz $s^A = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_i^A - \bar{z}^A)^2}$ jest odchyleniem standardowym.

Natomiast u jest kwantylem rozkładu normalnego standardowego rzędu 0.9995. Należy pamiętać, że powyższy przedział jest asymptotycznym przedziałem ufności dla nieznanego średniego błędu. Przybliżenie to jest tym lepsze, im liczniejsza jest badana próba. Rozpatrywane zbiory punktów kontrolnych z pewnością można uznać za liczne - najmniej elementów (311) liczy próba odnosząca się do metody C. Otrzymano następujące przedziały ufności dla średniego błędu w ramach poszczególnych metod (przy poziomie ufności 0.999):

$$\begin{aligned} \mu_A &\in [0.05 m; 0.07 m], \\ \mu_B &\in [0.16 m; 0.24 m], \\ \mu_C &\in [0.16 m; 0.20 m], \\ \mu_D &\in [0.32 m; 0.36 m]. \end{aligned}$$

Następnym etapem analizy jest weryfikacja hipotez parametrycznych. Schemat postępowania omówiono na przykładzie analizy błędów metod A i B, gdyż w przypadku pozostałych pięciu rozpatrywanych par metod zastosowano podejście analogiczne. Niech $z_1^A, z_2^A, \dots, z_n^A$ ($n=484$) oraz $z_1^B, z_2^B, \dots, z_m^B$ ($m=1636$) oznaczają zaobserwowane błędy poszczególnych punktów kontrolnych metod A oraz B, odpowiednio. Zweryfikowano hipotezę mówiącą, że $\mu_A = \mu_B$, czyli średnie błędy metod A i B są takie same. Przyjętą hipotezą alternatywną, przyjmowaną w przypadku odrzucenia rozważanej hipotezy, jest stwierdzenie, że $\mu_A < \mu_B$, czyli średni błąd metody A jest mniejszy niż metody B.

W celu weryfikacji hipotez użyto testów dla dwóch średnich z prób niezależnych [27]. W zakresie rozpatrywanych par metod (A-B, A-C, A-D, B-C, B-D, C-D) wyznaczono p -wartości testów, czyli najmniejsze poziomy istotności testów, przy których zaobserwowane wartości statystyk testowych prowadzą do odrzucenia weryfikowanych hipotez. W pięciu przypadkach są to liczby wielkości 10^{-21} , co spowodowało odrzucenie weryfikowanych hipotez na korzyść hipotez alternatywnych. Jednakże wynik porównania metod B i C nie dostarczył zdecydowanego rozstrzygnięcia i wskazania lepszej z nich. Średnie błędy próbkowe w tych przypadkach wynoszą: $\bar{z}^B = 0.20 m$ i $\bar{z}^C = 0.18 m$. Dla testu z weryfikowaną hipotezą $\mu_B = \mu_C$ p -wartość wyniosła 0.06, co „delikatnie” skłania nas do przyjęcia hipotezy alternatywnej ($\mu_B > \mu_C$). Aczkolwiek testując hipotezę $\mu_B = \mu_C$ przy $\mu_B \neq \mu_C$, czyli hipoteza alternatywna orzeka, iż błędy są różne, nie wskazując metody lepszej, otrzymujemy p -wartość równą 0.13, co nie daje nam dostatecznych

podstaw do odrzucenia hipotezy $\mu_B = \mu_C$. Dlatego w przypadku porównania metod B i C sytuację uznano za nierozstrzygniętą. W tablicy 1 przedstawiono uzyskane rezultaty (w komórkach umieszczono nazwę metody lepszej w parze, R oznacza brak rozstrzygnięcia).

Tablica 1. Wyniki testów dotyczących średnich błędów poszczególnych metod

	A	B	C	D
A		A	A	A
B			R	B
C				C

Wykonane testy parametryczne, jak i obliczone przedziały ufności pozwoliły jednoznacznie stwierdzić, że metoda A jest najdokładniejszą metodą wyznaczenia położenia sytuacyjnego szczegółów zgromadzonych w bazie mapy cyfrowej. Potwierdzono to w każdym z trzech testów dotyczących tej metody, jak również obliczonym przedziałem ufności szacującym średni błąd metody A liczbą z przedziału $[0.05 \text{ m}; 0.07 \text{ m}]$. Przeprowadzona analiza statystyczna nie wskazała lepszej z metod pomiędzy przypadkami B i C. Natomiast w ostatniej sytuacji nie ma wątpliwości, że mapa wykonana metodą przetworzenia graficzno-numerycznego map analogowych (metoda D) charakteryzuje się najniższą dokładnością, albowiem jej średni błąd przekracza 0.32 m oraz wszystkie testy (w zakresie rozpatrywanych par metod) zawsze oceniły ją jako mniej dokładną.

Należy podkreślić, że powyżej przedstawione wyniki uzyskano na podstawie asymptotycznych przedziałów ufności i testów dla średnich. Ponadto średnia z próby, odgrywająca w tym przypadku kluczową rolę, jest bardzo wrażliwa na obserwacje odstające. Dlatego wykonano także alternatywną analizę statystyczną, wywodzącą się z nurtu statystyki nieparametrycznej.

3.2. Nieparametryczne podejście do konstrukcji przedziałów ufności i weryfikacji hipotez dotyczących błędów metod

Schemat weryfikacji hipotez podobnie jak poprzednio przedstawiono na przykładzie. W porównaniu błędów metod A oraz B rozważano zbiory $z_1^A, z_2^A, \dots, z_n^A$ i $z_1^B, z_2^B, \dots, z_m^B$ będące zaobserwowanymi (stwierdzonymi) błędami poszczególnych punktów kontrolnych. Zaproponowano następujący model błędu pojedynczego punktu:

$$Z_i^A = \theta_A + \varepsilon_i^A, \quad i = 1, \dots, n$$

w którym błąd metody wyznaczenia położenia sytuacyjnego szczegółów rozłożono na sumę dwóch składowych. Na przykładzie metody A, składowa θ_A jest czynnikiem systematycznym błędu. Natomiast druga składowa jest czynnikiem losowym błędu, gdzie $\varepsilon_1^A, \varepsilon_2^A, \dots, \varepsilon_n^A$ są niezależnymi zmiennymi losowymi pochodzącymi z rozkładu ciągłego o medianie równej zero. Wówczas θ_A jest nieznaną medianą błędu wyznaczenia położenia sytuacyjnego punktu metodą A (model rozważany w podrozdziale 3.1 oparto na podobnym rozumowaniu, jednakże w odniesieniu do wartości oczekiwanej μ_A). W przypadku metod B, C i D zastosowano analogiczne rozwiązanie.

Rozważono weryfikację hipotezy postaci $\theta_A = \theta_B$, zakładającej, że błędy systematyczne obu metod są takie same. Hipoteza alternatywna orzeka, że składowa systematyczna błędu metody B przewyższa składową systematyczną błędu metody A ($\theta_A < \theta_B$). W weryfikacji hipotez zastosowano test Wilcozona dla dwóch prób niezależnych [26], często nazywany testem sumy rang. Stwierdzono, że p -wartości testów w zakresie wszystkich rozpatrywanych par metod (A-B, A-C, A-D, B-C, B-D, C-D) są bardzo małe - największa z nich, w przypadku weryfikacji hipotezy $\theta_B = \theta_C$, była rzędu 10^{-9} . Pozwoliło to jednoznacznie wybrać lepszą metodę w rozważanych sześciu parach. Wyniki tej analizy przedstawiono w tablicy 2 (a ich interpretacja jest analogiczna do rezultatów umieszczonych w tab. 1).

Tablica 2. Wyniki testów dotyczących składowych systematycznych błędów poszczególnych metod

	A	B	C	D
A		A	A	A
B			B	B
C				C

Następnie wyznaczono przedziały ufności dotyczące składowych systematycznych ($\theta_A, \theta_B, \theta_C$ oraz θ_D) błędów poszczególnych metod. W ich konstrukcji zastosowano metodę znaków, której szczegóły można znaleźć w monografii Hollandera i Wolfe [26]. W metodzie tej najpierw należy posortować zbiór danych, a następnie wybrać elementy stanowiące krańce przedziału ufności przy założonym poziomie ufności. W przeprowadzonej analizie, jako lewy i prawy koniec przedziału ufności (na poziomie ufności 0.999) wybrano elementy o numerach $(n+1-\beta)$ oraz β , gdzie β jest kwantylem rozkładu dwumianowego z parametrami n i 0.5 rzędu 0.9995. W przypadku rozważanych czterech metod, wyznaczenia położenia sytuacyjnego szczegółów zgromadzonych w bazach map cyfrowych, stwierdzono następujące przedziały ufności dla składowych systematycznych błędów poszczególnych metod (na poziomie ufności 0.999):

$$\begin{aligned}\theta_A &\in [0.04\text{ m}; 0.05\text{ m}], \\ \theta_B &\in [0.11\text{ m}; 0.13\text{ m}], \\ \theta_C &\in [0.13\text{ m}; 0.18\text{ m}], \\ \theta_D &\in [0.27\text{ m}; 0.30\text{ m}].\end{aligned}$$

Powyższe wyniki potwierdziły, że najwyższą dokładnością charakteryzuje się mapa cyfrowa wykonana z pomiarów tachimetrem elektronicznym (metoda A). Metoda ta wykazała swą przewagę w trakcie weryfikacji hipotez, a czynnik systematyczny jej błędu oszacowano przedziałem $[0.04\text{ m}, 0.05\text{ m}]$. Wysoką dokładność stwierdzono również w przypadku mapy opracowanej na podstawie wyników wcześniejszych pomiarów bezpośrednich (metoda B), której składowa systematyczna błędu jest nie większa niż 0.13 m. Nieznacznie niższą dokładnością charakteryzują się dane pozyskane w wyniku manualnej wektoryzacji punktów sytuacyjnych na rastrowym obrazie ortofotomapy cyfrowej ($\theta_C \leq 0.18\text{ m}$). Natomiast najniższą dokładność stwierdzono w przypadku mapy wykonanej metodą D. Wykonane testy wykazały największe wielkości składowej systematycznej błędu w przypadku mapy cyfrowej wykonanej poprzez manualną wektoryzację rastrowego obrazu map analogowych, wyznaczony przedział ufności oszacował czynnik systematyczny błędu na poziomie wielkości nie mniejszej niż 0.27 m.

Na końcu przedstawionej analizy dokładności, opracowania sytuacyjnego baz danych map wielkoskalowych, porównano rezultaty uzyskane metodami nieparametrycznymi z wynikami badań dotyczących ustalenia przedziałów ufności i weryfikacji hipotez w zakresie średnich błędów ocenianych baz danych. Analizując przedziały ufności, stwierdzono wyższe wielkości w przypadku oszacowania czynników systematycznych błędów punktów kontrolnych za pomocą średniej arytmetycznej, co wiąże się z jej dużą wrażliwością na wartości odstające, podczas gdy mediana z próby jest odporna na wartości skrajne. Ponadto użycie testów nieparametrycznych, w postaci (opartej na rangach) metody znaków, pozwoliło rozstrzygnąć kwestię porównania dokładności metod B i C. Ustalono mniejszą wielkość czynnika systematycznego błędu w zbiorach wcześniejszych danych pomiarowych ($\theta_B < \theta_C$).

4. Podsumowanie i wnioski końcowe

W niniejszej pracy przedstawiono propozycję automatyzacji analizy dokładności opracowania sytuacyjnego baz danych map wielkoskalowych. Analizę oparto na skonstruowaniu przedziałów ufności oraz parametrycznych i nieparametrycznych testów statystycznych dotyczących błędów metod wyznaczenia położenia szczegółów I grupy dokładnościowej (zgromadzonych w bazach danych). Badania wykonano na pierwotnych zbiorach punktów kontrolnych, zawierających także punkty

wykazujące błędy o wielkościach odstających (znacznie przewyższających błędy większości punktów zgromadzonych w próbie empirycznej). Należy zauważyć, iż zastosowane podejście dostarczyło istotnych informacji na temat dokładności ocenianych map cyfrowych. Albowiem w sposób ogólny uszeregowano dokładność opracowania sytuacyjnego poszczególnych metod pozyskania danych, a ponadto oszacowano charakteryzujące je składowe systematyczne błędów.

Literatura

- [1] JARZĄBEK J., Stan prac legislacyjnych Głównego Urzędu Geodezji i Kartografii, *Prezentacja na XIII Konferencji poświęconej ODGiK pt. Modelowania geodezji ciąg dalszy*, Elbląg, 28-29 kwietnia 2011.
- [2] SEJM RP, Ustawa z dnia 4 marca 2010 roku o infrastrukturze informacji przestrzennej (Dz. U. nr 76 poz. 489 z 2010 r.), *Ustawa dokonująca transpozycji dyrektywy Parlamentu Europejskiego i Rady nr 2007/2/WE z dnia 14 marca 2007 r. ustanawiającej infrastrukturę informacji przestrzennej we Wspólnocie Europejskiej (INSPIRE)*, Dz. Urz. UE L 108 z 25.04.2007, z późn. zm.
- [3] IWANIAK A., Infrastruktura informacji geodezyjnej i kartograficznej, *Roczniki Geomatyki* Tom V, Zeszyt 6, 2007, s. 63-71.
- [4] PIIT, Uzasadnienie projektu ustawy o infrastrukturze informacji przestrzennej, *Polska Izba Informatyki i Telekomunikacji*, 2008, 8 s., http://www.piit.org.pl/_gAllery/55/35/5535/Uzasadnienie_12.08.2008.pdf
- [5] PACHELSKI W., Działalność normalizacyjna w dziedzinie informacji geograficznej, *VIII Krajowa Konferencja Geodezyjna pt. Jakość, standaryzacja, normalizacja w geodezji i kartografii*, Pogorzela, 3-5 października 2002, http://www.geodezja-szczecin.org.pl/stara_strona/Konferencje/Konf2002/k09.html
- [6] DOSKOCZ A., Badanie dokładności wielkoskalowych map numerycznych wykonanych różnymi metodami. *Rozprawa doktorska*, UWM w Olsztynie 2002.
- [7] GAŹDZICKI J., Geoinformacja w samorządzie terytorialnym, *Referat wprowadzający Konferencji pt. GIS szansą rozwoju samorządów lokalnych*, Warszawa, 21-22 listopada 2006, http://forumgis.pl/images/content/geoinf_w_samorz_teryt.pdf
- [8] DĄBROWSKI W., DOSKOCZ A., MRÓWCZYŃSKI T., Miasto doceniło ortofoto. *Magazyn Geoinformacyjny GEODETA* nr 1/2007, s. 26-28.
- [9] BRZUCHOWSKA J., GIS w praktyce planistycznej - na przykładzie miasta Wrocław, *Warsztaty Instytutu Gospodarki Przestrzennej i Mieszkalnictwa*, Warszawa 2003, http://www.igpik.waw.pl/pdf/brzuchowska1_pl.pdf
- [10] GAŹDZICKI J., Leksykon geomatyczny, *Polskie Towarzystwo Informacji Przestrzennej*, Warszawa 2002, <http://www.ptip.org.pl>
- [11] GUPTILL S.C., MORRISON J.L. (ed.), Elements of spatial data quality, Opublikowane w imieniu Międzynarodowej Asocjacji Kartograficznej przez Elsevier Science, 1995.
- [12] RÖNSDORF C., Positional integration of geodata. Positional accuracy improvement: Impacts of improving the positional accuracy of GI databases, Related Papers - *EuroSDR Publication* No. 48/2004, http://www.euroedr.net/km_pub/no48/html/positional/related_papers.htm
- [13] LÓPEZ F.J.A., GORDO A.D.A., Analysis of Some Positional Accuracy Assessment Methodologies, *Journal of Surveying Engineering* No. 134(2)/2008, s. 45-54.
- [14] ZANDBERGEN P.A., Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy. *Transactions in GIS* No. 12(1)/2008, s. 103-130.
- [15] ZAKAREVIČIUS A., JONAUSKIENĖ I., Opportunity to use statistical control for quality assessment of land parcels cadastral data, (in Lithuanian), *Geodesy and Cartography* No. 33(4)/2007, s. 115-119.
- [16] BOGAERT P., DELINCÉ J., KAY S., Assessing the error of polygonal area measurements: a general formulation with applications to agriculture, *Measurement Science and Technology* No. 16(5)/2005, s. 1170-1178.

- [17] CROITORU A., DOYTSHER Y., Accounting for Discontinuities in Cadastral Data Accuracy: Toward a Patch Based Approach, *Paper of FIG Working Week*, Paris, France, 2003, http://www.fig.net/pub/fig_2003/TS_15/TS15_4_Croitoru_Doytsher.pdf
- [18] De BRUIN S., Modelling Positional Uncertainty of Line Features by Accounting for Stochastic Deviations from Straight Line Segments, *Transactions in GIS* No. 12(2)/2008, s. 165-177.
- [19] De BRUIN S., HEUVELINK G.B.M., BROWN J.D., 2008, Propagation of positional measurement errors to agricultural field boundaries and associated costs, *Computers and Electronics in Agriculture* No. 63(2)/2008, s. 245-256.
- [20] MAKOWSKI A., Standaryzacja we współczesnej kartografii, *VI Krajowa Konferencja Geodezyjna pt. Jakość i standaryzacja w geodezji i kartografii*, Pogorzela 28-30 września 2000, http://www.geodezja-szczecin.org.pl/stara_strona/Konferencje/Konf2000/k08.html
- [21] PACHELSKI W., PARZYŃSKI Z., Modele pojęciowe niektórych geodezyjnych składników infrastruktury danych przestrzennych, *Acta Sci. Pol. Geod. Descr. Terr.* No. 6(4)/2007, s. 23-37.
- [22] GUGiK, Opublikowane normy Polskiego Komitetu Normalizacyjnego w ramach prac Komitetu Technicznego nr 297 ds. Informacji geograficznej, *Serwis informacyjny Głównego Urzędu Geodezji i Kartografii*, 2005.
- [23] PN-EN ISO 19113:2005 Informacja geograficzna - Podstawy opisu jakości.
- [24] PN-EN ISO 19114:2005 Informacja geograficzna - Procedury oceny jakości.
- [25] Instrukcja techniczna O-2, Ogólne zasady opracowania map dla celów gospodarczych, *Główny Urząd Geodezji i Kartografii*, Warszawa 1987.
- [26] HOLLANDER M., WOLFE D. A., Nonparametric statistical methods, Wiley, New York 1999.
- [27] KORONACKI J., MIELNICZUK J., Statystyka dla studentów kierunków technicznych i przyrodniczych, WNT, Warszawa 2001.
- [28] DĄBROWSKI W., DOSKOCZ A., Dokładność sytuacyjnych danych numerycznych pozyskanych różnymi metodami, *Materiały I Ogólnopolskiej Konferencji Naukowo-Technicznej pt. Kartografia numeryczna i informatyka geodezyjna*, s. 29-38, Rzeszów - Polańczyk 2005.
- [29] DOSKOCZ A., The use of statistical analysis for estimation of positional accuracy of large-scale digital maps, *Geodezja i Kartografia*, Vol. 54, No 3, 2005, s. 131-150.
- [30] R DEVELOPMENT CORE TEAM, R: A Language and Environment for Statistical Computing, Vienna 2009, <http://www.R-project.org>

PROPOSITION OF AUTOMATIZATION OF ANALYSIS OF ACCURACY OF THE LARGE-SCALE DIGITAL MAPS DATABASES

Summary

The paper presents proposition of automatization of estimation of positional accuracy of large-scale maps data. The research was realized by means 4th various methods of producing digital map databases: new total station survey, re-calculation of previous direct measurements realised by orthogonal and polar surveys, manual vectorisation of a raster orthophotomap image and graphical-and-digital processing of analogue maps. Obtained results of accuracy estimation using proposed statistical analysis are coherent with conclusions developed on the basis of classical estimation of accuracy (in reference to the requirements for technical standards). It was confirmed high accuracy of the digital map produced on the basis of survey with an electronic tacheometer. The accuracy of a digital map produced on the bases of the past field surveys is also relatively satisfactory. The analysis indicated the high accuracy (in relation to well identified details of the 1st group, i.e. inspection chambers over underground installations) of a digital orthophotomap. The lowest accuracy was confirmed for a digital map produced by means of graphical-and-digital processing of analogue maps (by vectorization of raster maps).

According with the authors opinion, it is desirable to establish procedures of automatization of estimation of accuracy of digital maps and its application in the appropriate expert system. This will ensure the efficient verification of accuracy collected of large-scale digital maps databases in terms of expectations of their users and provide the characteristics to enable the creation of an extended set of metadata.

Key-words: database, large-scale map, analysis of accuracy