

# **AUTOREFERAT**

przedstawiający osiągnięcia w pracy naukowo – badawczej

dr inż. Marek Ślusarski

Kraków 2017

## Spis treści

1. Imię i nazwisko .....	3
2. Posiadane dyplomy, stopnie naukowe .....	3
3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych .....	3
4. Wskazanie osiągnięcia będącego podstawą do ubiegania się o stopień naukowy doktora habilitowanego .....	4
4.1. Tytuł osiągnięcia naukowego .....	4
4.2. Autor oraz tytuł publikacji .....	4
4.3. Omówienie celu naukowego i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.....	4
4.3.1. Wprowadzenie .....	4
4.3.2. Metodyka oceny dokładności danych gromadzonych w urzędowych bazach danych przestrzennych .....	5
4.3.3. Niepewność pola powierzchni działek ewidencyjnych .....	11
4.3.4. Analiza ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu .....	14
4.3.5. Model jakości danych gromadzonych w urzędowych bazach .....	21
4.3.6. Wizualizacja zjawiska niepewności w danych przestrzennych .....	28
4.3.7. Podsumowanie i wnioski końcowe .....	36
Literatura .....	38
5. Omówienie pozostałych osiągnięć naukowo – badawczych .....	39

## **1. Imię i nazwisko**

MAREK ŚLUSARSKI

## **2. Posiadane dyplomy, stopnie naukowe – z podaniem nazwy, miejsca i roku ich uzyskania oraz tytułu rozprawy doktorskiej**

- 1992 Tytuł zawodowy magistra inżyniera w dyscyplinie geodezja i kartografia, specjalność geodezja rolna, uzyskany na Wydziale Inżynierii Środowiska i Geodezji, Akademii Rolniczej w Krakowie (od 2008 r. Uniwersytet Rolniczy w Krakowie).  
Praca magisterska pt. "Zastosowanie precyzyjnej niwelacji trygonometrycznej i metody biegunowej do badania deformacji przestrzennych osuwisk".
- 2002 Stopień naukowy doktora nauk rolniczych w zakresie kształtowania środowiska, nadany przez Radę Wydziału Inżynierii Środowiska i Geodezji, Akademii Rolniczej w Krakowie.  
Tytuł rozprawy doktorskiej: "Metodyka pozyskiwania i doboru danych w celu utworzenia jednolitego systemu informacji o nieruchomościach".  
Promotor: prof. dr hab. inż. Mirosław Żak  
Recenzenci: prof. dr hab. inż. Karol Noga  
prof. dr hab. inż. Józef Czaja

## **3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych**

- 1992 – 2002 asystent w Katedrze Geodezji na Wydziale Inżynierii Środowiska i Geodezji Akademii Rolniczej w Krakowie
- 2002 – 2014 adiunkt w Katedrze Geodezji na Wydziale Inżynierii Środowiska i Geodezji Akademii Rolniczej w Krakowie (od 2008 r. Uniwersytet Rolniczy w Krakowie)
- od 2014 asystent w Katedrze Geodezji na Wydziale Inżynierii Środowiska i Geodezji Uniwersytetu Rolniczego w Krakowie

#### **4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.):**

##### **4.1. Tytuł osiągnięcia naukowego**

Osiągnięciem naukowym wynikającym z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki jest monografia pt. „Metody i modele oceny jakości danych przestrzennych”, wydana przez Wydawnictwo Uniwersytetu Rolniczego w Krakowie.

##### **4.2. Autor oraz tytuł publikacji**

Ślusarski M. 2017. Metody i modele oceny jakości danych przestrzennych. Zeszyty Naukowe Uniwersytetu Rolniczego w Krakowie, nr 537. Rozprawy z. 414, s. 207.

Recenzenci: prof. dr hab. inż. Karol Noga,

dr hab. inż. Piotr Parzych, prof. AGH.

##### **4.3. Omówienie celu naukowego i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania**

###### **4.3.1. Wprowadzenie**

Bazy danych przestrzennych gromadzą informacje o obiektach geograficznych, będących reprezentacją zjawisk świata rzeczywistego. Przechowywane i udostępniane zbiory danych nie są jednorodne, ponieważ pozyskiwane są na podstawie zróżnicowanych materiałów źródłowych. Rozproszone źródła danych przestrzennych charakteryzują się odmienną organizacją w zakresie kompletności, spójności oraz dokładności. Obecnie gwałtownie rośnie: dostępność, wymiana i wykorzystanie danych przestrzennych (np. wdrażanie dyrektywy INSPIRE, dzięki której tworzone są cyfrowe, urzędowe bazy) oraz zwiększa się dystans pomiędzy wytwórcami danych a użytkownikami danych; jakość danych nie jest jednak gwarantowana.

Jakość danych jest realnym, ogólnoswiatowym problemem (Deren i. in. 2012). Brakuje rozwiązań dotyczących opisu i raportowania jakości zestawów danych, gromadzących informacje o najwyższym stopniu szczegółowości. Do takich baz należą głównie rejestry urzędowe, które powinny charakteryzować się najwyższym stopniem zaufania. Kataster nieruchomości gromadzi dane o położeniu punktów granicznych i polu powierzchni działek ewidencyjnych; informacje te mają wymierne znaczenie podczas realizowania transakcji kupna-sprzedaży. Kataster sieci uzbrojenia terenu odpowiada za wiarygodność informacji dotyczących podziemnych rur i kabli; rzetelność tych danych wpływa bezpośrednio na bezpieczeństwo realizowanych zamierzeń budowlanych.

Przeprowadzone przez autora prace studialne i badawcze pozwoliły na postawienie następującej tezy: *Szybko rosnąca dostępność, wymiana i wykorzystanie danych przestrzennych oraz ich niejednorodność, powodują konieczność poszukiwania skutecznych metod i modeli oceny ich jakości.* Realizowane badania posłużyły do weryfikacji pięciu hipotez badawczych:

1. Niejednorodność źródeł pochodzenia danych przestrzennych powoduje, że analizy dokładności tych danych powinny być realizowane za pomocą metod odpornych na obserwacje odstające.
2. Z powodu błędów pomiarowych technik geodezyjnych, dokładność określenia pola powierzchni działek ewidencyjnych – odpowiadająca precyzji zapisu wielkości pola ( $1 \text{ m}^2$ ) – nie jest gwarantowana. Z związku z tym należy poszukiwać metod pozwalających na wyznaczenie wskaźników bezwzględnej niepewności pól powierzchni tych działek.

3. W ocenie ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu, którego źródłem jest jakość danych przestrzennych, należy zastosować techniki pozwalające na wyznaczenie poziomu ryzyka ilościowego i jakościowego oraz prezentacji kartograficznej tego ryzyka.
4. Model jakości danych przestrzennych jest zbiorem zasad standardowego podejścia do oceny i raportowania jakości tych danych. Model ten powinien być czytelny i zrozumiały, zarówno dla twórców zbiorów danych, jak i ich użytkowników.
5. Skutecznym sposobem przekazywania użytkownikom zbiorów danych przestrzennych informacji o ich niepewności mogą być techniki wizualizacji kartograficznej.

Głównym celem badań jest udowodnienie postawionej tezy poprzez opracowanie metod i modeli opisu oraz oceny jakości danych przestrzennych, gromadzonych w rejestrach publicznych. W badaniach zaproponowano oryginalne rozwiązania analityczne, dotyczące wybranych aspektów działania w zakresie jakości danych przestrzennych baz wielkoskalowych. Do opracowanych rozwiązań należą:

- odporna metoda oceny dokładności danych przestrzennych;
- procedura określania niepewności pola powierzchni działek ewidencyjnych;
- model i mapa ryzyka uszkodzenia sieci uzbrojenia podziemnego terenu;
- model procesu kontroli jakości danych;
- zbiór technik wizualizacji zjawiska niepewności w danych przestrzennych.

Rezultaty prowadzonych badań dotyczą baz przestrzennych gromadzących dane w najwyższym stopniu szczegółowości: ewidencji gruntów i budynków (EGiB), geodezyjnej ewidencji sieci uzbrojenia terenu (GESUT) oraz bazy danych obiektów topograficznych (BDOT500).

#### **4.3.2. Metodyka oceny dokładności danych gromadzonych w urzędowych bazach danych przestrzennych**

Dokładność położenia jest jednym z podstawowych wyznaczników jakości danych przestrzennych, który bezpośrednio wpływa na wartość wykonywanych analiz geoprzestrzennych. Podstawowym podejściem do analizy dokładności położenia obiektów baz danych przestrzennych jest metoda oparta na reprezentatywnych zbiorach współrzędnych punktów kontrolnych, służących do wyznaczenia wartości ich błędów. Metoda ta jest podstawą analiz prezentowanych w wielu pracach badawczych (De Bruin i in. 2008; Zandbergen 2008; Ratcliffe 2001; Liao i in. 2010; Ayhan i in. 2011; Duskocz 2013). Zastosowana jest również w krajowych i międzynarodowych standardach technicznych (FGDC 1998; ISO 2013).

Drugim podejściem do problemu analizy dokładności danych przestrzennych są badania dokładności położenia obiektów liniowych i powierzchniowych. W podejściu tym zaproponowano zastosowanie technik buforowania (Tveite i Langaas 1999; Seo i O'Hara 2009), złożonych technik symulacji stochastycznej (Zhang i Kirby 2000; Shi i Liu 2000) oraz geometrii fraktalnej (Duckham i Drummond 2000; Mozas i Ariza 2011). Opracowana została również metoda badania dokładności linii i powierzchni, oparta na wielowymiarowym rozkładzie normalnym (Zhang i Kirby 2000).

Prezentowane w pracy badania dokładności dotyczyły danych gromadzonych w urzędowych bazach: EGIB, GESUT i BDOT500. W bazach tych badano dane pochodzące z pięciu źródeł:

- geodezyjne pomiary terenowe (GPT);
- ortofotomapa (ORTO);
- mapa zasadnicza w skali 1:500 (M-500);
- mapa zasadnicza w skali 1:1000 (M-1000);
- mapa zasadnicza w skali 1:2000 (M-2000).

Reprezentatywną próbę badawczą stanowiły odchyłki współrzędnych  $dx$ ,  $dy$  oraz wektor  $Dl$  – przesunięcie położenia punktu próby badawczej względem położenia uznanego za bezbłędne. W zakresie jednego źródła danych wykorzystano cztery próby losowe o liczebności  $n = 500$  każda,

co oznacza, że po sprawdzeniu czy próby pochodzą z jednorodnych populacji, liczebność każdej próby wynosiła 2000 elementów. Analizę dokładności wykonano dwoma podejściami: pierwsze – na podstawie uciętych prób (zbiorów danych), z których usunięto obserwacje odstające, i drugi – z zastosowaniem statystyki odpornościowej.

Analizując dokładność danych na podstawie uciętych prób za pomocą parametrów i testów statystycznych, można wnioskować, że rozkład odchyłek opisujących dokładności danych przestrzennych pochodzących z geodezyjnych pomiarów terenowych i ortofotomapy, zbliżony jest do rozkładu log-normalnego. Rozkład błędów położenia obiektów pochodzących z map w skalach 1:500, 1:1000 i 1:2000 jest zbliżony do rozkładu Weibulla.

Stopień dopasowania rozkładów badano analizując ich dystrybuanty. Poziom dopasowania dystrybuanty empirycznej z dystrybuantą teoretycznego rozkładu oceniono za pomocą dwóch wskaźników: błędu średniego kwadratowego RMSE oraz błędu przeciętnego. Wartości obydwu wskaźników dopasowania dla wszystkich analizowanych przypadków zawierają się w przedziale od 0,02 do 0,03.

Wartości statystyk testowych oraz wielkości miar dopasowania rozkładu empirycznego do teoretycznego pokazują, że rozkład błędów opisujących dokładności danych pozyskanych na podstawie geodezyjnych pomiarów terenowych i ortofotomapy, najbliższy jest rozkładowi log-normalnemu. Realizując założenie o log-normalnym rozkładzie zmiennej  $Dl$ , obliczono dla prób badawczych estymatory skupienia i rozproszenia. Wyznaczoną dokładność porównano z dokładnością szacowaną, przy założeniu rozkładu normalnego (tab. 1).

Tabela 1. Wyznaczona dokładność danych przestrzennych według rozkładów log-normalnego i normalnego

Estymator	Źródło danych przestrzennych	
	GPT	ORTO
Rozkład log-normalny		
Mediana $Med_{Dl}$ [m]	0,16	0,23
Rozproszenie względne [m]	$0,11 < Dl < 0,24$	$0,16 < Dl < 0,34$
Rozkład normalny		
Wartość przeciętna ( $\hat{\mu}$ ) [m]	0,18	0,26
Współczynnik rozproszenia ( $\hat{\sigma}$ ) [m]	0,07	0,11

Wartości liczbowe estymatorów skupienia pokazują, że obliczenia na podstawie rozkładu normalnego prowadzą do przeszacowania dokładności położenia tych obiektów przestrzennych, które pochodzą z geodezyjnych pomiarów terenowych i ortofotomapy. Wielkość przeszacowania dokładności położenia w oparciu o rozkład normalny (dla obydwu przypadków: GPT i ORTO) wynosi ok. 13%. Otrzymane wartości świadczą o zasadności stosowania metody szacowania dokładności danych przestrzennych, opartej na własnościach rozkładu log-normalnego.

Drugie podejście do szacowania dokładności danych oparto na zastosowaniu statystyki odpornościowej, rozumianej jako odpornej na obserwacje odstające. W analizach dokładność danych przestrzennych próby nie zawierają wielokrotnych obserwacji pomiarowych tego samego obiektu. Poza oczywistymi pomyłkami, każda obserwacja pomiarowa jest źródłem informacji o dokładności zbioru danych. Dlatego szacowanie wielkości charakteryzujących dokładność takich danych powinno być oparte na metodach, w których modyfikowane są elementy próby statystycznej, a nie odrzucane obserwacje uznane za odstające.

W badaniach dokładności danych przestrzennych zastosowano cztery metody odporne na obserwacje odstające (trzy techniki zmodyfikowane oraz jedną autorską):

- metoda najmniejszych odchyłeń absolutnych (zmodyfikowana);
- metoda duńska (zmodyfikowana);
- ZWA (zasada wyboru alternatywy – zmodyfikowana);
- odporna metoda częstości względnej – rozwiązanie opracowane przez autora monografii.

W metodzie najmniejszych odchyłeń absolutnych, podobnie jak w pozostałych metodach, modyfikacja polegała na tym, że do diagnozowania obserwacji odstających użyto mediany ( $Med$ ) i znormalizowanego odchylenia bezwzględnego mediany ( $NMAD$ ) – zamiast wartości średniej i odchylenia standardowego. Zasadę przekształcania obserwacji prezentuje wzór:

$$x_i^* = \begin{cases} x_i & \text{dla } |x_i - Med| \leq c_1 \cdot NMAD \\ Med - c_1 \cdot NMAD & \text{dla } x_i < (Med - c_1 \cdot NMAD) \\ Med + c_1 \cdot NMAD & \text{dla } x_i > (Med + c_1 \cdot NMAD) \end{cases} \quad (1)$$

gdzie:  $x_i$  – element próby losowej,  $i = 1, \dots, n$

$Med$  – mediana,

$NMAD$  – znormalizowane odchylenie bezwzględne mediany,

$NMAD = 1,4826 \cdot \text{mediana}(|x_i - Med|)$ ,

$c_1$  – stała z przedziału  $\langle 1, 2 \rangle$ , zależna od frakcji obserwacji odstających rozkładu zmiennych losowych.

Wyznaczono odporny estymator wartości oczekiwanej parametr  $x_i^*$  (dla każdego źródła danych), który charakteryzuje dokładność położenia. Obliczone tą metodą wartości porównano z wartościami wyznaczonymi na podstawie uciętych prób losowych (tab. 2). Przyjęto założenie, że porównawczym kryterium odniesienia (dla możliwości zastosowania tej metody) jest statystyka klasyczna oparta na dużych ( $n > 500$ ), uciętych próbach losowych. Stwierdzono, że istnieje pełna zgodność w szacowaniu dokładności dla danych pochodzących z GPT, ORTO oraz M-500. Dla mniej dokładnych źródeł (M-1000 i M-2000) metoda najmniejszych odchyłeń absolutnych przeszacowuje wyniki od ok. 10% do 20%.

Tabela 2. Estymatory dokładności położenia obiektów przestrzennej bazy danych, szacowane metodą najmniejszych odchyłeń absolutnych

Estymatory	Źródło danych przestrzennych				
	GPT	ORTO	M-500	M-1000	M-2000
Statystyka odpornościowa (metoda najmniejszych odchyłeń absolutnych)					
Liczebność próby ( $n$ )	2000	2000	2000	2000	2000
Odporna wartość oczekiwana ( $x_i^*$ )	0,18	0,27	0,41	0,70	1,38
Odporny współczynnik rozproszenia ( $s_i^*$ )	0,07	0,13	0,17	0,28	0,55
Statystyka klasyczna (ucięte zbiory danych)					
Liczebność próby ( $n$ )	1893	1837	1807	1793	1848
Wartość przeciętna ( $\hat{\mu}$ )	0,18	0,26	0,39	0,64	1,16
Współczynnik rozproszenia ( $\hat{\sigma}$ )	0,07	0,11	0,12	0,21	0,31

W drugiej badanej technice – metodzie duńskiej – użyto mediany i odchylenia bezwzględnego mediany do diagnozowania obserwacji odstających oraz wyznaczono wartości parametrów sterujących. W celu wykorzystania optymalnej funkcji do badań dokładności danych przestrzennych, autor pracy analizował i testował różne warianty funkcji tłumienia, stosowane w metodzie duńskiej (Krarup, Kubik 1983). Ostatecznie zastosowano funkcję zapewniającą zbieżność procesu iteracyjnego do punktu  $\hat{X}$  oraz realizację procesu obliczeniowego w maksymalnie czterech cyklach iteracyjnych. Funkcję tą można określić wzorem:

$$f(u^*) = \begin{cases} 1 & \text{dla } u^* \in \langle -c_3; c_3 \rangle \\ \exp[-l(|u^*| - c_3)^g] & \text{dla } |u^*| > c_3 \end{cases} \quad (2)$$

gdzie:  $u^*$  – odporny estymator diagnozowania obserwacji odstających:

$$u_i^* = \frac{x_i - \text{Med}(x)}{\text{MAD}(x)}$$

$l$  – parametr sterujący ( $l = 0,1$ ),

$c_3$  – stała ( $c_3 = 3$ ),

$g$  – parametr sterujący ( $g = 3$ ).

Analogicznie jak w poprzedniej metodzie, obliczono odporne wartości oczekiwane  $X^*$ , które porównano z wartościami wyznaczonymi na podstawie uciętych dużych prób losowych (tab. 3). Stwierdzono, że istnieje pełna zgodność w szacowaniu dokładności danych dla źródeł: GPT, ORTO, M-500, M-1000. Dla M-2000 natomiast, zmodyfikowana metoda duńska przeszacowuje wyniki o ok. 20 %.

Tabela 3. Estymatory dokładności położenia obiektów przestrzennej bazy danych, szacowane metodą duńską

Estymatory	Źródło danych przestrzennych				
	GPT	ORTO	M-500	M-1000	M-2000
Statystyka odpornościowa (metoda duńska)					
Liczebność próby ( $n$ )	2000	2000	2000	2000	2000
Odporna wartość oczekiwana $X^*$	0,17	0,28	0,41	0,62	1,39
Odporny współczynnik rozproszenia ( $s_i^*$ )	0,06	0,12	0,15	0,22	0,51
Statystyka klasyczna (ucięte zbiory danych)					
Liczebność próby ( $n$ )	1893	1837	1807	1793	1848
Wartość przeciętna ( $\hat{\mu}$ )	0,18	0,26	0,39	0,64	1,16
Współczynnik rozproszenia ( $\hat{\sigma}$ )	0,07	0,11	0,12	0,21	0,31

Trzecim analizowanym podejściem była zasada wyboru alternatywy (ZWA) (Kadaj 1980). W analizie dokładności danych przestrzennych tą metodą, w celu zapewnienia zbieżności procesu iteracyjnego do punktu  $\hat{X}$ , zastąpiono wartości średniej i odchylenia standardowego odpornymi miarami położenia i rozrzutu: medianą oraz odchyleniem bezwzględnym mediany. Zastosowano funkcję  $p^*(v)$ , modyfikującą – w różnym stopniu – każdą wagę, w kolejnych cyklach iteracyjnych. Funkcja wyraża się wzorem:

$$p^*(v_i) = \exp\left(-\frac{[x_i - \text{Med}(x)]^2}{2 \cdot [\text{MAD}(x)]^2}\right) \quad (3)$$

Analogicznie jak w poprzednich metodach, obliczono odporne wartości oczekiwane  $X^*$ , które porównano z wartościami wyznaczonymi na podstawie uciętych dużych prób losowych (tab. 4). W metodzie ZWA, analizując odporną wartość oczekiwaną stwierdzono, że istnieje pełna zgodność w szacowaniu dokładności dla GPT, ORTO, M-500, a dla mniej dokładnych źródeł (M-1000 i M-2000) metoda ZWA przeszacowuje wyniki o ok. 10%.



Tabela 4. Estymatory dokładności położenia obiektów przestrzennej bazy danych, szacowane metodą ZWA

Estymatory	Źródło danych przestrzennych				
	GPT	ORTO	M-500	M-1000	M-2000
Statystyka odpornościowa (metoda ZWA)					
Liczebność próby ( $n$ )	2000	2000	2000	2000	2000
Odporna wartość oczekiwana ( $\hat{X}$ )	0,17	0,24	0,40	0,65	1,27
Odporny współczynnik rozproszenia ( $s_i^*$ )	0,05	0,08	0,10	0,14	0,29
Statystyka klasyczna (ucięte zbiory danych)					
Liczebność próby ( $n$ )	1893	1837	1807	1793	1848
Wartość przeciętna ( $\hat{\mu}$ )	0,18	0,26	0,39	0,64	1,16
Współczynnik rozproszenia ( $\hat{\sigma}$ )	0,07	0,11	0,12	0,21	0,31

Ostatnim podejściem do analizy dokładności danych przestrzennych jest „odporna metoda częstości względnej” – rozwiązanie opracowane przez autora pracy. Przyjęto założenie, że wszystkie dane źródłowe będą w różnym stopniu modyfikowane. Zakres modyfikacji danych będzie zależny od prawdopodobieństwa, że zmienna  $Dl$  przyjmie wartość z pewnego zbioru. Zmienna  $Dl$  jest długością wektora przesunięcia punktu próby badawczej względem położenia punktu referencyjnego. Prawdopodobieństwo występowania zmiennej  $Dl$  opisuje krzywa częstości względnej skumulowanej empirycznego rozkładu tej zmiennej  $F_n(Dl)$ . Krzywą częstości można aproksymować za pomocą innej krzywej, należącej do określonej klasy funkcji.

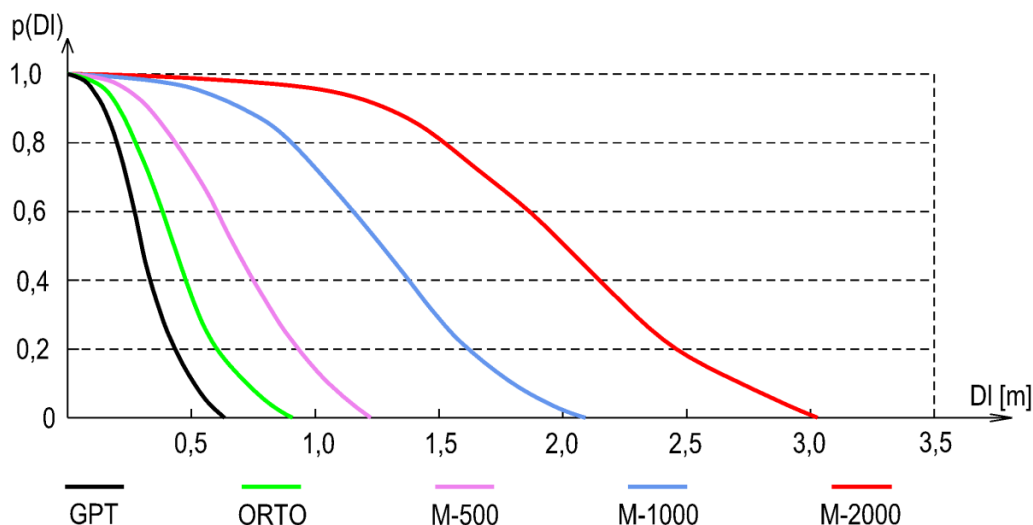
Zastosowano tutaj aproksymację krzywej częstości względnej skumulowanej ( $F_n$ ) empirycznego rozkładu zmiennej ( $Dl$ ) w postaci:  $1 - F_n(Dl)$ . Procedurę aproksymacji krzywej  $1 - F_n(Dl)$  zrealizowano metodą analityczną, poprzez wykorzystanie algorytmu Bermiana oraz zasady minimalizacji Gaussa-Seidla (Atkinson i Han 2009; Niemi 2013). Opracowaną przez autora pracy funkcję aproksymacyjną  $p(Dl)$  przedstawia poniższe równanie:

$$p(Dl) = e^{-l \cdot \left(\frac{Dl}{2 \cdot MAD \cdot k}\right)^p} \quad (4)$$

gdzie:  $MAD$  – odchylenie bezwzględne mediany,

$l, k, p$  – parametry sterujące ( $l = 0,88, k = 2,54, p = 2,12$ ) dla GPT; ( $l = 0,75, k = 2,54, p = 2,32$ ) dla ORTO; ( $l = 0,50, k = 2,54, p = 2,52$ ) dla M-500; ( $l = 0,45, k = 2,54, p = 3,40$ ) dla M-1000; ( $l = 1,34, k = 2,54, p = 4,25$ ) dla M-2000.

Wykres funkcji aproksymacyjnej dla odpornej metody częstości względnej – z pięciu źródeł pochodzenia danych przestrzennych – prezentuje rysunek 1.



Rys. 1. Funkcja wagowa dla odpornej metody częstości względnej ( $MAD = 0,07$  dla GPT,  $0,09$  – ORTO,  $0,12$  – M-500,  $0,23$  – M-1000,  $0,50$  – M-2000)

Błąd średniokwadratowy RMSE zastosowano jako miarę dopasowania funkcji aproksymacyjnej  $p(Dl)$  do krzywej częstości względnej w postaci  $1 - Fn(Dl)$ . Obliczone wartości błędu średniokwadratowego aproksymacji, dla danych pochodzących z GPT wynoszą: 0,024, dla ORTO: 0,028, dla M-500: 0,019, dla M-1000: 0,026 oraz dla M-2000: 0,029. Wartości błędów średnich pokazują, że funkcja aproksymacyjna  $p(Dl)$  dopasowana jest do krzywej częstości względnej z precyzją wystarczającą dla celów badania dokładności danych przestrzennych.

W procesie szacowania dokładności danych przestrzennych odporną metodą częstości względnej, funkcja  $p(Dl)$  jest funkcją wagową, która modyfikuje wielkości wektorów przesunięcia punktów próby badawczej ( $Dl$ ). Ekspotencjalna funkcja  $p(Dl)$  powoduje modyfikację – w różnym stopniu – każdej wartości wektorów  $Dl$ , według wzoru:

$$Dl_i^* = Dl_i \cdot p(Dl_i) \quad (5)$$

Następnie wyznaczany jest wskaźnik  $\widehat{Dl}^*$ , według wzoru:

$$\widehat{Dl}^* = \frac{\sum_{i=1}^n Dl_i^*}{\sum_{i=1}^n p(Dl_i)} \quad (6)$$

Współczynnik  $\widehat{Dl}^*$  jest odpornym na obserwacje odstające estymatorem wartości oczekiwanej, opisującym podstawowy element jakości danych, czyli dokładność położenia.

Analizę dokładności danych przestrzennych gromadzonych w bazach EGiB, GESUT i BDOT500, zrealizowano odporną metodą częstości względnej, poprzez oszacowanie dokładności położenia badanych obiektów. Obliczono parametr  $\widehat{Dl}^*$ , który jest odpornym estymatorem wartości oczekiwanej dokładności położenia. Podobnie jak w poprzednich metodach, przyjęto założenie, że porównawczym kryterium odniesienia dla możliwości zastosowania proponowanej metody odpornej jest statystyka klasyczna, oparta na dużych uciętych próbach losowych. Wartości liczbowe estymatorów pokazują (tab. 5), że istnieje pełna zgodność w szacowaniu dokładności położenia obiektów przestrzennych, pochodzących z pięciu analizowanych źródeł danych przestrzennych.

Tabela 5. Estymatory dokładności położenia obiektów przestrzennej bazy danych, szacowane odporną metodą częstości względnej

Estymatory	Źródło danych przestrzennych				
	GPT	ORTO	M-500	M-1000	M-2000
Statystyka odpornościowa (odporna metoda częstości względnej)					
Liczebność próby ( $n$ )	2000	2000	2000	2000	2000
Odporna wartość oczekiwana ( $\widehat{Dl}^*$ )	0,17	0,25	0,38	0,63	1,17
Odporny współczynnik rozproszenia ( $s_i^*$ )	0,05	0,08	0,10	0,15	0,28
Statystyka klasyczna (ucięte zbiory danych)					
Liczebność próby ( $n$ )	1893	1837	1807	1793	1848
Wartość przeciętna ( $\hat{\mu}$ )	0,18	0,26	0,39	0,64	1,16
Współczynnik rozproszenia ( $\hat{\sigma}$ )	0,07	0,11	0,12	0,21	0,31

Wykonano również dodatkowe testy badawcze, dotyczące oceny poprawności szacowania dokładności położenia obiektów przestrzennych za pomocą opracowanej przez autora metody odpornej na obserwacje odstające. Podstawą analizy były dwa zbiory danych kontrolnych (A i B) – dla dwóch obrębów ewidencyjnych, pochodzące z terenów zurbanizowanych.

Wykorzystując opracowaną przez autora funkcję wagową (4), obliczono dla zbiorów kontrolnych A i B współczynnik  $\widehat{Dl}^*$ , który jest odpornym estymatorem wartości oczekiwanej dokładności położenia (tab. 6). Wyznaczone wielkości odpornej wartości oczekiwanej ( $\widehat{Dl}^*$ ) oraz wartości przeciętnej ( $\hat{\mu}$ ) są podobne; różnice nie przekraczają 5%. Przedstawione wyniki testów badawczych, zrealizowanych na próbach kontrolnych, potwierdzają poprawność szacowania dokładności położenia obiektów przestrzennych za pomocą zaproponowanej odpornej metody częstości względnej.

Tabela 6. Estymatory dokładności położenia obiektów zbiorów kontrolnych (A, B), szacowane odporną metodą częstości względnej

Estymatory	Zbiory kontrolne	Źródło danych przestrzennych				
		GPT	ORTO	M-500	M-1000	M-2000
Statystyka odpornościowa (odporna metoda częstości względnej)						
Liczebność próby ( $n$ )	A	150	150	150	150	150
	B	150	150	150	150	150
Odporna wartość oczekiwana ( $\widehat{Dl}^*$ )	A	0,16	0,23	0,35	0,63	1,25
	B	0,14	0,22	0,36	0,53	1,15
Odporny współczynnik rozproszenia ( $s_i^*$ )	A	0,05	0,07	0,09	0,14	0,29
	B	0,05	0,07	0,09	0,12	0,27
Statystyka klasyczna (ucięte zbiory danych)						
Liczebność próby ( $n$ )	A	141	143	137	139	130
	B	144	138	140	130	137
Wartość przeciętna ( $\hat{\mu}$ )	A	0,17	0,24	0,37	0,60	1,20
	B	0,15	0,21	0,34	0,55	1,11
Współczynnik rozproszenia ( $\hat{\sigma}$ )	A	0,06	0,10	0,11	0,19	0,32
	B	0,05	0,09	0,10	0,17	0,29

Zróznicowane źródła danych przestrzennych – wyniki pomiarów terenowych (współczesnych i dawnych), analogowa mapa zasadnicza (w różnych skalach), ortofotomapa – oraz odmienne techniki pozyskiwania tych danych, determinują znaczącą niejednorodność jakościową urzędowych baz danych. Niejednorodność ta szczególnie dotyczy dokładności położenia obiektów, będącej głównym elementem jakości danych, który bezpośrednio wpływa na wartość wykonywanych analiz geoprzestrzennych.

Zmodyfikowane techniki odpornościowe oraz autorska „metoda częstości względnej” pozwalają na precyzyjne wyznaczenie dokładności danych przestrzennych. Metody te mogą być z powodzeniem wykorzystywane jako wydajne narzędzia oceny i raportowania jakości danych baz EGiB, GESUT i BDOT500, w procesach tworzenia i aktualizacji tych urzędowych baz. Wartości realnej dokładności danych gromadzonych w bazie danych, na poziomie całej bazy oraz poszczególnych jej obiektów, zapisane w zbiorach metadanych, będą kluczową informacją o jakości danych dla użytkowników instytucjonalnych oraz indywidualnych.

#### 4.3.3. Niepewność pola powierzchni działek ewidencyjnych

W wektorowych modelach danych przestrzennych (prostym i topologicznym) obiekty opisywane są za pomocą współrzędnych punktów. Reprezentacja obiektów przestrzennych za pomocą tych współrzędnych pozwala na wykonywanie operacji obliczeniowych, takich jak odległość czy powierzchnia. Standardową operacją obliczeniową narzędzi komputerowych typu CAD i GIS jest obliczenie pola powierzchni wielokąta na podstawie współrzędnych punktów wierzchołkowych. W przypadku obiektów geograficznych, komputerowe bazy danych nie gromadzą informacji o ich bezbłędnym położeniu. Ze względu na techniki pozyskiwania danych przestrzennych oraz na stosowaną generalizację, reprezentacja obiektów w bazie danych jest w pewnym stopniu przybliżona. Stąd, współrzędne punktów charakteryzujących te obiekty są również przybliżone. Niepewność wartości współrzędnych przenosi się na niepewność obliczanego pola powierzchni (Oort i in. 2005; Hejmanowska 2013).

Podstawowym obiektem powierzchniowym bazy EGiB jest działka ewidencyjna, a jej głównym atrybutem jest pole powierzchni. Wykonanie analizy bezwzględnej dokładności wyznaczenia pola powierzchni działki (podejście ściśle) metodą biegunową wymaga posiadania informacji o wartości kowariancji dla współrzędnych punktów osnowy oraz kowariancji dla mierzonych kątów i odległości. Błąd średni funkcji pola ma postać:

$$m_{\xi}^2 = \mathbf{f}_{\xi}^T \cdot [Cov(X, Y)_{osnowa} + Cov(X, Y)_{pomiar}] \cdot \mathbf{f}_{\xi} \quad (7)$$

gdzie:  $\mathbf{f}_S$  – macierz pochodnych cząstkowych względem współrzędnych punktów granicznych,  
 $Cov(X, Y)_{osnowa}$ ,  $Cov(X, Y)_{pomiar}$  – kowariancja osnowy geodezyjnej i kowariancja pomiarów.

Pozyskanie z Zasobu Geodezyjnego informacji o wartościach kowariancji osnowy geodezyjnej i kowariancji pomiarów, dla celów wykonania globalnej (np. dla całego obrębu) analizy niepewności pola powierzchni, jest bardzo trudne, a w wielu przypadkach praktycznie niemożliwe. Z tego powodu, ścisła analiza może być zastosowana w ograniczonym zakresie, praktycznie tylko dla nowych pomiarów.

Inne podejście (przybliżone) do szacowania niepewności pola powierzchni działek polega na obliczeniu tej niepewności na podstawie błędów średnich współrzędnych punktów granicznych, odniesionych bezpośrednio do poziomej osnowy geodezyjnej 1 klasy. W tym przypadku błąd średni funkcji pola ma postać:

$$m_S^2 = \mathbf{f}_S^T \cdot \mathbf{Q}_{X,Y} \cdot \mathbf{f}_S \quad (8)$$

gdzie:  $\mathbf{f}_S$  – macierz pochodnych cząstkowych względem współrzędnych,  
 $\mathbf{Q}_{X,Y}$  – macierz kowariancji dla współrzędnych punktów granicznych.

Macierz kowariancji  $\mathbf{Q}_{X,Y}$  ma postać:

$$\mathbf{Q}_{X,Y} = \begin{bmatrix} m_{X_n}^2 & cov(X_n, Y_n) \\ cov(Y_n, X_n) & m_{Y_n}^2 \end{bmatrix}$$

gdzie:  $m_X$ ,  $m_Y$  – błędy średnie określenia współrzędnych  $X$  i  $Y$  w odniesieniu do osnowy geodezyjnej 1 klasy,  
 $cov(X, Y)$  – kowariancja pomiędzy współrzędnymi.

Poprawność powyższego rozumowania zweryfikowano poprzez obliczenie różnic w wartościach szacowania pola powierzchni pomiędzy metodą ścisłą a przybliżoną. Obliczone maksymalne wartości (zależne od współczynników korelacji) różnic pomiędzy dokładnościami wyznaczenia pola powierzchni metodą ścisłą i przybliżoną nie są znaczące (tab. 7). Różnice te nie przekraczają wartości 5 m<sup>2</sup> dla działek dużych i wąskich ( $P = 5,00$  ha,  $k = 10$ ).

Tabela 7. Różnice w wartościach szacowania pola powierzchni działek pomiędzy metodą ścisłą a przybliżoną

Powierzchnia [ha]	Wartości obliczonych różnic pomiędzy metodą ścisłą a przybliżoną [m <sup>2</sup> ]			
	$k = 1$	$k = 2$	$k = 5$	$k = 10$
0,10	0	0	1	2
0,20	0	0	2	2
0,50	0	1	2	2
1,00	1	1	2	3
2,00	1	2	3	4
5,00	2	3	3	5

$k$  – współczynnikiem wydłużenia działki

Korekta pola powierzchni działki ewidencyjnej w bazie EGiB na podstawie dokładniejszych pomiarów nie rodzi wątpliwości natury prawnej i technicznej. Problemem jest zakres korekty pola powierzchni, wyznaczonego na podstawie ponownych pomiarów, wykonywanych równorzędnymi technikami (w sensie dokładności). W obecnie obowiązujących przepisach prawnych nie określono zasad korekty pola powierzchni działki ewidencyjnej w bazie EGiB na podstawie ponownych pomiarów, wykonywanych równorzędnymi technikami. Korekta powierzchni działki w bazie EGiB oraz w rejestrze ksiąg wieczystych, po każdym pomiarze jej granic (wznowienie znaków granicznych, podział nieruchomości), podważa wiarygodność EGiB jako rejestru publicznego.

W celu oceny dokładności wyznaczania pola powierzchni na podstawie współczesnych technik pozyskiwania danych (współrzędnych) wykonano badania eksperymentalne. Metodą biegunową oraz techniką kartometryczną na ortofotomapie pomierzono 24 działki o różnych polach powierzchni i kształtach. Badano działki o powierzchniach ok.: 0,10 ha; 0,20 ha; 0,50 ha; 1,00 ha; 2,00 ha i 5,00 ha, które charakteryzowały się współczynnikiem wydłużenia ( $k$ ): 1, 2, 5 i 10. Odległości pomiędzy punktami załamania linii granicznych nie były większe od ok. 50 m. Każda z 24 działek została pomierzona 20 razy.

Eksperymentalną ocenę dokładności wyznaczenia pola powierzchni oparto na analizie wyznaczonych estymatorów. Obliczono, dla każdej działki, estymatory pola powierzchni, m. in.: wartość średnią oraz stosunek poprawki maksymalnej ( $v_{max}$ ) do błędu średniego ( $m_0$ ). Zależność  $v_{max}/m_0$  dla żadnej z analizowanych działek nie przekroczył wartości 2,1. Stąd wniosek, że błąd graniczny wyznaczenia pola powierzchni nie przekracza dwukrotnej wartości błędu średniego.

Empiryczną analizę dokładności wyznaczenia pola powierzchni, opartą na pomiarach geodezyjnych, uzupełniono metodą wielokrotnego losowania danych pomiarowych. Położenie wierzchołków wielokątów zmieniano dodając do nich pseudolosowe wartości błędów przypadkowych, występujących w pomiarach terenowych oraz w technice kartometrycznej na ortofotomapie. Operację tę powtarzano wielokrotnie (1000 razy) obliczając ostatecznie średnią wartość pola powierzchni oraz jego charakterystyki. Zależność  $v/m_0$  (dla ponad 99% przypadków) jest mniejsza lub równa wartości 2,0. Z tego powodu można wnioskować, że błąd graniczny dokładności wyznaczenia pola powierzchni, nie przekroczy z prawdopodobieństwem  $P = 0,99$ , dwukrotnej wartości błędu średniego.

Przedstawione powyżej wyniki analiz mogą być podstawą sformułowania zasad zmiany pola powierzchni działki ewidencyjnej na podstawie nowych pomiarów. W przypadkach wykonywania nowych pomiarów, błąd graniczny korekty pola powierzchni w bazie EGiB równy jest wartości dwukrotnego błędu wyznaczenia pola, obliczonego na podstawie różniczkowania funkcji po obu stronach wzoru Gaussa. Jako błąd średni położenia punktu granicznego należy przyjąć jego dopuszczalną wartość dla danej techniki pomiarowej. Stąd wzór na maksymalną dopuszczalną różnicę pól powierzchni ma postać:

$$dS_{max} = 2,0 \cdot m_{pmax} \sqrt{\frac{1}{8} \sum_{i=1}^n d_i^2} \quad (9)$$

gdzie:  $m_{pmax}$  – dopuszczalny dla danej techniki pomiarowej błąd średni położenia punktu granicznego,

$d_i$  – najkrótsza przekątna znajdująca się naprzeciw punktu nr  $i$ ,

$n$  – liczba punktów załamania granic działki.

Dokładność wyznaczenia powierzchni nieruchomości, składającej się z działek ewidencyjnych, ma istotny wpływ na jej wartość. Wartością nieruchomości zainteresowany jest jej właściciel oraz potencjalny nabywca (transakcja kupna-sprzedaży), jak również administracja samorządowa (podatek od nieruchomości). Z powodu błędów pomiarowych technik geodezyjnych, dokładność wyznaczania pola powierzchni działek ewidencyjnych – odpowiadająca precyzji zapisu wielkości pola ( $1 \text{ m}^2$ ) – nie jest gwarantowana. W związku z tym w bazie EGiB powinna być informacja o niepewności pól powierzchni działek ewidencyjnych.

Proponowane podejście (przybliżone) do szacowania niepewności pól powierzchni działek pozwala na wykonania globalnej (np. dla całego obrębu) analizy tej niepewności. Wartości rzeczywistej niepewności pola powierzchni poszczególnych działek zapisane w zbiorach metadanych bazy EGiB będą przydatne dla ich użytkowników. Przykładowo, właściciel nieruchomości lub potencjalny nabywca gruntu, dzięki wskaźnikowi dokładności wyznaczonego pola posiada informację na temat możliwego niedoszacowania pola powierzchni na gruncie.

Opracowany empiryczny wzór na błąd graniczny korekty pola powierzchni w bazie EGiB może być uzupełnieniem brakujących zapisów w obecnie obowiązujących standardach technicznych. Zastosowanie tego wzoru ograniczy zakres korekty powierzchni działki w bazie EGiB oraz w rejestrze ksiąg wieczystych, po każdym pomiarze jej granic. Działania takie wpłyną pozytywnie na wiarygodność ewidencji gruntów i budynków jako rejestru publicznego.

#### 4.3.4. Analiza ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu

Pojęcie ryzyka oznacza niepewność co do możliwości wystąpienia zdarzenia, mogącego mieć negatywny wpływ na osiągnięcie zamierzonego celu (Gleim 2004). Definicja ta dotyczy jakościowych metod oceny zagrożeń, bazujących głównie na wiedzy i doświadczeniu eksperta. Definicję ryzyka w ujęciu ilościowym podają Kaplan i Garrick (1981). Ryzyko związane z pewnym zdarzeniem opisane jest jako trójelementowy zbiór:

$$R = \{S, P, C\}, \quad (10)$$

gdzie:  $S$  – scenariusz pewnego zdarzenia,

$P$  – prawdopodobieństwo zajścia pewnego zdarzenia,

$C$  – możliwe skutki zajścia pewnego zdarzenia.

Ocenę ryzyka można rozważać również w aspekcie analizy jakości danych przestrzennych. Dane przestrzenne wykorzystywane są do realizacji szerokiego zakresu przedsięwzięć. Jakość tych danych wpływa bezpośrednio na powodzenie takich projektów.

Ryzyko uszkodzenia sieci podziemnego uzbrojenia terenu jest nieodłącznym elementem procesu budowlanego, a jednym z jego źródeł jest jakość danych przestrzennych. Istnieją trzy główne przyczyny uszkodzeń podziemnych rur i kabli, w aspekcie oceny jakości zbiorów danych przestrzennych. Pierwsza przyczyna to obecność (na obszarze inwestycji) podziemnego uzbrojenia terenu, nie wykazanego na mapie do celów projektowych. Źródło tego ryzyka można zdefiniować jako brak kompletności zbiorów bazy danych. Druga przyczyna ryzyka dotyczy niewystarczającej dokładności określenia położenia poziomego rur i kabli, przedstawionych na mapie do celów projektowych, której podstawą opracowania są dane przestrzenne gromadzone w urzędowych bazach danych. Trzecią przyczyną ryzyka jest niewystarczająca dokładność pionowa określenia położenia rur i kabli podziemnych.

Analizę ryzyka wykonano dwoma metodami: punktową oraz z wykorzystaniem teorii zbiorów rozmytych. W metodzie punktowej, pierwszy etap postępowania dotyczy standaryzacji czynników ryzyka oraz określenia skutków oddziaływania tego ryzyka. Zdefiniowano funkcje standaryzacyjne dla trzech czynników ryzyka oraz skutki oddziaływania na podstawie prac własnych autora oraz wiedzy i doświadczenia ekspertów z branży budowlanej.

Funkcja standaryzacyjna dla czynnika ryzyka zdefiniowanego jako: *brak kompletnej informacji o uzbrojeniu podziemnym (BK)*, określona jest następująco:

$$f(BK) = \begin{cases} 1 & \text{dla } BK = I \\ 3 & \text{dla } BK = II \\ 6 & \text{dla } BK = III \end{cases} \quad (11)$$

gdzie:  $BK = I$  – brak kompletnej informacji o uzbrojeniu podziemnym w przedziale 0 – 5 %,

$BK = II$  – w przedziale 6 ÷ 10 %,

$BK = III$  – w przedziale > 10 %.

Funkcja standaryzacyjna dla czynnika ryzyka zdefiniowanego jako: *dokładność pozioma uzbrojenia podziemnego (DP)*, określona jest następująco:

$$f(DP) = \begin{cases} 1 & \text{dla } DP = I \\ 2 & \text{dla } DP = II \\ 4 & \text{dla } DP = III \\ 8 & \text{dla } DP = IV \end{cases} \quad (12)$$

gdzie:  $DP = I$  – źródło danych: geodezyjne pomiary terenowe,

$DP = II$  – źródło danych: geodezyjne pomiary kartometryczne na mapie analogowej w skali 1:500,

$DP = III$  – źródło danych: geodezyjne pomiary kartometryczne na mapie analogowej w skali 1:1000,

$DP = IV$  – źródło danych: geodezyjne pomiary kartometryczne na mapie analogowej w skali 1:2000.

Funkcja standaryzacyjna dla czynnika ryzyka zdefiniowanego jako: *dokładność pionowa uzbrojenia podziemnego (DV)*, określona jest następująco:

$$f(DV) = \begin{cases} 1 & \text{dla } DV = I \\ 3 & \text{dla } DV = II \end{cases} \quad (13)$$

gdzie:  $DV = I$  – istnieją w zbiorze danych informacje o położeniu wysokościowym,

$DV = II$  – brak informacji w zbiorze danych o położeniu wysokościowym.

Funkcja skutków oddziaływania określona jest następująco:

$$f(p) = \begin{cases} 8 & \text{dla } p = I \\ 4 & \text{dla } p = II \\ 2 & \text{dla } p = III \end{cases} \quad (14)$$

gdzie:  $p = I$  – bardzo duże oddziaływanie czynnika ryzyka: brak kompletnej informacji o uzbrojeniu podziemnym ( $BK$ ),

$p = II$  – duże oddziaływanie czynnika ryzyka: dokładność pozioma uzbrojenia podziemnego ( $DP$ ),

$p = III$  – średnie oddziaływanie czynnika ryzyka: dokładność pionowa uzbrojenia podziemnego ( $DV$ ).

Ostatni etap postępowania to obliczenie poziomu ryzyka, wyrażonego w postaci ilościowej i jakościowej. Ilościowa wielkość ryzyka określona jest liczbą, uzyskaną na podstawie wzoru:

$$R^2 = (bk \cdot p1)^2 + (dp \cdot p2)^2 + (dv \cdot p3)^2 \quad (15)$$

gdzie:  $R$  – ryzyko uszkodzenia sieci podziemnego uzbrojenia terenu,

$bk, dp, dv$  – znormalizowane wielkości czynników ryzyka,

$p1, p2, p3$  – wagi oddziaływania ryzyka dla czynników  $BK, DP$  i  $DV$ .

Obliczona wartość ryzyka należy do przedziału od 9,2 do 58,0 punktów. Oznacza to, że minimalne ryzyko uszkodzenia podziemnych rur i kabli, podczas wykonywania prac budowlanych wynosi 9,2 punktu, a ryzyko maksymalne jest ponad sześciokrotnie większe: 58,0 punktów. Dla ryzyka jakościowego zdefiniowano wartości liczbowe ryzyka małego, średniego i dużego (tab. 8).

Tabela 8. Zakres ryzyka uszkodzenie sieci podziemnego uzbrojenia terenu

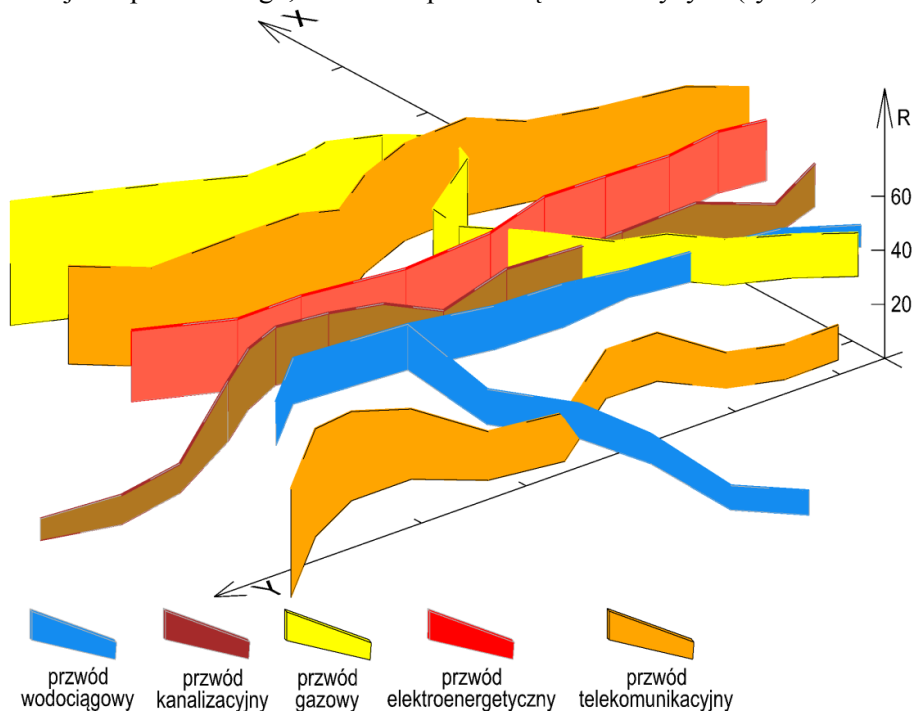
Ryzyko	
Ilościowe	Jakościowe
	RM – małe, RS – średnie, RD – duże
[9,2, ..., 58,0]	[9,2, ..., 12,8] – RM
	[18,0, ..., 18,9] – RS
	[24,1, ..., 58,0] – RD

Obliczone ryzyko jakościowe i ilościowe jest podstawą opracowania map (2D i 3D) ryzyka uszkodzenia podziemnych rur i kabli. Mapa 2D ryzyka uszkodzenia sieci podziemnego uzbrojenia jest wielkoskalową mapą tematyczną, prezentującą ryzyko projektowe w formie jakościowej (rys. 2). Mapa sporządzana jest zgodnie z zasadą warstwowego nakładania obszarów charakteryzujących się różnymi wartościami ryzyka jakościowego (od najniższej do najwyższej).



Rys 2. Mapa 2D ryzyka uszkodzenia podziemnego uzbrojenia (fragment)  
(kolor żółty – ryzyko małe, pomarańczowy – ryzyko średnie, czerwony – ryzyko duże)

Wydajnym narzędziem wspomagającym realizację bezpiecznego procesu budowlanego jest również mapa 3D ryzyka ilościowego. Na mapie tej w układzie  $X, Y$  przedstawiono położenie przewodów uzbrojenia podziemnego, a na osi  $R$  punktową wartość ryzyka (rys. 3).



Rys 3. Mapa ryzyka 3D uszkodzenia podziemnego uzbrojenia (fragment)



Drugim podejściem do analizy ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu jest wykorzystanie teorii zbiorów rozmytych. Logika rozmyta ma zastosowanie do opisu złożonych lub słabo zdefiniowanych zjawisk, trudnych do opisania za pomocą metod probabilistycznych. Punktem wyjścia wnioskowania opartego na danych jakościowych są zbiory rozmyte, rozumiane jako zbiory elementów, które w różnym stopniu do tego zbioru należą (Zadeh 1965).

Analogicznie jak w poprzednim podejściu, przyjęto założenie, że ryzyko ( $R_U$ ) jest wieloczynnikowe:

$$R_U = f(BK, DP, DV) \quad (16)$$

gdzie:  $BK$  – brak kompletnej informacji o uzbrojeniu podziemnym,

$DP$  – dokładność położenia poziomego uzbrojenia podziemnego,

$DV$  – dokładność położenia pionowego.

Schemat procesu wnioskowania rozmytego polega – w pierwszym etapie – na przetwarzaniu wejściowych danych ilościowych na pojęcia lingwistyczne (rozmywanie). Następnie wyznaczane są jakościowe wartości (zbiory) wyjściowe za pomocą bazy reguł (wnioskowanie). W ostatnim etapie, wyjściowe zbiory rozmyte przetwarzane są na zmienne ilościowe (wyostrzenie).

Dla każdego z czynników ryzyka określono 3 wartości lingwistyczne: niską, średnią i wysoką. Wartości zmiennych lingwistycznych określono sposobem eksperckim. Wykorzystano wiedzę i doświadczenie specjalistów z przedsiębiorstw budowlanych, realizujących projekty liniowe (sieci wodociągowe, kanalizacyjne, elektroenergetyczne oraz gazowe). Następnie określono rozkład wartości funkcji przynależności do zbiorów rozmytych (tab. 9), na podstawie którego zdefiniowano funkcję przynależności dla zmiennych lingwistycznych  $BK$ ,  $DP$  i  $DV$ , oraz opisujące te funkcje zbiory rozmyte (rys.4 – rys. 6).

Tabela 9. Rozkład wartości funkcji przynależności do zbiorów rozmytych

Brak kompletnej informacji o uzbrojeniu podziemnym ( $BK$ )	niski	$x$	0,0	5,0	10,0	15,0	16,0	-	-	-	-
		$u(x)$	1,00	0,86	0,12	0,02	0,00	-	-	-	-
	średni	$x$	5,0	10,0	15,0	21,0	25,0	30,0	35,0	37,0	-
		$u(x)$	0,02	0,16	0,82	1,00	0,82	0,16	0,02	0,00	-
	wysoki	$x$	25,0	30,0	35,0	41,0	-	-	-	-	-
		$u(x)$	0,00	0,12	0,88	1,00	-	-	-	-	-
Dokładność pozioma ( $DP$ ) oraz Dokładność pionowa ( $DV$ )	niska	$x$	0,0	5,0	10,0	15,0	20,0	25,0	-	-	-
		$u(x)$	1,00	1,00	1,00	0,88	0,12	0,00	-	-	-
	średnia	$x$	12,0	15,0	20,0	25,0	28,0	59,0	65,0	70,0	75,0
		$u(x)$	0,00	0,02	0,12	0,86	1,00	1,00	0,86	0,12	0,02
	wysoka	$x$	62,0	65,0	70,0	75,0	78,0	-	-	-	-
		$u(x)$	0,00	0,02	0,28	0,70	1,00	-	-	-	-

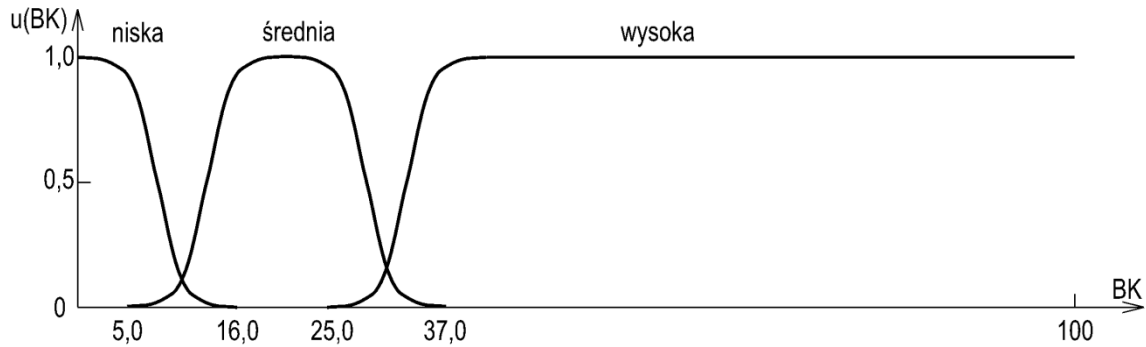
Funkcja przynależności dla zmiennej  $BK$ :

$$u(BK_n) = \begin{cases} 1 & \text{jeżeli } x = 0 \\ \frac{1}{1 + \exp(-a(x - c))} & \text{jeżeli } 0 < x \leq 16 \quad (a = -0,9; c = 8) \\ 0 & \text{jeżeli } x > 16 \end{cases} \quad (17)$$

$$u(BK_s) = \begin{cases} 0 & \text{jeżeli } x \leq 5 \\ \frac{1}{1 + \exp(-a(x - c))} & \text{jeżeli } 5 < x < 21 \quad (a = 0,9; c = 13) \\ 1 & \text{jeżeli } x = 21 \\ \frac{1}{1 + \exp(a(x - c))} & \text{jeżeli } 21 < x \leq 37 \quad (a = -0,9; c = 29) \\ 0 & \text{jeżeli } x > 37 \end{cases} \quad (18)$$

$$u(BK_w) = \begin{cases} 0 & \text{jeżeli } x \leq 25 \\ \frac{1}{1 + \exp(-a(x-c))} & \text{jeżeli } 25 < x \leq 41 \quad (a = 0,9; c = 33) \\ 1 & \text{jeżeli } 41 < x \leq 100 \end{cases} \quad (19)$$

gdzie:  $BK_n$ ,  $BK_s$ ,  $BK_w$  – brak kompletnej informacji o uzbrojeniu podziemnym: poziom niski, średni i wysoki.



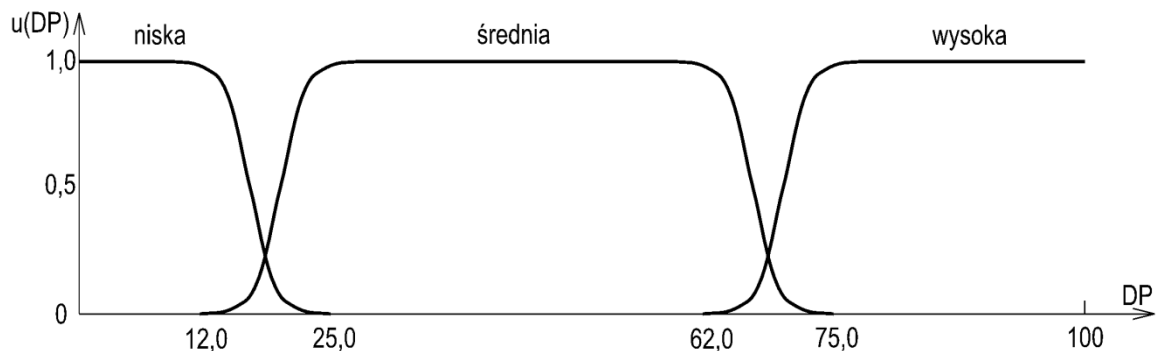
Rys. 4. Funkcja przynależności dla zmiennej lingwistycznej „brak kompletnej informacji o uzbrojeniu podziemnym” ( $BK$ )

Funkcja przynależności dla zmiennej  $DP$ :

$$u(DP_n) = \begin{cases} 1 & \text{jeżeli } 0 < x \leq 9 \\ \frac{1}{1 + \exp(a(x-c))} & \text{jeżeli } 9 < x \leq 25 \quad (a = -0,9; c = 17) \\ 0 & \text{jeżeli } x > 25 \end{cases} \quad (20)$$

$$u(DP_s) = \begin{cases} 0 & \text{jeżeli } x \leq 12 \\ \frac{1}{1 + \exp(-a(x-c))} & \text{jeżeli } 12 < x \leq 28 \quad (a = 0,9; c = 20) \\ 1 & \text{jeżeli } 28 < x \leq 59 \\ \frac{1}{1 + \exp(a(x-c))} & \text{jeżeli } 59 < x \leq 75 \quad (a = -0,9; c = 67) \\ 0 & \text{jeżeli } x > 75 \end{cases} \quad (21)$$

$$u(DP_w) = \begin{cases} 0 & \text{jeżeli } x \leq 62 \\ \frac{1}{1 + \exp(-a(x-c))} & \text{jeżeli } 62 < x \leq 78 \quad (a = 0,9; c = 70) \\ 1 & \text{jeżeli } 78 < x \leq 100 \end{cases} \quad (22)$$



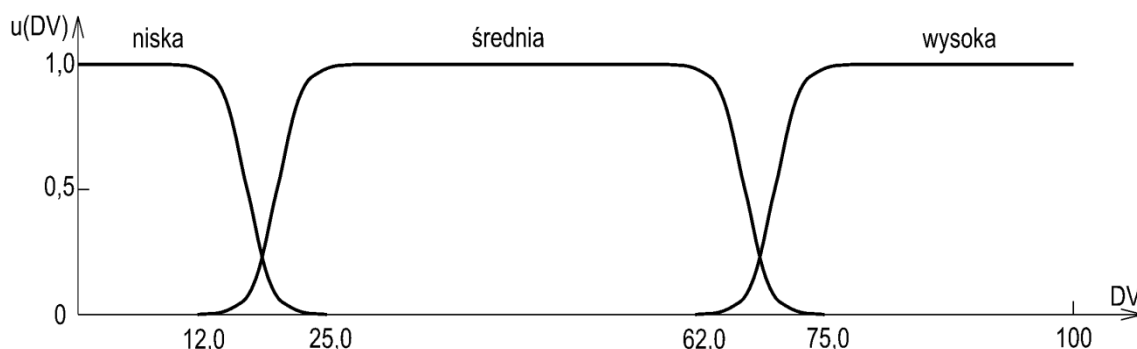
Rys. 5. Funkcja przynależności dla zmiennej lingwistycznej „dokładność pozioma” ( $DP$ )

Funkcja przynależności dla zmiennej  $DV$ :

$$u(DV_n) = \begin{cases} 1 & \text{jeżeli } 0 < x \leq 9 \\ \frac{1}{1 + \exp(a(x - c))} & \text{jeżeli } 9 < x \leq 25 \quad (a = -0,9; c = 17) \\ 0 & \text{jeżeli } x > 25 \end{cases} \quad (23)$$

$$u(DV_s) = \begin{cases} 0 & \text{jeżeli } x \leq 12 \\ \frac{1}{1 + \exp(-a(x - c))} & \text{jeżeli } 12 < x \leq 28 \quad (a = 0,9; c = 20) \\ 1 & \text{jeżeli } 28 < x \leq 59 \\ \frac{1}{1 + \exp(a(x - c))} & \text{jeżeli } 59 < x \leq 75 \quad (a = -0,9; c = 67) \\ 0 & \text{jeżeli } x > 75 \end{cases} \quad (24)$$

$$u(DV_w) = \begin{cases} 0 & \text{jeżeli } x \leq 62 \\ \frac{1}{1 + \exp(-a(x - c))} & \text{jeżeli } 62 < x \leq 78 \quad (a = 0,9; c = 70) \\ 1 & \text{jeżeli } 78 < x \leq 100 \end{cases} \quad (25)$$



Rys. 6. Funkcja przynależności dla zmiennej lingwistycznej „dokładność pionowa” ( $DV$ )

Ryzyko uszkodzenia podziemnych rur i kabli zdefiniowano jako trójelementowy zbiór, za pomocą zmiennych lingwistycznych: ryzyko niskie, średnie i wysokie. Zakres wartości zmiennych lingwistycznych opracowano w oparciu o stosowane wskaźniki analizy ryzyka w budownictwie.

Funkcja przynależności dla zmiennej  $R$  ma postać:

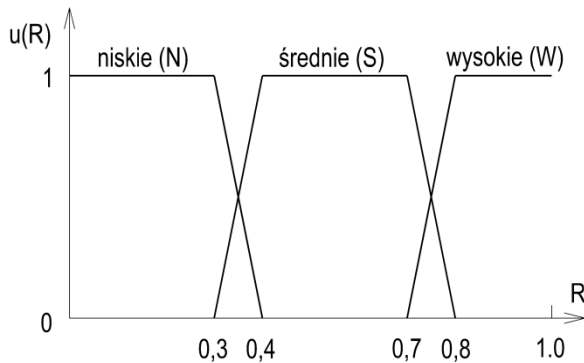
$$u(R_n) = \begin{cases} 1 & \text{jeżeli } x \leq 0,3 \\ \frac{d - x}{d - c} & \text{jeżeli } 0,3 < x \leq 0,4 \quad (c = 0,3; d = 0,4) \\ 0 & \text{jeżeli } x > 0,4 \end{cases} \quad (26)$$

$$u(R_s) = \begin{cases} 0 & \text{jeżeli } x \leq 0,3 \\ \frac{x - a}{b - a} & \text{jeżeli } 0,3 < x \leq 0,4 \quad (a = 0,3; b = 0,4) \\ 1 & \text{jeżeli } 0,4 < x \leq 0,7 \\ \frac{d - x}{d - c} & \text{jeżeli } 0,7 < x \leq 0,8 \quad (c = 0,7; d = 0,8) \\ 0 & \text{jeżeli } x > 0,8 \end{cases} \quad (27)$$

$$u(R_w) = \begin{cases} 0 & \text{jeżeli } x < 0,7 \\ \frac{x-a}{b-a} & \text{jeżeli } 0,7 < x \leq 0,8 \quad (a = 0,7; b = 0,8) \\ 1 & \text{jeżeli } 0,8 < x \leq 1 \end{cases} \quad (28)$$

gdzie:  $R_n, R_s, R_w$  – ryzyko: niskie, średnie i wysokie.

Opisujące tę funkcję zbiory rozmyte przedstawiono na rysunku 7.



Rys. 7. Funkcja przynależności dla zmiennej lingwistycznej „ryzyko”(R)

Wnioskowanie rozmyte zostało przeprowadzone dla pewnego modelu ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu. Model ten jest charakterystyczny dla współczesnej bazy danych przestrzennych, istniejącej głównie na obszarach miast. W granicach administracyjnych miast, dla całego terenu były zakładane i prowadzone mapy zasadnicze w skalach 1:500 i 1:1000, dlatego jakość materiałów państwowego zasobu geodezyjnego jest znacznie wyższa w miastach w porównaniu z terenami wiejskimi. W analizowanym modelu, kompletność danych, dotyczących podziemnego uzbrojenia, wynosi 91,0%. Dokładność pozioma, niezbędna do celów realizacji zamierzeń budowlanych, zapewniona jest w 70,0%. W kontekście wymagań prac budowlanych, dokładność pionowa danych zapewniona jest w 50,0%.

W pierwszym etapie wnioskowania rozmytego, wykonywana jest fuzyfikacja (Piegat 2003). Proces ten polega na obliczeniu stopnia przynależności wartości wejściowych do zbiorów rozmytych. Wykonując obliczenia uzyskano wartości stopni przynależności zmiennych wejściowych do zbiorów rozmytych (tab. 10).

Tabela. 10. Wartości przynależności zmiennych wejściowych do zbiorów rozmytych

Zmienna lingwistyczna	Oznaczenie zbioru		
	Niska	Średnia	Wysoka
Brak kompletnej informacji o uzbrojeniu podziemnym (BK)	0,29	0,03	0,00
Dokładność pozioma (DP)	0,00	0,06	0,50
Dokładność pionowa (DV)	0,00	1,00	0,00

W celu obliczenia stopni przynależności czynników ryzyka  $BK, DP$  i  $DV$ , do zbiorów rozmytych reprezentujących wartości zmiennych lingwistycznych, wykorzystywano ekspercką bazę reguł. W tym przypadku potrzebna jest znajomość 4 reguł:

- I. Jeżeli  $((BK = N) \text{ i } (DP = S) \text{ i } (DV = S))$  wtedy  $(R = S)$
- II. Jeżeli  $((BK = N) \text{ i } (DP = W) \text{ i } (DV = S))$  wtedy  $(R = N)$
- III. Jeżeli  $((BK = S) \text{ i } (DP = S) \text{ i } (DV = S))$  wtedy  $(R = S)$
- IV. Jeżeli  $((BK = S) \text{ i } (DP = W) \text{ i } (DV = S))$  wtedy  $(R = S)$

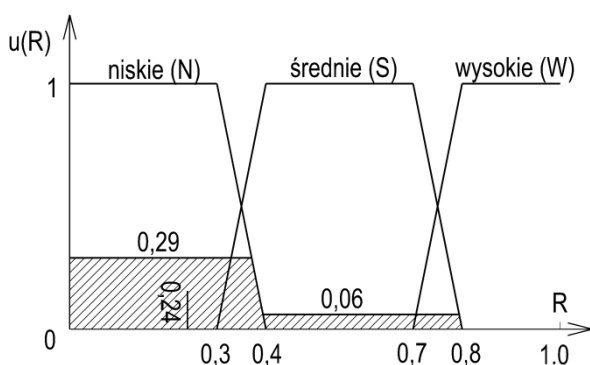
Określając wartość minimalną stopnia przynależności poszczególnych przesłanek danej reguły, zastosowano operator MIN (Yager i Filev 1995):

$$\begin{cases} 0,29/N \wedge 0,06/S \wedge 1/S = 0,06/RS \\ 0,29/N \wedge 0,06/S \wedge 1/S = 0,06/RS \\ 0,03/S \wedge 0,06/S \wedge 1/S = 0,03/RS \\ 0,03/S \wedge 0,50/W \wedge 1/S = 0,03/RS \end{cases} \quad (29)$$

Stosując operator MAX, wyznaczono rozmyty wynik wnioskowania:

$$0,06/RS \vee 0,29/RN \vee 0,03/RS \vee 0,03/RS = 0,29/RN \vee 0,06/RS \quad (30)$$

Wynik wnioskowania, czyli obliczone ryzyko uszkodzenia sieci podziemnego uzbrojenia należy do dwóch zbiorów (rys.8). Wynik ten jest przypisany do zbioru „ryzyko niskie” – w stopniu równym 0,29, natomiast do zbioru „ryzyko średnie” – w stopniu równym 0,06. Na ostatnim etapie wnioskowania rozmytego, wielkości wynikowe przekształcane są na wartości liczbowe. W analizowanym modelu, ryzyka otrzymano w wyniku defuzyfikacji (metodą środka ciężkości) wartość ryzyka  $R = 0,24$ . Obliczone ryzyko w pełni przynależy do zbioru „ryzyko niskie” (rys. 8).



Rys. 8. Ryzyko uszkodzenia podziemnego uzbrojenia terenu

Opracowana metoda pozwala na szacowanie ryzyka uszkodzenia rur i kabli dla innych klas jakości danych przestrzennych, np. dla danych istniejących na terenach wiejskich. Dysponując opracowanym modelem oceny ryzyka, można zmieniać rozkład wartości funkcji przynależności do zbiorów rozmytych, w celu oceny wpływu czynników na poziom ryzyka. Zaproponowana metoda analizy ryzyka, oparta na logice rozmytej, może być alternatywą dla konwencjonalnych metod oceny ryzyka, w przypadkach analizy zjawisk złożonych lub słabo zdefiniowanych, trudnych do opisanego za pomocą metod probabilistycznych. Zastosowanie wnioskowania rozmytego, analogicznie jak dla metody punktowej, pozwala na tworzenie wielkoskalowych map ryzyka.

Ryzyko uszkodzenia sieci podziemnego uzbrojenia terenu jest nieodłącznym elementem procesu budowlanego, a jednym z jego źródeł jest jakość danych przestrzennych. W razie wystąpienia ryzyka, jego oddziaływanie (skutek) na realizację zamierzenia budowlanego może być znaczące i dotyczyć problemu ochrony zdrowia i bezpieczeństwa ludzi, jak również powodować straty finansowe. Uszkodzenie podziemnych rur i kabli może doprowadzić do wypadków, opóźnień lub nawet zatrzymania budowy. Opracowane wielkoskalowe mapy ryzyka 2D i 3D mogą być wykorzystane przez przedsiębiorstwa budowlane i pełnić rolę wydajnych narzędzi wspomagających realizację bezpiecznego procesu budowlanego.

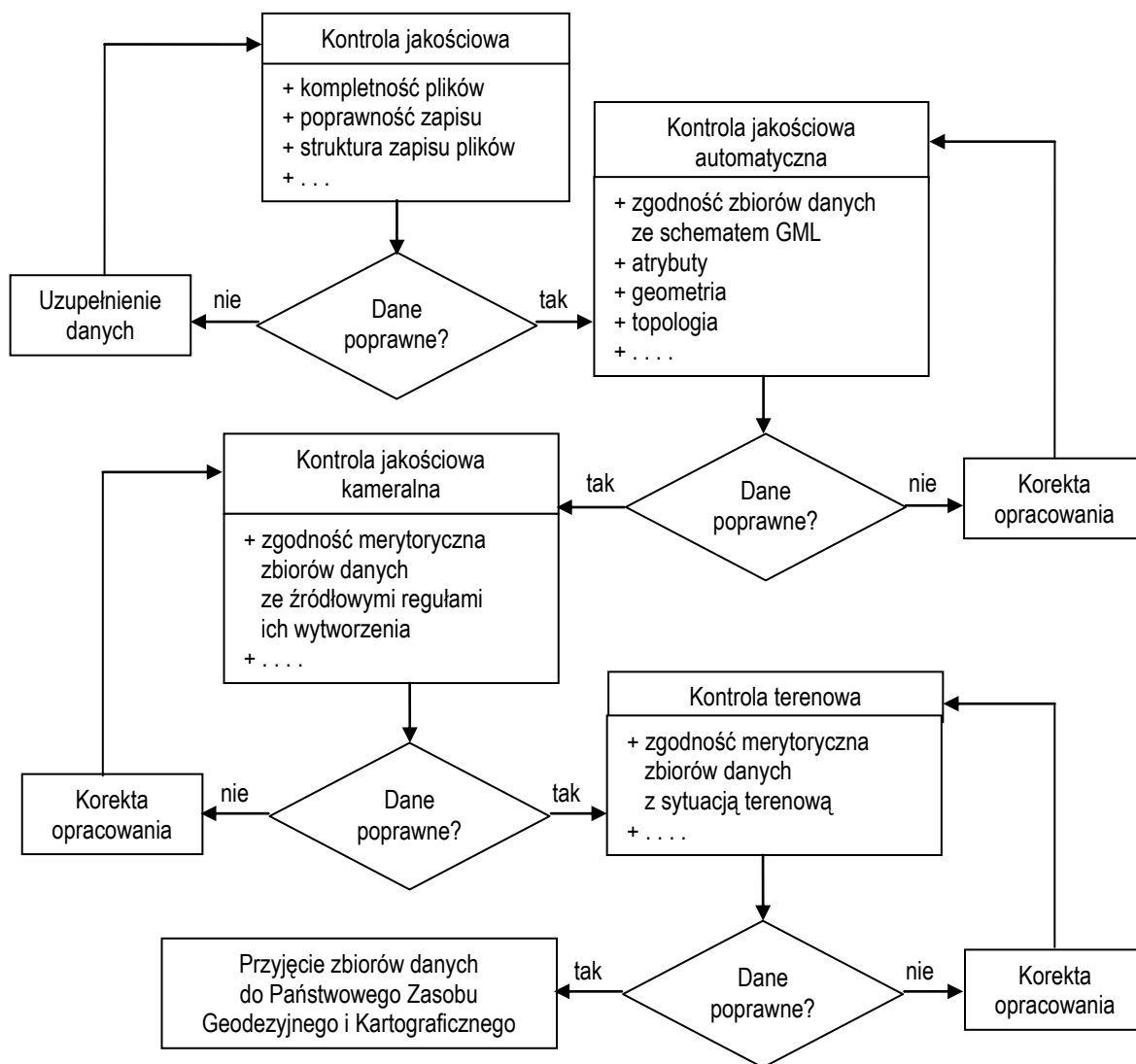
#### 4.3.5. Model jakości danych gromadzonych w urzędowych bazach

Model jakości danych jest zbiorem zasad służących do opisu jakości zbiorów danych (ESDIN 2010). Celem proponowanego modelu jakości danych jest zdefiniowanie standardowego podejścia do oceny i raportowania jakości danych przestrzennych baz EGIB, GESUT i BDOT500. Pierwszy etap takiej oceny zawiera kontrole automatyczne zbiorów danych EGIB, GESUT i BDOT500, i obejmuje pięć podstawowych składników: walidację danych ze schematem GML (ang. *geography markup*

language), kontrolę atrybutową w plikach GML, kontrolę geometryczną, kontrolę topologii oraz inne kontrole spójności logicznej.

Drugim etapem kontroli jakościowej jest kontrola kameralna. Celem tego etapu jest zbadanie szczegółowej zgodności merytorycznej przekazywanych danych. Weryfikowane są tutaj wyniki prac geodezyjnych, wykonywanych w celu wytworzenia zbiorów danych. Trzecim etapem kontroli jakościowej jest kontrola terenowa. Celem tego etapu jest zbadanie zgodności przekazywanych danych z rzeczywistą sytuacją w terenie. Kontroli podlegają wszystkie obiekty wprowadzane do baz EGIB, GESUT i BDOT500, na obszarze wybranej do kontroli próbki danych.

Kontrola ilościowa i kontrola jakościowa mogą być wykonywane w kilku iteracjach – do momentu uzyskania pozytywnych wyników. Pozytywny wynik oceny jakości danych nie oznacza wyeliminowania wszystkich błędów. Ewaluacja jakości danych baz EGIB, GESUT i BDOT500 (kameralna i terenowa) oparta jest na próbach statystycznych. W takich przypadkach zbiory elementów jakości danych zostaną pozytywnie zweryfikowane, jeżeli próbki danych będą charakteryzowały się akceptowalnym poziomem jakości. Usunięcie wszystkich błędów jakości danych możliwe jest tylko dla składników spójności logicznej zbiorów danych. Schemat procesu kontroli jakości danych przedstawiono na rysunku 9.



Rys. 9. Schemat procesu kontroli jakości danych baz EGIB, GESUT i BDOT500

Na podstawie własnych doświadczeń autora pracy, dotyczących badania jakości danych urzędowych baz przestrzennych, oraz innych opracowań (Greń 1984; ISO 1999), określono rodzaje weryfikacji jakościowej bazy danych przestrzennych oraz zaproponowano minimalne liczebności próbek statystycznych (tab. 11).

Tabela 11. Weryfikacja jakościowa danych baz EGIB, GESUT i BDOT500

Lp.	Rodzaj weryfikacji jakościowej bazy danych przestrzennych	Nazwa bazy danych		
		EGiB	GESUT	BDOT500
		Minimalna wielkość próby dla obrębu ewidencyjnego		
1	Wykorzystania operatorów technicznych z PZGiK	10% operatorów jednostkowych		
2	Poprawność ustalenia granic i pomiaru punktów granicznych	Całość opracowania	–	–
3	Poprawność wykonania geodezyjnych pomiarów sytuacyjnych	Całość opracowania		
4	Poprawności określenia atrybutów punktów granicznych	10% działek	–	–
5	Poprawności określenia danych opisowych dotyczących budynków	10% budynków	–	–
6	Poprawności określenia atrybutów charakteryzujących obiekty	–	10% obiektów, dla każdej z klas obiektów	
7	Dokładności geodezyjnych pomiarów terenowych oraz pomiarów fotogrametrycznych	100 punktów granicznych oraz 5% budynków	100 obiektów, dla każdej z klas obiektów	
8	Weryfikacja dokładności pomiarów kartometrycznych	200 punktów granicznych oraz 10% budynków	200 obiektów, dla każdej z klas obiektów	

Ewaluacja jakości danych baz EGIB, GESUT i BDOT500 według takich czynników, w przypadku których ocena oparta jest na próbach statystycznych, skutkuje przyjęciem lub odrzuceniem badanych elementów jakości dla całego zbioru danych. Uznanie, że populacja spełnia wymagane założenia, realizowane jest na podstawie dopuszczalnej liczebności elementów niezgodnych w próbce danych. Wyznaczenie wartości dystrybuanty rozkładu dwumianowego pozwala określić dopuszczalną liczebność elementów niezgodnych w próbce statystycznej. Wielkości tych elementów – określanych jako liczby kwalifikujące – obliczono dla przykładowych, dopuszczalnych poziomów niezgodności (tab. 12).

Tabela 12. Liczby kwalifikujące dla ewaluacji jakości danych baz EGIB, GESUT i BDOT500

Liczebność próby ( $n$ )	Poziom niezgodności ( $p_n$ )			
	0,01	0,02	0,05	0,10
40	2	3	5	9
80	3	5	9	15
120	4	6	12	20
160	5	8	15	25
200	6	10	16	28

Wielkości liczb kwalifikacyjnych (tab. 12) decydują o przyjęciu lub odrzuceniu badanych elementów jakości danych dla całego zbioru danych – w tych przypadkach, dla których ocena oparta była na próbach statystycznych. Przykładowo, dla zbioru danych, którego liczebność próby kontrolnej wynosi  $n = 120$  oraz dopuszczalny poziom niezgodności  $p_n = 0,02$ , liczba elementów niezgodnych (niepełniających kryteriów jakości danych) nie może być większa od sześciu, tak aby badany element jakości danych spełniał wymagane kryteria jakości danych.

Zdefiniowany model jakości danych może być z powodzeniem wykorzystywane jako wydajne narzędzia oceny i raportowania jakości danych baz EGIB, GESUT i BDOT500, w procesach tworzenia i aktualizacji tych urzędowych baz.

## Ocena jakości danych gromadzonych w bazach – badania eksperymentalne

Elementy jakości danych charakteryzują się różnym stopniem istotności, szczególnie z punktu widzenia użytkownika danych. Z tego względu, podczas oceny jakościowej baz danych przestrzennych należy przyporządkować im odpowiednie wagi. Wielkości wag wpływają bezpośrednio na wyniki prowadzonych analiz. Zaproponowano, aby wagi w badaniu jakości danych przestrzennych, skierowanych na potrzeby użytkownika, były określone metodą hierarchicznego procesu analitycznego – AHP (Saaty 1980).

Pierwszy etap w procesie wyznaczania wag w badaniu baz EGIB, GESUT i BDOT500, z wykorzystaniem metody AHP, obejmuje utworzenie macierzy porównań parami dla pięciu elementów jakości danych (kompletności, spójności logicznej, dokładności położenia, dokładności czasowej i dokładności tematycznej). Preferencje wyrażone są w skali ocen od 1 do 9, gdzie 1 oznacza równoważność porównywalnych elementów, a 9 – ekstremalną preferencję jednego elementu względem drugiego. Badania preferencji elementów jakości danych wykonano z poziomu potrzeb użytkownika (metodą ankietową) w wydziałach inwestycji i planowania urzędów gmin, w powiecie krakowskim. Do dalszej analizy wykorzystano 50 kwestionariuszy ankietowych, wypełnionych przez pracowników urzędów gmin. Wyniki porównania parami elementów jakości danych, wspólne dla baz EGIB, GESUT i BDOT500, przedstawiono w tabeli 13.

Tabela 13. Preferencje jakości danych EGIB, GESUT i BDOT500 z poziomu użytkownika

	Kompletność	Spójność logiczna	Dokładność położenia	Dokładność czasowa	Dokładność tematyczna
Kompletność	1	8/1	5/4	3/1	2/1
Spójność logiczna	1/8	1	1/8	1/5	1/8
Dokładność położenia	4/5	8/1	1	3/1	2/1
Dokładność czasowa	1/3	5/1	1/3	1	1/7
Dokładność tematyczna	1/2	8/1	1/2	7/1	1

Następnym etapem procesu określania wartości wag jest obliczenie wektora własnego macierzy preferencji (Wachstiel 2013). W rezultacie przeprowadzonych obliczeń uzyskano wektor wag. Wektor wag ( $w$ ), w zakresie pięciu elementów jakości danych reprezentujących preferencje użytkownika, przedstawiono poniżej:

$$w = \begin{bmatrix} 0,31 \\ 0,04 \\ 0,28 \\ 0,09 \\ 0,28 \end{bmatrix}$$

Poprawność zastosowania metody AHP do określania wag w badaniu jakości danych baz przestrzennych zweryfikowano za pomocą badań eksperymentalnych. Badania takie przeprowadzono analizując informacje zawarte w bazach EGIB, GESUT i BDOT500 z terenu gminy Michałowice w woj. małopolskim. Jest to dynamicznie rozwijająca się gmina, która bezpośrednio sąsiaduje z miastem Krakowem. Wartości punktowe oceny jakości danych, opisane za pomocą pięciu podstawowych elementów (Bielecka 2010), przedstawiono w tabeli 14. Wartości punktowe przedstawiono w przedziale od 1 do 100. Wagi poszczególnych kryteriów jakości danych wprowadzono na podstawie wektora  $w$ .



Tabela 14. Wartości punktowe oceny jakości danych i wartości obliczonych estymatorów

Kryteria jakości danych	EGiB	GESUT	BDOT500	Wagi	EGiB	GESUT	BDOT500
	Wartości punktowe oceny				Wartości punktowe oceny – ważone		
Kompletność	91	84	77	0,31	28	26	24
Spójność logiczna	98	97	94	0,04	4	4	4
Dokładność położenia	65	52	37	0,28	18	15	10
Dokładność czasowa	92	90	89	0,09	8	8	8
Dokładność tematyczna	73	69	81	0,28	20	19	23
Wartość średnia	84	78	76		78	72	69
Współczynnik zmienności	0,17	0,22	0,30		0,08	0,09	0,10

Wartość średnia, uzyskana ze wszystkich kryteriów jakości danych, pokazuje jaka część danych spełnia założone kryterium 100 punktów. Różnice wielkości oceny jakości danych dla baz EGiB, GESUT i BDOT500, wynoszą odpowiednio 6, 6 i 7 punktów (pomiędzy wartościami średniej i średniej ważonej). Wielkości tych różnic wskazują na zasadność stosowania współczynników ważności w badaniu jakości danych przestrzennych. Dodatkowo, wartości współczynników zmienności dla średniej i średniej ważonej (dla trzech badanych baz danych) wynoszą odpowiednio 0,17 i 0,08; 0,22 i 0,09; 0,30 i 0,10. Wielkości te wskazują, że istnieje spójność wyniku oceny jakości danych z wykorzystaniem wag. W idealnym modelu współczynnik zmienności średnich wynosi zero.

### **Harmonizacja modeli pojęciowych baz EGIB, GESUT i BDOT500 z bazą BDOT10k w aspekcie jakości i zasilania danych**

Dostęp do wiarygodnej i aktualnej informacji przestrzennej oraz możliwości wykorzystania tej informacji w procesach decyzyjnych jest kluczowym problemem w Polsce, jak również w innych krajach europejskich. Jednym ze sposobów rozwiązania tego problemu są działania, które pozwolą na utworzenie krajowych infrastruktur danych przestrzennych. Dzięki takim infrastrukturom możliwe jest osiągnięcie interoperatywności zbiorów danych, metadanych i usług danych przestrzennych (Baranowski 2012). Procesem stosowanym w celu doprowadzenia do zdolności współdziałania komponentów infrastruktury informacji przestrzennej jest harmonizacja. Działania takie w zakresie urzędowych zbiorów danych przestrzennych wpłyną m. in. na poprawę jakości informacji przestrzennej, na unikanie wielokrotnego gromadzenia tych samych danych przez różne instytucje oraz ułatwienie zasilania i wymiany danych pomiędzy rozproszonymi bazami danych.

EGiB, GESUT i BDOT500 są urzędowymi bazami gromadzącymi dane, charakteryzujące się najwyższym stopniem szczegółowości; z tego powodu wykorzystywane są do zasilania innych systemów w procesach ich tworzenia oraz aktualizacji. Szczególnie ważna jest wymiana danych pochodzących z tych rejestrów z bazami danych obiektów topograficznych (BDOT10k) i obiektów ogólnogeograficznych (BDOO). Bazy te gromadzą informacje o położeniu przestrzennym obiektów oraz ich atrybutach opisowych, odpowiadających szczegółowością mapie topograficznej w skali 1:10 000 i skalach mniejszych. Badanie poziomu harmonizacji zbiorów danych EGiB, GESUT i BDOT500 z bazą BDOT10k – w aspekcie jakości i zasilania danych – zrealizowano poprzez analizę modeli pojęciowych przedmiotowych rejestrów danych przestrzennych.

Badanie poziomu harmonizacji zbiorów danych wykonano przez kwalifikację analizowanych klas obiektów do jednej z trzech grup, charakteryzujących zgodność wartości atrybutów:

1. Pełna – wysoka zgodność w zakresie definicji i reprezentacji przestrzennej i opisowej. Pobieranie i zasilanie informacji (w zasadzie) realizowane jest automatycznie.
2. Częściowa – istnieje możliwość pozyskania geometrii obiektów oraz pewnej części informacji opisowych. Wymagane działanie manualne.

### 3. Brak – nie istnieje możliwość automatycznego pobierania i zasilania danych.

Analizując zakres harmonizacji modeli pojęciowych w aspekcie jakości danych przestrzennych, badano obiekty klasy budynki (baza EGiB) oraz obiekty zasilane z baz GESUT i BDOT500. Modele pojęciowe dotyczące jakości danych przestrzennych dla obiektów klasy budynki baz BDOT10k i EGiB, charakteryzują się istotnymi różnicami. Możliwość zasilania własności jakościowych z bazy EGiB do BDOT10k jest ograniczona. BDOT10k ma zdecydowanie bogatszą charakterystykę dotyczącą jakości danych. Przykładowo, w EGiB obiekt budynek nie posiada własności, która wprost charakteryzuje dokładność położenia, a istnieje tylko odwołanie do operatu technicznego (tab. 15). Pełną zgodność stwierdzono tylko dla daty utworzenia obiektu oraz dla nazw źródeł danych o pochodzeniu obiektów.

Tabela 15. Własności dotyczące jakości danych obiektów klasy „budynek” bazy BDOT10k zasilane z EGiB

Baza: BDOT10k		Baza: EGiB		Zgodność wartości
Klasa obiektów: Budynek (OT_BUBD_A)		Klasa obiektów: EGB_Budynek		
Własność - nazwa	Własność - opis	Własność - nazwa	Własność - opis	
Data utworzenia	Data utworzenia obiektu	Start życia obiektu	Data utworzenia obiektu	Pełna
Cykl życia	1. Data utworzenia nowej wersji obiektu oraz modyfikacji jego atrybutów 2. Data przeniesienia wcześniejszej wersji do archiwum	Start życia wersji obiektu	Data utworzenia wersji obiektu	Częściowa
		Koniec życia wersji obiektu	Data przeniesienia wersji obiektu do archiwum	
		Koniec życia obiektu	Data przeniesienia obiektu do archiwum	
Kategoria dokładności geometrycznej	1. Dokładny 2. Przybliżony 3. Niepewny	–	Odwołanie do operatu technicznego	Częściowa
Dokładność geometryczna	Średni błąd położenia obiektu w metrach	–	Odwołanie do operatu technicznego	Częściowa
Źródło danych geometrycznych	EGiB, ortofo, mapa: zasad., topo. i in.	EGiB	Baza EGiB	Pełna
Źródło danych atrybutowych	EGiB, ortofo, mapa: zasad., topo. i in.	EGiB	Baza EGiB	Pełna

Analizując modele pojęciowe dotyczące jakości danych przestrzennych dla obiektów pochodzących z GESUT i BOT500 w odniesieniu do BDOT10k, stwierdzono występowanie istotnych rozbieżności. Zasilanie własności jakościowych z bazy GESUT i BOT500 jest ograniczone. Przykładowo, własności dotyczące dokładności położenia mogą być wykorzystane, ale tylko po wykonaniu dodatkowych operacji uzgadniających. Do takich dodatkowych czynności należy określenie wzajemnych relacji pomiędzy dokładnością geometryczną (BDOT10k), a źródłem danych o położeniu obiektów (GESUT i BOT500) – (tab. 16). Pełną zgodność stwierdzono tylko dla daty utworzenia obiektu oraz dla nazw źródeł danych o pochodzeniu obiektów.

Tab. 16. Własności dotyczące jakości danych obiektów bazy BDOT10k zasilanych z baz GESUT i BDOT500

Baza: BDOT10k Klasa: Obiekt topograficzny (OT_ObjektTopograficzny)		Bazy: GESUT, BDOT500 Klasy: Obiekt GESUT, obiekt BDOT500 (GES_ObjektGESUT), (BDZ_ObjektBDOT500)		Zgodność wartości
Własność - nazwa	Własność - opis	Własność - nazwa	Własność - opis	
Data utworzenia	Data utworzenia obiektu	Start życia obiektu	Data utworzenia obiektu	Pełna
Cykl życia	1. Data utworzenia nowej wersji obiektu oraz modyfikacji jego atrybutów 2. Data przeniesienia wcześniejszej wersji do archiwum	Cykl życia obiektu	Data utworzenia wersji obiektu	Częściowa
		Koniec życia obiektu	Data przeniesienia obiektu do archiwum	
Kategoria dokładności geometrycznej	1. Dokładny 2. Przybliżony 3. Niepewny	Źródło danych o położeniu obiektu	Pomiar na osnowę, pomiar wykrywaczem przewodów, digitalizacja mapy, i in.	Częściowa
Dokładność geometryczna	Średni błąd położenia obiektu w metrach	Źródło danych o położeniu obiektu		Częściowa
Źródło danych geometrycznych	GESUT, BDOT500, ortofoto, mapa: zasad., topo. i. in.	GESUT BDOT500	Baza GESUT lub baza BDOT500	Pełna
Źródło danych atrybutowych	GESUT, BDOT500, ortofoto, mapa: zasad., topo. i. in.	GESUT BDOT500		Pełna

Analizując zakres harmonizacji modeli pojęciowych w aspekcie zasilania danych bazy BDOT10k danymi z EGiB i GESUT, badano dwie klasy obiektów: budynki oraz sieci uzbrojenia terenu.

Baza EGiB jest dla bazy BDOT10k podstawowym źródłem informacji o położeniu przestrzennym budynków i ich atrybutach. Wymagane dokładności położenia dla budynków w BDOT10k (1,5 m) gwarantowane są przez dane pozyskane praktycznie z każdej bazy EGiB. Badając modele pojęciowe dotyczące możliwości pozyskiwania danych dla bazy BDOT10k danymi dotyczącymi budynków z EGiB, stwierdzono występowanie istotnych rozbieżności. Zasilanie danymi z bazy EGiB jest ograniczone. W obu rejestrach zastosowano odmienne identyfikatory, a w EGiB nie istnieje atrybut nazwa własna obiektu. Automatyczne pozyskanie odpowiedniego zakresu budynków będzie ograniczone z powodu różnych definicji obiektów budowlanych – BDOT10k na podstawie prawa budowlanego, w EGiB według Polskiej Klasyfikacji Obiektów Budowlanych.

Baza GESUT stanowi dla bazy BDOT10k zasadnicze źródło informacji o położeniu przestrzennym sieci uzbrojenia terenu (linii napowietrznych i przewodów rurowych) oraz o ich atrybutach. Wymagane w BDOT10k dokładności położenia dla sieci uzbrojenia (1,5 m) zapewnione są przez dane pozyskane zasadniczo z każdej bazy GESUT. Badając modele pojęciowe dotyczące możliwości zasilania danych bazy BDOT10k danymi dotyczącymi sieci uzbrojenia z GESUT, stwierdzono występowanie pewnych rozbieżności. Zasilanie danymi z bazy GESUT nie będzie w pełni zautomatyzowane. W obu rejestrach zastosowano dla obiektów spójne identyfikatory, ale dla GESUT dopuszczono użycie identyfikatorów lokalnych, które nie mogą być wykorzystane przez BDOT10k.

Pozyskanie osi geometrycznych sieci uzbrojenia nie stwarza większych problemów technicznych. Potencjalne trudności będą wynikać m.in. z błędów topologicznych i geometrycznych, na przykład bliskiego sąsiedztwa wierzchołków lub odmiennej segmentacji. Pozyskanie informacji dotyczących wartości części atrybutów, takich jak rodzaj i położenie wymagać będzie wykonania dodatkowych operacji uzgadniających.

Doprowadzenie do wzajemnej spójności modeli pojęciowych urzędowych baz danych przestrzennych jest realnym problemem, a obecny poziom harmonizacji uniemożliwia efektywną wymianę informacji pomiędzy różnymi rejestrami. Bazy EGiB, GESUT i BDOT500 charakteryzują się, w odniesieniu do BDOT10k, niskim poziomem harmonizacji modeli pojęciowych. W wyniku przeprowadzonych badań stwierdzono, że pobieranie oraz zasilanie informacji w procesach tworzenia oraz aktualizacji BDOT10k jest ograniczone. Pozyskiwanie informacji o położeniu przestrzennym obiektów nie stwarza większych problemów technicznych. Niezgodności w modelach pojęciowych analizowanych rejestrów dotyczą głównie atrybutów opisowych. Baza BDOT10k charakteryzuje się większą liczbą klas obiektów oraz posiada bogatszą charakterystykę atrybutów.

Korekta modelu pojęciowego baz EGiB, GESUT i BDOT500 w celu osiągnięcia wysokiego poziomu harmonizacji z modelem BDOT10k jest możliwa do zrealizowania. Potrzebne są tylko odpowiednie, instytucjonalnie działania formalno-prawne.

#### **4.3.6. Wizualizacja zjawiska niepewności w danych przestrzennych**

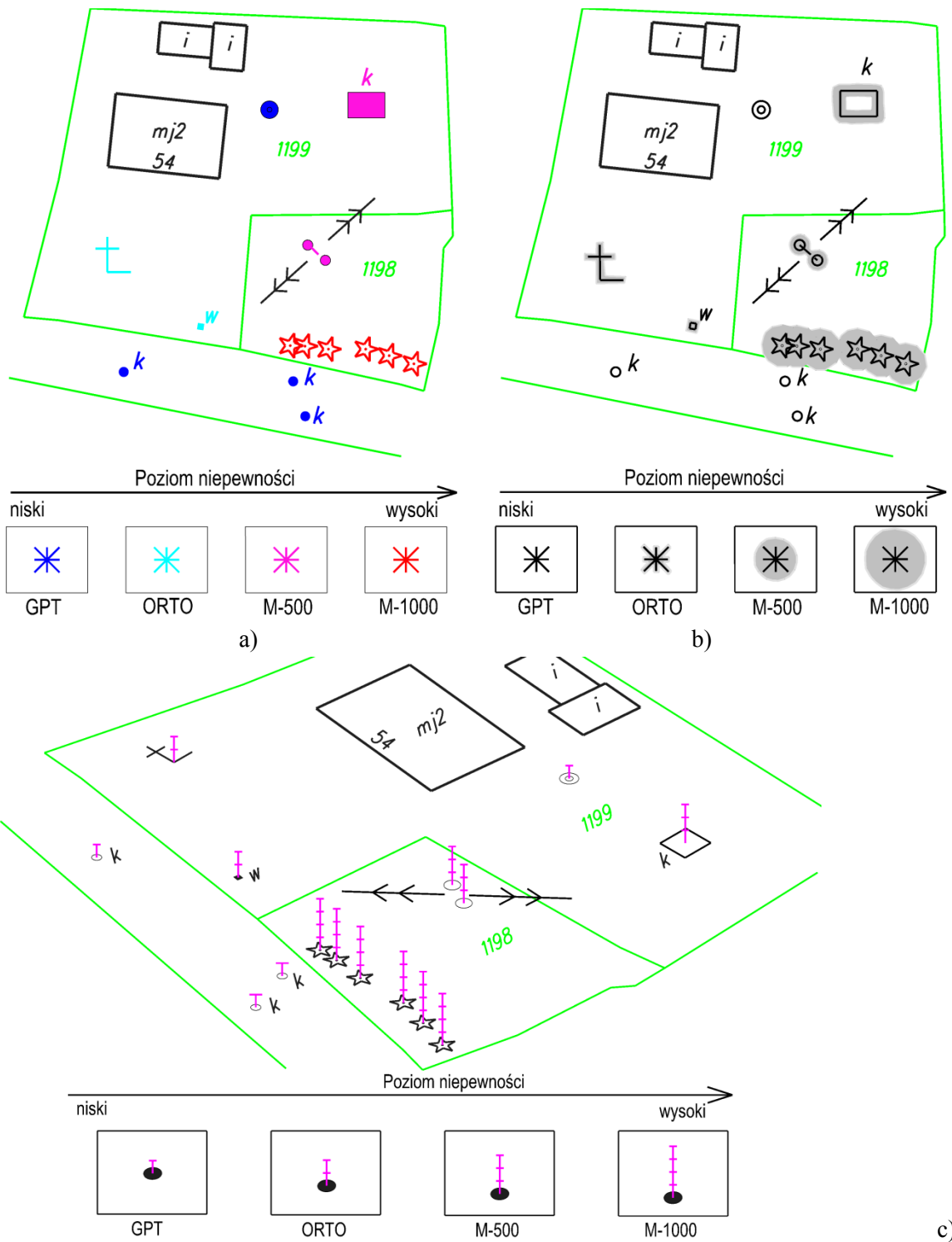
Skutecznym podejściem do problemu przekazywania użytkownikom zbiorów danych przestrzennych, informacji o ich niepewności, są techniki wizualizacji kartograficznej. Techniki takie będą szczególnie efektywne podczas przekazywania informacji o niepewności dla dużych zestawów danych. Kanał wzrokowy człowieka jest czuły w rozpoznawaniu struktury i relacji; grafika oraz prezentacja kartograficzna są naturalnymi środkami wyrażania struktury przestrzennej. Wizualne postrzeganie prezentowanych zjawisk jest skutecznym narzędziem komunikacji, będącym w stanie obsłużyć duże ilości informacji (Goodchild i Clarke 2002).

Na podstawie analizy przedmiotowej literatury (Pang 2001; Kadros 2005; Prassni i in. 2010) oraz własnych doświadczeń i prac badawczych, dotyczących badania jakości danych urzędowych baz przestrzennych, zdefiniowano zbiór metod wizualizacji niepewności danych baz EGiB, GESUT i BDOT500. Zbiór ten zawiera techniki wizualizacji przeznaczone do prezentowania trzech rodzajów niepewności: położenia, wartości atrybutów oraz niepewności czasowej. Niepewność położenia została zdefiniowana, dla obiektów punktowych, liniowych i powierzchniowych, za pomocą kilku (od trzech do pięciu) zmiennych wizualnych. Niepewność wartości atrybutów oraz niepewność czasowa, opisujące (przykładowo) kompletność lub aktualność zbiorów, prezentowane są za pomocą trzech zmiennych graficznych.

Formy zastosowanych gradacji (stopniowania) dla wykorzystanych zmiennych wizualnych zależne są od hierarchii ważności źródeł pochodzenia danych przestrzennych. Prezentowane poniżej (rys. 10 - rys.15) wizualizacje niepewności odnoszą się do czterech źródeł danych: GPT, ORTO oraz M-500 i M-1000.

Niepewność położenia dla obiektów punktowych przedstawiono za pomocą następujących zmiennych wizualnych:

1. Wypełnienie obiektów przy użyciu kolorów i jasności barw: zastosowane wypełnienie jest wprost proporcjonalne do średniego błędu położenia obiektów punktowych (rys. 10a);
2. Rozmycie tła znaków kartograficznych obiektów punktowych: wielkość rozmycia jest wprost proporcjonalna do średniego błędu położenia tych obiektów (rys. 10b);
3. Glify (symbole) umieszczane w punktach wstawiania znaków kartograficznych: wysokości glifów są wprost proporcjonalne do średniego błędu położenia obiektów punktowych (rys. 10c).

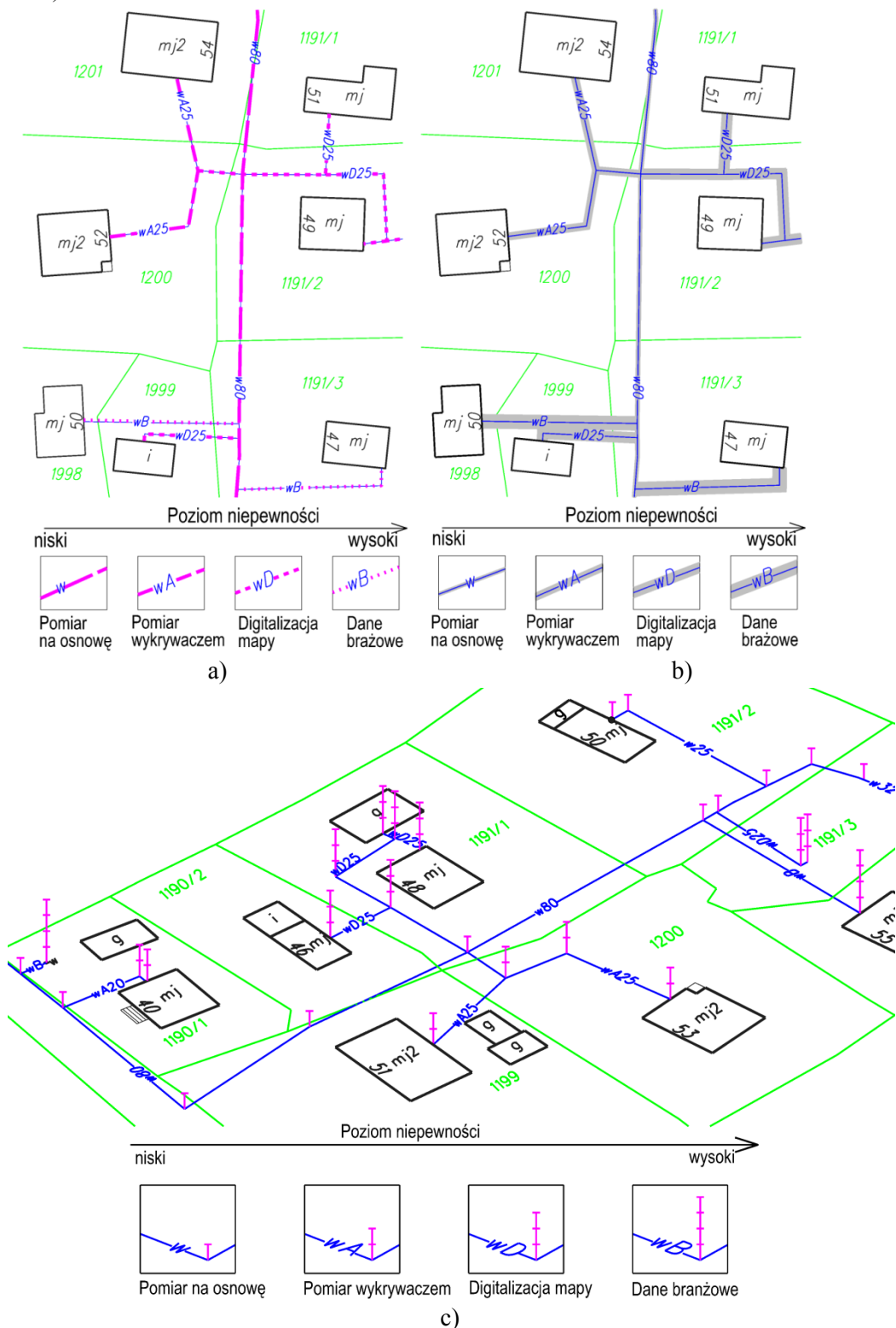


Rys. 10. Niepewność położenia dla obiektów punktowych: a) wypełnienie obiektów za pomocą kolorów i jasności barw, b) rozmycie tła znaków kartograficznych, c) glyfy.

Niepewność położenia dla obiektów liniowych przedstawiono za pomocą następujących zmiennych wizualnych:

1. Konturowanie obiektów liniowych za pomocą linii przerywanych: style linii przerywanych zależne są od średniego błędu położenia punktów definiujących linię (rys. 11a);
2. Bufory rozmycia konturów obiektów liniowych: szerokość bufora rozmycia jest wprost proporcjonalna do średniego błędu położenia punktów charakterystycznych (punkty węzłowe, punkty załamania), (rys. 11b);

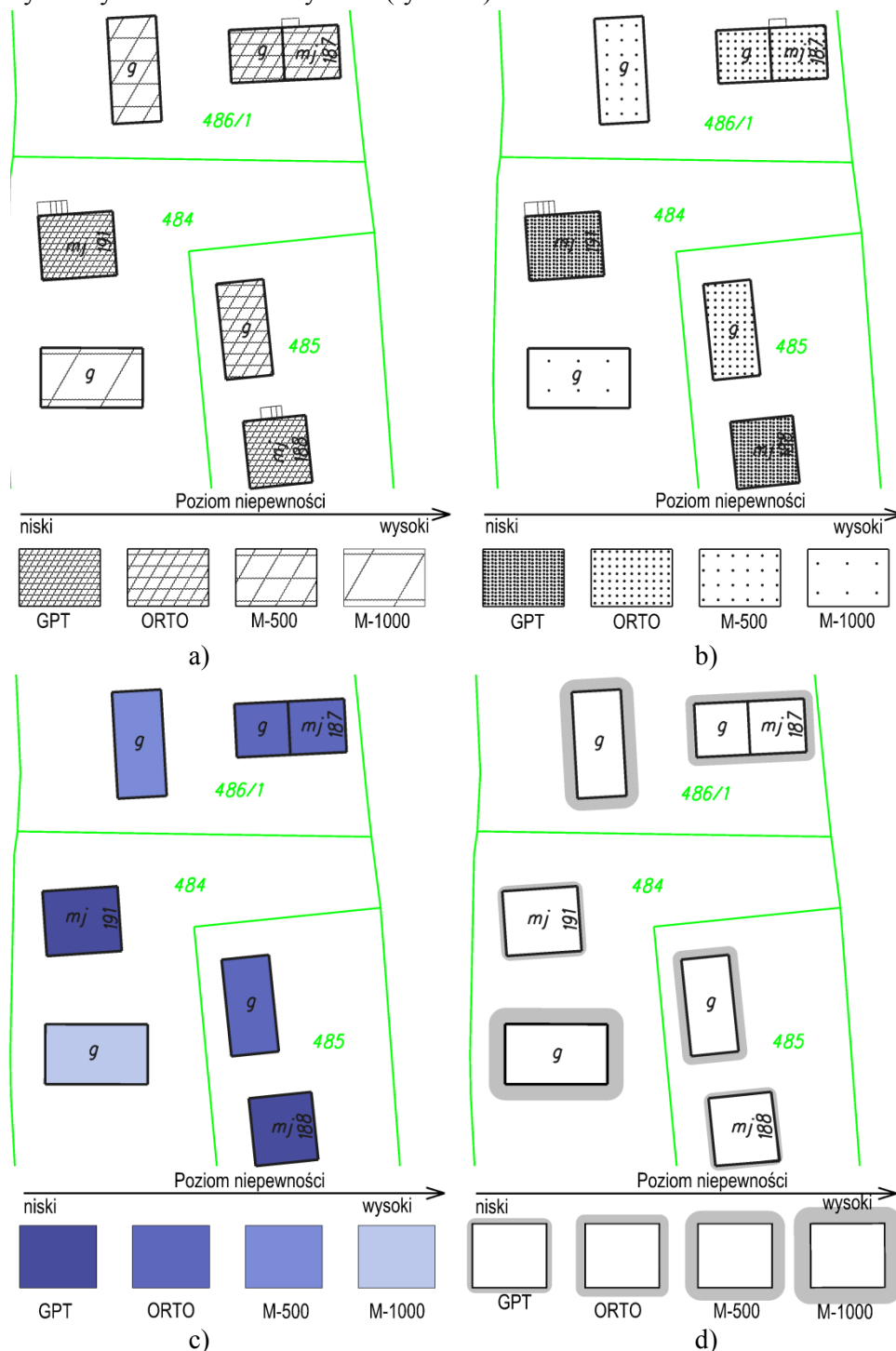
3. Glify (symbole) umieszczone w punktach charakterystycznych: wysokości symboli są wprost proporcjonalne do średniego błędu położenia punktów charakterystycznych obiektów (rys. 11c).



Rys. 11. Niepewność położenia dla obiektów liniowych: a) konturowanie obiektów liniowych, b) bufory rozmycia konturów, c) glify.

Niepewność położenia dla obiektów klasy „budynki” przedstawiono za pomocą następujących zmiennych graficznych:

1. Wypełnienie konturów budynków zagęszczeniem deseni: zastosowano dwa rodzaje deseni, których zagęszczenie zgodne jest z zasadą hierarchicznej struktury danych (rys. 12a-b);
2. Wypełnienie konturów budynków z użyciem waloru koloru (rys. 12c);
3. Bufory rozmycia konturów budynków (rys. 12d).

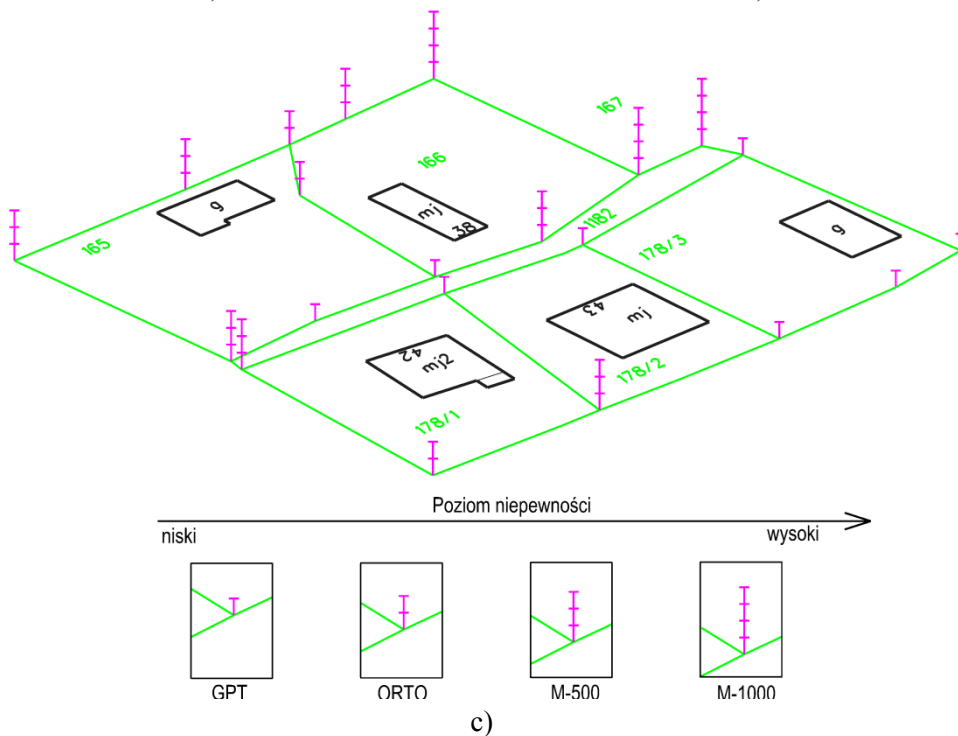
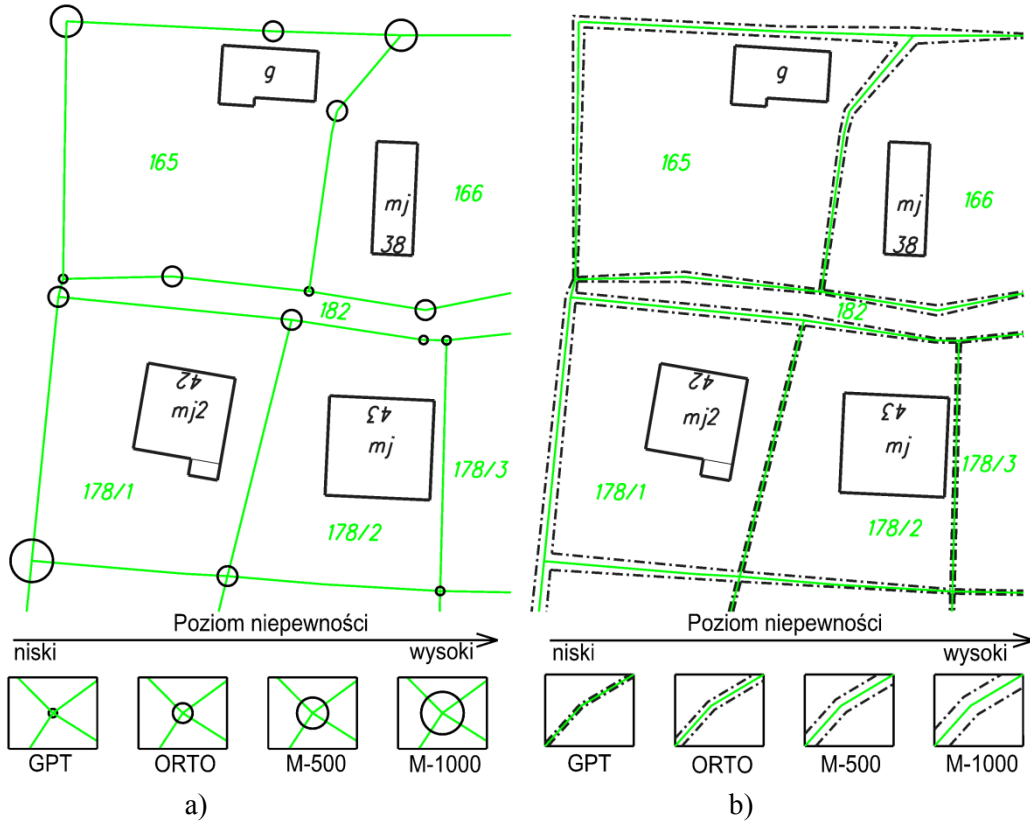


Rys. 12. Niepewność położenia dla obiektów klasy „budynki”: a) i b) wypełnienie konturów budynków zagęszczeniem deseni, c) wypełnienie z użyciem waloru koloru, d) bufory rozmycia konturów.

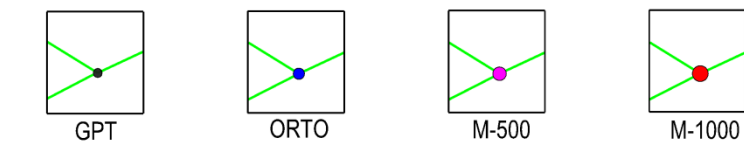
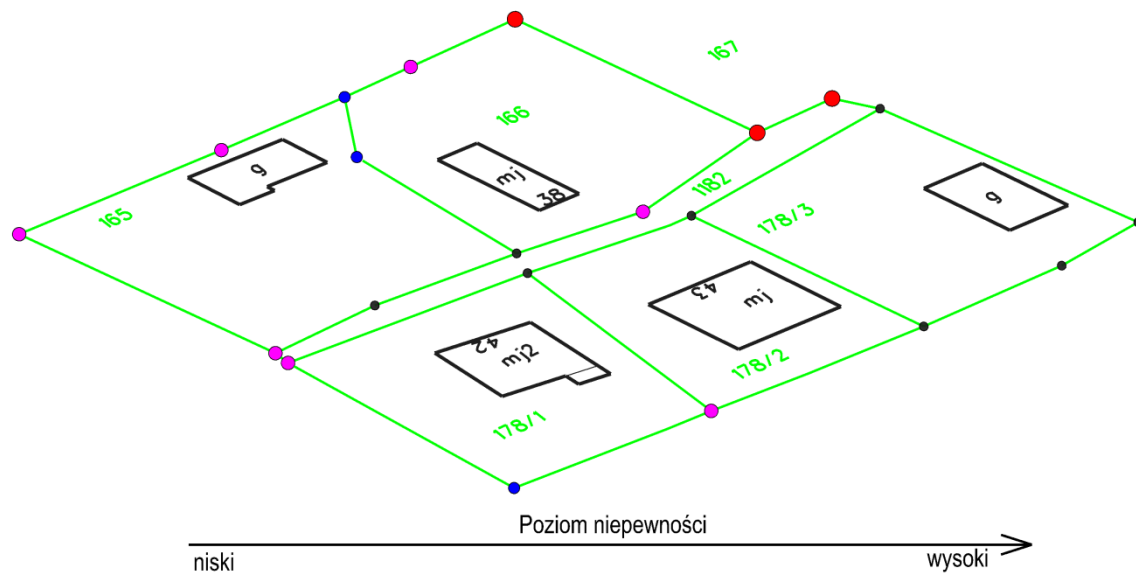
Niepewność położenia dla obiektów klasy „działki ewidencyjne” przedstawiono za pomocą następujących zmiennych graficznych:

1. Okręgi umieszczone w punktach granicznych: wielkości promieni okręgów są wprost proporcjonalne do średniego błędzie położenia punktów granicznych (rys. 13a);

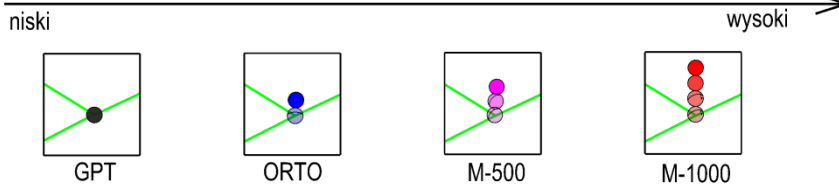
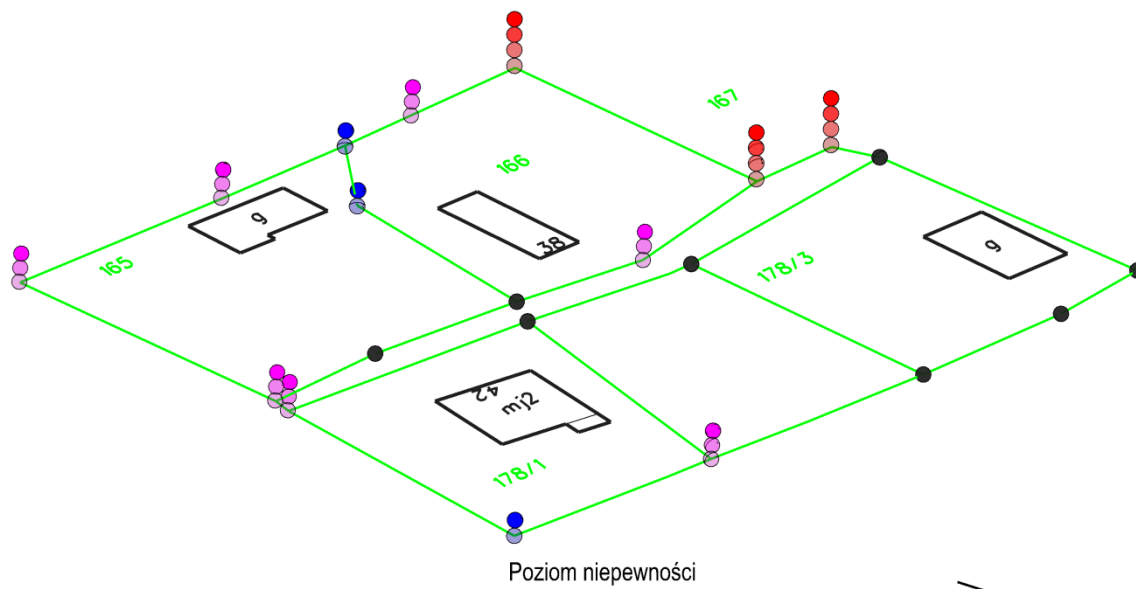
- Odcinki buforowe linii granicznych: zakres buforów linii jest wprost proporcjonalny do średniego błędu położenia punktów granicznych (rys. 13b);
- Glify (symbole) umieszczone w punktach granicznych: wysokości symboli są wprost proporcjonalne do średniego błędu położenia punktów granicznych (rys. 13c);
- Animacja w formie migających pikseli w punktach granicznych: szybkość zmian jest wprost proporcjonalna do wielkości niepewności (rys. 13d);
- Animacja w formie ruchomych pikseli w płaszczyźnie pionowej, zlokalizowanych w punktach granicznych: szybkość zmian jest wprost proporcjonalna do wielkości niepewności (rys. 13e).







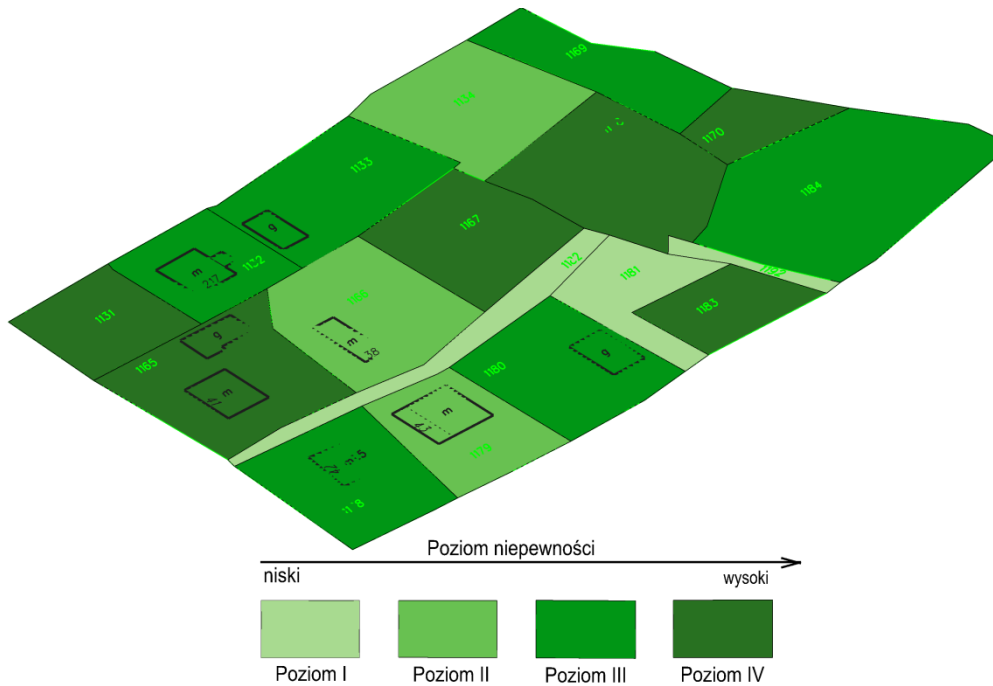
d)



e)

Rys. 13. Niepewność położenia dla obiektów klasy „działki ewidencyjne”: a) okręgi umieszczone w punktach granicznych, b) odcinki buforowe linii granicznych, c) glify, d) i e) animacje w formie migających i ruchomych pikseli.

Niepewność pola powierzchni działki ewidencyjnej przedstawiono przy użyciu następującej zmiennej wizualnej: wypełnienie konturów działek ewidencyjnych za pomocą waloru koloru, proporcjonalnie do dokładności wyznaczonego pola powierzchni działek (rys. 14).

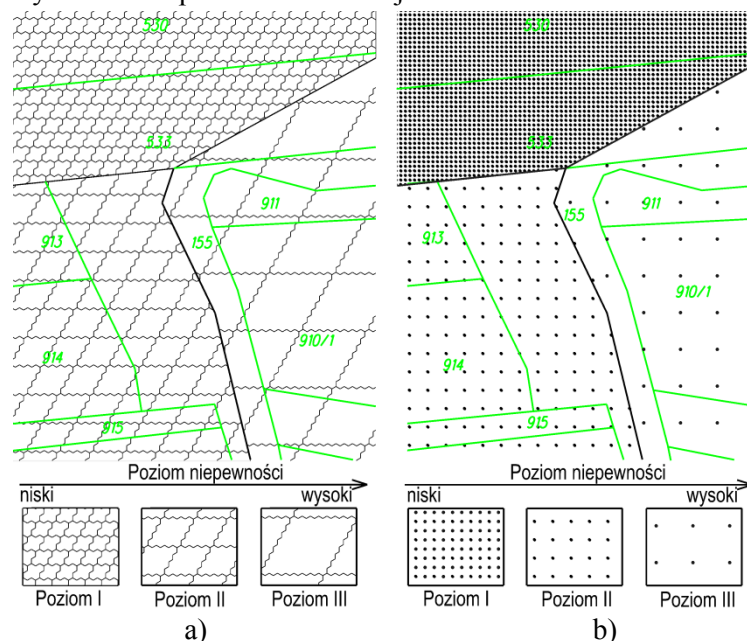


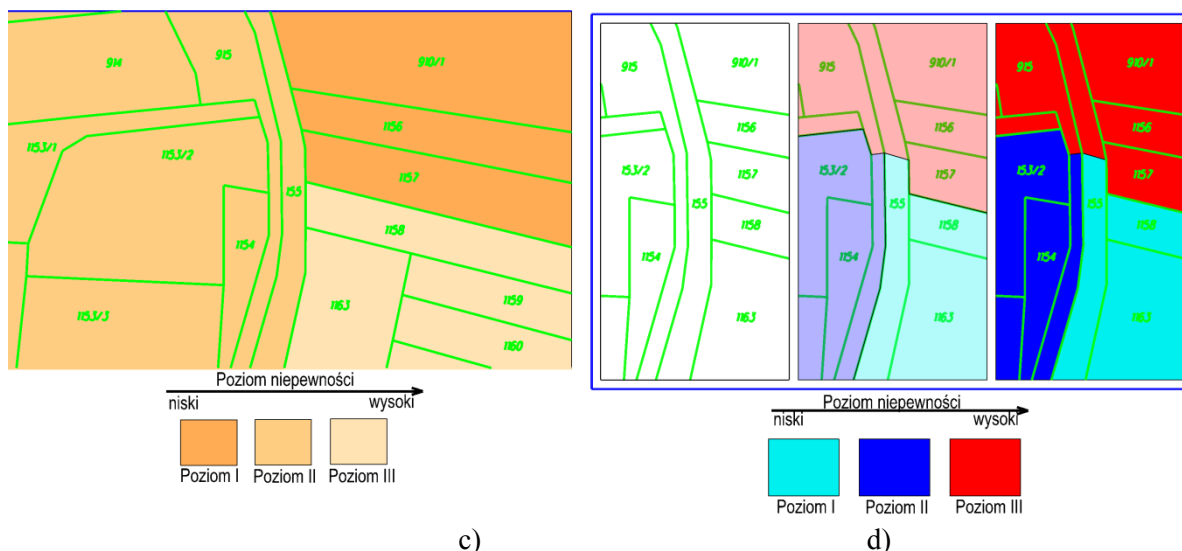
Rys. 14. Niepewność pola powierzchni działki ewidencyjnej przedstawiona przy użyciu wypełnienia konturów działek ewidencyjnych za pomocą waloru koloru.

Niepewność wartości atrybutów oraz niepewność czasową (przykładowo: kompletność lub aktualność zbiorów danych) przedstawiono za pomocą następujących zmiennych wizualnych:

1. Wypełnienie walorem desenia konturów obszarów charakteryzujących się różnymi wielkościami niepewności co do wartości atrybutów oraz niepewności czasowej: zagęszczenie deseni zgodne jest z zasadą hierarchicznej struktury danych; zastosowano dwa rodzaje deseni (rys. 15a-b);
2. Wypełnienie konturów obszarów z użyciem waloru koloru proporcjonalnie do wielkości niepewności atrybutów oraz niepewności czasowej (rys. 15c);
3. Animacja w formie migających obszarów: szybkość zmian jest wprost proporcjonalna do wielkości niepewności atrybutów oraz niepewności czasowej (rys. 15d).

Formy zastosowanych gradacji (stopniowania) zmiennych wizualnych zależne są od hierarchii ważności niepewności atrybutów i niepewności czasowej.





Rys. 15. Niepewność wartości atrybutów i niepewność czasowa: a) i b) wypełnienie konturów zagęszczeniem deseni, c) wypełnienie obszarów z użyciem waloru koloru, d) animacja w formie migających obszarów.

Kompleksową estymację metod wizualizacji wykonano techniką ankietową, przeprowadzoną wśród specjalistów korzystających z baz danych przestrzennych. Zaproponowany zbiór metod wizualizacji niepewności danych baz EGİB, GESUT i BDOT500 testowało i wzajemnie porównywało 75 specjalistów, wykorzystujących zbiory danych przestrzennych w realizowanych pracach badawczych, projektowych i decyzyjnych. Stosując trzy kategorie oceny (liczba klas niepewności danych, preferowana hierarchia ważności oraz preferowana metoda wizualizacji), wybrano najbardziej skuteczne metody wizualizacji (tab. 17).

Tabela 17. Najwyżej ocenione przez specjalistów techniki wizualizacji niepewności danych przestrzennych

Rodzaj obiektu	Rodzaj niepewności	Preferowana metoda wizualizacji
Obiekt punktowy	Położenie	1. Glyfy (symboli), umieszczanych w punktach wstawiania znaków kartograficznych.
Obiekt liniowy	Położenie	1. Konturowanie obiektów za pomocą linii przerywanych. 2. Bufory rozmycia konturów obiektów.
Budynek	Położenie	1. Wypełnianie konturów budynków walorem koloru. 2. Bufory rozmycia konturów budynków.
Działka ewidencyjna	Położenie	1. Okręgi umieszczane w punktach granicznych. 2. Glyfy umieszczane w punktach granicznych.
Obszar opracowania	Niepewność wartości atrybutów i niepewność czasowa	1. Wypełniania konturów obszarów z użyciem waloru koloru.

Ocena metod wizualizacji niepewności danych pokazuje, że techniki oparte na wypełnianiu obszarów zagęszczeniem deseni nie otrzymały najwyższych wskaźników oceny punktowej. Nie zostały docenione również metody dynamiczne: animacje w formie migających pikseli oraz animacje z ruchomymi pikselami w płaszczyźnie pionowej. Techniki wizualne wykorzystujące desenie oraz animacje przeznaczone są do ich użytkowania na nośnikach cyfrowych (monitory komputerów i tableatów). Skuteczna (jakościowa i merytoryczna) wizualizacja niepewności danych baz EGİB, GESUT i BDOT500 przy użyciu tych metod, dla obszarów o dużych powierzchniach (np. obręb ewidencyjny), wymaga zastosowania wysoce wydajnego sprzętu i oprogramowania komputerowego. W związku z tym, brak odpowiednio efektywnych narzędzi informatycznych był główną przyczyną niskiej oceny technik dynamicznych oraz metod wypełniania obszarów zagęszczeniem deseni.

Najczęściej stosowaną obecnie metodą, służącą do przekazywania użytkownikom informacji o niepewności danych przestrzennych są zbiory metadanych. Innym, skutecznym podejściem do przekazywania użytkownikom zbiorów danych informacji o ich niepewności, są techniki wizualizacji kartograficznej. Opracowany przez autora pracy zbiór metod wizualizacji zjawiska niepewności w danych przestrzennych może być z powodzeniem wykorzystywany przez użytkowników (specjalistów z różnych branż), nie mających świadomości o ograniczonej jakości danych przestrzennych. Techniki te będą szczególnie efektywne podczas przekazywania informacji o niepewności dla dużych zestawów danych.

#### **4.3.7. Podsumowanie i wnioski końcowe**

Jakość danych gromadzonych w urzędowych bazach danych przestrzennych ma kluczowe znaczenie przy podejmowaniu decyzji strategicznych, jak również przy realizacji prac planistycznych i projektowych. Świadomość poziomu jakości tych danych istotna jest również dla indywidualnych użytkowników urzędowych danych przestrzennych. W pracy postawiono hipotezę badawczą: *szybko rosnąca dostępność, wymiana i wykorzystanie danych przestrzennych oraz ich niejednorodność powodują konieczność poszukiwania skutecznych metod i modeli oceny ich jakości*. Opracowane przez autora metody i modele opisu oraz ewaluacji jakości danych przestrzennych, gromadzonych w rejestrach publicznych, są dowodem na postawioną tezę.

Na podstawie wyników autorskich badań z zakresu analizowanej tematyki badawczej można sformułować następujące wnioski końcowe:

#### **Wnioski z zakresu oceny dokładności danych gromadzonych w urzędowych bazach danych przestrzennych:**

- W zakresie określania dokładności danych przestrzennych metodami statystyki klasycznej (po usunięciu obserwacji odstających) należy stosować własności rozkładu log-normalnego, dla danych pochodzących z geodezyjnych pomiarów terenowych i ortofotomapy. Rozkład błędów położenia obiektów pochodzących z map w skalach 1:500, 1:1000 i 1:2000 zbliżony jest do rozkładu Weibulla.
- W procesie analizy dokładności danych przestrzennych metodami statystyki odpornościowej z powodzeniem można stosować zmodyfikowane metody: najmniejszych odchyłeń absolutnych, duńską oraz zasadę wyboru alternatywy (ZWA).
- Odporna metoda częstości względnej – opracowana przez autora pracy – stanowi skuteczną technikę określania dokładności danych przestrzennych, których źródłem są geodezyjne pomiary terenowe, ortofotomapa oraz mapy zasadnicze w skalach 1:500, 1:1000 i 2000.

#### **Wnioski z zakresu niepewności pola powierzchni działek ewidencyjnych:**

- W bazie ewidencji gruntów i budynków powinny znajdować się informacje o bezwzględnej niepewności pól powierzchni działek ewidencyjnych. Niepewność tą, dla celów wykonania globalnej analizy, można określić proponowaną metodą przybliżoną.
- W przypadku wykonywania ponownych pomiarów punktów granicznych działek ewidencyjnych, realizowanych równorzędnymi technikami (w aspekcie dokładności), błąd graniczny korekty pola powierzchni w bazie EGiB równy jest wartości dwukrotnego błędu wyznaczenia pola, obliczonego na podstawie różniczkowania wzoru Gaussa. Jako błąd położenia punktu granicznego należy przyjąć jego dopuszczalną wartość dla danej techniki pomiarowej.

### **Wnioski z zakresu analizy ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu:**

- Istnieją trzy główne czynniki wpływające na wartość ryzyka uszkodzenia sieci podziemnego uzbrojenia terenu, spowodowanego niską jakością danych przestrzennych. Czynniki te to: brak kompletności zbiorów danych przestrzennych oraz niewystarczająca dokładność określenia położenia poziomego i położenia pionowego elementów podziemnego uzbrojenia.
- Punktowa metoda oceny ryzyka projektowego oraz metoda oparta na wnioskowaniu rozmytym zapewniają odpowiednie określenie poziomu ryzyka ilościowego i jakościowego przy realizacji inwestycji budowlanych (liniowych i kubaturowych).
- Opracowane wielkoskalowe mapy ryzyka 2D i 3D są wydajnymi narzędziami wspomagającymi realizację bezpiecznego procesu budowlanego.

### **Wnioski z zakresu modelu jakości danych gromadzonych w urzędowych bazach:**

- Dane zasilaające urzędowe bazy danych przestrzennych powinny podlegać kontroli ilościowej oraz jakościowej (automatycznej, kameralnej i terenowej). Dla celów kontroli jakości danych baz EGIB, GESUT i BDOT500 określono rodzaje weryfikacji jakościowej, minimalne liczebności prób oraz dopuszczalne liczebności elementów niezgodnych w próbach statystycznych.
- W zakresie potrzeb użytkowników baz danych przestrzennych, najważniejszymi kryteriami jakości danych są: kompletność zbiorów oraz dokładność położenia i dokładność tematyczna. Dokładność czasowa i spójność logiczna nieznacznie wpływają na ocenę jakościową tych baz danych.
- W przypadku oceny jakości danych z punktu widzenia użytkownika stopnie istotności elementów jakości danych optymalnie można określić metodą AHP (hierarchicznego procesu analitycznego).

### **Wnioski z zakresu wizualizacji zjawiska niepewności w danych przestrzennych:**

- Skutecznym podejściem w przekazywaniu użytkownikom zbiorów danych przestrzennych, informacji o niepewności danych, są techniki wizualizacji kartograficznej.
- Techniki statyczne należą do wydajnych metod dla obszarów o dużych powierzchniach, takich jak np. obręb ewidencyjny.
- Glify (symbole) najlepiej prezentują niepewność dla obiektów punktowych. Konturowanie za pomocą linii przerywanych oraz buforów rozmycia konturów obiektów są skuteczne dla obiektów liniowych. Dla budynków do optymalnych metod wizualizacji należą: wypełnianie konturów walorem koloru oraz buforów rozmycia konturów. Dla działek, wydajnymi technikami są okręgi oraz glify umieszczane w punktach granicznych.

### Za najważniejsze osiągnięcia związane z badaniami prezentowanymi w monografii można uznać:

1. Opracowanie odpornej na obserwacje odstające metody oceny dokładności danych przestrzennych nazwanej „Odporną metodą częstości względnej”.
2. Określenie wzoru na wyznaczenie błędu granicznego dla korekty pola powierzchni w bazie EGIB – w przypadku wykonywania ponownych pomiarów punktów granicznych działek ewidencyjnych, realizowanych równorzędnymi technikami (w aspekcie dokładności).
3. Opracowanie dwóch metod oceny ryzyka projektowego (punktowej oraz opartej na wnioskowaniu rozmytym), zapewniających odpowiednie określenie poziomu ryzyka przy realizacji inwestycji budowlanych, wraz z wielkoskalowymi mapami 2D i 3D tego ryzyka.
4. Zdefiniowanie modelu jakości danych jako zbioru zasad służących do oceny i raportowania jakości danych przestrzennych baz urzędowych.
5. Opracowanie zbioru metod wizualizacji zjawiska niepewności w danych przestrzennych.

## Literatura

- Atkinson K., Han W. 2009. *Theoretical Numerical Analysis*, Springer Science + Business Media, New York.
- Ayhan E., Akar O. E., Uzun S., Dilaver A., Kansu O. 2011. Analysis of digital data obtained from raster and vector maps. *Journal of Surveying Engineering* vol. 137, 65–69.
- Baranowski M. 2012. *Infrastruktura informacji przestrzennej w ujęciu systemowym*. Wyd. IGIK Warszawa, Seria monograficzna nr 15.
- Bielecka E. 2010. Zasady oceny danych przestrzennych oraz ich zastosowanie do oceny jakości danych gromadzonych w TBD, *Rocznik Geomatyki*. T. VIII, z. 4(40), 53–66.
- 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*, 63, 245–256.
- Deren L, Jingxiong Z, Huayi W. 2012. Spatial data quality and beyond. *International Journal of Geographical Information Science* Vol. 26, No. 12, 2277–2290.
- Doskocz A. 2013. *Metodyka oceny dokładności wielkoskalowych map cyfrowych*. Wyd. UWM w Olsztynie.
- ESDIN. 2010. D8.4 ESDINQuality Final Report – Part A. [http://www.esdin.eu/sites/esdin.eu/files/D8-4\\_ESDIN\\_Quality\\_Final\\_Report.pdf](http://www.esdin.eu/sites/esdin.eu/files/D8-4_ESDIN_Quality_Final_Report.pdf).
- FGDC. 1998. Federal Geographic Data Committee. *Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy*.
- Gleim I. N. 2004. *CIA Review – Part I – The Internal Audit Activity’s Role in Governance, Risk and Control*, Certified Internal Auditor - Gleim Publications Inc., Gainesville FL.
- Goodchild M.F., Clarke K. 2002. Data quality in massive datasets, in: *Handbook of Massive Datasets*, Abello J., Pardalos P.M. and Resende M.G.C. (red.), Kluwer Academic Publishers, Norwell, 643–659.
- Greń J. 1984. *Statystyka matematyczna. Modele i zadania*. PWN, Warszawa.
- Hejmanowska B. 2013. *Zastosowanie rozkładu Laplace’a do określania niepewności danych przestrzennych na przykładzie NMT i systemu IACS*. Uczelniane Wyd. Naukowo-Dydaktyczne AGH w Krakowie.
- ISO. 1999. ISO 2859-1: Sampling procedures for inspection by attributes - Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection.
- ISO. 2013. ISO 19157: Geographic information – Data quality.
- Kadaj R. 1980. Rozwinięcie koncepcji niestandardowej metody estymacji. *Geodezja i Kartografia*, t. XXIX, z. 3/4, 185–196.
- Kaplan S., Garrick B.J. 1981. On the quantitative definition of risk. *Risk Analysis*, Vol. 1, No 1, 11–27.
- Kardos J., Benwell G., Moore A. 2005. The Visualisation of Uncertainty for Spatially Referenced Census Data Using Hierarchical Tessellations. *Transactions in GIS*. Vol. 9/1, 19–34.
- Krupp T., Kubik K. 1983. The Danish method. Experience and philosophy. *Deutsche Geodätische Kommission. Reihe A*, 98, 131–134.
- Liao S.B., Bai Y. 2010. A new grid-cell-based method for error evaluation of vector-to-raster conversion. *Computational Geosciences*, 14, 539–549.
- Niemi W. 2013. *Symulacje stochastyczne i metody Monte Carlo*, Wyd. UW w Warszawie.
- Oort P., Stein A., Bregt A.K., de Bruin S., Kuipers, J. 2005. A variance and covariance equation for area estimates with a GIS. *Forest Science* 51(4), 347–356.
- Pang A. 2001. Visualizing Uncertainty in Geo-spatial Data. In: *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*.
- Piegat A. 2003. *Modelowanie i sterowanie rozmyte*. EXIT Warszawa.

- Prassni, J., Ropinski, T., Hinrichs, K. 2010. Uncertainty-aware guided volume segmentation. IEEE transactions on visualization and computer graphics, 16(6), 1358–1365.
- Ratcliffe J.H., 2001. On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. International Journal of Geographical Information Science, 15, 473–485.
- Wachstiel Ł. 2013. Zastosowanie metody AHP do wyboru optymalnego zintegrowanego systemu informatycznego wspomagającego zarządzanie uczelnią. Studia Ekonomiczne. Nr 159, 109–123.
- Yager R.R., Filev D. P. 1995. Podstawy modelowania i sterowania rozmytego, WNT Warszawa.
- Zadeh L.A. 1965. Fuzzy sets. Information and Control, vol. 8, 338–353.
- Zandbergen P.A. 2008. Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy. Transactions in GIS, 12(1), 103–130.

## 5. Omówienie pozostałych osiągnięć naukowo - badawczych

Badania prowadzone przed obroną pracy doktorskiej dotyczyły głównie dwóch zakresów tematycznych:

1. Metodyki precyzyjnego określania współrzędnych przestrzennych punktów kontrolowanych w badanych przemieszczeń.
2. Tworzenia systemu informacyjnego opisującego nieruchomości.

W zakresie pierwszego tematu badawczego wykonano analizę dokładności wyznaczania współrzędnych przestrzennych badanych punktów, porównując dotychczas stosowane metody pomiaru deformacji przestrzennych osuwisk, z proponowaną, nową technologią takiego pomiaru. Pomiaru badawcze na obiektach testowych (kopalnie odkrywkowe gliny ceramicznej) wykazały, że precyzyjna niwelacja trygonometryczna, połączona z równoczesnym wyznaczaniem współrzędnych płaskich X i Y metodą biegunową stanowi technologię konkurencyjną dla geometrycznej niwelacji precyzyjnej i metody wcięć liniowo-kątowych. W celu automatyzacji prac kameralnych zaprogramowano moduły obliczeniowe, służące do wyznaczania współrzędnych przestrzennych badanych punktów. Technologia została wdrożona na terenie kopalni odkrywkowej „Zesławice” w Krakowie-Nowej Hucie.

Wybrane publikacje z tego tematu:

- Borowiecki I., Ślusarski M. 1993. Wpływ refrakcji na pomiary różnic wysokości metodą niwelacji trygonometrycznej. Zeszyt. Nauk. AR w Krakowie, Nr 289, Seria Geodezja Z. 14, Kraków, 3-11.
- Borowiecki I., Ślusarski M. 1993. Warunki atmosferyczne jako czynniki determinujące dokładność pomiaru odległości dalmierzem elektrooptycznym. Zeszyt. Nauk. AR w Krakowie, Nr 289, Seria Geodezja Z. 14, Kraków, 13-23.
- Borowiecki I., Ślusarski M. 1994. Przestrzenna metoda biegunowa w zastosowaniu do badania przemieszczeń mas ziemnych. Przegląd Geodezyjny Nr 4, 17-19.
- Borowiecki I., Ślusarski M. 1994. Wyznaczenie współrzędnych przestrzennych punktów kontrolowanych metodą tachimetrii precyzyjnej. Zeszyt. Nauk. AR w Krakowie, Nr 293, Sesja Naukowa Z. 42, Kraków, 45-54.

W ramach drugiego tematu badawczego opracowano system informacyjny, opisujący nieruchomości w zakresie przestrzennym i atrybutowym. Opracowano nową koncepcję sposobu wykorzystania danych o nieruchomościach gromadzonych w różny sposób i w różnych miejscach (integracja rozproszonych baz danych). Przeprowadzono badania zasobu informacyjnego dla rozproszonych baz danych opisujących nieruchomości. Badania doprowadziły do opracowania optymalnego rejestru atrybutów i danych opisujących nieruchomości dla przyszłego systemu informacji o nieruchomościach.

W następnym etapie realizacji projektu badawczego opracowano mechanizmy przesyłania i przetwarzania danych do tworzonego systemu zarządzania informacjami przestrzennymi i

opisowymi o nieruchomościach. Zrealizowano relacyjną bazę danych INFOLAND, jako samodzielną aplikację komputerową.

Prace badawcze z tego zakresu realizowane były w ramach dwóch grantów finansowanych przez Komitet Badań Naukowych; autor jako kierownik projektu (Nr 9 T 12E 016 07) oraz główny wykonawca (Nr 54 T12 96 11). Opracowanie koncepcji pozyskiwania i doboru danych w celu zasilania systemu informacji o nieruchomościach było podstawowym celem rozprawy doktorskiej, obronionej w 2002 r. przed Radą Wydziału Inżynierii Środowiska i Geodezji, Akademii Rolniczej w Krakowie.

Wybrane publikacje z tego tematu:

- Ślusarski M. 1995. Numeryczna mapa atrybutów nieruchomości gruntowych i budowlanych. Zeszyt. Nauk. AR w Krakowie, Nr 296, Sesja Naukowa Z. 44, Kraków, 183-186.
- Ślusarski M. 1997. Teoretyczne podstawy budowy systemu informacji o nieruchomościach dla celów powszechnej taksacji nieruchomości – modelowanie koncepcyjne. Archiwum Fotogrametrii Kartografii i Teledetekcji. Vol. 6, Kraków s. 245-254.
- Borowiecki I., Ślusarski M. 1997. Propozycja tworzenia systemu informacyjnego o rynku nieruchomości. ZN AR w Krakowie, Nr 325, Seria Geodezja Nr 16, Kraków, 25-30.

Po uzyskaniu stopnia doktora kierunki zainteresowań autora dotyczyły głównie trzech zakresów tematycznych:

1. Ewaluacji systemów geoinformacyjnych;
2. Wykrywania i korekty błędów topologicznych obiektów powierzchniowych;
3. Budowy modelu katastru wielowymiarowego.

W zakresie ewaluacji systemów geoinformacyjnych przedstawiono propozycję metodyki kompleksowego badania własności urzędowych systemów, gromadzących informacji o przestrzeni, oraz wykonano badania wartości użytkowej geoportali internetowych. Zaproponowano metodę ogólnej oceny systemu informacji o terenie, budowanego na poziomie powiatu. Ewaluację systemu oparto na czterech własnościach: dokładności, kompletności, aktualności i zgodności. Ocenę funkcjonowania systemu (jako całości) przeprowadzano poprzez obliczenie wartości przeciętnej absolutnej oraz współczynnika zmienności średnich. Informacje o jakości danych, prezentowane za pomocą proponowanych estymatorów, stanowią wiarygodny komponent baz metadanych systemów geoinformacyjnych zróżnicowanych pod względem organizacyjnym i w różnym stopniu z informatyzowanych.

Opracowano również metodę szczegółowej ewaluacji systemów informacji o terenie, opartą na normie PN-EN-ISO 19113 – umożliwiającą sporządzenie precyzyjnego i wiarygodnego raportu oceny. Raport taki powinien zawierać atrybuty opisujące: przeznaczenie, pochodzenie, wykorzystanie, kompletność, spójność i dokładność danych. Dodatkowo, należy określać: zakres i miarę jakości danych, procedurę oceny jakości oraz typ i jednostkę wartości jakości. Wynikiem oceny jakości danych powinny być współczynniki, takie jak średnia absolutna oraz współczynnik zmienności średnich. Opracowana metoda pozwala na porównywanie informacji zgromadzonych w bazach geoinformacyjnych zróżnicowanych pod względem organizacyjnym i utrzymywanych przez odmiennie zorganizowane systemy informatyczne.

Badano również wartości użytkowe lokalnych geoportali internetowych na terenie całego kraju. Zakres ocenianych informacji został podzielony na trzy grupy: przejrzystość, interaktywność oraz dostępność danych geoprzestrzennych. Wykonane prace doświadczalne pokazały znaczące różnice w wartościach globalnej oceny poszczególnych geoportali. Proponowaną metodę zastosowano z powodzeniem w ewaluacji witryn internetowych miast wojewódzkich z zakresu informacji przestrzennej.



Wybrane publikacje z tego zakresu tematycznego:

- Ślusarski M. 2005. Propozycja ogólnej oceny systemu informacji o terenie budowanego na poziomie powiatu. *Geodezja, Kartografia i Aerofotogrametria* Z. 66. Lwów, 82-85.
- Ślusarski M. 2009. Metodyka opisu jakości danych przestrzennych w aspekcie norm ISO. *Geodezja, Kartografia i Aerofotogrametria* Z. 71. Lwów, 136-139.
- Ślusarski M. 2012. Doskonalenie jakości informacji w powiatowych systemach geoinformacyjnych. *Infrastruktura i ekologia terenów wiejskich*. Nr 1/III, 41-47.
- Siejka M., Ślusarski M. 2012. Próba oceny informacji geoportalowych na przykładzie wybranych miast polski. *Czasopismo Techniczne*. 2-Ś/2012, 227-236.
- Siejka M., Ślusarski M. 2013. Analiza jakościowa danych Geodezyjnej Ewidencji Sieci uzbrojenia Terenu. *Infrastruktura i ekologia terenów wiejskich*. Nr 2/III, 119-126.
- Siejka., Ślusarski M. 2014. Ocena geoportali internetowych powiatów, na wybranych przykładach oraz według ustalonych kryteriów. *Infrastruktura i ekologia terenów wiejskich*. Nr 2/II, 545-555.

Jednym z podstawowych elementów jakości danych jest spójność logiczna, którą opisuje się m. in. za pomocą spójności topologicznej. Wykrywanie i korekta błędów topologicznych obiektów powierzchniowych jest aktualnym problemem badawczym z powodu permanentnego tworzenia i aktualizowania baz danych gromadzących informacje przestrzenne. Opracowano technologię wykrywania i likwidacji błędów topologicznych, dotyczącą urzędowych baz danych (kataster, LPIS). Opracowano metodykę pracy na obiektach wirtualnych, w środowisku CAD software. Obiekty wirtualne to poprawnie utworzone zbiory jednostek elementarnych: odcinków (jako reprezentantów granic obiektów powierzchniowych) oraz identyfikatorów obszarów.

Technologia może być z powodzeniem stosowana dla przestrzennych baz danych, gdzie opracowywane są duże ilości obiektów, takich jak bazy EGIB, GESUT, BDOT500 oraz LPIS. Zaproponowany algorytm kontroli poprawności topologicznej obiektów powierzchniowych został wdrożony przy aktualizacji baz danych systemu LPIS, prowadzonego przez Ministerstwo Rolnictwa od 2012 roku. Technologia ta została pozytywnie zweryfikowana podczas prowadzonych prac.

W zakresie omawianej tematyki badawczej zawiera się również technologia wykrywania i automatycznej likwidacji błędów topologicznych, dotyczących braku spójności baz danych systemu LPIS. Wykorzystano metodykę pracy na obiektach wirtualnych w środowisku CAD software. Opracowana metodyka umożliwiła automatyczną korektę błędów wynikających z niespójności dwóch baz: bazy pól zagospodarowania zawierającej informacje o sposobie użytkowania gruntów, względem bazy referencyjnej (katastralnej). Metodyka ta jest częścią technologii podnoszącej jakość i wydajność kontroli systemów baz danych LPIS.

Wybrane publikacje z tego tematu:

- Siejka M., Ślusarski M., Zygmunt M. 2013. Correction of topological errors in geospatial databases. *International Journal of Physical Sciences*. Vo.8(12),498-507.
- Siejka M., Ślusarski M., Zygmunt M. 2014. Verification technology for topological errors in official databases, with the case study in Poland. *Survey Review*, Vol. 46 (No334), 50-57.
- Zygmunt M., Siejka M., Ślusarski M., Siejka Z., Piech I., Bacor S. 2015. Database inconsistency errors correction, on example of LPIS databases in Poland. *Survey Review*, Vol. 47 (No343), 256-264.

Od kilku lat prowadzone są w Polsce rozważania na temat potrzeb i możliwości wprowadzenia katastru wielowymiarowego. Na podstawie prowadzonych autorskich badań zaproponowano sposób przejścia z katastru 2D na system „3D + czas”. Zadanie to zostało zrealizowane za pomocą dostępnych narzędzi informatycznych oraz w oparciu o istniejące urzędowe rejestry danych przestrzennych. Kluczowe, dla realizacji przedsięwzięcia w systemie „3D + czas”, jest wykorzystanie informacji o

rzeczywistym ukształtowaniu terenu. Budowa modelu uproszczonego (na powierzchni płaskiej), bez uwzględniania ukształtowania terenu, nie daje możliwości praktycznego wykorzystania systemu 3D.

Jednym ze sposobów uzyskania informacji o ukształtowaniu terenu jest numeryczny model terenu (DTM) pochodzący ze skaningu laserowego, niezbędny do poprawnej budowy przestrzennego rejestru sieci uzbrojenia terenu. Tworzony obecnie GESUT umożliwia z wystarczającą dokładnością budowę przestrzennego katastru sieci uzbrojenia terenu i może być wykorzystany do celów obrotu prawami do dysponowania danym urządzeniem, a także do celów projektowych, podatkowych i innych. Powstanie przestrzennego GESUT daje możliwość do utworzenia urzędowego rejestru właścicieli poszczególnych odcinków urządzeń infrastruktury podziemnej.

Prawny obowiązek zapisu daty pierwszej rejestracji i wszystkich kolejnych zmian (sprzedaż, podział działki, budowa obiektu i innych zmian w zagospodarowaniu przestrzeni) w odniesieniu do działki, budynku i urządzeń infrastruktury, dają punkt wyjścia do utworzenia katastru „3D + czas”. Do tego celu mogą być wykorzystane współczesne informatyczne narzędzia bazodanowe i CAD lub GIS. Tak zbudowany kataster może być skutecznym narzędziem wykorzystywanym w zarządzaniu przestrzenią: planowanie przestrzenne, ochrona środowiska, ochrona zasobów naturalnych, rozwój urbanizacji i infrastruktury.

Wybrane publikacje z tego tematu:

- Siejka M., Ślusarski M., Zygmunt M. 2014. 3D + time Cadastre, possibility of implementation in Poland. Survey Review, Vol. 46 (No335), 79-89.
- Cienciała A., Florek R., Dawidowicz A., Żróbek R., Hopfer A., Pietrzak L., Cegielski S., Ślusarski M., Siejka M., Zygmunt M., Śliwiński Ł., Parzych P., Bydłosz J., Karabin M. 2016. Kierunki rozwoju polskiego katastru na tle rozwiązań światowych. Monografia. Wyd. Politechnika Warszawska.
- Jurkiewicz M., Ślusarski M. 2017. Selected aspects of creation of a multi-dimensional cadastre in Poland. Infrastruktura i Ekologia Terenów Wiejskich. Nr 2017/ IV (1), 1483-1491.

*Marcel Ślusarski*