Beata Hejmanowska

Zastosowanie rozkładu Laplace'a do określania niepewności danych przestrzennych na przykładzie NMT i systemu IACS

Application of Laplace distribution in spatial data uncertainty estimation DTM and IACS examples

WYDAWNICTWA AGH

KRAKÓW 2013

KU 0523 publikacja Wydawnictw Akademii Górniczo-Hutniczej
im. Stanisława Staszica w Krakowie

© Wydawnictwa AGH, Kraków 2013
ISBN 978-83-7464-649-9

Redaktor Naczelny Wydawnictw AGH: *Jan Sas*

Komitet Naukowy Wydawnictw AGH:
Zbigniew Kąkol (przewodniczący),
Marek Cała,
Borys Mikułowski,
Tadeusz Sawik,
Mariusz Ziółko

Recenzenci: *dr hab. inż. Anna Barańska, prof. AGH*
dr hab. inż. Robert Olszewski, prof. PW

Afiliacje autorki monografii:
AGH Akademia Górniczo-Hutnicza

Wydrukowano z materiałów i składu dostarczonych przez autorkę

Projekt okładki i strony tytułowej: *mgr inż. Piotr Kramarczyk*

Redakcja Wydawnictw AGH
al. Mickiewicza 30, 30-059 Kraków
tel. 12 617 32 28, tel./faks 12 636 40 38
e-mail: redakcja@wydawnictwoagh.pl
http://www.wydawnictwa.agh.edu.pl

Spis treści

Spis tabel **5**

Spis rysunków **7**

Symbole, oznaczenia, skróty **13**

Streszczenie **17**

Summary **21**

1 Wprowadzenie **25**

2 Sposoby wyrażania dokładności danych przestrzennych **29**

2.1. Wyrażanie niepewności pomiaru w metrologii 32

2.2. Określanie niepewności danych przestrzennych 37

2.3. Model koncepcyjny jakości danych przestrzennych 40

2.4. Dane przestrzenne 42

2.5. Dane opisowe 49

2.6. Dane przetworzone - NMT 50

2.7. Przedstawienie problemu określania niepewności danych . 51

3 Metodyka **57**

4 Określanie niepewności w systemie IACS **67**

4.1. Pomiar referencyjny TS 71

4.2. Pomiar walidacyjny GNSS 79

4.3. Pomiar walidacyjny IKONOS 83
4.4. Kontrola jakości LPIS 1 89
4.5. Kontrola jakości LPIS 2 94

5 Określanie niepewności numerycznych modeli terenu 101
5.1. Analiza jakości NMT ALS 104
5.2. Analiza jakości NMT LPIS 114
5.3. Analiza jakości NMT ISOK 123
5.4. Porównanie NMT LPIS i ISOK 130

6 Podsumowanie i wnioski końcowe 137

Bibliografia 143

Spis tablic

1 Baza reguł 47

2 Zestawienie powierzchni i obwodów referencyjnych działek (Kramarczyk, Warchoł 2011) 73

3 Zestawienie powierzchni i obwodów referencyjnych działek (Hanower) 81

4 Parametry opisujące rozbieżność pomiędzy NMT a chmurą punktów [m] (Hejmanowska, Kay 2011) 105

5 Błąd średni kwadratowy określający wpasowanie rozkładu Gaussa, rozkładu Laplace'a i funkcji odpornej do histogramu odchyłek [m] (Hejmanowska, Kay 2011) 109

6 Odwrotność dystrybuanty jako wskaźnik dokładności względnej, [m] (Hejmanowska, Kay 2011) 110

Spis rysunków

1 Interpretacja graficzna różnicy pomiędzy niepewnością a) a błędem b) (Główny Urząd Miar 1999) 33

2 Model koncepcyjny niepewności danych przestrzennych (zmodyfikowany za Fisher 1999, 2005, zmodyfikowany za Klir i Yuan 1995) 38

3 Model koncepcyjny jakości danych przestrzennych 41

4 Przykład granicy działki rolnej dobrze zdefiniowanej i rozmytej (JRC MARS 2013 b) 42

5 Przykład niejednoznacznej granicy działki (JRC MARS 2013 b) 43

6 Rozmyty proces decyzyjny 45

7 Etapy klasyfikacji dla działki o nieokreślonej granicy . . 48

8 Przykład histogramu odchyłek NMT i wysokości pomierzonych *in-situ* wraz z różnymi rozkładami prawdopodobieństwa (zmodyfikowany za Höhle, Höhle 2009) 55

9 Wykresy funkcji gęstości prawdopodobieństwa: f_1, f_2, f_3 a), dla danych wygenerowanych symulatorem liczb losowych, zakładając rozkład normalny N(0,1), parametry statystyczne b) oraz wykres normalności: Q-Q c); przebieg funkcji f_1 i f_2 jest prawie identyczny, niewielkie różnice są widoczne dla wartość odciętych bliskich 0 65

10 Przykładowy zestaw działek wykorzystywanych w walidacji GPS (Błonia, Kraków) (Kramarczyk, Warchoł 2011) . 72

11 Histogram odchyłek powierzchni dla pomiaru TS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) . . 74

12 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni z pomiaru TS 75

13 Histogram odchyłek względnych powierzchni dla pomiaru TS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 76

14 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek względnych powierzchni z pomiaru TS 77

15 Przykładowy zestaw działek wykorzystywanych w walidacji pomiaru powierzchni działek z wykorzystaniem urządzeń GNSS (Hanower, Niemcy) 80

16 Histogram odchyłek powierzchni dla pomiaru walidacyjnego GNSS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 82

17 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni walidacyjnego GNSS 83

18 Przykładowy zestaw działek wykorzystywanych w walidacji obrazu satelitarnego (IKONOS) (Milenov i Gabalova 2009) 84

19 Histogram odchyłek powierzchni dla pomiaru walidacyjnego IKONOS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 85

20 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni w pomiarze walidacyjnym IKONOS - wszystkie odchyłki 87

21 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni w pomiarze walidacyjnym IKONOS - 90% odchyłek, a), b); 80% odchyłek, c), d) 88

22 Ilustracja przebiegu granicy działki w systemie LPIS (kolor biały) i rzeczywistej granicy na ortofotomapie (kolor szary) 90

23 Histogram odchyłek powierzchni dla pomiaru kontrolnego LPIS 1 wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 91

24 Wykres Q-Q wraz z błędami wpasowania, analiza wszystkich odchyłek powierzchni z pomiaru kontrolnego LPIS 1 92

25 Wykres Q-Q wraz z błędami wpasowania, analiza 90% odchyłek powierzchni z pomiaru kontrolnego LPIS 1 93

26 Histogram odchyłek bezwzględnych powierzchni dla pomiarów kontrolnych z wykorzystaniem RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 96

27 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek bezwzględnych [ha] powierzchni z pomiaru kontrolnego RTK 97

28 Histogram odchyłek względnych powierzchni dla pomiarów kontrolnych z wykorzystaniem RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 98

29 Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek względnych powierzchni z pomiaru kontrolnego RTK . . 99

30 Lokalizacja obszarów testowych JRC, Ispra, Włochy, (Hejmanowska, Kay 2011) 105

31 Rozkład przestrzenny rozbieżności pomiędzy NMT i chmurą punktów. Ilustracja u góry przedstawia obszar nr 5, o urozmaiconej rzeźbie, a ilustracja dolna obszar nr 2, płaski (Hejmanowska, Kay 2011) 106

32 Histogram odchyłek - obserwowany, funkcja gęstości prawdopodobieństwa Gaussa, funkcji odpornej i funkcji Laplace'a, wykres rozbieżności,wykres Q-Q dla obszaru numer 5 (Hejmanowska, Kay 2011) 108

33 Dokładność zewnętrzna/bezwzględna NMT ALS, wykres Q-Q, histogram rozbieżności obserwowanych, modelowane funkcje gęstości prawdopodobieństwa (Hejmanowska, Kay 2011) 110

34 Analiza dokładności wewnętrznej NMT ALS okolic Błoń (Kraków), histogram empiryczny w skali logarytmicznej i liniowej, wykres Q-Q, parametry statystyczne 111

35 NMT ALS okolic Błoń (Kraków), mapa różnicowa NMT ALS minus chmura punktów 112

36 Analiza parametrów statystycznych mapy różnicowej uzyskanej w wyniku odjęcia modeli NMT wykonanych dwoma różnymi algorytmami, histogram empiryczny w skali logarytmicznej, wykres Q-Q, parametry statystyczne 113

37 Lokalizacja obszaru testowego Mstów, fragment mapy topograficznej (geoportal.gov.pl) 115

38 Lokalizacja obszaru testowego Mstów, fragment ortofotomapy (geoportal.gov.pl) 116

39 Histogram różnic wysokości pomiędzy NMT LPIS i chmurą punktów wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 118

40 Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i chmurą punktów . . . 119

41 Histogram różnic wysokości pomiędzy NMT LPIS i RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 120

42 Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i RTK 121

43 Odchyłki pomiędzy NMT LPIS i "chmurą punktów" w zakresie jednego odchylenia standardowego, (z lewej, kolor zielony (-0,250 m; 0,250 m), kolor niebieski poniżej -0,250 m, kolor czerwony powyżej 0,250 m) oraz punktami RTK (z prawej, kolor zielony (-0,456 m ; 0,456 m), kolor niebieski poniżej - 0,456 m, kolor czerwony powyżej 0,456 m) 122

44 Histogram różnic wysokości pomiędzy NMT ISOK i chmurą punktów wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 123

45 Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT ISOK i chmurą punktów . . . 124

46 Histogram różnic wysokości pomiędzy NMT ISOK i RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 125

47 Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT i RTK 126

48 Rozkład przestrzenny odchyłek pomiędzy NMT ISOK a chmurą punktów z progiem rozszerzenia niepewności równym 0,07m 128

49 Rozkład przestrzenny odchyłek bezwzględnych NMT ISOK RTK 129

50 Histogram różnic wysokości pomiędzy NMT LPIS i NMT ISOK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3) 130

51 Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i NMT ISOK - wszystkie punkty 131

52 Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i NMT ISOK - 80% punktów 132

53 Rozkład przestrzenny różnic pomiędzy NMT LPIS i NMT ISOK - rozkład Gaussa; kolorem zielonym przedstawione są obszary na poziomie ufności 95% (-1,294 m ; 1,294 m), kolorem niebieskim obszary poniżej -1,294 m, a w kolorze czerwonym powyżej 1,294 m 134

54 Rozkład przestrzenny różnic pomiędzy NMT LPIS i NMT ISOK - rozkład Laplace'a; kolorem zielonym przedstawione są obszary na poziomie ufności 95% (-1,090 m ; 1,090 m), kolorem niebieskim obszary poniżej -1,090 m, a w kolorze czerwonym powyżej 1,090 m 135

55 Rozkład przestrzenny różnic pomiędzy NMT LPIS i NMT ISOK - metoda odporna; kolorem zielonym przedstawione są obszary na poziomie ufności 95% (-0,590 m ; 0,590 m), kolorem niebieskim obszary poniżej -0,590 m, a w kolorze czerwonym powyżej 0,590 m 136

Symbole, oznaczenia, skróty

Poniżej wyjaśnione zostały wykorzystywane w monografii skróty, oznaczenia i pojęcia.

ALS – Airborne Laser Scanning - lotniczy skaning laserowy

ASPRS – American Society of Photogrammetry and Remote Sensing - Amerykańskie Towarzystwo Fotogrametrii i Teledetekcji

GNSS – Global Navigation Satellite System - globalny system nawigacji satelitarnej

GPS – Global Positioning System - globalny system pozycjonowania

CODGIK – Centralny Ośrodek Dokumentacji Geodezyjnej i Kartograficznej

DN – Digital Number - liczbowa wartość zapisana dla pojedynczego piksela obrazu

IACS – Integrated Administration Control System - Zintegrowany System Zarządzania i Kontroli (ZSZIK)

IKONOS – satelita obrazujący powierzchnię Ziemi z rozdzielczością 1 m w wersji panchromatycznej i 4 m w wersji multispektralnej - w monografii obraz satelitarny z tego satelity

WV2 – WorldView-2 - satelita obrazujący powierzchnię Ziemi, o rozdzielczości w nadirze, w wersji panchromatycznej 0,46 m i multispektralnej 1,8 m - w monografii obraz satelitarny z tego satelity

ISOK – Informatyczny System Osłony Kraju

INSPIRE – INfrastructure for SPatial InfoRmation in the European Community (pol. Infrastruktura Informacji Przestrzennej we Wspólnocie Europejskiej) – dyrektywa 2007/2/WE Parlamentu Europejskiego Unii Europejskiej z dnia 14 marca 2007, opublikowana w Dzienniku Urzędowym 25 kwietnia 2007 i obowiązująca od 15 maja 2007 roku, ustanawiająca jednolitą infrastrukturę informacji przestrzennej

JRC – Joint Research Centre - Wspólne Centrum Badawcze - Dyrekcja Generalna Komisji Europejskiej, z misją zapewnienia wsparcia naukowego i technicznego dla koncepcji, rozwoju, wdrażania i monitorowania polityki Unii Europejskiej, w tym systemu IACS i dyrektywy INSPIRE, w monografii oddział JRC w Isprze, we Włoszech

LPIS – Land Parcel Identification System - System Identyfikacji Działek Rolnych, referencyjna baza danych dla systemu IACS

NMT – Numeryczny Model Terenu

NMPT – Numeryczny Model Pokrycia Terenu

NMT GRID – Numeryczny Model Terenu w modelu siatkowym

NMT TIN – Numeryczny Model Terenu w postaci trójkątowej

OKI – Ośrodki Koordynacyjno-Informacyjne przy RZGW zajmujące się miedzy innymi tworzeniem map zagrożenia i ryzyka powodziowego

RTK – Real Time Kinematic GPS - metoda pomiaru różnicowego GPS wyznaczania pozycji odbiornika ruchomego poprawianej w czasie rzeczywistym z wykorzystaniem danych obserwacyjnych ze stacji bazowej

RZGW – Regionalny Zarząd Gospodarki Wodnej

TS – Total Station

Przewodnik – opublikowane w 1993 roku, w wersji angielskiej, wytyczne dla potrzeb metrologii: "Wyrażanie niepewności pomiaru - przewodnik"

rozkład Gaussa – rozkład normalny Gaussa-Laplace'a, nazwa skrócona, przyjętą z powodu porównywania z rozkładem Laplace'a

$\boldsymbol{x}$ – odchyłka pomiędzy wartością pomierzoną a referencyjną, w różnych sytuacjach oznacza również błąd lub niepewność

liczba punktów – całkowita liczba obserwacji branych pod uwagę w analizie dokładności

liczba wartości odstających – liczba obserwacji usuniętych z dalszych analiz

min – minimalna zaobserwowana wartość odchyłki

max – maksymalna zaobserwowana wartość odchyłki

μ – błąd średni, średnia wartość odchyłki, w rozdziale 4.3 również funkcja przynależności

σ – odchylenie standardowe

$1,64 \cdot \sigma$ – rozszerzenie odchylenia standardowego (połowa szerokości przedziału ufności, na poziomie ufności 90%), w rozkładzie Gaussa

$1,96 \cdot \sigma$ – rozszerzenie odchylenia standardowego (połowa szerokości przedziału ufności, na poziomie ufności 95%), w rozkładzie Gaussa

RMSE – Root Mean Square Error - błąd średni kwadratowy

$1,96 \cdot RMSE$ – rozszerzenie błędu średniego kwadratowego (połowa szerokości przedziału ufności, na poziomie ufności 95%), w rozkładzie Gaussa

k – miara skupienia

s – miara asymetrii

m – mediana

P – percentyl

P(0,683) – percentyl 68,3%

P(0,95) – percentyl 95%

NMAD – znormalizowane odchylenie bezwzględne mediany

b – parametr funkcji Laplace'a

$2^{0,5} \cdot b$ – pierwiastek z wariancji dla rozkładu Laplace'a

F^{-1} – odwrotność dystrybuanty

$F_1^{-1}(0,975)$ – odwrotność dystrybuanty rozkładu Gaussa odpowiadająca funkcji f_1

$F_2^{-1}(0,975)$ – odwrotność dystrybuanty rozkładu Laplace'a odpowiadająca funkcji f_2

$F_3^{-1}(0,975)$ – odwrotność dystrybuanty rozkładu Gaussa odpowiadająca funkcji f_3

$RMSE_{f1}$ – błąd średni kwadratowy dopasowania histogramu do rozkładu Gaussa zakładając μ i σ

$RMSE_{f2}$ – błąd średni kwadratowy dopasowania histogramu do rozkładu Laplace'a zakładając medianę i b

$RMSE_{f3}$ – błąd średni kwadratowy dopasowania histogramu do rozkładu Gaussa zakładając medianę i NMAD

p-value – powierzchnia rzeczywistego obszaru krytycznego w teście normalności, porównana z założonym poziomem istotności

Streszczenie

Inspiracją badań, zaprezentowanych w niniejszej monografii, były obserwacje poczynione w trakcie prac w ramach projektu dotyczącego przetwarzania danych z lotniczego skaningu laserowego (Hejmanowska i in. 2008 a). Produktem końcowym był NMT (Numeryczny Model Terenu), którego jakość próbowano oszacować. Dobierano różne oczka siatki i analizowano różnice pomiędzy NMT i chmurą punktów ze skaningu laserowego. Obserwowano między innymi histogramy odchyłek pomiędzy NMT a punktami pomiarowymi ALS. Okazało się, że kształt wszystkich histogramów tych odchyłek znacznie odbiega od kształtu funkcji rozkładu normalnego. Histogramy były bowiem silnie skoncentrowane wokół zera i miały "długie ogony", co kojarzyło się "na pierwszy rzut oka" z rozkładem Laplace'a.

W wyniku późniejszych studiów literaturowych znaleziono potwierdzenie powyższego zjawiska oraz propozycję modyfikacji rozkładu normalnego dla potrzeb analiz dokładność NMT z pomiarów ALS i lotniczej fotogrametrii (Höhle i Höhle 2009). W publikacji tej została zaproponowana tzw. metoda odporna (*robust method*), jako alternatywa w stosunku do powszechnie wykorzystywanego podejścia, czyli charakteryzowania dokładności za pomocą odchylenia standardowego, czy błędu średniego kwadratowego bazującego na rozkładzie normalnym. Po przeanalizowaniu wyników prezentowanych w publikacji, czyli histogramów

odchyłek, kształtu odpowiadającej im krzywej Gaussa i przebiegów krzywych ilustrujących proponowaną, odporną funkcję gęstości prawdopodobieństwa, nasuwa się wniosek, że krzywa Gaussa znacznie przeszacowuje błąd NMT, ale metoda odporna go niedoszacowuje. Wniosek ten potwierdził sensowność poszukiwań innej funkcji rozkładu gęstości prawdopodobieństwa. Zaproponowany został rozkład Laplace'a, początkowo jedynie na podstawie analizy wizualnej.

Badania nad jakością NMT, realizowane głównie w kontekście wpływu jego dokładności na wynik analiz przestrzennych, prowadzone były od pewnego czasu (Hejmanowska 2005, Hejmanowska i in. 2008 a i b). Wpływ ten jest możliwy do uwzględnienia w analizach przestrzennych, jednak, żeby o nim mówić trzeba wiedzieć, w jaki sposób wiarygodnie określić jakość NMT. Badania dotyczące określania jakości modelu NMT doprowadziły do publikacji (Hejmanowska i in. 2008 b), w której przedyskutowano różne definicje i podejścia do określania jakości, wiarygodności, dokładności, niedokładności, błędu i niepewności NMT. Za interesującą można uznać klasyfikację zaproponowaną przez Fishera (1999), wychodzącą od pojęcia **niepewność**, którą w przypadku dobrze zdefiniowanego NMT można byłoby określić jako **błąd**. Jeśli NMT nie został dobrze zdefiniowany, to niepewność mogłaby być opisana takimi pojęciami jak: nieokreśloność i niejednoznaczność.

Równocześnie w środowisku geodezyjnym zaczęły pojawiać się informacje na temat zaleceń dotyczących nieużywania pojęcia **błąd**, zamiast którego zaproponowano właśnie określenie **niepewność**. W roku

1993 opublikowano w wersji angielskiej Przewodnik (polska wersja - Główny Urząd Miar 1999), w którym zawarto założenia dotyczące wyrażania niepewności pomiaru. Główny nacisk w Przewodniku jest położony na warunki pomiaru, które bardzo rzadko bywają w pełni kontrolowane. Błąd jest zarezerwowany dla przypadku idealnych warunków pomiaru oraz zakładając, że obiekt jest "mierzalny" i znamy wartość referencyjną cechy obiektu, która jest mierzona. Wydaje się, że w różnych sytuacjach pomiar geodezyjny nie jest wykonywany w kontrolowanych warunkach laboratoryjnych, a obiekt bywa czasem bardzo trudny do zdefiniowania i pomiaru referencyjnego. Dobrym przykładem może być NMT z jego naturalnym mikro- i makroreliefem, zmieniającym się pod wpływem pór roku, czy zabiegów agrotechnicznych.

W obu wymienionych materiałach źródłowych (Fisher 1999, 2005, Główny Urząd Miar 1999) pojawia się pojęcie niepewności, przy czym jest ono nieco inaczej rozumiane. W związku z tym dokonano próby połączenia tych dwóch podejść w model koncepcyjny jakości danych przestrzennych, w którym:

1. Jeśli warunki pomiaru i obiekt są możliwe do poznania i idealnie opisane to stosujemy pojęcie **błąd**.
2. Jeśli warunki pomiaru i obiekt są dobrze określone, ale nie wiemy, czy jest to sytuacja idealna moglibyśmy używać pojęcia **niepewności**.
3. W przypadku gdy obiekt i/lub warunki pomiaru nie są wystarczająco zdefiniowane, mamy do czynienia z **nieokreślonością**, w przy-

padku której stosuje się zwykle analizę z wykorzystaniem zbiorów rozmytych. Jako parametr jakości danych można by przyjąć szerokość przedziału wyostrzania.

4. Ostatnim przypadkiem jest sytuacja, w której obiekt i/lub warunki pomiaru są niejednoznacznie zdefiniowane. Nie oznacza to, że takie dane są całkowicie bezużyteczne, a jedynie tyle, że nie znamy parametrów opisujących ich jakość i taka informacja powinna być zamieszczona jako metadane.

W monografii skoncentrowano się na dwóch pierwszych, "mierzalnych" przypadkach, czyli na sytuacji idealnej, kiedy jakość danych możemy określić pojęciem **błąd** oraz kiedy pomiar wykonywany jest w warunkach naturalnych i mamy do czynienia z **niepewnością**. W obu przypadkach, jako parametr charakteryzujący jakość danych przyjmuje się, zgodnie z Przewodnikiem, odchylenie standardowe błędu/niepewności. Zakłada się przy tym milcząco, że rozkład błędów/niepewności ma charakter rozkładu normalnego. W monografii poddano pod dyskusję zasadność takiego założenia, przeanalizowano publikowaną w literaturze metodę odporną i jednocześnie zaproponowano parametry rozkładu Laplace'a dla potrzeb określania niepewności danych charakteryzujących się wąskimi histogramami odchyłek od wartości referencyjnych, z "długimi ogonami". W monografii przeprowadzono dyskusję na temat sposobów określania jakości danych na podstawie dwóch typów danych przestrzennych: wektorowych obiektów powierzchniowych i NMT w postaci regularnej siatki.

Summary

Research presented in the monograph was inspired by project concerned processing of airborne laser scanning data (Hejmanowska et al. 2008 a). Quality of Digital Terrain Model (DTM) was analyzed as project final product. Differences between DTM and heights of the points from the points' cloud were analyzed, histograms of the differences were observed. Shape of the histograms significantly differs form normal distribution. Histograms were concentrated around zero and so called "long tiles" appeared. Laplace distribution function was a first association. This phenomenon of specific histograms was confirmed in after performed literature studies. In the paper (Höhle, Höhle 2009) robust method was proposed as an alternative to the Gauss approach. Analysis of the empirical histograms, Gauss distribution and curve based on proposed, robust method one can come to the conclusion that normal distribution actually overestimate the DTM error however robust method underestimate it. The conclusion confirms sense of looking for other distribution function. Author of the monograph proposed Laplace distribution function, at the beginning only basing on the visual analysis. Researches concerning DTM quality were performed until now mainly focusing the influence of the DTM error on results of spatial analysis (Hejmanowska B. 2005, Hejmanowska et al. 2008 a, b). The error propagation is possible to be considered in spatial analysis. However firstly reliable method for DTM

quality assessment is needed. Research in this domain was summarized in the paper (Hejmanowska et al. 2008 b), where different approaches to DTM quality, reliability, accuracy, inaccuracy, uncertainty and error was defined and discussed. Interesting approach was proposed by Fisher (1999). He stared from uncertainty, as general. In the case of well defined DTM uncertainty can be described by error. If DTM is not well defined uncertainty can be expressed by vagueness or ambiguity. Parallel in surveying environment some suggestions appeared concerning not use expression: error but just uncertainty. In 1993 assumptions and recommendations for the expression of uncertainty in measurements were published in English, as important for metrology guideline: "Evaluation of measurement data — Guide to the expression of uncertainty in measurement". Polish version was published in 1999 by Central Office of Measures. Main pressure is placed in the Guideline on measurements conditions, rarely fully controlled. Expression: error is reserved for ideal measurement situation, assuming the object is measurable and reference value is available. It seems the surveying is mainly not performed in control laboratory conditions and the object is often difficult to defining and reference measurement. DTM is a good example with his natural micro- and macro relief varying in a year and according agriculture treatment. In two mentioned source papers (Fisher 1999, 2005 and Guideline) expression: uncertainty appears but understood in different ways. Therefore synthetic approach was proposed of the in the monograph, as a conceptual model for expression of spatial data quality where:

1. If the measurement conditions and the object are possible to ideal

description we can use expression: error.

2. If the measurement conditions and the object are well defined, but we are not sure the situation is possible to ideal description we should use expression: uncertainty.
3. In the case the object and/or measurement conditions are not well defined, we can talk about vagueness, when usually fuzzy sets are applied. Width of defuzzification range was proposed as quality indicator.
4. In last case (ambiguity) object and/or measurement conditions are not unequivocal. It does not mean the data are useless, but only any parameters describing their quality are available and this information should be placed in metadata.

   Monograph is focused on the two first, "measurable" cases, ideal situation when data quality can be determined by error and the measurements of natural objects in natural conditions when we talk about uncertainty. In both cases as error/uncertainty standard deviation is taken as quality indicator, according Guideline. Assumption of normal error/uncertainty distribution is tacit made. Legitimacy of the assumption was in monograph discussed. Known from the literature alternative robust method and Laplace distribution were tested as well. Two groups of the examples were in presented. One concerns uncertainty analysis of vector polygon objects and second DTM uncertainty in grid form.

Rozdział 1
Wprowadzenie

Inspiracją do badań, których wyniki są przedstawione w niniejszej monografii był projekt dotyczący przetwarzania danych z lotniczego skaningu laserowego (Hejmanowska i in. 2008 a), w ramach którego z chmury punktów tworzony był numeryczny model terenu (NMT) oraz numeryczny model pokrycia terenu (NMPT). Gotowe produkty zostały poddane analizie dokładności. Norma PN-EN ISO 19113 podaje parametry opisujące jakość NMT, między innymi: dokładność położenia (dokładność bezwzględną lub zewnętrzną oraz dokładność względną lub wewnętrzną). Dokładność zewnętrzną można wyznaczyć w oparciu o porównanie NMT z położeniem punktów referencyjnych lub uznanych za takie. Dokładność wewnętrzną można zdefiniować jako rozbieżność pomiędzy punktami pomiarowymi wykorzystanymi do budowy NMT a modelem NMT. Analizując wizualnie histogramy tych różnic można zauważyć, że odbiegają one od kształtu rozkładu normalnego, są bardzo skupione wokół zera i mają tzw. "długie ogony". Analiza różnych funkcji rozkładów prawdopodobieństwa skłaniała do wyboru rozkładu Laplace'a jako modelu opisującego to zjawisko, jednakże początkowo nie przeprowadzono bardziej szczegółowej analizy charakteru tego rozkładu w aspekcie błędów NMT. Dalszej inspiracji dostarczyła publikacja (Höhle, Höhle 2009), w której również zauważony został odmienny niż normalny charakter rozkładu błędów NMT.

W cytowanej publikacji zaproponowano modyfikację funkcji rozkładu prawdopodobieństwa Gaussa stosowanej powszechnie do oceny dokładności wyników pomiaru. Generalnie głównym zarzutem w stosunku do rozkładu normalnego jest przeszacowywanie błędu (ASPRS Lidar Committee 2004, Aguilar i in. 2010, Atkinson i in. 2005, Darnel i in. 2008, Zandbergen 2008, Höhle, Höhle 2009), a zaproponowana metoda (Höhle, Höhle 2009), nazwana w publikacji metodą odporną (*robust method*) lepiej opisuje rozkład empiryczny błędów NMT. Analizując wizualnie zaprezentowane w publikacji histogramy empiryczne oraz stopień dopasowania funkcji gęstości prawdopodobieństwa rozkładu Gaussa i proponowanej funkcji odpornej, można zauważyć w pierwszym przypadku przeszacowywanie błędów NMT, a w drugim pewne ich niedoszacowanie.

Bazując jednak na własnych, wcześniejszych analizach postanowiono zaproponować metodę alternatywną opartą o funkcję gęstości rozkładu Laplace'a (Hejmanowska i Kay 2011). W publikacji znajdują się wyniki porównania estymacji błędów NMT z wykorzystaniem funkcji gęstości rozkładu Gaussa, funkcji odpornej (Höhle, Höhle 2009) i funkcji gęstości rozkładu Laplace'a. Przeprowadzone analizy potwierdziły opisywane w publikacjach zjawisko przeszacowania błędu zakładając rozkład normalny. Potwierdziły również założoną przez Autorkę monografii hipotezę a'priori, że z kolei metoda Höhle'a niedoszacowuje błędu NMT. Natomiast krzywa gęstości prawdopodobieństwa, bazująca na funkcji Laplace'a, znajduje się pomiędzy tymi funkcjami. Wyniki tych badań opisano w rozdziale 7.1.

W związku z tym, że wyniki badań nad NMT były obiecujące ale dotyczyły tylko jednego przypadku, analiz dokładności NMT ze skaningu laserowego (Hejmanowska, Kay 2011), postanowiono przeanalizować rozkłady błędów w innych sytuacjach. Po pierwsze wykonano testy symulacyjne w celu zbadania natury badanych rozkładów. Następnie przeprowadzono dalsze badania na danych rzeczywistych. Dla potrzeb analiz kształtu histogramów potrzebna jest stosunkowo duża liczba danych a proces ich pozyskania jest dość pracochłonny.

W związku ze stażem naukowym Autorki opracowania w JRC w latach 2009-2011, zaistniała możliwość wykorzystania innego rodzaju danych. W JRC opracowuje się standardy dla systemu dopłat bezpośrednich do rolnictwa IACS. W ramach tego systemu został, w każdym kraju Unii Europejskiej, opracowany system identyfikacji działek rolnych (LPIS), który jest szczególnie liczebną bazą danych. Działki w systemie LPIS stanowią podstawę corocznej deklaracji rolnika, dotyczącej wielkości uprawianej powierzchni. Każdego roku wykonuje się również pomiary kontrolne powierzchni wybranych działek. JRC opracowuje standardy dotyczące zarówno kontroli jakości systemu LPIS, jak też procedur kontrolnych pomiarów powierzchni. Kontrole powierzchni wykonuje się metodą pomiarów bezpośrednich, najczęściej z wykorzystaniem GPS lub metodą wektoryzacji ortofotomap. Ortofotomapy są wykorzystywane również w kontroli jakości danych LPIS. Oba te procesy są w JRC badane, a instytucja ta wykonuje też różnego rodzaju pomiary w celu testowania wprowadzanych standardów. W związku z generalnie dużą liczebnością wyników pomiarów, co nie jest zjawiskiem powszechnym w zwykłych

pomiarach geodezyjnych, można z powodzeniem przeprowadzać analizy histogramów oraz wpasowywać w nie różne funkcje teoretyczne.

Analizując dane literaturowe jak i badania własne pod kątem wyjaśnienia źródła zjawiska odbiegania empirycznego rozkładu błędów od rozkładu normalnego postawiono wstępną hipotezę:

W systemach GIS mamy często do czynienia z ogromną liczbą danych, szczególnie jeśli mówimy o danych rastrowych lub o dużych bazach wektorowych o zasięgu krajowym. W takiej sytuacji większość danych jest poprawna, czyli odchyłki od wartości referencyjnych są skoncentrowane wokół zera. Jest jednak również wiele przypadków, w których dane w różnym stopniu odbiegają od wartości referencyjnych. Mnogość takich sytuacji powoduje występowanie "długich ogonów", dla których nie powinno się stosować rozkładu Gaussa do szacowania jakości danych. Natomiast w przypadku kiedy w systemie znajdują się dane pomiarowe wykonane zgodnie ze sztuką pomiarów geodezyjnych, można analizy dokładności przeprowadzać tak, jak robi się to do tej pory, czyli w oparciu o rozkład Gaussa.

Struktura monografii jest następująca. Najpierw zostały omówione zagadnienia związane z dokładnością, błędami, niepewnością itd. W dalszej kolejności zostały opisane badane funkcje rozkładu prawdopodobieństwa oraz metodyka ich porównywania. Następnie zaprezentowano wyniki dla testów teoretycznych i praktycznych, wykonanych w oparciu o pomiary powierzchni działek rolnych oraz analizy NMT.

Rozdział 2

Sposoby wyrażania dokładności danych przestrzennych

W systemach GIS gromadzi się dane pochodzące z różnych źródeł i w różnym stopniu przetworzone. Mogą to być punkty, linie i obszary powierzchniowe w modelu wektorowym lub piksele w modelu rastrowym. Ich cechą charakterystyczną jest położenie, zdefiniowane poprzez współrzędne punktów, punktów załamania linii lub poligonu oraz współrzędne piksela w rastrze. Drugą cechą charakterystyczną danych przestrzennych są atrybuty opisowe do nich przypisane. Dane mogą pochodzić z pomiarów, w tym z pomiarów geodezyjnych np. działki ewidencyjne. Mogą też być wynikiem modelowania. Przykładem może być numeryczny model terenu, ale także np. wszelkiego rodzaju mapy geologiczne czy glebowe. Świadome korzystanie z danych nieuchronnie prowadzi do pytania o ich jakość. Oczywiście chcielibyśmy mieć raczej do czynienia z produktami wysokiej, a nie niskiej jakości, ponieważ mamy do nich większe zaufanie. Można by w tym momencie zadać pytanie o definicję jakości i sposoby jej określania. Pojęcie to wywodzi się od Platona, który po raz pierwszy go użył *(gr. poiotes)* mówiąc, że "jakość konkretnych rzeczy to stopień osiągniętych przez nie doskonałości" (Bielawa 2011). Oznacza to, że rzeczy o wyższej jakości są bardziej doskonałe od tych o mniejszej jakości, dlatego też są bardziej pożądane. Angielskie słowo *"quality"* pochodzi

natomiast z języka łacińskiego *(qualitas)* i jest tłumaczeniem Cycerona, który przy okazji określił w ten sposób właściwości przedmiotu. Istnieje wiele sposobów rozumienia pojęcia ”jakość”, poza filozofią można oczywiście wymienić dziedzinę zajmującą się ekonomią, zarządzaniem i normę ISO:9001. Definicja słowa jakość w słowniku Doroszewskiego jest trochę w tym kontekście zaskakująca ”właściwość, rodzaj, gatunek, wartość; zespół cech stanowiących o tym, że dany przedmiot jest tym przedmiotem, a nie innym”. Więcej informacji na temat różnych podejść do jakości znajduje się między innymi w publikacji (Bielawa 2011). Podsumowując można zacytować za słownikiem języka polskiego, że jakość to jest ”1. wartość czegoś, 2. istotne cechy przedmiotu wyróżniającego go spośród innych”. Oznacza to, że w analizowanym kontekście właściwe jest użycie tej pierwszej definicji, jeśli mówimy o jakości danych przestrzennych, to mamy na myśli określenie ich wartości, a nie opis tych ich cech, które określają, że są to właśnie dane przestrzenne, a nie coś innego. Oczywiste zatem jest, że chcielibyśmy, żeby dane były najwyższej jakości w tym pierwszym rozumieniu. Równocześnie nasuwa się pytanie jakie są parametry, które określają jakość danych i w jaki sposób można je wyznaczyć. Nawet bez wiedzy specjalistycznej można zapytać o powszechnie rozumiane parametry jakości danych: choćby **dokładność** czy **aktualność**. W związku z tym, że pomimo wprowadzania różnych unormowań, nie ma jasno sprecyzowanych obowiązujących parametrów opisu jakości, warto sięgnąć do definicji. Z pojęciem jakości łączą się bowiem inne określenia: wiarygodność, dokładność, niedokładność, błąd czy niepewność. **Wiarygodny** oznacza godny zaufania i to pojęcie jest stosowane np.

w statystyce. Można znaleźć go również w definicji wiedzy, którą określa ”ogół wiarygodnych informacji o rzeczywistości wraz z umiejętnością ich wykorzystywania” (Słownik języka polskiego). **Dokładność** (przeciwieństwo **niedokładności**) jako rzeczownik zwykle jest definiowana poprzez znaczenie przymiotnika: dokładny - ” odznaczający się starannością, uwagą, dbałością o szczegóły w wykonywaniu czegoś, ścisły, precyzyjny, skrupulatny” (Doroszewski 2013). Kontynuując, **błąd** to ”mylne mniemanie, fałsz; niezgodność z istotnym stanem rzeczy, ale również odstępstwo on normy; pomyłka, czy miara odchylenia wartości przybliżonej od dokładnej”. I ostatecznie **niepewność**: ”brak pewności, bezpieczeństwa; wahanie”, która jest w znacznie mniejszym zakresie opisana niż pojęcie pewności. **Pewność**: ”niezachwiane i uzasadnione przekonanie o istnieniu czego lub o tym, że rzecz ma się w określony sposób”, ale także ”odznaczanie się cechami gwarantującymi możność polegania na czym, na kim, wiary w co” i ”gwarancja, że coś na pewno nastąpi; zabezpieczenie, bezpieczeństwo” oraz ”rzecz, fakt, zjawisko całkowicie pewne, nie budzące wątpliwości, udokumentowane, niewątpliwe”. W życiu, w zależności od różnych sytuacji, używamy tych wszystkich pojęć.

W geodezji natomiast jesteśmy przyzwyczajeni do wykorzystywania teorii błędów dla opisu jakości pomiaru. To tradycyjne podejście, wykorzystywane również w dziedzinach zajmujących się metrologią, zostało w obecnych czasach zmodyfikowane. Najważniejszą zmianą jest wprowadzenie pojęcia **niepewności** zamiast powszechnie stosowanego pojęcia **błąd**.

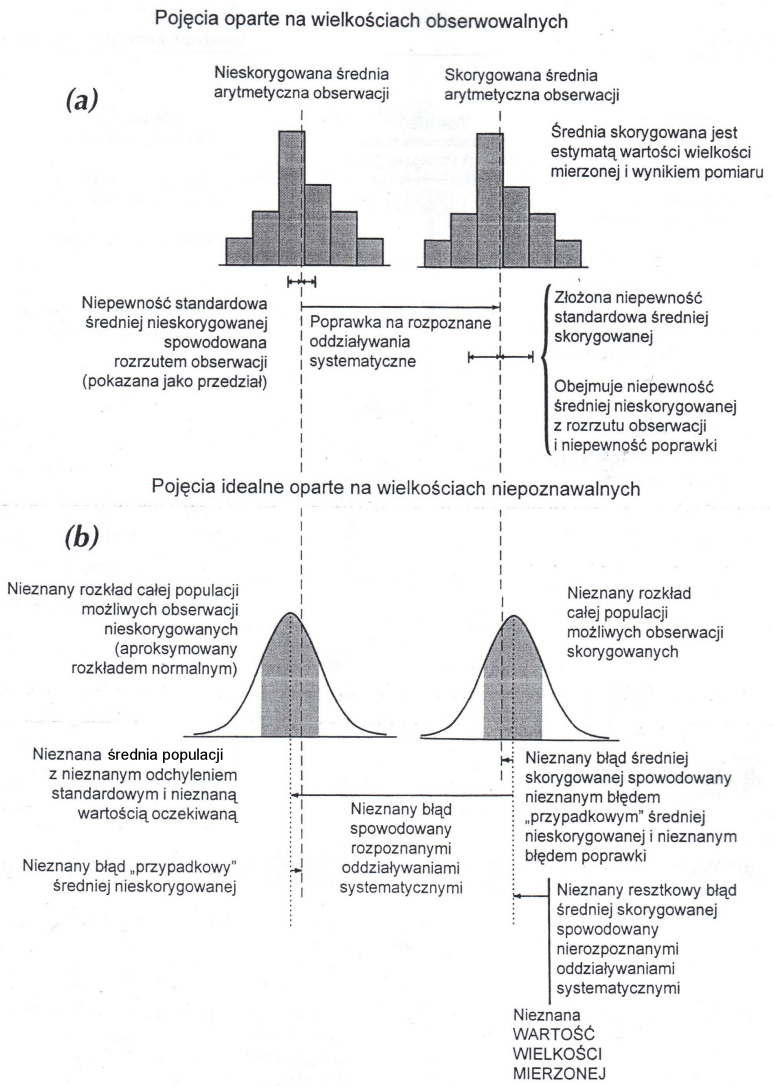
2.1. Wyrażanie niepewności pomiaru w metrologii

Stopniową zmianę w podejściu do opisu jakości wyników pomiarów można zaobserwować jako wpływ opublikowanego w 1993 roku Przewodnika dotyczącego wyrażania niepewności pomiaru. Przewodnik został opracowany przez grupę roboczą składającą się z ekspertów wytypowanych przez Międzynarodowe Biuro Miar, Międzynarodową Komisję Elektrotechniczną, Międzynarodową Organizację Normalizacyjną i Międzynarodową Organizację Metrologii Prawnej. Główny Urząd Miar wydał Przewodnik w wersji polskiej w 1999 roku. Zgodnie z Przewodnikiem:

> "Słowo 'niepewność' oznacza wątpliwość i stąd w szerokim znaczeniu 'niepewność pomiaru' oznacza wątpliwość co do wartości wyniku pomiaru. Z powodu braku różnych słów na oznaczenie pojęcia ogólnego niepewności i specyficznych wielkości będących miarami ilościowymi tego pojęcia, na przykład odchylenia standardowego, koniecznym jest używanie słowa 'niepewność' w tych dwóch różnych znaczeniach."

W przypisach do wydania polskiego znajduje się uwaga prof. Jaworskiego, że stwierdzenie to nie jest zupełnie ścisłe:

> "bo przecież można nazwę niepewność odnosić tylko do pojęcia ogólnego (tj. cechy pomiaru), a ilościowe miary niepewności nazywać parametrami niepewności."



Rysunek 1: Interpretacja graficzna różnicy pomiędzy niepewnością a) a błędem b) (Główny Urząd Miar 1999)

Rysunek 1 przedstawia graficzną interpretację pojęć: **wartość, błąd, niepewność** i pokazuje różnicę pomiędzy błędem a niepewnością. Wykresy z lewej strony przedstawiają wartości niepewności (Rys. 1 a) i błędu

(Rys. 1 b) nieskorygowane ze względu na oddziaływania systematyczne, z prawej przedstawiają wartości skorygowane. Rysunek 1 a) z lewej ilustruje histogram rozkładu wartości uzyskanych w wyniku pomiaru odniesione do wartości średniej. Pokazana jest również niepewność średniej nieskorygowanej. Po uwzględnieniu poprawki na oddziaływania systematyczne uzyskujemy wykres na rysunku 1 a) po prawej stronie. Zaznaczono na nim złożoną niepewność standardową średniej skorygowanej, uwzględniając niepewność średniej nieskorygowanej i niepewność poprawki. Rysunek 1 b) przedstawia sytuację idealną, dla której zarezerwowane jest pojęcie błąd. Podstawową różnicą pomiędzy tymi dwoma omawianymi przypadkami jest znajomość wartości referencyjnej. Przy czym jest to sytuacja idealna dlatego też na rysunku 1 b) używa się pojęć "nieznany/a". Nieznana jest średnia populacji i jej odchylenie standardowe. Nieznany jest także błąd spowodowany oddziaływaniem systematycznym. Można powiedzieć, że w Przewodniku **błąd** jest rozumiany jako błąd prawdziwy. Natomiast pojęcie **błąd** w ujęciu klasycznym, kiedy nie znamy wartości mierzonej, nazywanej też wartością prawdziwą powinno być zastąpione pojęciem **niepewność**.

Zgodnie z Przewodnikiem:

"Niepewność wyniku pomiaru ogólnie składa się z szeregu składników, które można zgrupować w dwie kategorie, zgodnie ze sposobem obliczania ich wartości liczbowych:

- A. takie, które zostały obliczone metodami statystycznymi;

– B. takie, które zostały obliczone innymi metodami.

Nie zawsze zachodzi prosta zależność pomiędzy klasyfikacją na kategorie A i B, a wcześniej stosowaną klasyfikacją na niepewności ”przypadkowe” i niepewności ”systematyczne”. Termin ”niepewności systematyczne” może być mylący i należny go unikać.”

Ogólnie rzecz biorąc obie te niepewności, zgodnie z Przewodnikiem, powinny być charakteryzowane przez estymaty wariancji. Wśród różnych pojęć definiowanych w Przewodniku występuje pojęcie niepewności standardowej:

”niepewność wyniku pomiaru wyrażona w formie odchylenia standardowego”

Obliczanie, a w zasadzie wyznaczanie niepewności metodą B może być oparte o inne dane, niż uzyskane w wyniku bezpośrednio wykonanego eksperymentu. Mogą to być (według Przewodnika):

– ”poprzednie dane pomiarowe;
– posiadane doświadczenie wraz z ogólną znajomością zjawisk i właściwości odpowiednich materiałów odniesienia i przyrządów;
– specyfikacje wytwórców;

– dane uzyskane z wzorcowania i certyfikacji;

– niepewności przypisane danym odniesienia zaczerpniętym z podręczników."

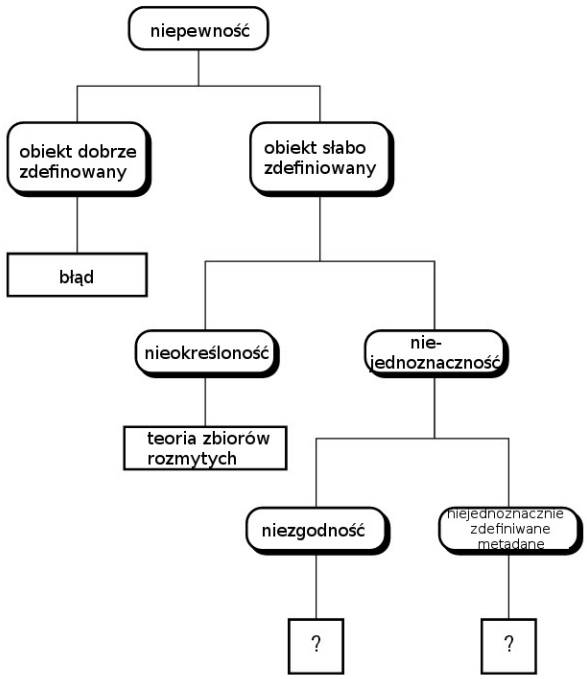
Przykład mieszczący się w powyższym wyliczeniu ma miejsce powszechnie w pomiarach geodezyjnych, gdzie jako parametr opisujący jakość wyników pomiaru przyjmuje się dopuszczalne wartości błędów, zakładając wypełnienie warunków specyfikacji technicznych, instrukcji geodezyjnych, czy innych wytycznych. Na przykład w pomiarach kontrolnych powierzchni działek rolnych w systemie IACS przyjmuje się wyniki walidacji jako punkt odniesienia do określenia dokładności pomiaru.

Pomimo, ze minęło już ponad 10 lat od opublikowania Przewodnika pojęcie niepewności nie jest powszechnie stosowane w dziedzinie geodezji i kartografii. Wynika to prawdopodobnie z tradycji, używania coraz dokładniejszych urządzeń pomiarowych i przeświadczenia, że obecnie mierzymy już tak dokładnie, jak nigdy dotychczas i błąd jest na tyle mały, że uzyskiwana dokładność jest wyższa od oczekiwanej. Nie przywiązujemy nawet tak dużej jak w przeszłości, wagi do analiz dokładności i wpływu różnych czynników na wynik pomiaru. Inną przyczyną jest odrębność dziedzin **geodezja i kartografia** i **metrologia**. Nawet klasycznie używane w metrologii pojęcie budżetu niepewności (obejmujące możliwie wszystkie czynniki wpływające na niepewność pomiaru), bywa nieznane studentom geodezji, pomimo omawiania tych zagadnień na zajęciach z rachunku wyrównawczego. Dokładność pomiarów geodezyjnych determinują branżowe instrukcje i wytyczne techniczne, nato-

miast dokładność urządzeń geodezyjnych oraz procedurę ich sprawdzania określają normy ISO. Z drugiej strony dane pomiarowe coraz częściej są wprowadzane i wykorzystywane w systemach GIS, dla których opracowano normy ISO 19113:2002 Informacja Geograficzna – Podstawy Opisu Jakości oraz komplementarną dla niej ISO 19114:2003 Informacja Geograficzna – Procedury Oceny Jakości. W tych normach jakość danych definiuje się za pomocą następujących parametrów: kompletność, spójność logiczna, dokładność położenia, dokładność czasowa i dokładność tematyczna. Zatem w tym przypadku mamy do czynienia z pojęciem dokładność.

2.2. Określanie niepewności danych przestrzennych

Problem definiowania jakości danych w ogóle, w tym danych przestrzennych, był wielokrotnie poruszany w literaturze. Oprócz różnych definicji błędu oraz bardziej opisowych pojęć, o zabarwieniu optymistycznym - dokładność, czy pesymistycznym - niedokładność, pojawiają się określenia takie jak niepewność, niejednoznaczność czy nieokreśloność. Z tych wszystkich pojęć najbardziej znane i najlepiej zdefiniowanie jest pojęcie **błąd**, który często bywa rozumiany jako precyzja (błąd wewnętrzny, powtarzalność) i dokładność (błąd bezwzględny, odtwarzalność). Nie można powiedzieć, że używanie pojęcia **błąd** do określania jakości danych jest błędne, czy anachroniczne. Należy natomiast mieć świadomość ograniczeń jego stosowania. Umiejscowienie określenia **błąd** w szerszym kontekście innych pojęć można znaleźć w publikacji (Longley i inni 2005).



Rysunek 2: Model koncepcyjny niepewności danych przestrzennych (zmodyfikowany za Fisher 1999, 2005, zmodyfikowany za Klir i Yuan 1995)

Mamy coraz dokładniejsze urządzenia pomiarowe, których dokładność jako taką można określać z pewnością za pomocą pojęcia **błąd**, jeśli będziemy wykonywali pomiar obiektów dobrze zdefiniowanych w kontrolowanych warunkach laboratoryjnych. W publikacjach (Fisher 1999, 2005) został zaproponowany podział niepewności ze względu na obiekt, a jako ogólne pojęcie charakteryzujące jakość przyjęto niepewność (Rys.2). Dalej w zależności od obiektu możemy go scharakteryzować za pomocą błędu, jeżeli jest on dobrze zdefiniowany. W przypadku trudności w zdefiniowaniu obiektu możemy mieć do czynienia z nieokreślonością i niejednoznacznością. Problem z identyfikacją obiektu nie jest abstrakcyjny

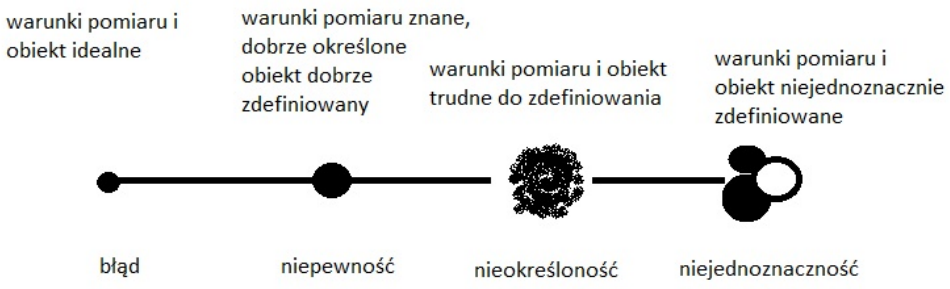
w geodezji, szczególnie jeśli weźmiemy pod uwagę obecnie wykorzystywane coraz dokładniejsze urządzenia pomiarowe. Pomimo, iż wzrosła ich dokładność, obiekty, które mierzymy nie zmieniły się. Pojawia się zatem pytanie, co tak naprawdę mierzymy? Przykładem może być skaning laserowy: mamy gęstą chmurę punktów, ale możemy nie mieć tych punktów, które "chcielibyśmy" mieć. W praktyce nie interesuje nas chmura punktów, ale obiekt i pewne ściśle określone produkty: przekroje czy miary. Tendencję zapominania o obiekcie można również zaobserwować w przepisach dotyczących walidacji pomiarów powierzchni działek rolnych w systemie dopłat bezpośrednich do rolnictwa IACS. Walidację metody pomiaru działek rolnych wykonuje się na polach testowych o sztucznie sygnalizowanych granicach. Uzyskany w ten sposób parametr określający dokładność pomiaru wykorzystujemy później do pomiaru innych obiektów, czyli działek rolnych o naturalnych granicach. Dawniej, kiedy używano nawigacyjnych urządzeń GPS oraz obrazów satelitarnych IKONOS o rozdzielczości przestrzennej 1 m, to zalecenia przyjmowania tolerancji pomiarów proponowane przez ośrodek UE w JRC na podstawie takiej walidacji obejmowały zarówno błąd spowodowany urządzeniem pomiarowym, jak również pochodzący od obiektu. Wynikało to z faktu, że samo urządzenie pomiarowe w tym przypadku było na tyle niedokładne, że niedokładność obiektu "uwzględniała się sama". Rozwój technologii doprowadził do absurdalnej sytuacji, w której obecnie obowiązujące tolerancje pomiaru powierzchni, wyznaczane w warunkach "laboratoryjnych" okazały się niemożliwe do uzyskania w praktyce (Kramarczyk i in. 2012, Hejmanowska i in. 2012). Wynika to z faktu, że walidację prze-

prowadza się w łatwym terenie na punktach sygnalizowanych, uzyskując tym samym maksymalną dokładność pomiaru powierzchni daną metodą. Uzyskaną w ten sposób tzw. tolerancję pomiaru wykorzystuje się podczas pomiarów kontrolnych rzeczywistych działek rolnych. Dokładność pomiaru działki rolnej, wykonywanego w warunkach naturalnych, w terenie urozmaiconym ze względu na morfologię i pokrycie jest z reguły mniejsza niż w warunkach "laboratoryjnych". Dopuszczalną różnicę pomiędzy zadeklarowaną przez rolnika powierzchnią działki rolnej a powierzchnią uzyskaną podczas pomiaru oblicza się w oparciu o uzyskaną podczas walidacji wartość tolerancji. Warunki pomiaru walidacyjnego i kontrolnego mogą znacznie różnic się od siebie. Kontroler musi jednak zastosować daną tolerancję pomiaru, co może mieć konsekwencje finansowe dla rolnika. Rozbieżność pomiędzy warunkami pomiaru walidacyjnego i kontrolnego były powodem sprzeciwu zarówno w ramach JRC jak i wśród krajów członkowskich (Hejmanowska i in. 2012, Jahn i in. 2012). Niestety pomimo wielu burzliwych dyskusji, problem ten nie został do dziś rozwiązany. Jest to przykład, jak trudno jest zmienić istniejące procedury, które opracowano bez uwzględnienia wszystkich źródeł błędów.

2.3. Model koncepcyjny jakości danych przestrzennych

Po przeanalizowaniu powyższych podejść do szacowania jakości danych, można by spróbować dokonać ich syntezy, szczególnie, że Przewodnik skupia się na pomiarach, a specjaliści GIS w ogólności na danych. Zamiast struktury hierarchicznej można spróbować zaproponować strukturę liniową (Rys. 3). W sytuacji idealnej, w której mamy dokładnie zdefi-

niowany obiekt mierzony, a także są znane wszystkie warunki pomiaru i budżet niepewności jest kompletny, możemy używać terminu **błąd**. Taka sytuacja może wystąpić w kontrolowanych warunkach laboratoryjnego pomiaru obiektu, jeśli jest znana dla niego wartość referencyjna. W przypadku pomiarów, dla których nie mamy wartości referencyjnych oraz nie jesteśmy całkowicie pewni co do budżetu niepewności, powinniśmy używać pojęcia **niepewność**. Oba wymienione terminy dotyczą sytuacji "mierzalnej" tzn. że jest możliwość pomiaru obiektu, czy zjawiska. Idąc w kierunku trudności w definiowaniu obiektu czy zjawiska oraz warunków pomiaru dochodzimy do przypadku nieokreśloności, w której tradycyjnie stosuje się zbiory i logikę rozmytą. Natomiast z najbardziej skomplikowaną sytuacją spotykamy się, kiedy zarówno warunki pomiaru jak i obiekt, czy zjawisko są niejednoznaczne.



Rysunek 3: Model koncepcyjny jakości danych przestrzennych

2.4. Dane przestrzenne

Można wymienić wiele przykładów ilustrujących opisane powyżej cztery przypadki. Weźmy pod uwagę: działkę zasygnalizowaną granicznikami, działkę rolną o dobrze widocznych naturalnych granicach (miedza pomiędzy różnymi uprawami), działkę o słabo widocznej, lub trudnej do interpretacji granicy (pastwisko, granica porośnięta drzewami) oraz działkę o niejednoznacznej granicy naturalnej, niemożliwej do zinterpretowania bez dodatkowych informacji (Rys. 4 i Rys. 5). W przypadku zasygnalizowanej granicy działki i znajomości jej referencyjnego położenia, można by dokładność pomiaru określić za pomocą błędu.

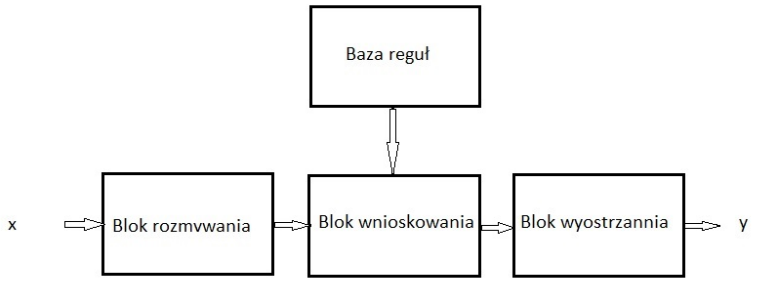
Rysunek 4: Przykład granicy działki rolnej dobrze zdefiniowanej i rozmytej (JRC MARS 2013 b)

Rysunek 5: Przykład niejednoznacznej granicy działki (JRC MARS 2013 b)

Na Rys. 4, lewe górne zdjęcie przedstawia działkę o dobrze zdefiniowanej granicy, kontrastującą z otoczeniem. Digitalizacja przebiegu granicy jest w tym przypadku łatwa, również rozpoznanie granicy tej działki w terenie powinno być bezproblemowe. Jednakże jest to działka o naturalnych granicach, kiedy zawsze występuje konieczność interpretacji granicy. Ponadto w tym przypadku nie mamy możliwości określenia wartości referencyjnej położenia granicy. W związku z tym, zgodnie z Przewodnikiem, powinniśmy do określania dokładności pomiaru użyć pojęcia niepewności. Pozostałe dwie działki na Rys. 4 mają granice rozmyte, trudniejsze w interpretacji, z różnych powodów. W jednym jest to spowodowane tym samym typem użytkowania, a w drugim drzewami, które rosną wzdłuż granicy. W takim przypadku nieokreśloności zwykle stosuje się teorię zbiorów rozmytych i/lub rozmyte reguły decyzyjne. Te dwa ostatnie przypadki nieokreśloności i niejednoznaczności nie

będą szerzej omawiane w dalszej części monografii. Ponadto nie da się do nich wprost zastosować odchylenia standardowego błędu, czy niepewności w celu określenia ich jakości. W związku z tym zagadnienie to zostało w skrócie omówione w niniejszym podrozdziale. Przedstawiono koncepcję wyznaczania niepewności danych dla przypadku nieokreśloności. Opisane poniżej podejście należy rozumieć jako bardzo wstępną propozycję. W największym skrócie, postępowanie w przypadku wykorzystania zbiorów rozmytych wywodzi się z niemożności zdefiniowania ostrej granicy pomiędzy zjawiskami (jest zimno, ciepło, gorąco) czy obiektami, jak na (Rys 4). Zamiast ostrej granicy możemy podać pewien zakres, obszar występowania jakiegoś zjawiska lub obecności jakiegoś obiektu. Klasycznym przykładem jest definicja subiektywnego odczucia zimna i ciepła. Nie można jednoznacznie powiedzieć, że zimno jest wtedy, kiedy temperatura wynosi 10 stopni Celsjusza, ciepło kiedy 20 a gorąco w przypadku 30 stopni Celsjusza. Można spróbować podać pewne zakresy, kiedy czujemy, że jest zimno: 2-18 stopni C, ciepło: 12-28 stopni C i gorąco: 22-38. Wykorzystanie zbiorów rozmytych przebiega zgodnie z ideowym schematem (Rys. 6).

Dane wejściowe x podlegają rozmyciu, następnie mamy do czynienia z etapem wnioskowania na podstawie zdefiniowanych reguł. Ostatecznie konieczne jest wykonanie tzw. wyostrzenia w celu uzyskania jednej wartości y. W przypadku rozróżniania typu pokrycia działek rolnych z wykorzystaniem danych teledetekcyjnych, można wziąć pod uwagę różne parametry obrazu np. jasność i teksturę. Zdefiniujmy jasność w sposób opisowy. Przyjmijmy, że obraz powierzchni działki może być "ciemny"



Rysunek 6: Rozmyty proces decyzyjny

lub ”jasny”. W związku z tym nie przypisujemy im pojedynczych wartości jasności (DN równe 150 i 200), a pewien zakres wartości, o tak zwanym różnym stopniu przynależności, na przykład w skali od 0 do 1. Stopień przynależności opisuje funkcja przynależności, $\mu(x)$, która może mieć różny przebieg. W najprostszym przypadku jest to funkcja liniowa, zwykle najpierw rosnąca, potem po uzyskaniu maksimum, malejąca. Często, dla poszczególnych klas, przykładowo: ”ciemny”, ”jasny” funkcje przynależności nachodzą na siebie. Oznacza to, że w pewnych przypadkach można zaklasyfikować daną jasność do różnych klas, z różnym prawdopodobieństwem. Przeanalizujmy funkcję przynależności na Rys. 7 a. Dla klasy ”ciemny” jej wartość wynosi 1 dla wartości 150 i maleje do zera dla wartości mniejszych i większych od 150. Podobnie jest zbudowana funkcja przynależności dla jasności: ”jasny”. W pewnym zakresie jasności funkcje te nachodzą na siebie. Drugim analizowanym parametrem jest tekstura, dla której funkcję przynależności można określić jak na

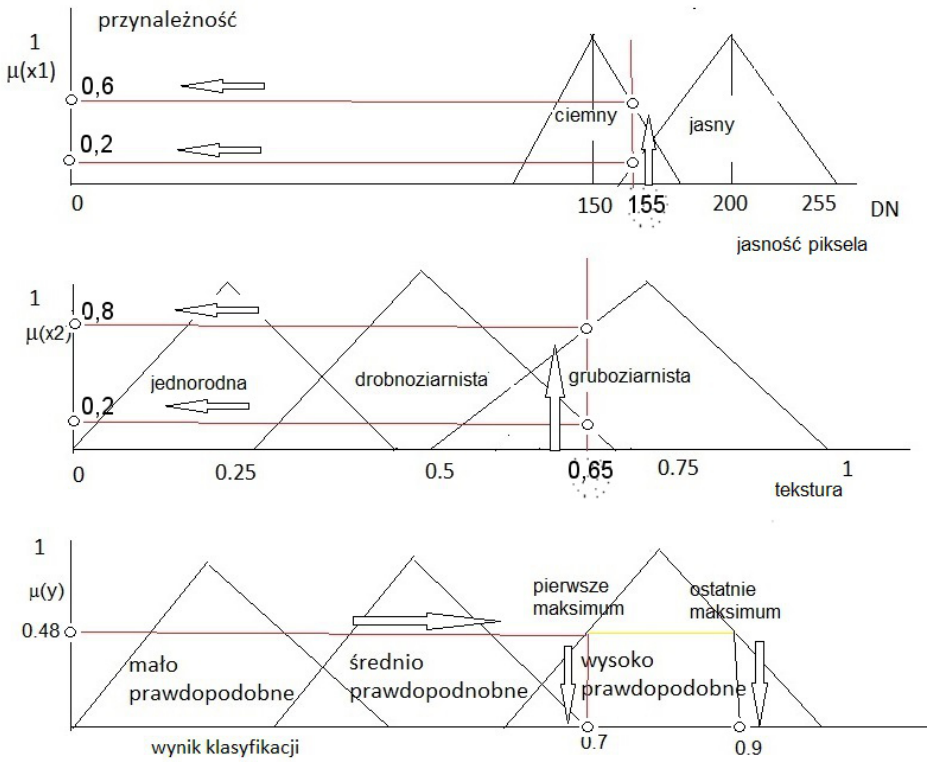
Rys. 7 b. Załóżmy, że chcemy dokonać klasyfikacji działki o wartości jasności równej 155 i teksturze 0,65. Z wykresów na Rys. 7 a i b można odczytać wartości funkcji przynależności do odpowiednich klas. W tym przypadku jasność działki (równa 155) można określić jako "ciemny", a odpowiadająca jej wartość funkcji przynależności $\mu(x_1)$ wynosi 0,6. Jasność działki można określić również jako "jasny". Wartość funkcji przynależności $\mu(x_1)$ wynosi w tym przypadku 0,2. Podobnie można odczytać wartości funkcji przynależności dla tekstury. Tekstura równa 0,65 może oznaczać, że jest to tekstura gruboziarnista i wartość funkcji przynależności $\mu(x_2)$ wynosi 0,8. Tekstura ta również może być teksturą drobnoziarnistą, a wartość funkcji przynależności $\mu(x_2)$ wynosi 0,2. Zgodnie ze zdefiniowanymi funkcjami przynależności tekstura ta nie może jednak być określona jako jednorodna. W następnej kolejności budujemy bazę reguł (Tab. 1). Załóżmy, że jeśli jasność: "ciemny" i tekstura: "gruboziarnista" to jest wysokie prawdopodobieństwo (0,75), że piksel należy do działki leśnej, jeśli jasność: "jasny" i tekstura "jednorodna" to istnieje wysokie prawdopodobieństwo (0,75), że piksel należy do działki rolnej pokrytej użytkiem zielonym. Pierwsza kolumna tabeli określa prawdopodobieństwo, że działka jest działką leśną. Czyli na przykład jeśli jasność działki przyjmuje wartość "ciemny" ale tekstura jest "jednorodna" to jest małe prawdopodobieństwo, że jest to działka leśna. W drugiej kolumnie znajdują się wartości prawdopodobieństwa dla działki rolnej, pokrytej użytkiem zielonym. Prawdopodobieństwa w Tabeli 1 są wartościami przyjętymi arbitralnie przez osobę wykonującą klasyfikację. Podsumowując, mając wartości x_1=155 i x_2=0,65 obliczamy wartości przynależności

$\mu(x_1)$ i $\mu(x_2)$. Następnie w oparciu o bazę reguł stworzoną podczas etapu wnioskowania ustalamy wartość $\mu(y)$, która pozwala w procesie wyostrzania wyznaczyć wartość y. Załóżmy, że $\mu(y)$ równa się 0,48 (0,6 razy 0,8), spróbujmy znaleźć dla niej wartość y. Istnieją różne metody wyostrzania; w omawianym przykładzie zastosowano metodę pierwszego maksimum, uzyskując wartość y=0,7 (wysokie prawdopodobieństwo, że działka jest działką leśną). W przypadku ostatniego maksimum mamy y=0,9, co również oznacza wysokie prawdopodobieństwo, że działka jest działką leśną. W obu przypadkach mamy do czynienia z wysokim prawdopodobieństwem, przy czym wynik może nie być jednoznaczny. Abstrahując od innych metod wyostrzania, można by w tym momencie zaproponować parametr określający jakość wyniku analizy rozmytej, jako różnicę poziomów ufności odpowiadających poszczególnym przedziałom, czyli w tym przypadku 20% (0,9 minus 0,7).

Tabela 1: Baza reguł: prawdopodobieństwo, że działka jest lasem lub użytkiem zielonym, dla kombinacji parametrów jasności i tekstury

Tekstura/Jasność	ciemny	jasny
jednorodna	0,25	0,75
drobnoziarnista	0,50	0,50
gruboziarnista	0,75	0,25

Praktyczna implementacja teorii zbiorów rozmytych w różnych procesach decyzyjnych w oparciu o systemy GIS wymaga każdorazowo przeprowadzenia całego procesu, jak na Rys.6. Powyżej przestawiony został jedynie przykładowy, ideowy schemat postępowania podczas określania granicy działki rolnej. Ostatni do omówienia pozostaje przypadek niejed-



Rysunek 7: Etapy klasyfikacji dla działki o nieokreślonej granicy

noznaczności. Rysunek (Rys.5) przedstawia sytuację, kiedy granica jest niejednoznaczna, tzn. patrząc na obraz nie jesteśmy w stanie jednoznacznie pokazać przebiegu granicy działki. W jednym przypadku gęstość obszaru zadrzewienia jest na tyle mała, że drzewa nie geometryzują wystarczająco jednoznacznie granicy działki. W drugim przypadku identyfikacja granicy pomiędzy użytkami zielonymi zasygnalizowana w górnej części elementem liniowym, który mógłby być rowem melioracyjnym, nie ma kontynuacji wzdłuż wschodniej granicy. Również wydzielenie działki nie pokrytej roślinnością ze strony zachodniej jest niejednoznaczne. Bez in-

formacji dodatkowych nie ma możliwości wyznaczenia granicy tej działki. Nie mamy możliwości określenia jakości granicy, ponieważ nie jesteśmy w stanie w ogóle określić jej granicy. Informacja o problemie związanym z identyfikacją granicy powinna jednak znaleźć się w zbiorze metadanych charakteryzujących poszczególne obiekty.

2.5. Dane opisowe

Innym rodzajem danych są dane opisowe, dla których również mogą wystąpić wszystkie omówione powyżej cztery przypadki. W sytuacji idealnej możemy mówić o błędzie, w warunkach naturalnych o niepewności oraz nieokreśloności i niejednoznaczności. Nieokreśloność jest najczęściej w takiej sytuacji przedstawiana na tle tzw. paradoksu stosu. Chodzi o stos piasku, z którego usuwane są ziarna, aż zostanie jedno ziarenko. Czy to jeszcze będzie stos, kiedy stos przestaje być stosem? Albo kiedy las jest iglasty? Jeśli nie ma w nim żadnego drzewa innego niż iglaste, czy nawet malutka samosiejka liściasta powoduje, że las nie jest iglasty, jaka jest dopuszczalna liczba drzew liściastych, żeby las uznać za iglasty? Są to typowe zagadnienia z zakresu logiki rozmytej. Również niejednoznaczność występuje w przypadku danych opisowych. Dobrym przykładem podanym w podręczniku GIS (Longley i inni 2005) są założenia dotyczące klasyfikacji gleb obowiązujące w różnych krajach. Dane dotyczące gleb pochodzące z różnych krajów są w związku z tym wzajemnie niejednoznaczne i ich porównanie jest bardzo trudne, o ile w ogóle możliwe.

2.6. Dane przetworzone - NMT

W przypadku danych przetworzonych, np. modeli NMT możemy mieć również do czynienia z wyżej wymienionymi czterema przypadkami.
W przypadku modelu NMT wykonanego w warunkach "laboratoryjnych", dla którego istnieje model referencyjny (np. model zdefiniowany matematycznie) możemy mówić o błędzie. Sytuacja taka praktycznie nigdy nie istnieje w warunkach naturalnych, zatem powinniśmy mówić o niepewności, nieoznaczoności lub nawet niejednoznaczności. Numeryczny model terenu jest wdzięcznym przykładem pokazania pewnej rozmytej charakterystyki podziału sposobów określania jakości danych na poszczególne przypadki jak na Rys.3. Nieokreśloność bardzo dobrze uwidacznia się w lotniczym skaningu laserowym, za pomocą którego wyznaczana jest wysokość każdego punktu na ziemi, na który pada promień lasera, o ile punkt jest w stanie odbić promieniowanie. Chmura punktów w oczywisty sposób nie jest numerycznym modelem terenu. Musi ona, po pierwsze, zostać poddana klasyfikacji w celu odfiltrowania wszystkich punktów, które nie należą do powierzchni terenu, a po drugie, trzeba wykonać interpolację w celu uzyskania ciągłej powierzchni. Nawet wtedy pozostaje pewne naturalne rozmycie w postaci tzw. mikro i makroreliefu. Więcej na ten temat można znaleźć w publikacji (Oksansen 2006). Chmura przeklasyfikowana do wartości terenowych jest swego rodzaju warstwą o różnej grubości w zależności od szorstkości terenu, a gdzieś pośrodku powinien znajdować się NMT. Czyli można sobie wyobrazić NMT jako ciągłą powierzchnię przykrytą z obu stron chmurą punktów pomiarowych. W trakcie procesu klasyfikacji chmury punktów możemy mieć trudności

w określeniu, czy punkty należą do powierzchni terenu, czy są już niską roślinnością, jest to zatem przypadek nieokreśloności.

2.7. Przedstawienie problemu określania niepewności danych w systemach GIS

W miejsce danych źródłowych często mamy do czynienia nie z nimi bezpośrednio ale z produktami ich przetworzenia. Jednym z najczęściej wykorzystywanych tego rodzaju danych jest właśnie NMT. Jakość danych w systemach GIS powinna być opisywana zgodnie z wymienionymi powyżej normami ISO. Każdy z parametrów charakteryzujących jakość danych jest ważny, niemniej jednak jako przedmiot zainteresowania przyjęto w niniejszych rozważaniach **dokładność**. W związku z tym, że w geodezji pozyskujemy dane do systemów GIS naturalnym jest określanie również ich dokładności. Umiejscowienie metod geodezyjnych wśród innych metod pomiarowych w oczywisty sposób kieruje uwagę na podejście opisane w Przewodniku. Jednocześnie należy zwrócić uwagę na badania prowadzone przez specjalistów z dziedziny systemów GIS, w szczególności z uwagi na fakt, że w tych bazach znajdują się nie tylko dane pomiarowe. Wynikiem syntezy opisanych powyżej podejść (Fisher 1999, 2005, Przewodnik) jest zaproponowany na Rys.3 schemat modelu koncepcyjnego, dotyczącego pojęć opisujących jakość danych przestrzennych.

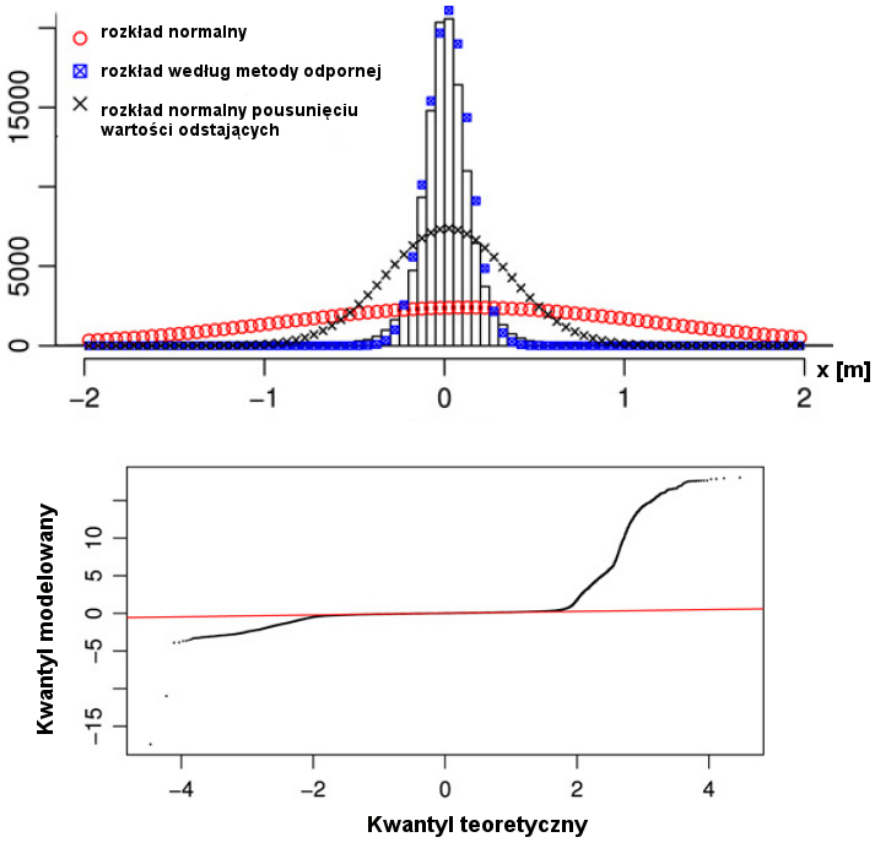
Koncentrując się jedynie na dwóch wymienionych typach danych: dane pomiarowe i opisowe, w obu przypadkach można mówić o ich jakości. Pomiar może być wykonany w warunkach idealnych i wtedy dokładność określimy za pomocą **błędu**, a w warunkach naturalnych użyjemy

określenia **niepewność**. W obu jednak przypadkach możemy określić dokładność wykorzystując wartość odchylenia standardowego. Różnica jest natury pojęciowej. Jeżeli mierzone obiekty są nieokreślone, to możemy zaimplementować algorytmy logiki rozmytej automatycznego "pomiaru", ponieważ trudno sobie wyobrazić praktyczne manualne aplikacje wykorzystujące logikę rozmytą. Niełatwo jest podać gotowy parametr dokładności w tym przypadku; wydaje się, że jest to pole do dalszych badań. W przypadku obiektów potencjalnie "mierzalnych" niejednoznaczność może oznaczać, że pomiaru w ogóle nie można wykonać, a zatem nie można podać żadnego parametru jego dokładności.

Uogólniając, istnieje rozmyta granica pomiędzy sytuacją idealną i warunkami naturalnymi. Bo co to właściwie znaczy, że obiekt i warunki pomiaru są idealnie zdefiniowane, w stosunku do czego? I co to właściwie znaczy, że mamy wartość referencyjną? Wzorzec metra mierzymy coraz dokładniej i to co było kiedyś referencją obecnie już nią nie jest. Z kolei jeśli będziemy mówili o warunkach naturalnych, czy o obiektach naturalnych, musimy wziąć pod uwagę to, że charakteryzują się one zawsze jakimś rozmyciem. Kiedy zatem powinniśmy zastosować pojęcie **niepewność**, a kiedy pojęcie **nieokreśloność** i zbiory rozmyte? Jeśli powiększymy działki na Rys.4, to będziemy mieli rozmycie: mniejsze (rzędu kilku pikseli) lub większe (rzędu kilkunastu, kilkudziesięciu pikseli). Omawiane zagadnienie jest trudne do przeanalizowania w jednej pracy. Dlatego też wybrany został jeden element, mianowicie określanie jakości danych z wykorzystaniem pojęcia niepewności. Parametrem charakteryzującym niepewność, zgodnie z Przewodnikiem,

jest odchylenie standardowe niepewności, podobnie zresztą jak w idealnej sytuacji kiedy estymatorem błędu jest również odchylenie standardowe błędu. Inną sprawą jest określenie błędu granicznego, czy niepewności rozszerzonej, bo jest zależne od podejścia do tego zagadnienia. Niemniej jednak bazując na odchyleniu standardowym domyślnie zakładamy, że mamy do czynienia z rozkładem normalnym błędów pomiarowych, co się przekłada na obliczenie błędu granicznego w oparciu o rozkład gęstości prawdopodobieństwa Gaussa. Wcześniejsze prace Autorki oraz wyniki badań prezentowane w literaturze, skłaniają do refleksji na temat rozkładów błędów, z którymi mamy do czynienia obecnie. Rozkład odbiegający od normalnego zaobserwowano analizując odchyłki pomiędzy chmurą punktów a NMT w 2005 r. w ramach projektu (Hejmanowska i in. 2005). Podobne wyniki opublikowano równolegle w raportach z innych badań (ASPRS Lidar Committee 2004, Aguilar i in. 2010, Atkinson i in. 2005, Darnel i in. 2008, Zandbergen 2008). Został również zaproponowany alternatywny wskaźnik dokładności NMT w stosunku do odchylenia standardowego (ASPRS Guidelines Vertical Accuracy Reporting for Lidar Data, Höhle, Höhle 2009). Głównym zarzutem w stosunku do szacowania dokładności NMT w oparciu o rozkład normalny i odchylenie standardowe było przeszacowywanie błędu. To znaczy błąd obliczony z wykorzystaniem odchylenia standardowego wydawał się być zbyt duży w stosunku do empirycznego rozkładu odchyłek (Rys. 8). Jednakże analizując histogramy z publikacji (Höhle, Höhle 2009) wydaje się, że o ile metoda wykorzystująca odchylenie standardowe przeszacowuje błąd NMT, o tyle metoda zaproponowana w wyżej wymienionej publikacji niedo-

szacowuje błędu. Metoda proponowana w publikacji opiera się również o rozkład normalny. Różnica polega jedynie na modyfikacji jego parametrów: zamiast błędu średniego przyjmuje się medianę, a zamiast odchylenia standardowego - NMAD. Analiza uzyskiwanego w rzeczywistości histogramu odchyłek, który jest bardzo wąski i charakteryzuje się tzw. "długimi ogonami", inspiruje do podjęcia próby znalezienia innego typu rozkładu. Wstępny przegląd różnych rozkładów doprowadził do wyboru rozkładu Laplace'a i dalszych badań w tym kierunku (Hejmanowska, Kay 2011). Wybór funkcji rozkładu prawdopodobieństwa Laplace'a okazał się trafnym rozwiązaniem. Krzywa Laplace'a znalazła się pomiędzy krzywą Gaussa i krzywą z metody odpornej. W związku z tym przeprowadzono dalsze próby aplikacji rozkładu Laplace'a w innych przypadkach analiz jakości danych przestrzennych.



Rysunek 8: Przykład histogramu odchyłek NMT i wysokości pomierzonych *in-situ* wraz z różnymi rozkładami prawdopodobieństwa (zmodyfikowany za Höhle, Höhle 2009)

Rozdział 3
Metodyka

Wziąwszy pod uwagę rozważania przeprowadzone w poprzednim rozdziale będziemy koncentrowali się na przypadkach 1 i 2, tzn. sytuacji idealnej oraz przypadku, w którym zarówno obiekt, jak i warunki pomiaru są wyczerpująco i dobrze zdefiniowane, czyli będziemy mówili o błędzie i niepewności pomiaru. Analizę dokładności możemy przeprowadzić zatem w oparciu o wartość prawdziwą, wartość średnią lub przyjmując za referencyjną wartość nominalną, np. powierzchnię działki w systemie LPIS. Założeniem przeprowadzonych badań jest analiza histogramu odchyłek uzyskanych w wyniku pomiarów i wybranie jak najlepszego parametru opisującego niepewność pomiaru, a w ogólności niepewność danych. Powszechnie w geodezji wykorzystuje się błąd średni kwadratowy, rzadziej średnią odchyłkę i odchylenie standardowe. W Przewodniku również parametrem dokładności pomiaru jest odchylenie standardowe i jego rozszerzenie na zadanym poziomie ufności. Tradycyjnie przyjmuje się założenie normalności rozkładu błędów, czy niepewności. O ile stosuje się do szacowania dokładności pomiaru metodę statystyczną opartą o wyniki pomiaru (metodę A), nie bazując na wiedzy, doświadczeniu, czy normach, to automatycznie przyjmuje się rozszerzenie odchylenia standardowego mnożąc go przez 2 (1.96) lub rzadziej przez 3. Pytanie co zrobić jeśli rozkład nie ma charakteru rozkładu normalnego? Pytanie to rzadko jest

stawiane w geodezji w powodu zbyt małej liczby pomiarów, która nie daje podstaw do sporządzenia wiarygodnego histogramu i analizy normalności rozkładu. Jeśli nawet jest postawione takie pytanie, to analiza dokładności przebiega tak jakby rozkład był normalny niezależnie od odpowiedzi. Taki przypadek występuje na przykład podczas klasyfikacji nadzorowanej, kiedy zakładamy, że rozkład jasności w polach testowych ma rozkład normalny i nawet jeśli tak nie jest i tak klasyfikacja jest wykonywana. Jeśli jednak mamy wystarczającą liczbę danych warto sprawdzić czy nasz rozkład ma charakter rozkładu Gaussa, czy nie. Może się bowiem okazać, że szacowanie dokładności, zakładając, że mamy do czynienia z rozkładem normalnym w rzeczywistości nie odzwierciedla niepewności pomiaru. Zjawisko to zostało zaobserwowane w przypadku określania dokładności NMT. Okazało się, że odchylenie standardowe przeszacowuje błąd/niepewność NMT. Amerykańskie Towarzystwo Fotogrametrii i Teledetekcji (ASPRS) zaleca zbadanie rozkładu i w przypadku kiedy nie jest on normalny, określanie dokładności z wykorzystaniem percentyli. Przy czym percentyl ma być obliczony z wartości bezwzględnych z odchyłek, co w pewnym sensie uwzględnia błąd systematyczny i może być on porównany z RMSE. Jednocześnie w publikacji (Höhle, Höhle 2009) zaproponowano metodę odporną bazującą na zmodyfikowanym rozkładzie normalnym. Problem związany z histogramami odchyłek NMT, jak wspomniano wcześniej, polega na tym, że histogramy te są bardzo wąskie i charakteryzują się "długimi ogonami". Jak wiadomo, rozkład normalny jest bardzo czuły na wartości odstające, a zatem "długie ogony" są dla niego problemem. W przypadku ciągłości danych w partii "ogonów" nie

mamy możliwości jednoznacznej detekcji i usunięcia wartości odstających. W tej sytuacji można próbować dopasować inną funkcję rozkładu prawdopodobieństwa. Przegląd różnych rozkładów doprowadził do rozkładu Laplace'a, który to rozkład, jak się okazało z badań literaturowych, był stosowany w dziedzinie nauk ekonomicznych (Gel 2010). Badanie dopasowania funkcji modelującej prawdopodobieństwo do rozkładów praktycznych musi się opierać na jakieś ocenie stopnia dopasowania. Można w tym zakresie przytoczyć publikację (Shunn i Wallach 2005). Oprócz informacji na temat braku jednoznacznych parametrów oceny dopasowania modelu do danych pomiarowych, wśród różnych parametrów tej oceny podają oni, jako rekomendowany, RMSD (*root mean squared deviation*) lub RMSSD (*root mean squared scaled deviation*). Jedną z popularnych metod jest test *chi-kwadrat*, który pozwala powiedzieć na ile model jest dopasowany do danych rzeczywistych. Jeśli tzw. *p-value* (prawdopodobieństwo popełnienia błędu polegającego na tym, że odrzucimy prawdziwą hipotezę o dobrym dopasowaniu modelu do danych) nie przekracza założonej wartości poziomu istotności testu, zwykle 0,05, to możemy uznać, że model jest źle dopasowany do danych rzeczywistych. Jest to jeden z testów badania dopasowania modelu. Krytyczne uwagi na jego temat można znaleźć w cytowanej powyżej publikacji. Generalnie jako dobry parametr uznano błąd RMSE i on też stał się podstawą oceny dokładności wpasowania modelu w dane pomiarowe. Ściślej mówiąc obliczane, były odchyłki pomiędzy histogramem empirycznym i przyjętą funkcją gęstości prawdopodobieństwa, z których następnie obliczany był błąd RMSE.

Zdefiniujmy x jako błąd, odchyłkę pomiędzy wartością pomierzoną i referencyjną:

$$x = x_{REF} - x_{obs} \tag{3.1}$$

gdzie:

x_{REF}– wartość referencyjna x_{obs} – wartość obserwowana

Gęstość prawdopodobieństwa rozkładu błędu f_1, zakładając rozkład normalny można zapisać następującym wzorem:

$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{3.2}$$

gdzie:

μ – średnia odchyłka

σ - odchylenie standardowe odchyłek

Średnią odchyłkę (μ) i odchylenie standardowe (σ)można oszacować następująco:

$$\mu = \frac{\sum_{x=1}^{n} x_i}{n} \tag{3.3}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-1}} \tag{3.4}$$

gdzie:

n - liczba obserwacji.

Dodatkowo powszechnie używany jest błąd średni kwadratowy (*RMSE-Root Mean Squared Error*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n-1}} \tag{3.5}$$

Dla rozkładów definiuje się również miary asymetrii (s - skośność) i skupienia (k - kurtoza), które można obliczyć z następujących wzorów:

$$s = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^3 \tag{3.6}$$

$$k = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{3.7}$$

Jako alternatywną funkcję gęstości prawdopodobieństwa można zaproponować funkcję gęstości rozkładu Laplace'a.

$$f_2(x) = \frac{1}{2b} \exp(-\frac{\mid (x-m) \mid}{b}) \tag{3.8}$$

przy czym

$$b = \frac{1}{n} \sum_{i=1}^{n} \mid x_i - m \mid \tag{3.9}$$

gdzie:

m - mediana.

W rozkładzie Laplace'a wariancja jest obliczana jak poniżej:

$$var = 2b^2 \tag{3.10}$$

Można podać funkcję odwrotną dystrybuanty, która umożliwia obliczenie wartości błędu występującej z określonym prawdopodobieństwem (p):

$$F_2^{-1}(p) = m - b\ sign(p - 0.5)\ ln(1 - 2 \mid p - 0.5 \mid) \tag{3.11}$$

Trzecia testowana funkcja, proponowana w publikacji (Höhle, Höhle 2009) bazuje na rozkładzie Gaussa, przy czym zamiast wartości średniej błędu przyjmuje się medianę (m), a zamiast odchylenia standardowego, znormalizowane odchylenie bezwzględne mediany (NMAD).

$$NMAD = 1.4826\ mediana(\mid x_i - m \mid) \tag{3.12}$$

W dalszej kolejności można zapisać gęstość prawdopodobieństwa zgodnie z formułą Gaussa, dla wartości mediany i NMAD:

$$f_3(x) = \frac{1}{\sqrt{2}\pi\ NMAD^2} \exp(-\frac{(x-m)^2}{2\ NMAD^2}) \tag{3.13}$$

Funkcja f_3 jest podstawą zaproponowanej w publikacji (Höhle, Höhle 2009) metody odpornej jako alternatywy do metody Gaussa.

Funkcje gęstości rozkładów prawdopodobieństwa umożliwiają podanie parametrów opisujących błąd, czy niepewność. Błąd średni i odchylenie standardowe są takimi parametrami w rozkładzie normalnym, mediana i NMAD w metodzie odpornej oraz mediana i $\sqrt{2}\cdot b$ w rozkładzie Laplace'a.

Percentyle: 68,5% i 95% oblicza się dla wartości bezwzględnych błędu, x i porównuje z odchyleniem standardowym σ i $1,96 \cdot \sigma$.

Celem prowadzonych analiz statystycznych było określenie każdorazowo najbardziej wiarygodnego parametru opisującego niepewność danych. We wszystkich przypadkach dostępne były wartości referencyjne, przy czym ich jakość była różna. Za każdym razem zbiór danych był na tyle duży, że można było analizować histogram odchyłek pomiędzy wartościami referencyjnymi, a danymi.

Wszystkie zamieszczone w monografii praktyczne przykłady obliczeniowe zostały przedstawione zgodnie z poniższym schematem. Sporządzane były na ogół dwa rysunki. Na pierwszym z nich prezentowano następujące wyniki analiz, oznaczone na rysunku literami:

- a - wykres, na którym znajduje się histogram empiryczny, oraz wykresy funkcji gęstości prawdopodobieństwa: f_1, f_2, f_3,
- b - wykres różnic pomiędzy wartościami funkcji gęstości rozkładów prawdopodobieństwa a wartościami empirycznymi w zależności od wartości odchyłki, x,
- c - tabela zawierająca odwrotności dystrybuant: $F_1^{-1}(0,975)$, $F_2^{-1}(0,975)$ i $F_3^{-1}(0,975)$,
- d - tabela zawierająca parametry dotyczące analizowanych odchyłek oraz poszczególnych rozkładów,
- e - wykres błędów RMSE dopasowania poszczególnych funkcji rozkładów prawdopodobieństwa do histogramu empirycznego.

Na rysunku w pozycji: c), w tabeli podawano następujące wielkości: liczba wszystkich danych, liczba danych usuniętych z analizy, wartość mi-

nimalna i maksymalna odchyłki, średnia odchyłka, odchylenie standardowe, rozszerzenia odchylenia standardowego, błąd średni kwadratowy i jego rozszerzenie, kurtoza, skośność, mediana, percentyl 68,3%, percentyl 95%, NMAD, b oraz pierwiastek z wariancji dla rozkładu Laplace'a. Należy w tym miejscu zauważyć, że w zasadzie nie usuwane były prawie żadne dane pomiarowe w klasycznym rozumieniu wartości odstających, a jedynie oczywiste pomyłki. W przypadku przedmiotowych histogramów mamy do czynienia z tak zwanymi "długimi ogonami", które świadczą o dużej liczbie wartości znacznie mniejszych i znacznie większych niż wartość średnia, przy czym histogram jest ciągły, a ogony poza tym, że "długie" są także relatywnie "grube". Fakt ten właściwie uniemożliwia obcięcie wartości odstających, bo trudno stwierdzić, od którego momentu mamy do czynienia z wartościami odstającymi. Reguła 3 odchyleń standardowych wydaje się tutaj mało wiarygodna.

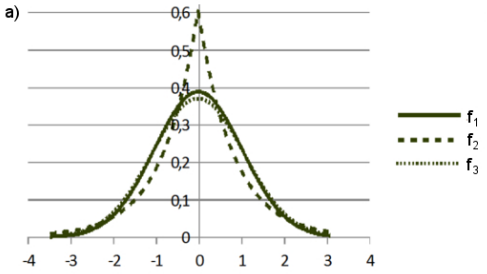
Drugi z rysunków zawiera wykresy kwantylowe Q-Q, przedstawiające zależność kwantyli teoretycznych w zależności od kwantyli empirycznych/modelowanych wraz z błędami RMSE ilustrującymi odstępstwo kwantyli empirycznych w stosunku do teoretycznych obliczonych zgodnie z funkcjami f_1, f_2, f_3.

Normalność rozkładu może być analizowana z wykorzystaniem testów statystycznych: Kołmogorowa-Smirnowa, Lillieforsa czy Shapiro-Wilka. Wyniki tych testów okazały się negatywne jedynie w przypadkach rozkładów o idealnym Gaussowskim kształcie. W innym przypadku zawsze dawały wynik zmuszający do odrzucenia hipotezy o normalno-

ści rozkładu. Dlatego też, praktycznie wyniki tych testów zamieszczono jedynie w pewnych rzadkich sytuacjach, kiedy rozkład był prawie idealny.

W celu przetestowania różnych, przyjętych do analiz parametrów dokładności przeprowadzono test na danych symulowanych. Za pomocą generatora liczb losowych wygenerowano w sposób przypadkowy, wykorzystując standardowy rozkład normalny Gaussa N(0,1) trzy różnoliczne zbiory (13, 61, 1000 liczb).

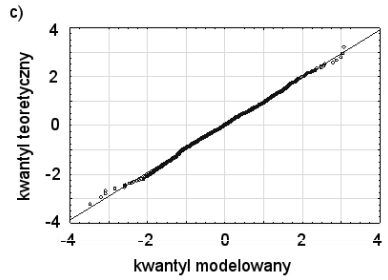
a)



b)

liczba punktów	13	61	1000
min	-1,261	-2,107	-3,487
max	1,659	2,387	3,079
μ	0,025	0,091	-0,021
σ	0,976	0,970	1,025
1,96 σ	1,914	1,900	2,009
$F_1^{-1}(0.975)$	1,940	1,967	1,980
m	0,124	0,208	-0,024
2^0,5 b	1,104	1,072	1,165
$F_2^{-1}(0.975)$	2,462	2,477	2,445
m	0,124	0,208	-0,024
NMAD	1,126	0,934	1,075
P(0,683)	1,034	1,007	1,072
$F_3^{-1}(0.975)$	2,330	2,038	2,083
P(0,95)	1,654	1,975	1,907

c)



Rysunek 9: Wykresy funkcji gęstości prawdopodobieństwa: f_1, f_2, f_3 a), dla danych wygenerowanych symulatorem liczb losowych, zakładając rozkład normalny N(0,1), parametry statystyczne b) oraz wykres normalności: Q-Q c); przebieg funkcji f_1 i f_2 jest prawie identyczny, niewielkie różnice są widoczne dla wartość odciętych bliskich 0

Na rysunku 3 b) w tabeli są pogrupowane wartości dla poszczególnych funkcji, kolejno: f_1, f_2, f_3. Średnia odchyłka jest zbliżona do

mediany i wynosi około 0. Odchylenie standardowe, NMAD, percentyl 68.5% i odchylenie standardowe dla funkcji Laplace'a są bliskie wartości 1. Rozszerzenia odchylenia standardowego na poziomie ufności 95%, zakładając rozkład Gaussa i dla funkcji odpornej, są podobne. Wartości 1,96 σ, F_1^{-1} (0,975), F_3^{-1} (0,975) i percentyl 95% wynoszą ok. 2. Natomiast dla rozkładu Laplace'a odpowiadająca im wartość $F_2^{-1}(0{,}975)$ wynosi około 2,5. Wartość ta wynika z istotnie różnego kształtu analizowanej funkcji rozkładu Laplace'a. Konkluzja wydaje się oczywista: funkcje bazujące na rozkładzie Gaussa opisują parametry statystyczne danych, które mają właśnie taki rozkład. Rozkład Laplace'a nie pasuje do tych danych i w tym przypadku przeszacowuje wartość rozszerzenia odchylenia standardowego na poziomie ufności 0.95. Dla wszystkich trzech zbiorów funkcja Gaussa i odporna są ze sobą zbieżne, przy czym im więcej liczb tym bardziej są one podobne. Wszystkie odpowiadające sobie parametry rozkładu Gaussa i rozkładu odpornego są zbliżone w przypadku dużego zbioru (μ i m; σ, NMAD i P(0,683) oraz 1,96 σ, $F_1^{-1}(0{,}975)$, F_2^{-1} (0,975) i P(0,95)). Niektóre parametry dla zbioru 13 liczb znacznie odbiegają od wartości dla zbiorów liczących 61 i 1000 wartości. Rozkład Laplace'a nie pasuje do rozkładu modelowanego za pomocą rozkładu Gaussa; parametry tego rozkładu różnią się znacząco od parametrów rozkładu Gaussa i rozkładu z metody odpornej.

W następnych rozdziałach zostały zaprezentowane dwa przykłady praktyczne. Jeden dotyczy szacowania błędów w pomiarach powierzchni, a drugi analizy dokładności pionowej NMT.

Rozdział 4

Określanie niepewności w systemie IACS

Kraje UE przeznaczają około połowę budżetu na rolnictwo. Można dyskutować nad zasadnością tych dopłat w ogóle, nie leży to jednak w zakresie tematycznym niniejszej monografii. UE wspiera sektor rolniczy w różnych dziedzinach, między innymi poprzez tzw. dopłaty bezpośrednie, które są wypłacane rolnikom w zależności od zadeklarowanej powierzchni upraw. Rozporządzenie Komisji Europejskiej (WE) Nr 1122/2009 reguluje działania w tym zakresie między innymi ustanawiając system kontroli, tzw. Zintegrowany System Zarządzania i Kontroli, częściej znany w wersji angielskiej jako IACS (*Integrated Administration Control System*). Zgodnie z rozporządzeniem, kraje UE są zobowiązane do prowadzenia przestrzennej bazy danych o działkach rolnych, które są deklarowane do dopłat. Jest to tzw. System Identyfikacji Działek Rolnych (*LPIS - Land Parcel Identification System*). W Polsce system LPIS bazuje na danych ewidencyjnych; dodatkowym źródłem danych jest ortofotomapa. Korzyścią związaną z wejściem Polski do UE i koniecznością zbudowania systemu LPIS było zarówno ujednolicenie danych ewidencyjnych do takiej postaci, żeby wszystkie były równocześnie dostępne dla całego kraju, czego przez wiele lat wcześniej nie udało się sfinalizować, jak i wykonanie ortofotomapy o wysokiej rozdzielczości przestrzennej. Warto

też wspomnieć o produkcie "ubocznym" ortofotomapy czyli o numerycznym modelu terenu, który został utworzony przy okazji i został czasowo wykorzystany w innym dużym systemie informacji przestrzennej, wdrożonym w Ośrodkach Koordynacyjno-Informacyjnych (OKI) przy Regionalnych Zarządach Gospodarki Wodnej (RZGW), zastępując nieudany produkt tzw. NMT OKI. W związku z dużymi środkami finansowymi przeznaczanymi na system dopłat bezpośrednich, w UE obowiązuje szereg procedur kontrolnych i audytów. Między innymi kontroli poddaje się wybraną liczbę deklaracji mierząc powierzchnię działki rolnej i określając typ uprawy. Wyżej wymienione rozporządzenie podaje pewne ograniczenia i zawiera opis restrykcji związanych z procedurą dopłat bezpośrednich. Artykuł 34 określa maksymalną dopuszczalną tolerancję pomiaru kontrolnego powierzchni działki rolnej, która wynosi 1,5 m. Przez tolerancję mnoży się obwód działki rolnej, uzyskując tym samym dopuszczalną odchyłkę pomiędzy powierzchnią deklarowaną a pomierzoną podczas kontroli. Artykuły 55 i 58 niniejszego rozporządzenia podają również inną wartość graniczną: 3% powierzchni, której przekroczenie powoduje restrykcje w postaci obniżenia płatności. Na oficjalnych stronach UE znajduje się ponadto dokument (Komisja Europejska 2007), w którym zawarto również informacje na temat procedur kontrolnych w systemie IACS. Podano tam reguły jakimi kieruje się UE w tzw. korektach finansowych, co w praktyce oznacza zmniejszenie kwoty dopłaty pochodzącej z UE, czyli *de facto* określa kwoty do zwrotu. Należy w tym miejscu zauważyć, że nigdzie nie zostały podane żadne podstawy teoretyczne przyjęcia wyżej wymienionych wartości granicznych. Jednostką

odpowiadającą za szczegółowe wdrożenia zasad funkcjonowania systemu IACS jest ośrodek badawczy JRC w Isprze, we Włoszech. Na stronach internetowych JRC (JRC MARS 2013 a) znajduje się szereg informacji oraz zaleceń, którymi muszą kierować się wszystkie kraje członkowskie. Z geodezyjnego punktu widzenia istotne są dwa działy, dotyczące procedur kontrolnych pomiarów powierzchni działek rolnych oraz kontroli jakości systemu LPIS. W przypadku pomiarów kontrolnych zdefiniowano procedurę walidacji metod pomiaru powierzchni działek rolnych oraz sposób pomiaru kontrolnego. Jeśli chodzi o kontrolę jakości systemu LPIS, to zastały opracowane szczegółowe, skomplikowane i rozbudowane procedury kontrolne. Należy w tym miejscu zwrócić uwagę na pewne sprzeczności. W obu przypadkach wykonuje się pomiar kontrolny powierzchni działki rolnej. W pierwszym przypadku pomiary wykonuje się różnymi instrumentami geodezyjnymi; może to być zarówno tachymetr jak i taśma, ale najczęściej wykorzystuje się tanie odbiorniki GPS w trybie autonomicznym lub coraz częściej z poprawkami DGPS oraz pomiar na ortofotomapie. Dopuszczalną odchyłkę oblicza się w jednostkach powierzchni poprzez pomnożenie obwodu działki przez tolerancję (maksymalne 1.5 m jak wspomniano wyżej). Tą procedurę nazywa się kontrolą na miejscu. W przypadku kontroli LPIS, pomiaru dokonuje się na ortofotomapie (często tej samej, którą wykorzystuje się podczas kontroli deklaracji) lub rzadziej z wykorzystaniem RTK. Dopuszczalne granice podawane są w procentach: 3% powierzchni działki (zgodnie z rozporządzeniem) oraz 5% i 7% dla małych działek (JRC MARS 2013 d). Wydaje się, że te procesy (kontroli na miejscu i kontroli LPIS), mogą przebiegać równolegle

i doprowadzać do niespójnych sytuacji. Może się bowiem zdarzyć dla tej samej działki, że wynik kontroli na miejscu będzie pozytywny, a wynik kontroli w systemie LPIS negatywny.

Jeśli chodzi o procedury kontrolne powierzchni deklarowanej, to składają się one z dwóch etapów. W pierwszym etapie wykonuje się walidację metody pomiarowej, a w drugim przeprowadza się już standardowe pomiary kontrolne. Procedury walidacji mają również szereg niespójności, co było przedmiotem innych publikacji (Kramarczyk i in. 2012, Hejmanowska i in. 2012). Głównym zarzutem dotyczącym walidacji było wykonywanie jej w warunkach "laboratoryjnych", na punktach sygnalizowanych i bezpośrednie przenoszenie uzyskanej tolerancji do pomiarów rzeczywistych działek rolnych o niesygnalizowanej granicy. Wydaje się to być niezrozumiałe. Jedynym wyjaśnieniem źródła tego problemu może być fakt, że JRC nigdy wcześniej, aż do 2011 roku nie zweryfikowało zaproponowanej przez siebie metody walidacji w rzeczywistych warunkach, pomimo wprowadzenia całej procedury w roku 2007 (Kay, Sima 2006). Jednakże nie to zagadnienie będzie przedmiotem poniższych rozważań. Podczas walidacji wykonuje się wiele powtórzeń pomiaru powierzchni każdej z działek testowych. Pozwala to ostatecznie na obliczenie odchylenia standardowego odtwarzalności, które określa niepewność pomiaru powierzchni wykonanego tą samą metodą przez różnych operatorów (JRC MARS 2013 b). W dalszej kolejności oblicza się tolerancję pomiaru dzieląc odchylenie standardowe odtwarzalności przez obwód działki. Taka standaryzacja jest konieczna ze względu na to, że odchylenie standardowe jest wyznaczane w m^2 i jest to wartość zależna

od powierzchni działki. Niepewność powierzchni w m^2 rośnie wraz ze wzrostem powierzchni. Można by obliczyć odchylenie względne, jako pewien rodzaj standaryzacji tej niepewności. Niestety w tym przypadku odchyłka względna bardzo rośnie wraz ze zmniejszaniem się powierzchni działki. Ostatecznie więc, w wyniku walidacji, po podzieleniu odchylenia przez obwód uzyskujemy tolerancję w metrach.

Wszystkie analizy prowadzące do obliczenia niepewności pomiaru powierzchni działki, w tym przypadku nazywanej tolerancją pomiaru, są oparte o założenie rozkładu normalnego. W przypadku analiz statystycznych odchyłek uzyskanych w kontroli systemu LPIS, również zakłada się milcząco normalny ich rozkład. W następnej części monografii zostaną zaprezentowane rozważania w oparciu o dane praktyczne dotyczące walidacji i kontroli LPIS, wraz z dyskusją zasadności takiego podejścia.

4.1. Pomiar referencyjny TS

Przykład zamieszczony w niniejszym rozdziale dotyczy pomiarów referencyjnych wykonywanych w ramach walidacji metod pomiaru powierzchni działek rolnych, w systemie dopłat bezpośrednich do rolnictwa IACS. Walidację przeprowadza się na wybranych działkach wykonując wielokrotny pomiar GNSS powierzchni działek przez różnych operatorów. Jako wartość referencyjną przyjmuje się powierzchnie uzyskane w wyniku pomiaru o większej dokładności, zwykle RTK lub TS. Schemat walidacji można znaleźć na stronach JRC (JRC MARS 2013 a). Milcząco zakłada się, że tak wykonany jednokrotny pomiar referencyjny jest wystarcza-

jąco dokładny w stosunku do pomiaru GNSS, co niekoniecznie musi być słuszne, szczególnie jeśli walidacja dotyczy wysokiej jakości urządzeń GNSS, a pomiar jest wykonywany metodą różnicową. Problem ten nie jest w ogóle poruszany w dokumentach JRC, opisujących obowiązujące procedury walidacjach. Niemniej jednak dokładność pomiaru referencyjnego jest uwzględniana w wykonywanych praktycznie walidacjach (Kramarczyk i in. 2012). Poniżej zamieszczono przykładową analizę dokładności pomiaru referencyjnego przeprowadzonego za pomocą TS.

Przykładowy zestaw działek wykorzystywanych podczas walidacji urządzeń GNSS (Kramarczyk, Warchoł 2011) znajduje się na rysunku (Rys.10).

Rysunek 10: Przykładowy zestaw działek wykorzystywanych w walidacji GPS (Błonia, Kraków) (Kramarczyk, Warchoł 2011)

Działki są sygnalizowane palikami, nie tylko w narożnikach ale

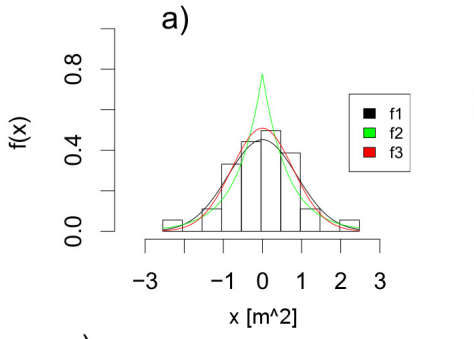
także wzdłuż granicy, tak, że jeden punkt przypada na około 25 m. Pomiar referencyjny wykonuje się tachymetrem lub RTK, mierząc każdy z punktów. W omawianym przykładzie powierzchnie działek były mierzone z wykorzystaniem tachymetru elektronicznego. Każda działka była mierzona 3 razy w dwóch położeniach lunety, z różnych punktów wewnątrz działki. Wartości średnie powierzchni i obwodów działek znajdują się w tabeli (Tab.2).

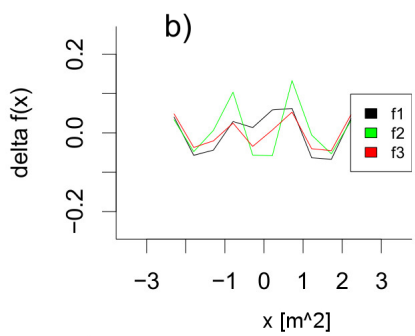
Następnie obliczono różnice pomiędzy każdym pomiarem, a wartością średnią. Uzyskane w ten sposób różnice poddano analizie statystycznej, wygenerowano histogram różnic oraz parametry analizowanych rozkładów (Rys. 11 i 12). Podobną analizę wykonano dla względnych różnic (Rys. 13 i 14).

Tabela 2: Zestawienie powierzchni i obwodów referencyjnych działek (Kramarczyk, Warchoł 2011)

Symbol działki	Powierzchnia $[m^2]$	Obwód $[m]$
A	3425,27	238,07
B	3429,62	237,87
C	7510,58	564,91
D	10046,76	400,76
E	10918,38	474,59
F	15548,93	564,05

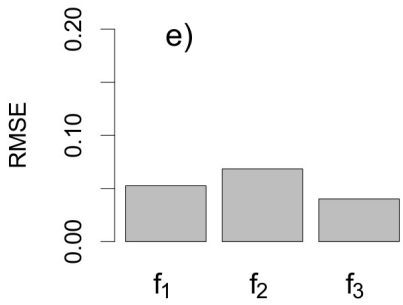
Kształt histogramu (Rys. 11 a) jest najbardziej zbliżony do funkcji gęstości rozkładu normalnego. Wykres Q-Q (Rys. 12) oraz wartości statystyk testów normalności wskazują na rozkład normalny (Kołmogorowa-Smirnowa d=0,098, p>0.20; Lillieforsa p>0.20; Shapiro-Wilka W=0,998, p=0,468). Dla porównania wartości testów dla symulowanego rozkładu





c)

$F_1^{-1}(0.95)$	$F_2^{-1}(0.95)$	$F_3^{-1}(0.95)$
1,729 m^2	1,921 m^2	1,534 m^2



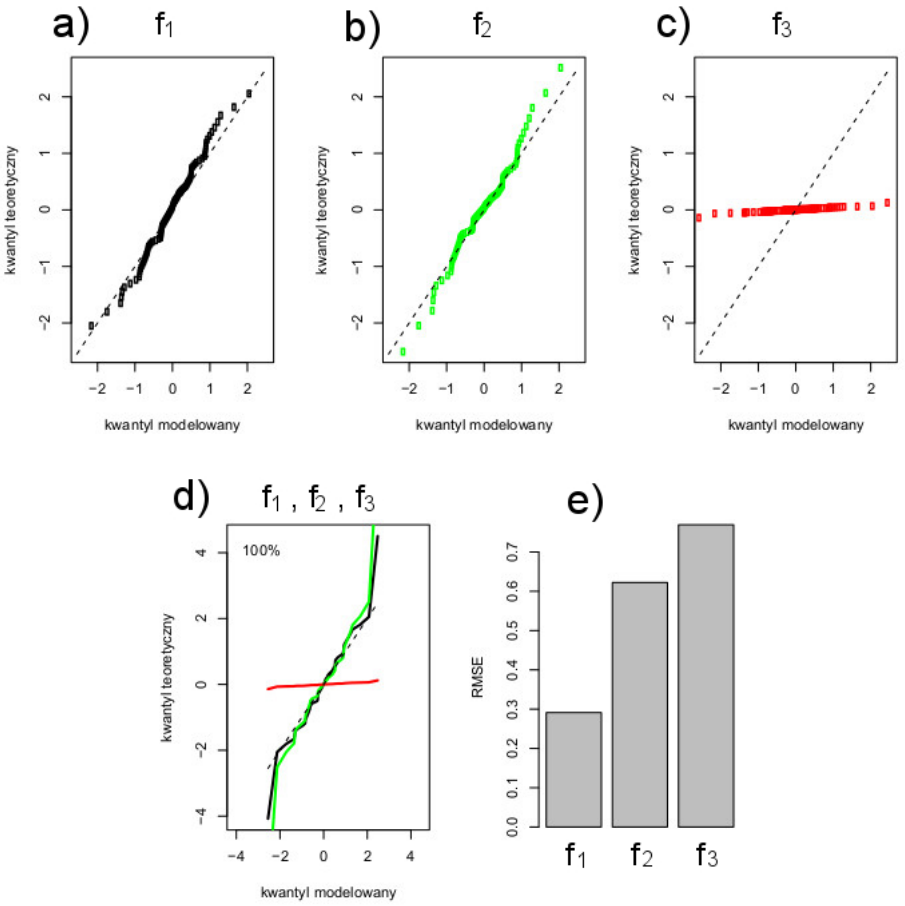
d)

Liczba punktów	36
Liczba wartości odstających	0
Min	-2,550 m^2
Max	2,470 m^2
μ	0,000 m^2
σ	0,882 m^2
1,64 σ	1,446 m^2
1,96 σ	1,729 m^2
RMSE	0,882 m^2
1,96*RMSE	1,729 m^2
k	2,050
s	-0,076
m	-0,002 m^2
P 0,683	0,744 m^2
P 0,95	2,477 m^2
NMAD	0,784 m^2
b	0,642 m^2
2^0,5*b	0,908 m^2

Rysunek 11: Histogram odchyłek powierzchni dla pomiaru TS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

Gaussa są przykładowo następujące: Kołmogorowa-Smirnowa d=0,019, p>0.20; Lilieforsa p>0.20; Shapiro-Wilka W=0,969, p=0,406.

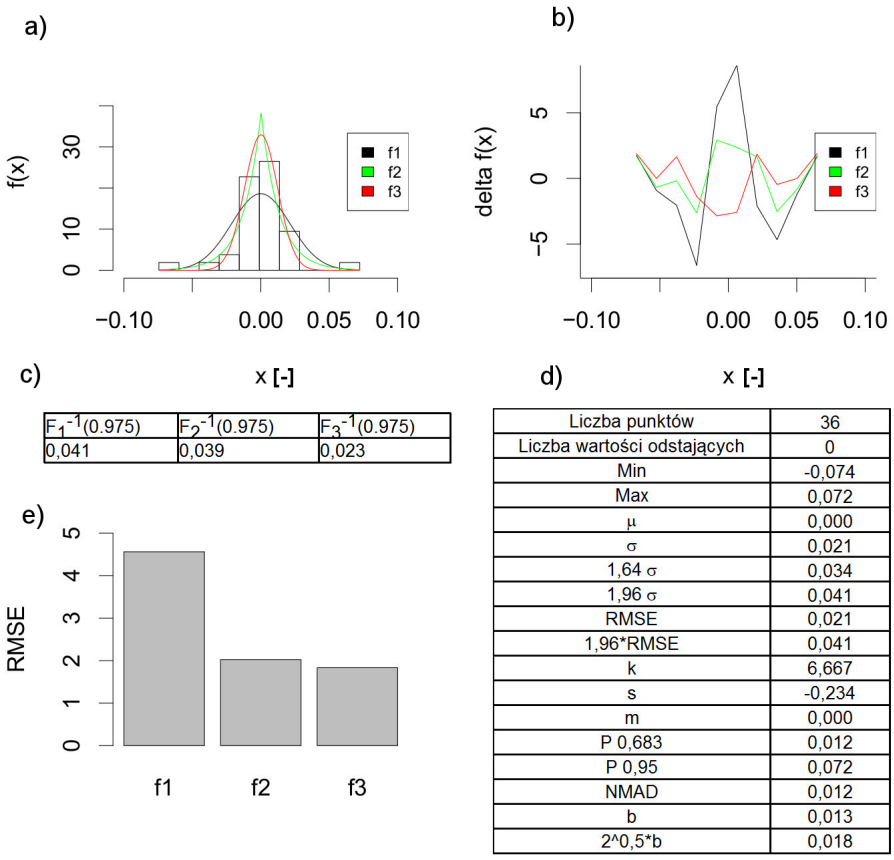
Odchylenie standardowe wynosi około 0.9 m^2 (Rys.11 d) i jest zbliżone do odchylenia standardowego w rozkładzie Laplace'a. NMAD i percentyl P(0,683) są nieco mniejsze i wynoszą około 0,8 m^2. Rozszerzenie



Rysunek 12: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni z pomiaru TS

na poziomie ufności 0,95 wynosi około 1,7 m^2 dla przypadku rozkładu Gaussa i 1,5 m^2 dla metody odpornej. W przypadku rozkładu Laplace'a, odpowiadająca temu rozszerzeniu wartość funkcji odwrotnej do dystrybuanty jest większa i wynosi 1,9 m^2.

Przyjmijmy rozszerzenie odchylenia standardowego z rozkładu nor-

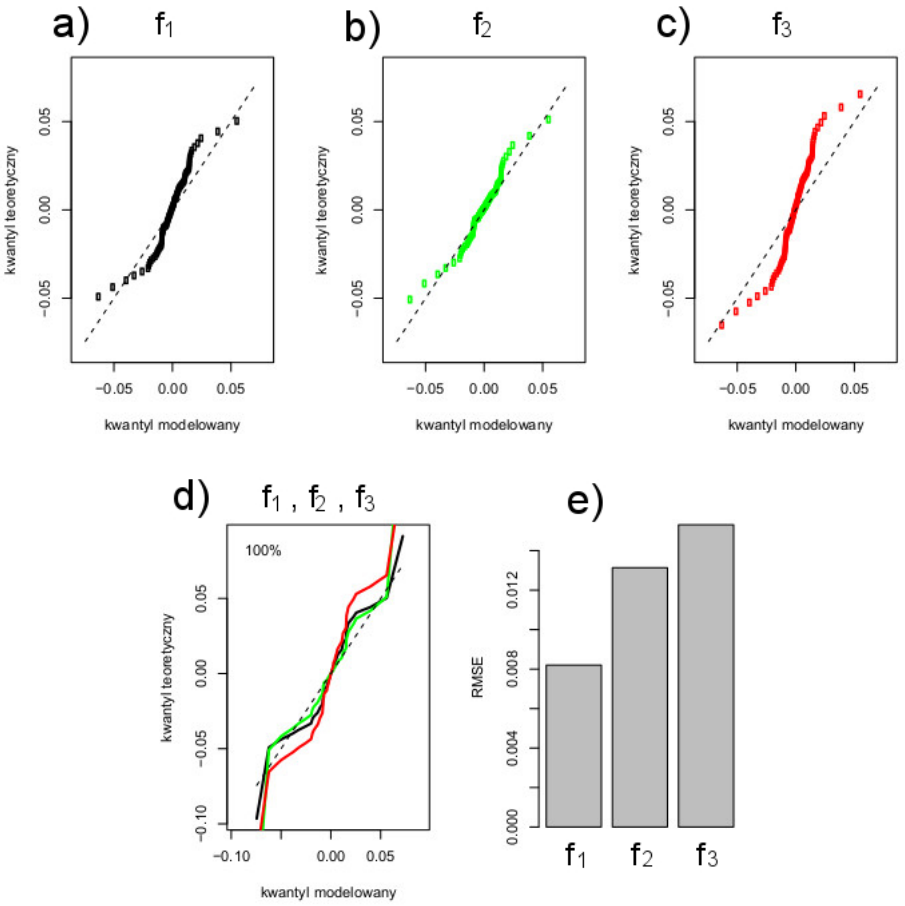


$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
0,041	0,039	0,023

Liczba punktów	36
Liczba wartości odstających	0
Min	-0,074
Max	0,072
μ	0,000
σ	0,021
1,64 σ	0,034
1,96 σ	0,041
RMSE	0,021
1,96*RMSE	0,041
k	6,667
s	-0,234
m	0,000
P 0,683	0,012
P 0,95	0,072
NMAD	0,012
b	0,013
2^0,5*b	0,018

Rysunek 13: Histogram odchyłek względnych powierzchni dla pomiaru TS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

malnego (1,7 m^2) jako najbardziej wiarygodny wskaźnik niepewności pomiaru referencyjnego. Wartość tę można przeliczyć na tolerancję, którą zwykle uzyskuje się podczas walidacji i wyraża w metrach, wynosi ona w tym przypadku 0,007 m. Można ją zatem jeszcze uznać za wartość referencyjną dla innych pomiarów, bowiem w wyniku walidacji najdokładniejszych metod pomiaru powierzchni z wykorzystaniem różnicowego



Rysunek 14: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek względnych powierzchni z pomiaru TS

pomiaru GPS wartości tolerancji wynoszą kilka centymetrów. W materiałach JRC nie ma żadnej informacji na temat tolerancji pomiaru tachimetrycznego. Uzyskany wynik można uznać w tym kontekście za wartościową informację na temat dokładności pomiarów referencyjnych wykonywanych za pomocą TS. Należy również w tym momencie nadmienić, że właściwym pojęciem charakteryzującym pomiar referencyjny jest nie-

pewność, ze względu na to, że wykonując pomiar referencyjny nie znamy wartości "referencyjnych", czyli znacznie dokładniejszych. Pomiar tachymetryczny jest bowiem obecnie najdokładniejszą metodę pomiarową. W kontroli na miejscu wykorzystuje się tolerancję w metrach, a po pomnożeniu przez obwód działki uzyskuje się niepewność powierzchni w metrach kwadratowych. Względne różnice nie są brane pod uwagę ani w walidacji i ani w kontroli na miejscu. Niemniej jednak w celu poznania również charakteru rozkładu odchyłek względnych przeprowadzono analizę, której wyniki zamieszczono poniżej.

Analizując histogram odchyłek względnych (Rys. 13) można zauważyć, że jest on bardziej smukły (k=6,667) oraz, że funkcja gęstości rozkładu Gaussa nie opisuje go poprawnie. Lepiej dopasowana jest funkcja odporna i funkcja gęstości rozkład Laplace'a, co potwierdzają wartości błędów wpasowania (Rys. 13 e). Jednakowoż odwrotności dystrybuanty dla funkcji Laplace'a i Gaussa są niemal takie same (0,04%) i zbieżne z 1,96 σ i 1,96 RMSE. Należy zauważyć ponadto, że percentyl P(0,95), zarówno dla odchyłek bezwzględnych, jak i względnych jest znacznie większy od odpowiednich odwrotności dystrybuant.

W przypadku odchyłek względnych zarówno wykres Q-Q (Rys. 14), jak i wartości testów (Kołmogorowa-Smirnowa, d=0,180, p<0.20; Lilieforsa, p<0.01, Shapiro-Wilka W=0,841, p=0,000) wskazują na przebieg histogramu empirycznego nieco odbiegający od rozkładu normalnego. Analizując wizualnie wykresy Q-Q wydaje się, że najlepiej dopasowany jest rozkład Laplace'a (Rys. 14 b). Jednak porównanie błędów RMSE

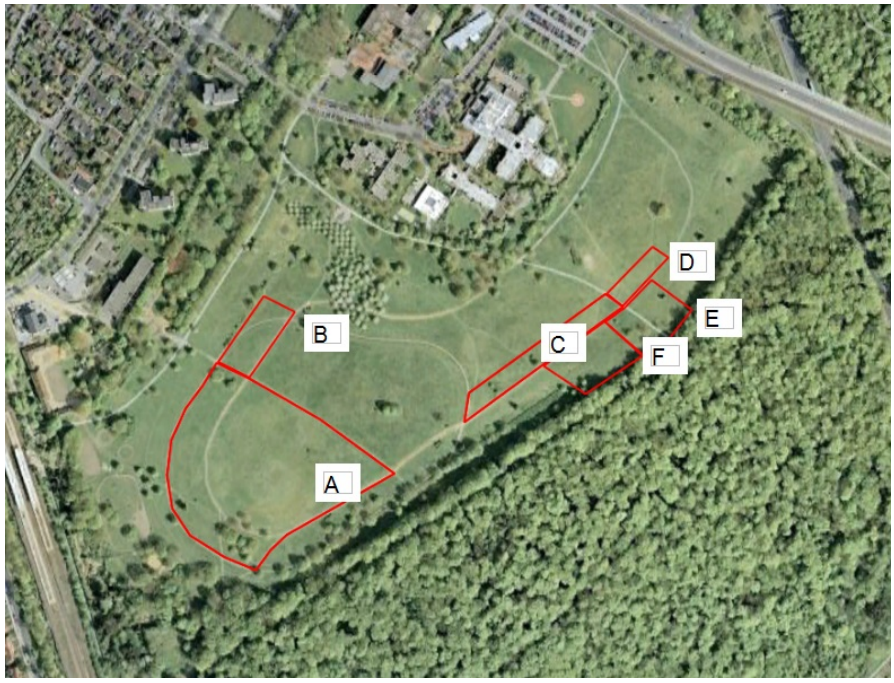
(Rys. 14 e) sugeruje, że najlepiej dopasowany jest rozkład Gaussa. Należy zwrócić jednak uwagę na stosunkowo małe wartości błędów RMSE dla wszystkich funkcji. Nasuwa się przy okazji wniosek, że być może jednak lepszym wskaźnikiem dopasowania byłby parametr związany z analizą odległości empirycznego wykresu kwantylowego od linii prostej kwantyla teoretycznego.

Podsumowując, rozkład odchyłek bezwzględnych ma charakter bardziej zbliżony do rozkładu normalnego niż rozkład odchyłek względnych. W związku z tym, podobnie jak w przypadku rozkładu symulowanego Gaussa, najbardziej wiarygodny wynik uzyskuje się dla rozkładu normalnego 1,729 m^2 (Rys. 11 c). W tym przypadku zastosowanie rozkładu Laplace'a nieco zawyża wartość niepewności (1,921 m^2). Natomiast w przypadku odchyłek względnych rozkład odbiega od przebiegu rozkładu normalnego, niemniej jednak rozszerzona niepewność pomiaru wynosi zarówno w przypadku rozkładu Gaussa jak i Laplace'a 0,04% (Rys. 14 c). Metoda odporna w obu przypadkach dała zaniżone wyniki.

4.2. Pomiar walidacyjny GNSS

Walidacja metod pomiaru powierzchni działek rolnych polega na wielokrotnym pomiarze powierzchni działek testowych sygnalizowanych w terenie. Punkty pośrednie znajdują się nie tylko w narożnikach, ale także wzdłuż granicy, maksymalnie co 25 m. Walidacja jednego urządzenia GNSS trwa zwykle 2,5-3 dni, trzech operatorów wykonuje pomiar 5 lub 6 działek w trzech seriach, o różnej porze dnia (rano, w południe, po po-

łudniu). Każda seria polega na przejściu wzdłuż granicy działek 4 razy. W niniejszym rozdziale przeanalizowane zostały wyniki walidacji przeprowadzonej w Hanowerze (Kramarczyk 2011, Kramarczyk i inni 2012). Wartości powierzchni i obwodów działek testowych znajdują się w tabeli 3.



Rysunek 15: Przykładowy zestaw działek wykorzystywanych w walidacji pomiaru powierzchni działek z wykorzystaniem urządzeń GNSS (Hanower, Niemcy)

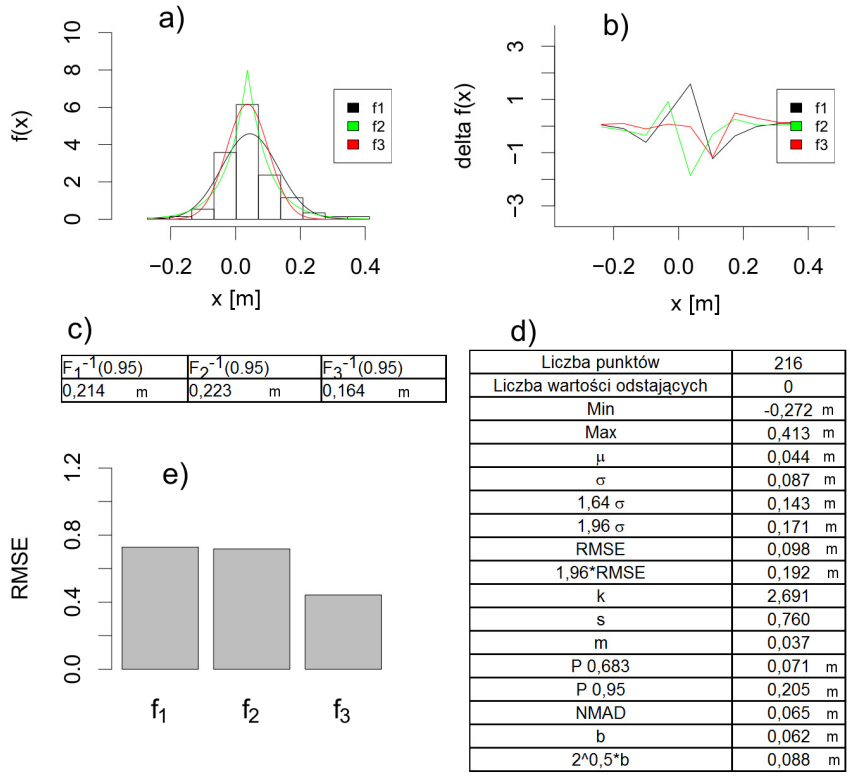
Dla każdego pomiaru obliczano odchyłkę powierzchni w stosunku do wartości referencyjnej, wyznaczonej w tym przypadku w oparciu o pomiar RTK, którą dzielono przez obwód uzyskując wartości tolerancji. Tak zestandaryzowane odchyłki poddano analizie statystycznej ze względu na kształt histogramu, który wskazuje na bardzo niewielkie odchylenie od

Tabela 3: Zestawienie powierzchni i obwodów referencyjnych działek (Hanower)

Symbol działki	Powierzchnia $[m^2]$	Obwód $[m]$
A	37027,79	769,48
B	5278,23	319,78
C	10355,21	522,56
D	5039,73	308,84
E	1573,43	170,18
F	7228,77	382,44

przebiegu funkcji gęstości rozkładu normalnego. Przeanalizujmy parametry dopasowanych rozkładów (Rys. 16 i 17). Wartość średnia zestandaryzowanej odchyłki w metrach wynosi: 0.044 m, wartość mediany jest zbliżona i wynosi: 0.037 m. Odchylenie standardowe Gaussa, NMAD, P(683) i odchylenie standardowe w rozkładzie Laplace'a są podobne i odpowiednio równe: 0.087 m, 0.065 m, 0.071 m i 0.088 m. Rozszerzenie niepewności na poziomie 95% jest zbieżne dla wszystkich parametrów i wynosi około 0.2 m. Błąd średni kwadratowy wpasowania dla histogramu jest najmniejszy dla funkcji odpornej (Rys. 16 e), a wykresu Q-Q dla rozkładu Gaussa (Rys. 17 e).

W przypadku pomiaru walidacyjnego GNSS, nie obserwuje się znaczącej zmiany wartości niepewności w zależności od dopasowywanego rozkładu. We wszystkich przypadkach rozszerzenie niepewności wynosiło około 0,20 m, nieznacznie mniejszą wartość uzyskano dla metody odpornej: 0,16 m. Analiza dokładności, przeprowadzona w oparciu o analizę odchyłek, jak powyżej opisano, jest zbieżna z wynikiem walidacji, który wyniósł w tym przypadku 0,21 m (Kramarczyk 2011). Widać, że uzy-

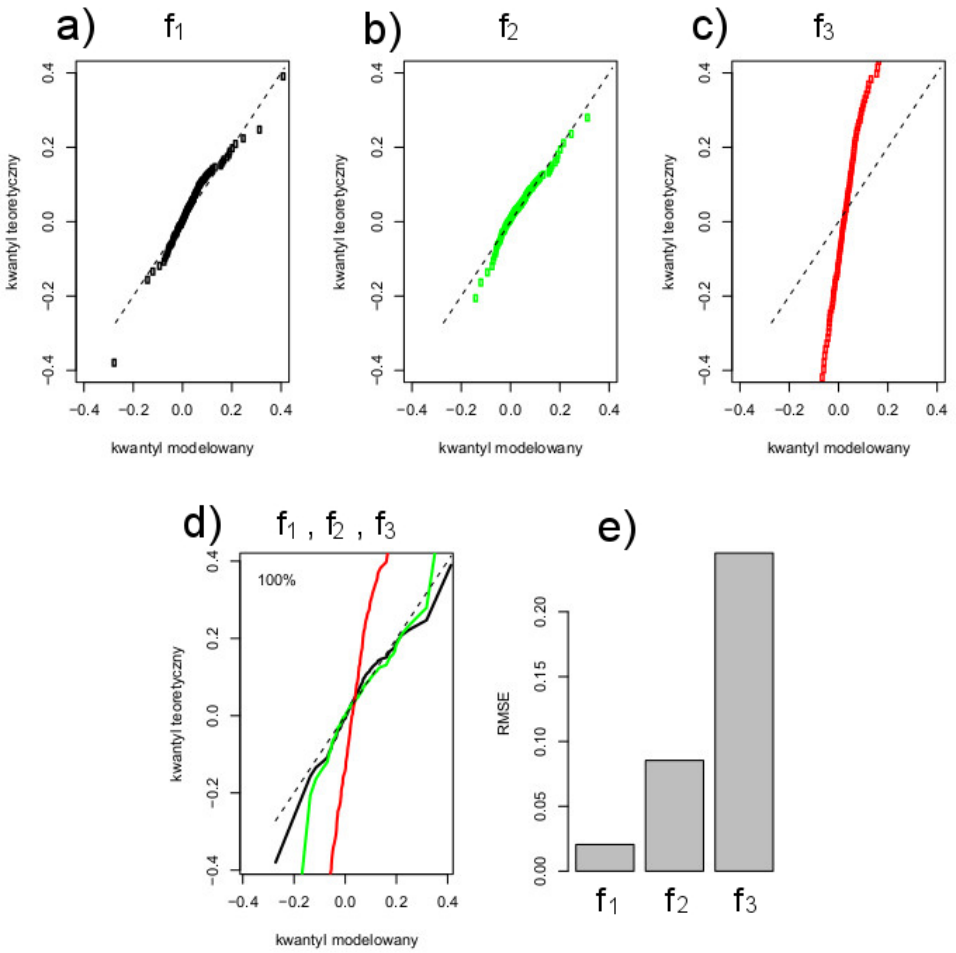


Liczba punktów	216
Liczba wartości odstających	0
Min	-0,272 m
Max	0,413 m
μ	0,044 m
σ	0,087 m
1,64 σ	0,143 m
1,96 σ	0,171 m
RMSE	0,098 m
1,96*RMSE	0,192 m
k	2,691
s	0,760
m	0,037
P 0,683	0,071 m
P 0,95	0,205 m
NMAD	0,065 m
b	0,062 m
2^0,5*b	0,088 m

Rysunek 16: Histogram odchyłek powierzchni dla pomiaru walidacyjnego GNSS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

skane wyniki są wzajemnie spójne i założenie normalności rozkładu w tym przypadku jest zasadne.

Ze względu na kontrolowane warunki pomiaru, dobra definicję obiektu oraz dostępność wartości referencyjnych można by w tym przypadku do określania tolerancji pomiaru powierzchni użyć pojęcia błąd.

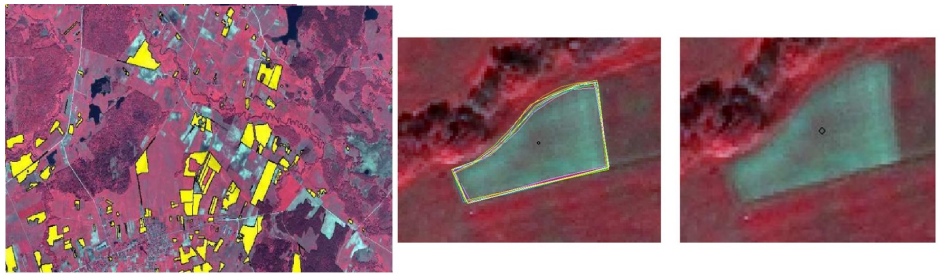


Rysunek 17: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni walidacyjnego GNSS

4.3. Pomiar walidacyjny IKONOS

Do kontroli powierzchni działek rolnych zadeklarownych przez rolnika wykorzystuje się także ortofotomapy. Proces walidacji obowiązuje nie tylko pomiary bezpośrednie, ale także pomiary powierzchni polegające na digitalizacji działki w oparciu o ortofotmapy lotnicze i satelitarne.

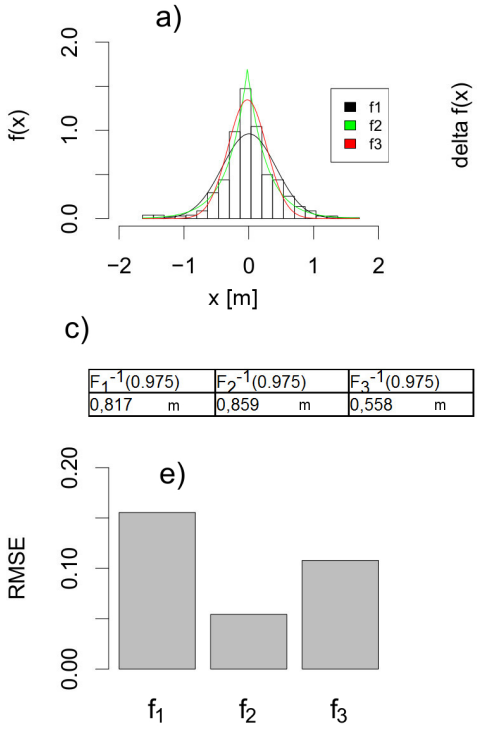
Walidacja pomiarów z wykorzystaniem ortofotomap przeprowadzana jest na podstawie wielokrotnego pomiaru powierzchni wykonywanego przez różnych obserwatorów, przy czym zestaw danych i schemat pomiaru jest nieco inny niż w przypadku walidacji urządzeń GNSS (JRC MARS 2013 a, b).



Rysunek 18: Przykładowy zestaw działek wykorzystywanych w walidacji obrazu satelitarnego (IKONOS) (Milenov i Gabalova 2009)

W niniejszym rozdziale zostały przedstawione wyniki analiz jednej z walidacji, obrazu satelitarnego IKONOS, wykonanej w JRC (Milenov i Gabalova 2009). W omawianej walidacji czterech operatorów wykonało 3-krotny pomiar powierzchni 56 działek. Ostatecznie po odrzuceniu wartości odstających, zgodnie z procedurą walidacji (JRC MARS 2013 b), do analizy statystycznej przyjęto 613 pomiary. Podobnie jak w przypadku walidacji GNSS, analizie poddano odchyłki w postaci wartości tolerancji, dzieląc różnice powierzchni przez obwód. Analiza histogramu i wykresu Q-Q (Rys. 19 i Rys. 20) wskazuje na pewną niezgodność histogramu empirycznego z rozkładem normalnym, głównie z lewej strony, choć generalnie rozkład empiryczny nie odbiega znacząco od przebiegu rozkładu Gaussa. Wartość średnia i mediana wynoszą odpowiednio: 0,004 m i -0,022 m,

co świadczy praktycznie o nieobecności błędu systematycznego. Wielkości: odchylenie standardowe w rozkładzie Gaussa, NMAD, P(0,683), odchylenie standardowe w rozkładzie Laplace'a są zbliżone i wynoszą odpowiednio (Rys. 19 d): 0,415 m, 0,296 m, 0,342 m, 0,416 m. Natomiast odwrotności dystrybuant wynoszą odpowiednio: 0,817 m, 0,859 m, 0,558 m (Rys. 19 c). W tym przypadku metoda odporna dała znacząco różny wynik, tzn. jak zwykle niedoszacowała niepewności.



$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
0,817 m	0,859 m	0,558 m

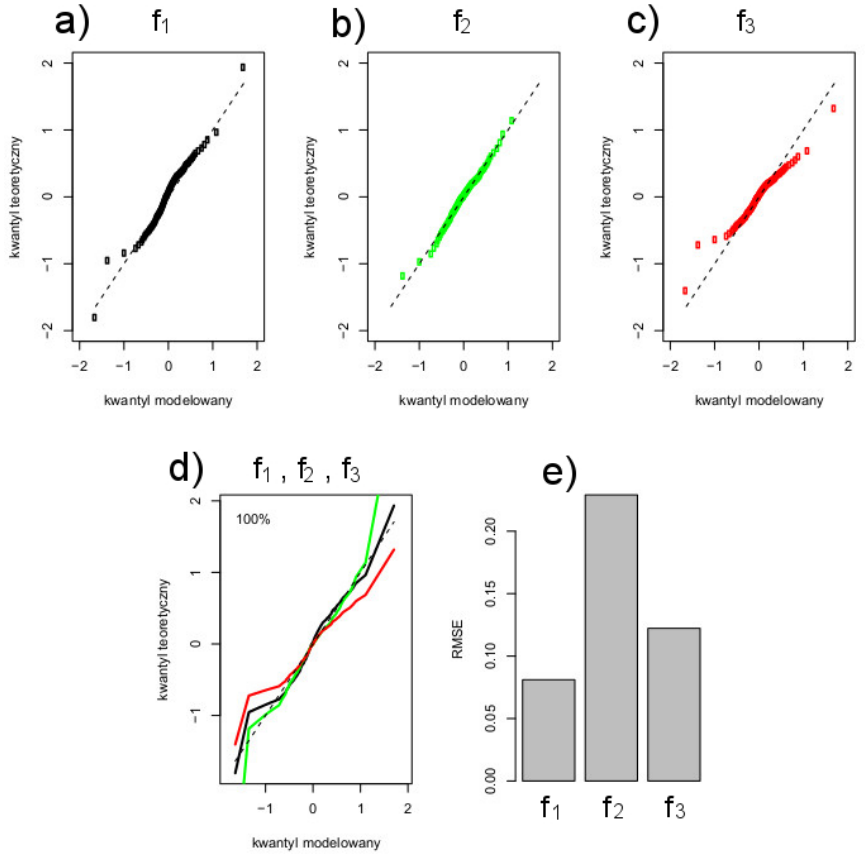
Liczba punktów	613
Liczba wartości odstających	0
Min	-1,638 m
Max	1,706 m
μ	0,004 m
σ	0,415 m
1,64 σ	0,681 m
1,96 σ	0,813 m
RMSE	0,415 m
1,96*RMSE	0,813 m
k	2,425 m
s	-0,065 m
m	-0,022 m
P 0,683	0,342 m
P 0,95	0,881 m
NMAD	0,296 m
b	0,294 m
2^0,5*b	0,416 m

Rysunek 19: Histogram odchyłek powierzchni dla pomiaru walidacyjnego IKONOS wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

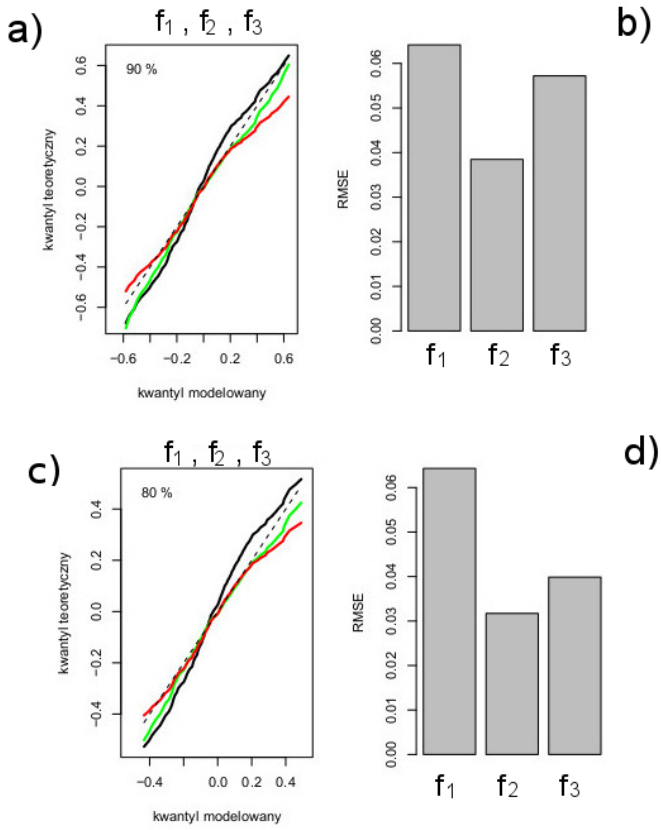
Analiza przebiegu histogramów i wykresów Q-Q pozwala zaobserwować, iż o ile błąd wpasowania funkcji gęstości w histogram empiryczny jest najmniejszy dla rozkładu Laplace'a, o tyle jeśli chodzi o wykresy Q-Q - ma on wartość największą. Nawet drobne odchylenie krzywej Q-Q od linii prostej powoduje relatywnie duży w tym przypadku błąd wpasowania (Rys.20 e). Spada on gwałtownie po usunięciu zaledwie 10% danych (Rys.20 i Rys.21). Ponadto również tutaj, podobnie jak w przypadku walidacji urządzeń GNSS nasuwa się wniosek o nie najlepszym wyborze błędu RMSE jako wskaźnika dopasowania wykresu Q-Q.

Niezależnie od tego, podsumowując można stwierdzić, że na poziomie odchylenia standardowego wszystkie rozkłady dały podobny wynik. Rozszerzenie z rozkładu Gaussa i Laplace'a jest zbliżone, a analogiczne z metody odpornej jest o 30% mniejsze.

W wyniku walidacji omawianego obrazu IKONOS uzyskano wartość bufora równą 0,93 m (Milenov, Gabalova 2009) co powoduje przydzielenie go do klasy 1 m (JRC MARS 2013 b). Powyższa analiza statystyczna potwierdza wartość bufora uzyskaną w wyniku walidacji inną metodą, obliczania dla każdej działki oddzielnie odchylenia standardowego powtarzalności i odtwarzalności, podobnie jak dla walidacji GNSS. Niezalecane jest natomiast w tym przypadku zastosowanie metody odpornej, której wynik powoduje konieczność przyjęcia w trakcie pomiarów kontrolnych wartości 0,75 m.



Rysunek 20: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni w pomiarze walidacyjnym IKONOS - wszystkie odchyłki

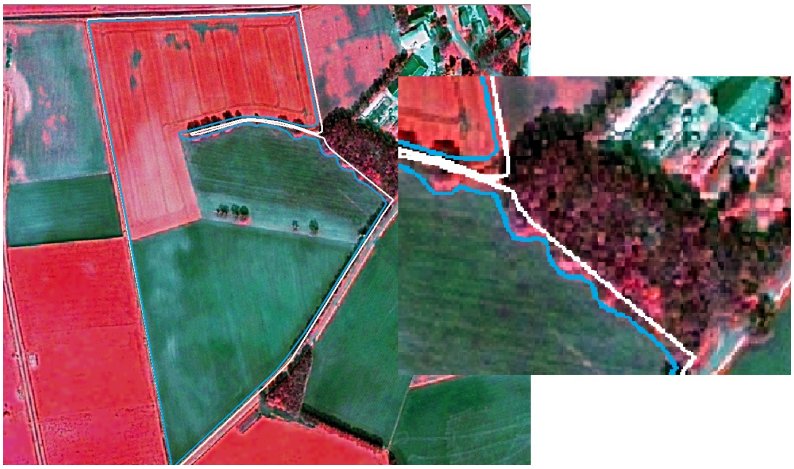


Rysunek 21: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek powierzchni w pomiarze walidacyjnym IKONOS - 90% odchyłek, a), b); 80% odchyłek, c), d)

4.4. Kontrola jakości LPIS 1

Oprócz kontroli powierzchni działek w trybie tak zwanej kontroli na miejscu wykonuje się również kontrolę referencyjnej bazy danych systemu IACS, czyli Systemu Identyfikacji Działek Rolnych LPIS. Procedury kontroli systemu LPIS są stosunkowo niedawno wprowadzone, stąd też JRC wykonało dwa projekty pilotażowe, których wyniki zostały przeanalizowane statystyczne w poniższych podrozdziałach. Projekty nazwano w monografii, ze względu na chronologię ich wykonania: LPIS 1 i LPIS 2. W kontroli systemu LPIS analizowane są odchyłki względne powierzchni działek (inaczej niż w kontroli na miejscu, gdzie analizowane są odchyłki w metrach kwadratowych). Z projektu LPIS 1 dostępne były jedynie odchyłki względne. W dokumentacji z projektu LPIS 2 znajdowały się również odchyłki bezwzględne i powierzchnie działek, co pozwoliło na analizę zarówno odchyłek bezwzględnych i względnych. W kontroli systemu LPIS wykonuje się pomiary kontrolne wykorzystując wektoryzację ortofotomap lub pomiary RTK. Dokładność wyznaczania powierzchni zależy między innymi od jakości ortofotomapy, a głównie od jej rozdzielczości. W ramach kontroli jakości LPIS porównuje się powierzchnię pomierzoną na ortofotomapie i powierzchnię referencyjną znajdującą się w systemie LPIS. W ramach projektu pilotażowego LPIS 1 (Milenov i in. 2009) przeprowadzono wektoryzację ortofotomapy, przykładowa działka znajduje się na Rys.22, gdzie w kolorze białym widać granicę działki w systemie LPIS, a w kolorze szarym wynik wektoryzacji ortofotomapy. Widoczna jest interpretacja granicy w trakcie wektoryzacji, oraz generalizacja w systemie LPIS, co również znajduje odzwierciedlenie w odchyłkach powierzchni.

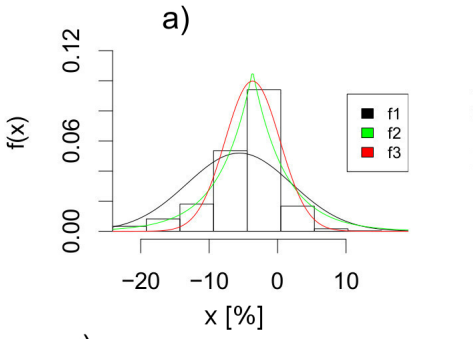
W projekcie zwektoryzowano około 800 działek. Zidentyfikowano w przybliżeniu 50 błędów grubych (brak działek w systemie LPIS).

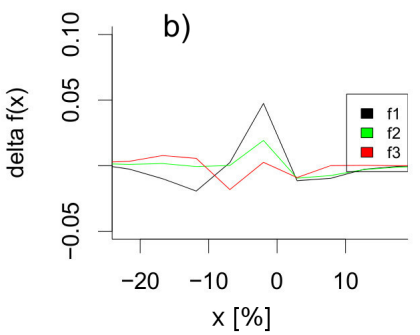


Rysunek 22: Ilustracja przebiegu granicy działki w systemie LPIS (kolor biały) i rzeczywistej granicy na ortofotomapie (kolor szary)

Charakter rozkładu odchyłek odbiega od normalnego (Rys.23). Widoczny jest również znaczący błąd systematyczny: średnia odchyłka wynosi -5,6%, mediana -3,7%. Odchylenie standardowe, percentyl 68,3% oraz odchylenie standardowe w rozkładzie Laplace'a wynosi około 7%, podczas gdy NMAD jest znacząco mniejszy, tylko 4,7% (Rys.23 d). Większa rozbieżność pojawia się na poziomie ufności 95%. Błąd systematyczny powoduje, że nie jest możliwe porównanie wszystkich parametrów rozkładów.

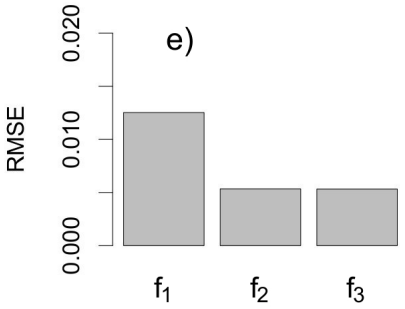
Wartość 1,96 σ (15%) można porównać z percentylem 0,95 (20%) ale te dwie wielkości są również znacząco różne. Odwrotności wartości skumulowanych funkcji rozkładów F_1^{-1}, F_2^{-1} i F_3^{-1} dla argumentu 0,975 wynoszą odpowiednio: 9,5%, 10,4% i 4,2% (Rys.23 c). Podejścia oparte





c)

$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
9,469 %	10,373 %	4,180 %



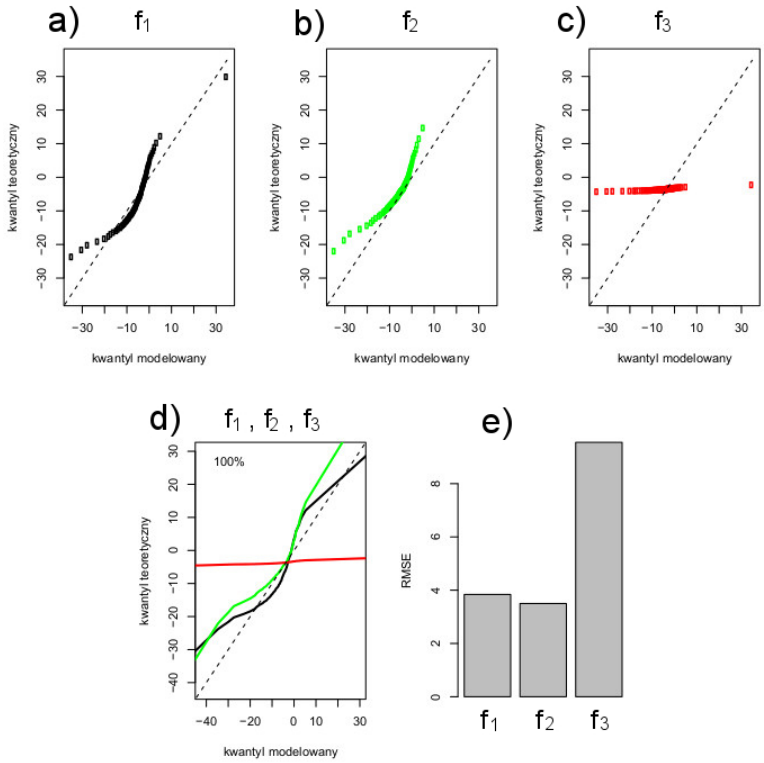
d)

Liczba punktów	745
Liczba wartości odstających	0
Min	-63,384 %
Max	34,850 %
μ	-5,585 %
σ	7,681 %
1,64 σ	12,597 %
1,96 σ	15,055 %
RMSE	9,499 %
1,96*RMSE	18,618 %
k	9,869
s	-2,178
m	-3,656
P 0,683	6,156 %
P 0,95	20,449 %
NMAD	3,998 %
b	4,683 %
2^0,5*b	6,623 %

Rysunek 23: Histogram odchyłek powierzchni dla pomiaru kontrolnego LPIS 1 wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

o rozkład Gaussa i o rozkład Laplace'a są zbliżone, natomiast metoda odporna daje wynik znacząco mniejszy. Porównując błąd (RMSE) wpasowania funkcji gęstości rozkładu do histogramów (Rys. 23 e), można zauważyć największy błąd dla funkcji gęstości rozkładu Gaussa, a dla rozkładu Laplace'a i metody odpornej jest on porównywalny i znacząco mniejszy niż dla rozkładu Gaussa. Jednocześnie błąd dla metody odpor-

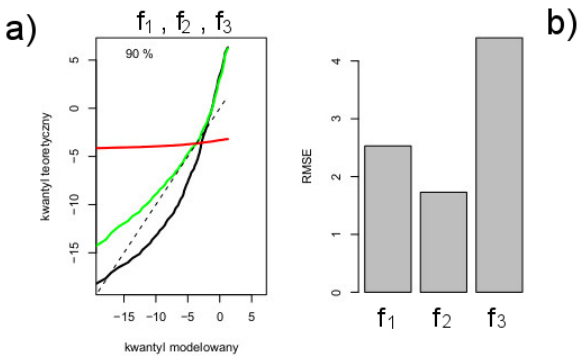
nej w dopasowaniu wykresów Q-Q (Rys. 24 e) wydaje się ją dyskwalifikować.



Rysunek 24: Wykres Q-Q wraz z błędami wpasowania, analiza wszystkich odchyłek powierzchni z pomiaru kontrolnego LPIS 1

W związku ze znaczną odchyłką systematyczną, związaną z inną interpretacją granicy w systemie LPIS i podczas wektoryzacji (przykład na Rys. 22) postanowiono skorygować wyniki o wartość średnią odchyłki. W wyniku tego uzyskano wartości odwrotności dystrybuant równe odpowiednio: 15%, 14% i 8%. Analizując błędy wpasowania do histogramu i wykresy Q-Q dla 100% i 90% odchyłek, wydaje się, że najbardziej wia-

rygodnym rozszerzeniem niepewności tych pomiarów po skorygowaniu o składową systematyczną jest wartość z rozkładu Gaussa i Laplace'a: 15%. Jeśli natomiast chcielibyśmy oszacować niepewność oryginalnych wyników pomiarów, z uwzględnieniem błędu systematycznego to powinniśmy wziąć pod uwagę rozszerzenie RMSE czyli około 19% lub percentyl P(0,95) równy około 20% (Rys.23 d).



Rysunek 25: Wykres Q-Q wraz z błędami wpasowania, analiza 90% odchyłek powierzchni z pomiaru kontrolnego LPIS 1

4.5. Kontrola jakości LPIS 2

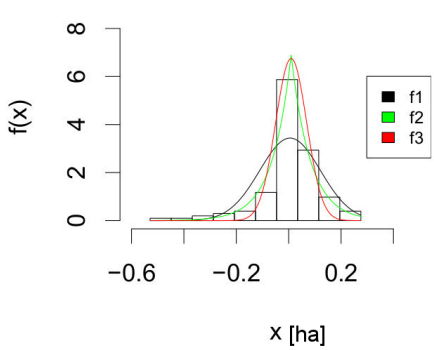
W poniższym przypadku wykorzystano wyniki pomiarów powierzchni działek wykonanych za pomocą RTK i uzyskanych w wyniku wektoryzacji ortofotomapy. W trakcie opracowywania procedur kontrolnych LPIS w JRC, wykonano pomiary na ortofotomapie WV2 i tym razem bezpośrednio w terenie z wykorzystaniem technologii RTK, w ramach projektu pilotażowego przeprowadzonego w Szkocji (Devos i in. 2012). Do analizy, po usunięciu błędów spowodowanych niespójną interpretacją granicy działki rolnej przyjęto, 146 wyników pomiarów, z czego usunięto 4 wyniki nie mieszczące się w przedziale ufności o rozszerzeniu 3σ. Rozkład odchyłek nie ma, również w tym przypadku, charakteru rozkładu Gaussa. Analizując wyniki tego eksperymentu pomiarowego (Rys. 26 i 27) nie zaobserwowano błędu systematycznego: średnia odchyłka wynosi 0,005 ha, mediana 0,001 ha. Odchylenie standardowe w rozkładzie Gaussa przyjmuje wartość 0,116 ha, percentyl 68,3% 0,078 ha, odchylenie standardowe w rozkładzie Laplace'a około 0,102 ha, podczas gdy NMAD jest znacząco mniejszy i równy 0,059 ha. W związku z brakiem błędu systematycznego wartość $1,96\sigma$ (0,227 ha) można porównać z percentylem 0,95 (0,270 ha). Wartości odwrotności dystrybuanty funkcji rozkładów F_1^{-1}, F_2^{-1}, F_3^{-1} wynoszą odpowiednio: 0,232 ha, 0,217 ha, 0,117 ha. Podejścia z zastosowaniem rozkładu Gaussa i Laplace'a dają w tym przypadku podobny wynik, a metoda odporna, oparta o NMAD daje wartość około 2 razy mniejszą. Porównując błąd wpasowania (RMSE) funkcji gęstości rozkładów do histogramu najmniejszy błąd identyfikujemy dla funkcji gęstości rozkładu Laplace'a. Jest on zauważalnie mniejszy niż dla funkcji gęstości

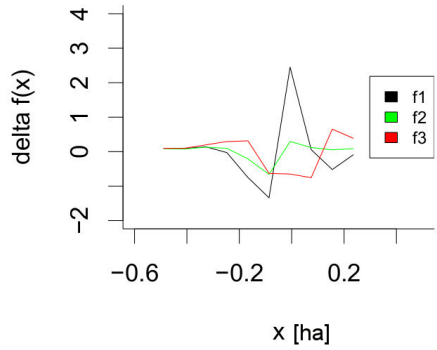
rozkładu Gaussa i metody odpornej. Wynika to ze znacznej rozbieżności funkcji gęstości rozkładu Gaussa i funkcji gęstości rozkładu w metodzie odpornej dla wartości w pobliżu zera. Na poziomie ufności 95% występuje zgodność pomiędzy rozkładami Laplace'a i Gaussa i niepewność można w tym przypadku określić na poziomie: $\pm 0,2$ ha (Rys.27 d).

Inaczej sytuacja wygląda jeśli porównujemy odchyłki względne. W tym przypadku mają one rozkład zdecydowanie inny niż normalny. Analizując rysunki 28 i 29, obserwujemy nieznaczny błąd systematyczny, μ= 1,049%, m=0,316%. Odchylenie standardowe Gaussa, NMAD, percentyl (0,683) i odchylenie standardowe Laplace'a znacznie różnią się od siebie i wynoszą odpowiednio: 8,661%, 1,250%, 1,643% i 3,941%. Natomiast rozszerzenia $1,96\sigma$, F_1^{-1}, F_2^{-1}, F_3^{-1} przyjmują wartość: 16,976%, 18,024%, 8,488% i 2,806%. Również błąd wpasowania funkcji teoretycznych do rozkładu empirycznego jest w tym przypadku bardziej zróżnicowany i wyraźnie najmniejszy dla funkcji Laplace'a. Ten przypadek pokazuje, że wykorzystanie odchylenia standardowego rozkładu normalnego do oszacowania niepewności odchyłek względnych znacznie ją przeszacowuje. Zakładając rozkład Gaussa na poziomie ufności 95 % odchyłki zwierają się w przedziale $\pm$18%. Wykorzystując rozkład Laplace'a, zakres ten wynosi $\pm$8%. Ma to kluczowe znaczenie podczas wykonywania analiz jakości LPIS w kontekście dopuszczalnych limitów (JRC MARS 2013 d), gdzie podawane są wartości względne odchyłek: 3%, 5% i 7%.

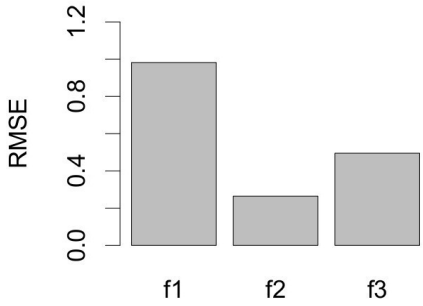
Jeśli dla danego obszaru kontrolnego przeprowadzimy analizę statystyczną odchyłek pomiędzy pomiarem RTK a wartościami z wektoryzacji

ortofotomapy i nie analizując charakteru rozkładu podamy, że na poziomie 95% nasze odchyłki są w zakresie ±18% to, po pierwsze znacznie przekroczymy dopuszczalne odchyłki, a po drugie podamy wartość niepewności tych danych niepotrzebnie zawyżoną około 2 razy.



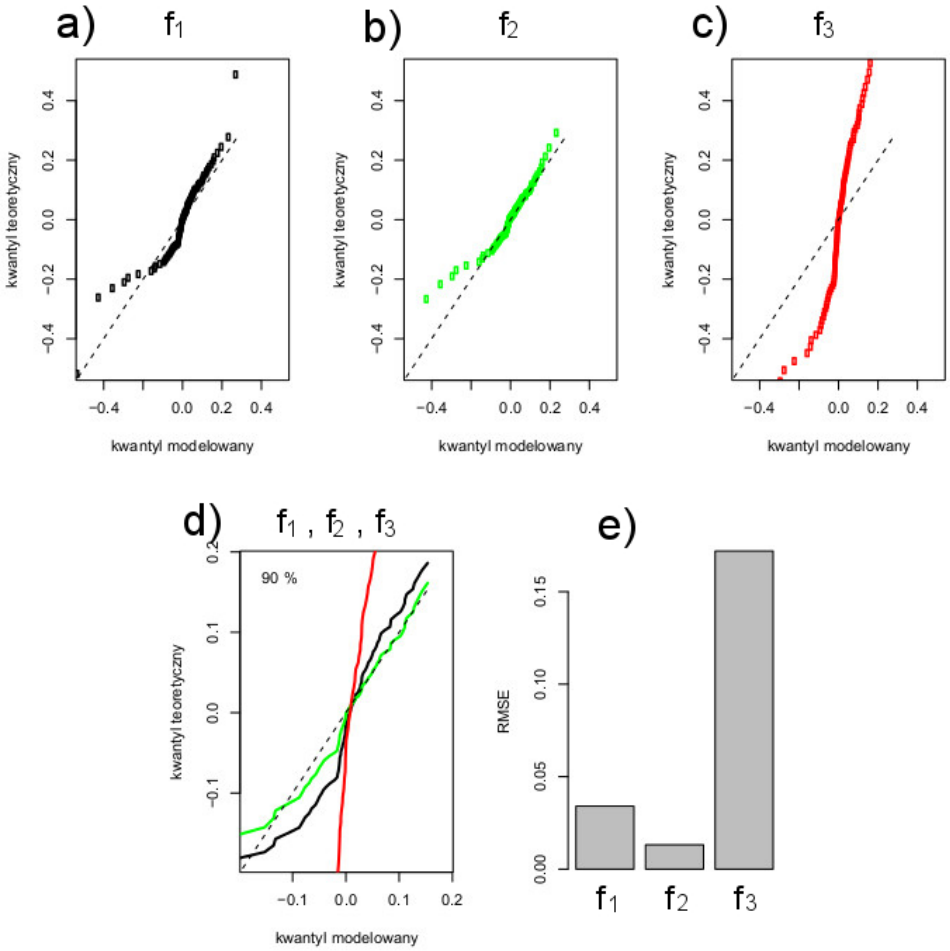


$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
0,232 [ha]	0,217 [ha]	0,117 [ha]

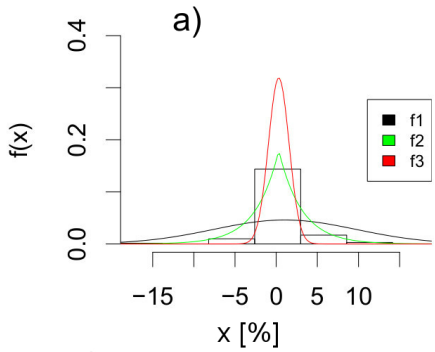


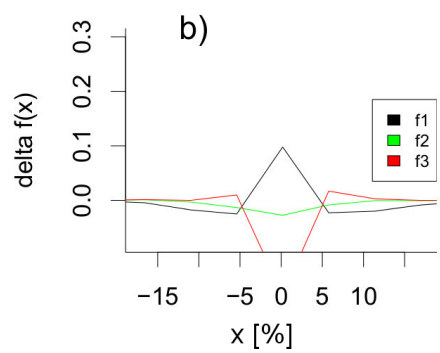
Liczba punktów	127
Liczba wartości odstających	0
Min	-0,529[ha]
Max	0,059[ha]
μ	0,005[ha]
σ	0,116[ha]
1,64 σ	0,190[ha]
1,96 σ	0,227[ha]
RMSE	0,116[ha]
1,96*RMSE	0,227[ha]
k	5,594
s	-1,640
m	0,001[ha]
P 0,683	0,078[ha]
P 0,95	0,270[ha]
NMAD	0,059[ha]
b	0,072[ha]
2^0,5*b	0,102[ha]

Rysunek 26: Histogram odchyłek bezwzględnych powierzchni dla pomiarów kontrolnych z wykorzystaniem RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)



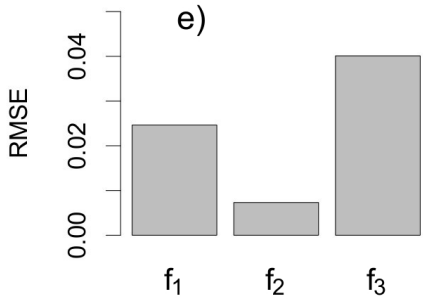
Rysunek 27: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek bezwzględnych [ha] powierzchni z pomiaru kontrolnego RTK





c)

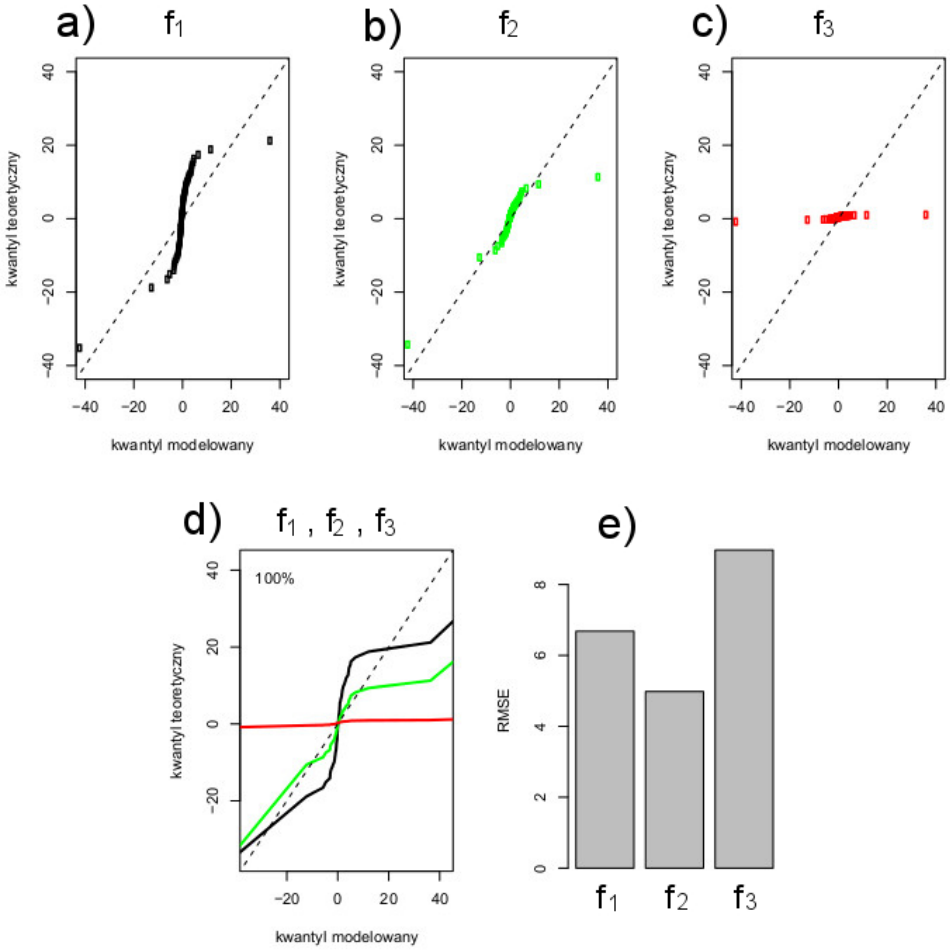
$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
18,024 %	8,488 %	2,806 %



d)

Liczba punktów	127
Liczba wartości odstających	0
Min	-42,740 %
Max	70,033 %
μ	1,049 %
σ	8,661 %
1,64 σ	14,204 %
1,96 σ	16,976 %
RMSE	8,725 %
1,96*RMSE	17,101 %
k	40,778
s	4,097
m	0,316
P 0,683	1,643 %
P 0,95	6,016 %
NMAD	1,250 %
b	2,787 %
2^0,5*b	3,941 %

Rysunek 28: Histogram odchyłek względnych powierzchni dla pomiarów kontrolnych z wykorzystaniem RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)



Rysunek 29: Wykres Q-Q wraz z błędami wpasowania, analiza odchyłek względnych powierzchni z pomiaru kontrolnego RTK

Rozdział 5

Określanie niepewności numerycznych modeli terenu

Numeryczne modele terenu są wykonywane obecnie z wykorzystaniem skaningu laserowego lub fotogrametrii lotniczej. Rozporządzenie ministra spraw wewnętrznych i administracji z 2011 roku w sprawie baz danych dotyczących zobrazowań lotniczych oraz satelitarnych i numerycznego modelu terenu (Rozporządzenie MSWiA 2011), reguluje czynności związane między innymi z tworzeniem NMT oraz kontrolą ich jakości. Pewne informacje na temat dokładności NMT można również znaleźć w wytycznych technicznych wykonywania ortofotomap w skali 1:10000 (GUGIK 2000). Generalnie jakość NMT określa się na podstawie pomiarów kontrolnych, za pomocą błędu średniego (błędu średniego kwadratowego). Ponadto w powyższym rozporządzeniu podaje się procentową liczbę obserwacji, które muszą się znaleźć w odpowiednich zakresach, będących odpowiednimi wielokrotnościami błędu średniego kwadratowego. Nie ma w cytowanych przepisach żadnych zastrzeżeń dotyczących rozkładu błędów. Analizując publikację Amerykańskiego Towarzystwa Fotogrametrii i Teledetekcji (ASPRS Lidar Committee 2004) na temat dokumentowania pionowej dokładności danych lidarowych, można w niej znaleźć zalecenia dotyczące sprawdzania rozkładów odchyłek NMT oraz rozważnego stosowania błędu średniego kwadratowego do określania jego

dokładności. Zagadnienie kształtu rozkładu błędów NMT było również przedmiotem badań naukowych (Aguilar i in. 2010, Atkinson i in. 2005, Darnel i in. 2008, Zandbergen 2008, Höhle, Höhle 2009). W tej ostatniej publikacji zaproponowana została tzw. metoda odporna bazująca na zmodyfikowanym rozkładzie normalnym, gdzie zamiast wartości średniej odchyłki przyjmuje się medianę, a zamiast odchylenia standardowego NMAD, co zostało już omówione w poprzednich rozdziałach.

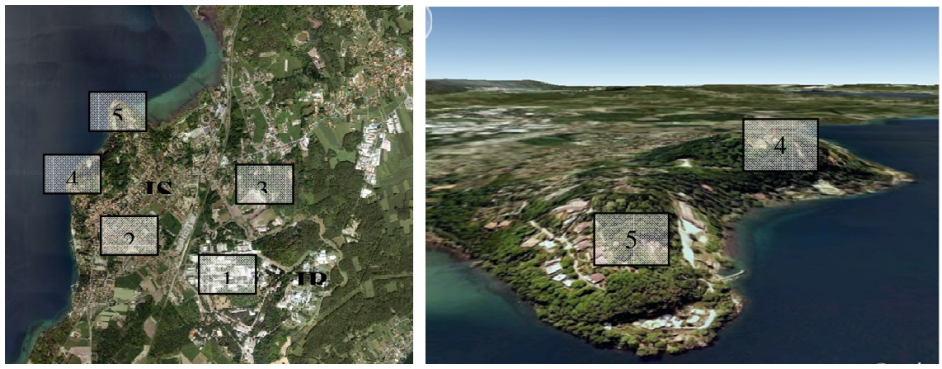
W przypadku NMT mamy do czynienia nie z wynikami pomiarów, ale z wynikiem ich przetworzenia. NMT ma postać dyskretną i w celu uzyskania wysokości w dowolnym punkcie konieczne jest zastosowanie jakiegoś algorytmu interpolacyjnego. Zatem łącznie mamy do czynienia z danymi pomiarowymi i pewnym modelem. Możemy więc mówić o rozbieżności pomiędzy wartościami modelowanymi (NMT) a wartościami zmierzonymi, co w przypadku klasycznej interpolacji można było określić za pomocą dokładności (błędu) interpolacji. W przypadku NMT zwykle nie podaje się tego typu informacji. Trudno powiedzieć właściwie z jakiego powodu. Można również obliczyć rozbieżność pomiędzy modelem, a położeniem punktu wyznaczonym w oparciu o bardziej dokładne metody. Sytuacja ta ma miejsce podczas kontroli NMT, gdzie mierzy się wysokości w liniach profilowych lub płaszczyznach kontrolnych. Warto również w tym przypadku zwrócić uwagę na fakt, że trudno jest wykonać pomiar kontrolny położenia konkretnego punktu pomiarowego ALS. Właściwie jest to niemożliwe. Stąd też koncepcja pomiaru płaszczyzn kontrolnych w wyżej wymienionym rozporządzeniu (Rozporządzenie MSWiA 2011).

W monografii zaproponowano bardziej kompleksową analizę dokładności niż to jest wymagane w rozporządzeniu, gdzie porównuje się model NMT z wynikiem pomiaru bezpośredniego, czyli tylko dokładność zewnętrzną. Zaproponowano analizę odchyłek pomiędzy pomiarem ALS (chmurą punktów) i NMT, co pozwala na określenie tzw. dokładności wewnętrznej, czy względnej albo inaczej mówiąc: dokładności interpolacji. Porównanie z NMT z wynikiem dokładniejszych pomiarów *in-situ*, pozwala natomiast na określenie dokładności bezwzględnej. W przypadku pomiarów ALS wykonywanych obecnie, czyli coraz bardziej dokładnych, problematyczny staje się pomiar bezpośredni. Pojawia się pytanie jaką wysokość porównujemy. Po pierwsze, musimy wiedzieć jaka jest powierzchnia odniesienia NMT i najczęściej musimy skorygować naziemne pomiary GPS w oparciu o repery. Po drugie, w analizie dokładności musimy uwzględnić typ pokrycia terenu. Niepewność wysokościowa dla obiektów antropogenicznych będzie z pewnością inna niż dla obiektów naturalnych, dla których wystąpi mniejsze lub większe rozmycie powierzchni terenu, powodując trudności w interpretacji, pomiarze i porównaniach. W następnych podrozdziałach zaprezentowano wyniki badań jakości NMT uzyskanych technologią ALS. Jeden przykład dotyczy analiz przeprowadzonych w 2005 roku w ramach projektu (Hejmanowska i in. 2008 a), które zostały opublikowane w pozycji (Hejmanowska, Kay 2011). Drugi przykład dotyczy aktualnie przeprowadzonych badań nad modelem NMT z projektów: LPIS oraz ISOK. Metodyka badań różni się nieco dla obu tych przypadków, a szczegółowe informacje można znaleźć w następnych podrozdziałach.

5.1. Analiza jakości NMT ALS

W niniejszym podrozdziale zamieszczone zostały wyniki wstępnych badań na temat wpływu rozkładu błędów na wartość estymowanych parametrów dokładności NMT. Chronologicznie rzecz biorąc, prezentowane tutaj wyniki pochodzą z pierwszych badań na ten temat (Hejmanowska i in. 2008 a), które jednak zostały zaprezentowane nieco później (Hejmanowska, Kay 2011). Obszarem testowym była wschodnia część jeziora Lago Maggiore w pobliżu ośrodka badań naukowych Unii Europejskiej, Joint Research Centre w Isprze, we Włoszech. Dane z lotniczego skaningu laserowego zostały pozyskane w 2005 r. z rozdzielczością 1 punkt na 1 metr kwadratowy. Wykonano również w 2008 r., pomiar bezpośredni RTK w wybranych punktach. Do szczegółowych analiz wybrano 5 obszarów testowych, charakteryzujących się zróżnicowanym typem powierzchni. Dwa obszary znajdowały się w terenie płaskim, jeden w terenie o nieco bardziej urozmaiconej rzeźbie i zabudowanym oraz dwa w terenie o znacznych nachyleniach, w pobliżu jeziora (Rys.30). Wysokość terenu zmieniała się od 240 do 350 metrów nad poziom morza, a w strefie o najbardziej zróżnicowanej rzeźbie, na ponad połowie powierzchni występowały nachylenia powyżej 45 stopni.

W pierwszej kolejności porównano NMT z chmurą punktów, biorąc pod uwagę NMT w siatce 1 x 1 m, a w drugiej kolejności w siatce 0.1 x 0.1 m. Analizy wykonano dla dwóch, znacząco różniących się wielkością oczka siatki, żeby zbadać czy ma to jakiś wpływ na wyniki analizy. Obliczenia wykonano z wykorzystaniem algorytmów rastrowych, czyli



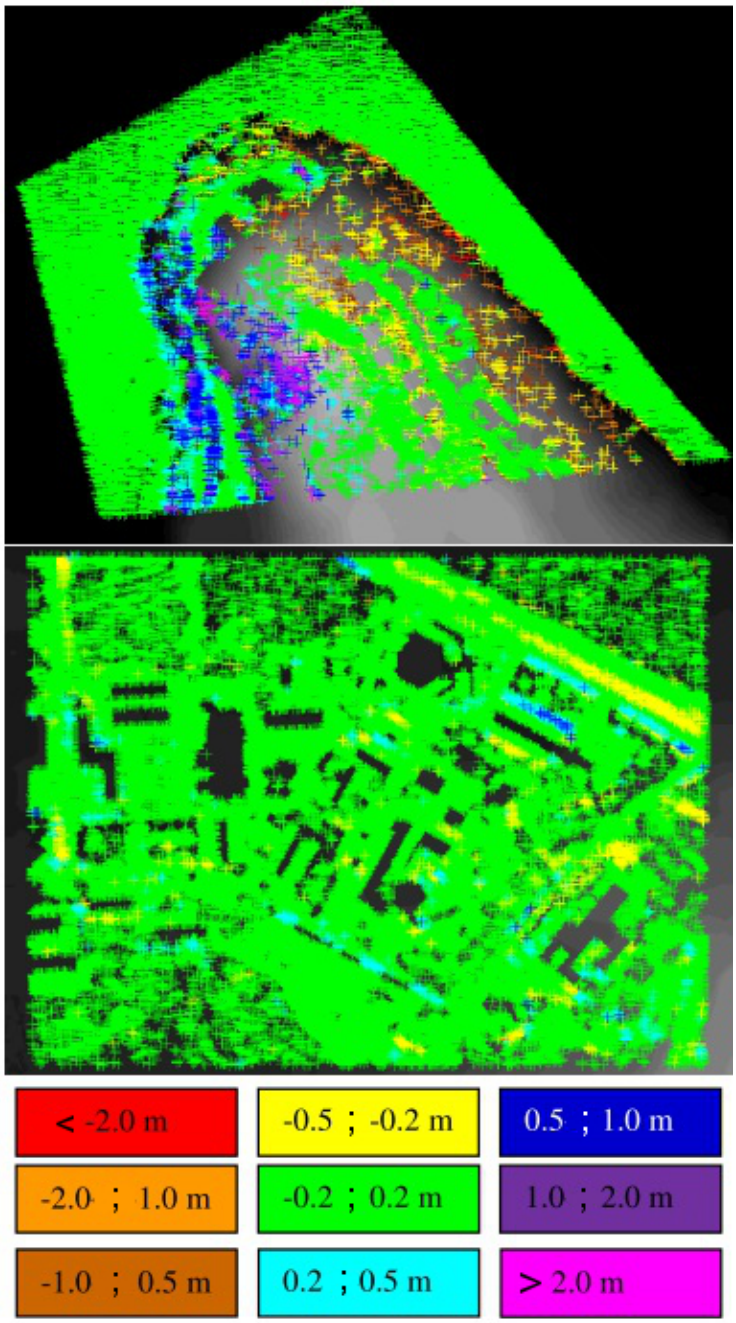
Rysunek 30: Lokalizacja obszarów testowych JRC, Ispra, Włochy,(Hejmanowska, Kay 2011)

przecinano NMT GRID wektorem zawierającym punkty pomiarowe.

Tabela 4: Parametry opisujące rozbieżność pomiędzy NMT a chmurą punktów [m] (Hejmanowska, Kay 2011)

Obszar testowy	1	2	3	4	5
liczba	94323	80227	69231	36369	26424
$Min\ \Delta H$	-1,05	-2,08	-3,63	-3,93	-7,81
$Max\ \Delta H$	1,41	2,93	2,5	9,86	10,64
μ	-0,007	0,001	-0,007	-0,017	-0,049
σ	0,073	0,083	0,129	0,317	0,271
m	0,0	0,0	0,0	0,0	-0,010
NMAD	0,044	0,030	0,044	0,044	0,059
k	17,14	106,65	59,54	87,67	228,25
s	0,192	1,581	-1,875	2,321	-1,476

Rozkład przestrzenny różnic pomiędzy modelem NMT i pomiarem ALS, czyli dokładności wewnętrznej, czy inaczej mówiąc względnej, zamieszczono na rysunku (Rys. 31). Punkty przedstawione w kolorach: czerwonym, pomarańczowym, brązowym i żółtym, zgodnie z pomiarem ALS, leżą powyżej NMT, a w kolorach: niebieskozielony, ciemno niebieski, fiole-

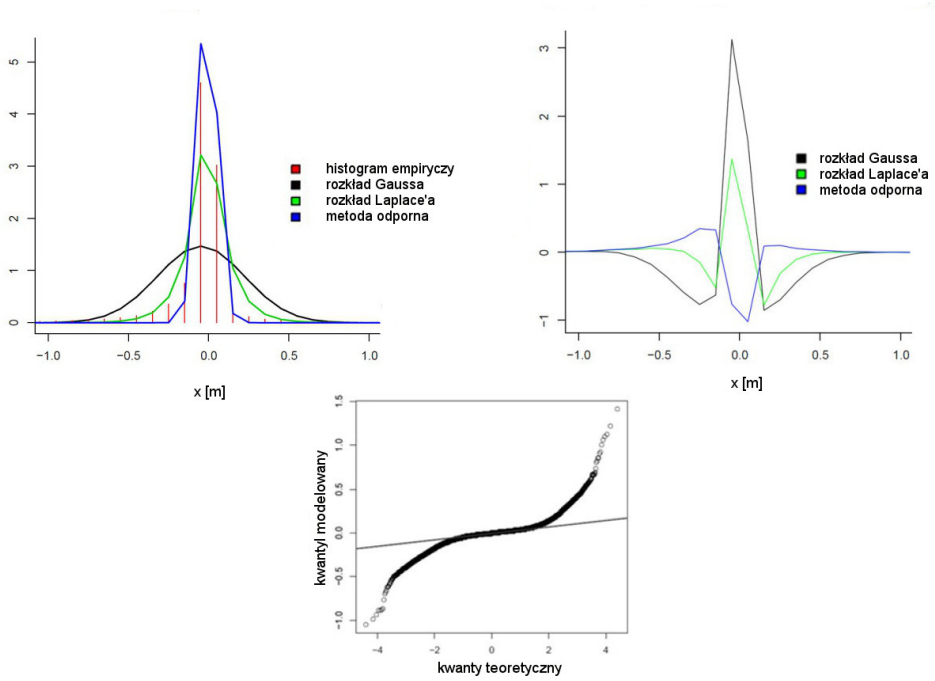


Rysunek 31: Rozkład przestrzenny rozbieżności pomiędzy NMT i chmurą punktów. Ilustracja u góry przedstawia obszar nr 5, o urozmaiconej rzeźbie, a ilustracja dolna obszar nr 2, płaski (Hejmanowska, Kay 2011)

towy i magenta poniżej NMT. Punkty w kolorze zielonym leżą mniej więcej na powierzchni NMT (różnice mieszczą się w zakresie od -0,2 m do 0,2 m). Rysunek 31 nawet bez dalszej analizy statystycznej, ilustruje pewną prawidłowość: w przypadku terenu płaskiego rozbieżności są jednorodne przestrzennie i w większości przypadków niewielkie. Jedynie w niektórych przypadkach pojawiają się błędy większe niż 0,20 m. Inaczej w terenie o urozmaiconej rzeźbie, gdzie widoczne jest przechylenie modelu: z jednej strony punkty ALS leżą powyżej, a z drugiej poniżej NMT. Wydaje się, że jest to argument za tym, żeby podczas tworzenia NMT wziąć pod uwagę takie mapy rozbieżności oraz zamieścić informacje o dokładności jako metadane charakteryzujące NMT.

Innym zagadnieniem niż analiza rozkładu przestrzennego odchyłek jest analiza histogramu, czyli częstości ich występowania oraz charakteru ich rozkładu. Przykładowy wykres Q-Q oraz histogram empiryczny wraz z trzema analizowanymi funkcjami gęstości prawdopodobieństwa znajdują się na (Rys. 32 a i c). Widoczny jest silny nie Gaussowski charakter rozkładu odchyłek. Rozbieżności pomiędzy histogramem empirycznym a przebiegiem funkcji teoretycznych są przedstawione na rysunku (Rys.32 b).

Krzywe funkcji gęstości z metody odpornej i opartej o rozkład Laplace'a pasują do obserwacji znacznie lepiej, o czym również świadczą niższe wartości RMSE (Tab.5). Jednocześnie nie obserwuje się, w tym przypadku, mniejszych wartości błędu RMSE dla rozkładu Laplace'a w porównaniu z metodą odporną. Natomiast, tak jak we wszystkich



Rysunek 32: Histogram odchyłek - obserwowany, funkcja gęstości prawdopodobieństwa Gaussa, funkcji odpornej i funkcji Laplace'a, wykres rozbieżności,wykres Q-Q dla obszaru numer 5 (Hejmanowska, Kay 2011)

innych przypadkach, daje się zaobserwować charakter trzech analizowanych funkcji. Mianowicie funkcja gęstości rozkładu Gaussa przeszacowuje błąd, podczas gdy metoda odporna znacznie go niedoszacowuje, a funkcja Laplace'a znajduje się pomiędzy nimi. Ma to oczywiście wpływ na parametry dokładności (Tab.6).

Analizując tę tabelę zestawieniu z tabelą (Tab.4) należy wziąć pod uwagę błąd systematyczny, szczególnie widoczny w obszarze 5 (-0,049 m). Niezależnie jednak od tego, stosując tradycyjną metodę obliczeniową uzyskamy w terenie płaskim około 0,15 m, a w terenie o urozmaiconej

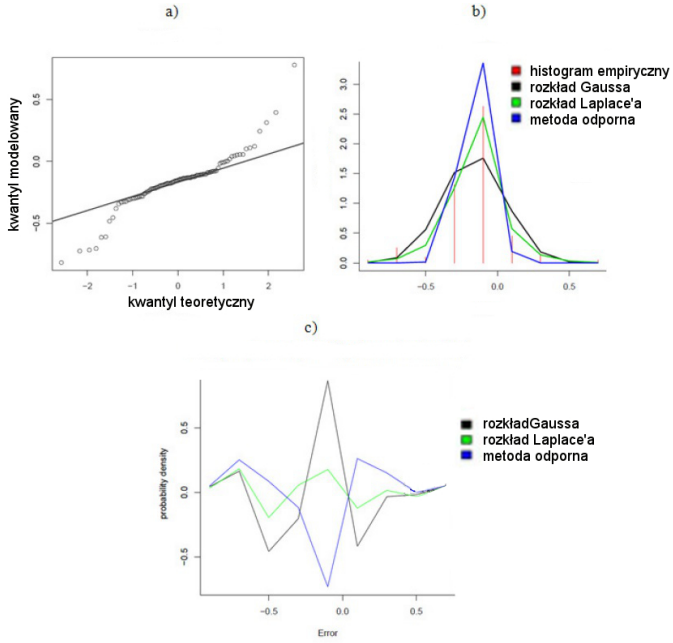
rzeźbie około 0,50 m. Stosując metodę odporną powiemy, że błąd wynosi we wszystkich przypadkach około 0,10 m (Tab.6).

Tabela 5: Błąd średni kwadratowy określający wpasowanie rozkładu Gaussa, rozkładu Laplace'a i funkcji odpornej do histogramu odchyłek [m] (Hejmanowska, Kay 2011)

	1	2	3	4	5
$RMSE_{f1}$	0,756	0,719	0,624	0,362	0,289
$RMSE_{f2}$	0,332	0,249	0,272	0,202	0,128
$RMSE_{f3}$	0,426	0,257	0,217	0,146	0,103

Natomiast wykorzystując rozkład Laplace'a uzyskuje się niepewność na poziomie od 0,15 m do 0,35 m (Tab.6). Powstaje pytanie, którą z tych wartości przyjąć jako najbardziej adekwatną do rzeczywistości. Wydaje się, że nie można na to pytanie odpowiedzieć jednoznacznie bez każdorazowej analizy histogramu odchyłek.

W dalszej kolejności przeanalizowano odchyłki pomiędzy NMT a wysokościami pomierzonymi RTK. Wyniki przedstawiono, analogicznie jak dla dokładności wewnętrznej, na rysunku (Rys.33). Wykres Q-Q świadczy o innym niż normalny charakterze rozkładu odchyłek. W tym przypadku stopień dopasowania rozkładu Laplace'a jest najlepszy: błędy dopasowania wynoszą odpowiednio: 0,389 m, 0,125 m i 0,289 m. Ponadto zaobserwowany został błąd systematyczny około - 0,17 m, często obecny przy pomiarach lidarowych. Oznacza to, że pomiary RTK są poniżej NMT. Dystrybuanty wynoszą odpowiednio: F_1^{-1}, F_2^{-1}, F_3^{-1} (0,975) 0,259 m, 0,305 m, 0,049 m, dla porównania percentyl P(0,95) wynosi 0,705 m, a 1,96 RMSE równa się 0,535 m (Hejmanowska, Kay 2011). W tym przy-



Rysunek 33: Dokładność zewnętrzna/bezwzględna NMT ALS, wykres Q-Q, histogram rozbieżności obserwowanych, modelowane funkcje gęstości prawdopodobieństwa (Hejmanowska, Kay 2011)

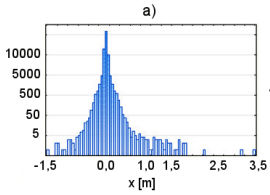
padku czynnikiem zaburzającym jest błąd systematyczny, który świadczy albo o błędnie wykonanym pomiarze RTK lub o nieprawidłowościach podczas tworzenia NMT.

Tabela 6: Odwrotność dystrybuanty jako wskaźnik dokładności względnej, [m] (Hejmanowska, Kay 2011)

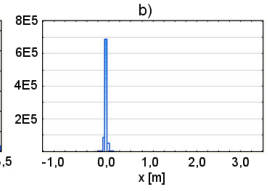
	1	2	3	4	5
F_1^{-1} $(0,975)$	0,136	0,164	0,245	0,605	0,480
F_2^{-1} $(0,975)$	0,132	0,120	0,189	0,380	0,310
F_3^{-1} $(0,975)$	0,086	0,084	0,086	0,086	0,106
P $(0,95)$	0,150	0,130	0,200	0,630	0,180

W celu weryfikacji przeszacowywania niepewności w rozkładzie Gaussa i niedoszacowywania jej w rozkładzie z metody odpornej przyprowadzono analizy na innym zestawie danych. Były to pomiary ALS zrealizowane w 2006 roku w Krakowie (Jędrychowski 2007). Skaning był wykonywany z różną gęstością od 12 do 20 punktów na metr kwadratowy. Chmury punktów były klasyfikowane w celu wyseparowania powierzchni terenu, a ostatecznie generowano model NMT w postaci *gridowej*. Testowano różne algorytmy tworzenia modelu. Oceniając wynikowy model NMT można go porównać z chmurą punktów (dokładność wewnętrzna), ale można również analizować odchyłki pomiędzy modelami uzyskanymi za pomocą różnych algorytmów. Przykładowe wyniki porównania NMT z chmurą punktów znajdują się na (Rys. 34).

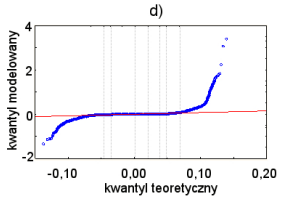
a)



b)



d)



e)

$F_1^{-1}(0.95)$	$F_2^{-1}(0.95)$	$F_3^{-1}(0.95)$
0,058 m	0,035 m	0,029 m

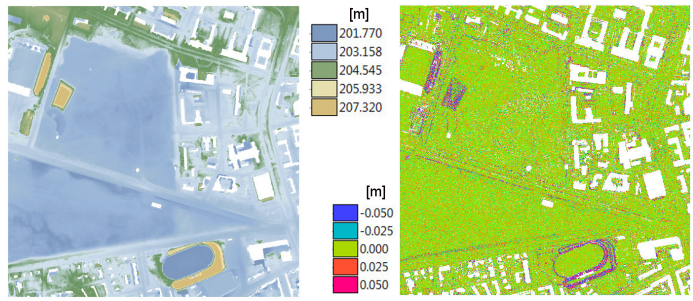
c)

Liczba punktów	839066
Liczba wartości odstających	0
Min	-1,350 m
Max	3,400 m
μ	0,001 m
σ	0,029 m
1,64 σ	0,048 m
1,96 σ	0,057 m
RMSE	0,029 m
1,96*RMSE	0,057 m
k	1115,439
s	18,451
m	0,000 m
P 0,683	0,010 m
P 0,95	0,040 m
NMAD	0,015 m
b	0,012 m
2^0,5*b	0,017 m

Rysunek 34: Analiza dokładności wewnętrznej NMT ALS okolic Błoń (Kraków), histogram empiryczny w skali logarytmicznej i liniowej, wykres Q-Q, parametry statystyczne

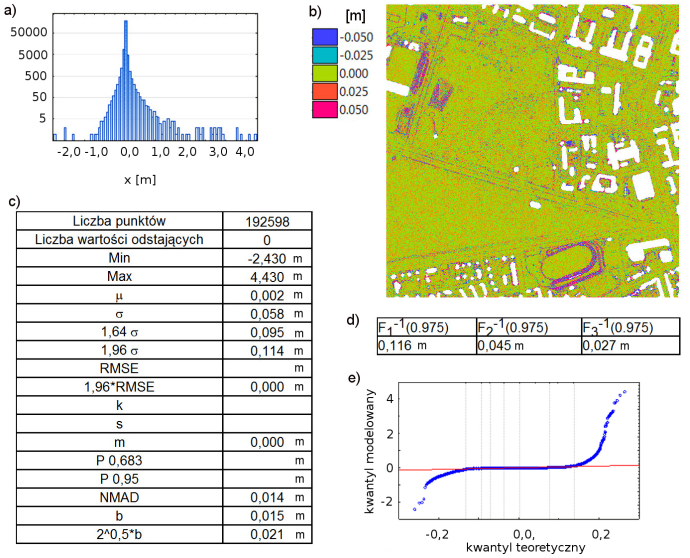
Na rysunku przedstawiono histogram empirycznyy w skali logarytmicznej i liniowej, wykres Q-Q oraz parametry statystyczne. Otrzymanego rozkładu absolutnie nie można uznać za zgodny z rozkładem normalnym. Oszacowanie niepewności interpolacji NMT za pomocą rozkładu Gaussa dostarcza wyniku około 5,8 cm, za pomocą rozkładu Laplace'a 3,5 cm, a metodą odporną 2,9 cm.

Na rysunku 35 można zobaczyć NMT okolic Błoń oraz mapę odchyłek pomiedzy NMT, a chmurą punktów. W przeciwieństwie do poprzedniego przykładu analizowany obecnie obszar jest prawie płaski, o deniwelacji maksymalnie 5 m. W przeważającej większości tego obszaru, odchyłki są skupione wokół zera, co widać również na histogramie (Rys.34). "Długie ogony" znajdują się w pewnych szczególnych miejscach, wokół stadionów (Cracowii i Wisły), górki w Parku Jordana, szpalerów drzew wokół Błoń oraz w bezpośredniej bliskości budynków. Wydaje się, że w tym przypadku rozkład Gaussa przeszacowuje niepewność NMT.



Rysunek 35: NMT ALS okolic Błoń (Kraków), mapa różnicowa NMT ALS minus chmura punktów

W przetwarzaniu danych przestrzennych bardzo często tworzy się mapy różnicowe, które następnie poddaje się analizie statystycznej. Przy-



Liczba punktów	192598
Liczba wartości odstających	0
Min	-2,430 m
Max	4,430 m
μ	0,002 m
σ	0,058 m
1,64 σ	0,095 m
1,96 σ	0,114 m
RMSE	m
1,96*RMSE	0,000 m
k	
s	
m	0,000 m
P 0,683	m
P 0,95	m
NMAD	0,014 m
b	0,015 m
2^0,5*b	0,021 m

$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
0,116 m	0,045 m	0,027 m

Rysunek 36: Analiza parametrów statystycznych mapy różnicowej uzyskanej w wyniku odjęcia modeli NMT wykonanych dwoma różnymi algorytmami, histogram empiryczny w skali logarytmicznej, wykres Q-Q, parametry statystyczne

kład takiej mapy znajduje się na (Rys. 36). Pokazano na niej różnice pomiędzy NMT utworzonym w postaci *grid* metodą interpolacji i NMT utworzonym początkowo jako *tin*, a potem zamienionym na postać *gridową*. Histogram różnic jest również bardzo odbiegający od rozkładu normalnego, a parametry oszacowania niepewności za pomocą odwrotności dystrybuant wynoszą: 0,116 m, 0,045 m i 0,027 m. Również w tym przypadku funkcja rozkładu Gaussa podaje największą wartość niepewności. Porównując rysunki 35 i 36 zauważa się te same newralgiczne obszary, dla których występują największe odchyłki.

Podsumowując omówione dwa przypadki analiz niepewności modeli

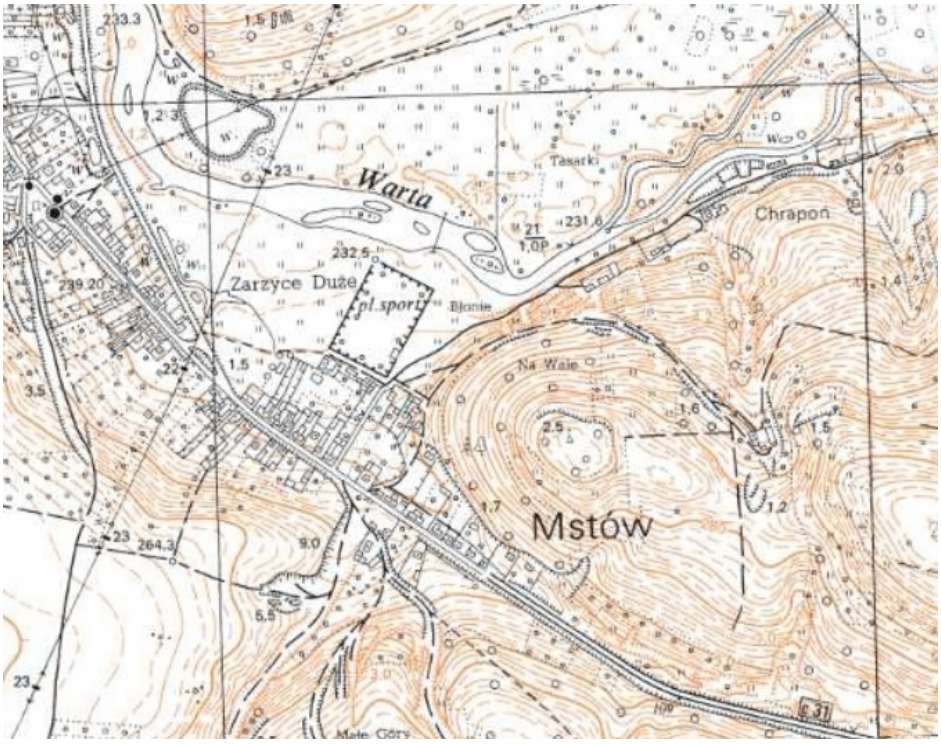
NMT można stwierdzić, że:

- rozkład odchyłek pomiędzy NMT, a chmurą punktów nie miał charakteru rozkładu normalnego,
- funkcja gęstości rozkładu Gaussa przeszacowywała niepewność wewnętrzną NMT, metoda odporna ją niedoszacowywała, a rozkła Laplace'a znajdował się pomiędzy nimi,
- określenie jakości zewnętrznej NMT było utrudnione ze względu na systematyczne przesunięcie modelu i pomiarów RTK,
- odchyłki z map różnicowych uzyskanych w wyniku porównania NMT utworzonego z wykorzystaniem różnych algorytmów, również nie miały rozkładu normalnego; charakter wartości niepewności uzyskanych za pomocą badanych funkcji gęstości prawdopodobieństwa był taki sam jak w przypadku porównania z chmurą punktów: z rozkładu Gaussa uzyskano 11,6 cm, z rozkładu Laplace'a 4,5 cm a z metody odpornej 2,7 cm,
- wyniki uzyskane z siatki 0,1 x 0,1 m nie różniły się od wyników z siatki 1 m x 1 m.

5.2. Analiza jakości NMT LPIS

Niniejszy podrozdział zawiera wyniki badań nad przykładowym NMT z tzw. projektu LPIS, w ramach którego wykonano ortofotomapę dla całego kraju, a NMT był konieczny jej wykonania. Do testowania wybrano jeden arkusz 1:10 000 w okolicy Mstowa (województwo śląskie, nieda-

leko Częstochowy). Model NMT obejmuje obszar pomiędzy 230 a 300 m n.p.m. W części północnej znajduje się Warta i teren jest mało urozmaicony, natomiast w części południowo-zachodniej znajduje się porośnięte lasem wzgórze (Rys.37 i Rys.38).



Rysunek 37: Lokalizacja obszaru testowego Mstów, fragment mapy topograficznej (geoportal.gov.pl)

Podobnie jak w poprzednim podrozdziale przeprowadzono analizę dokładności wewnętrznej i zewnętrznej modelu NMT LIPS. Model wykonywany był metodą fotogrametryczną, z ośrodka GODGIK pozyskano dane źródłowe do modelu i sam model w postaci *gridowej*. Porównanie wysokości NMT w punktach pomiarowych przeprowadzono w oparciu

Rysunek 38: Lokalizacja obszaru testowego Mstów, fragment ortofotomapy (geoportal.gov.pl)

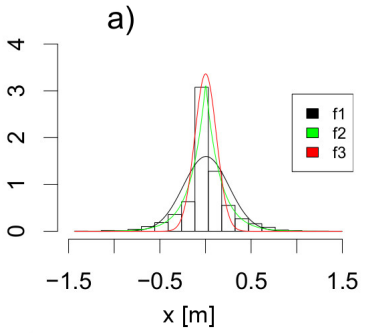
o interpolację biliniową wykorzystując najbliższe punkty siatki NMT. Odchyłki przeanalizowano zgodnie z opisaną wcześniej metodyką.

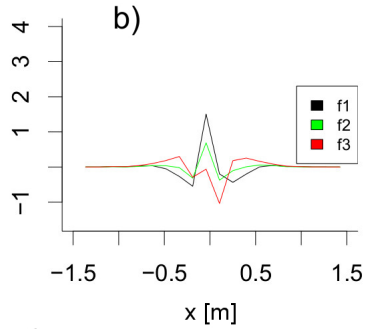
Rozkład odchyłek pomiędzy punktami pomiarowymi a NMT odbiega od rozkładu normalnego, co widać na histogramie i wykresie Q-Q (Rys.39 i Rys.40). Widoczna jest ta sama tendencja przeszacowywania niepewności z wykorzystaniem rozkładu Gaussa, niedoszacowanie metodą odporną oraz wyniki pośrednie z rozkładu Laplace'a. W związku z brakiem systematycznego przesunięcia pomiędzy NMT a chmurą punktów wartości 1,96σ, 1,96 RMSE i odwrotność dystrybuanty rozkładu

Gaussa są sobie bliskie. Analiza wizualna pozwala stwierdzić najlepsze dopasowanie funkcji gęstości rozkładu Laplace'a do histogramu w całym przebiegu. Dopasowanie funkcji gęstości rozkładu Gaussa wizualnie jest najgorsze. Potwierdzają oba błędy RMSE. Natomiast funkcja odporna jest dobrze wpasowana w okolicy zera, po czym jej ramiona gwałtownie spadają, przecinają histogram empiryczny, co skutkuje zaniżeniem błędu. W tym przypadku, w związku z brakiem przesunięcia systematycznego wszystkie parametry rozszerzonej dokładności wewnętrznej są bliskie 50 cm. Jedynie metoda odporna niedoszacowuje błędu o około 50% w stosunku do pozostałych (F_3^{-1} (0,975)=0,235 m).

W dalszej kolejności przeprowadzono analizę dokładności bezwzględniej NMT LPIS. Wykorzystano do tego wyniki pomiarów RTK, które zostały wykonane w siatce o wymiarach około 60 m na 60 m, z wykorzystaniem kalibracji na repery wysokościowe. Szczegółowe informacje na ten temat można znaleźć w pracy (Kulesza 2013). Zaobserwowano tam przesunięcie NMT w górę w stosunku do pomiaru RTK ($\mu = -0,254\ m$ i $m = -0,242\ m$, co czytamy na Rys.41). Wielkości odwrotności dystrybuanty wynoszą dla rozkładu Gaussa 0,657 m dla rozkładu Laplace'a 0,591 m a dla metody odpornej 0,303 m. Dopasowanie funkcji gęstości i wykresu Q-Q jest najlepsze dla rozkładu Laplace'a (Rys.41 i Rys.42).

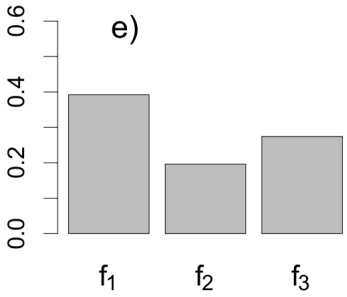
W przypadku występowania błędu systematycznego wydaje się, że RMSE i percentyl (0,95) są najbardziej wiarygodne, chociaż z pewnością przeszacowują niepewność. Powinniśmy zatem unikać takiej sytuacji, w której występuje błąd systematyczny albo wyeliminować go z pomia-





c)

$F_1^{-1}(0.95)$	$F_2^{-1}(0.95)$	$F_3^{-1}(0.95)$
0,500 m	0,475 m	0,235 m

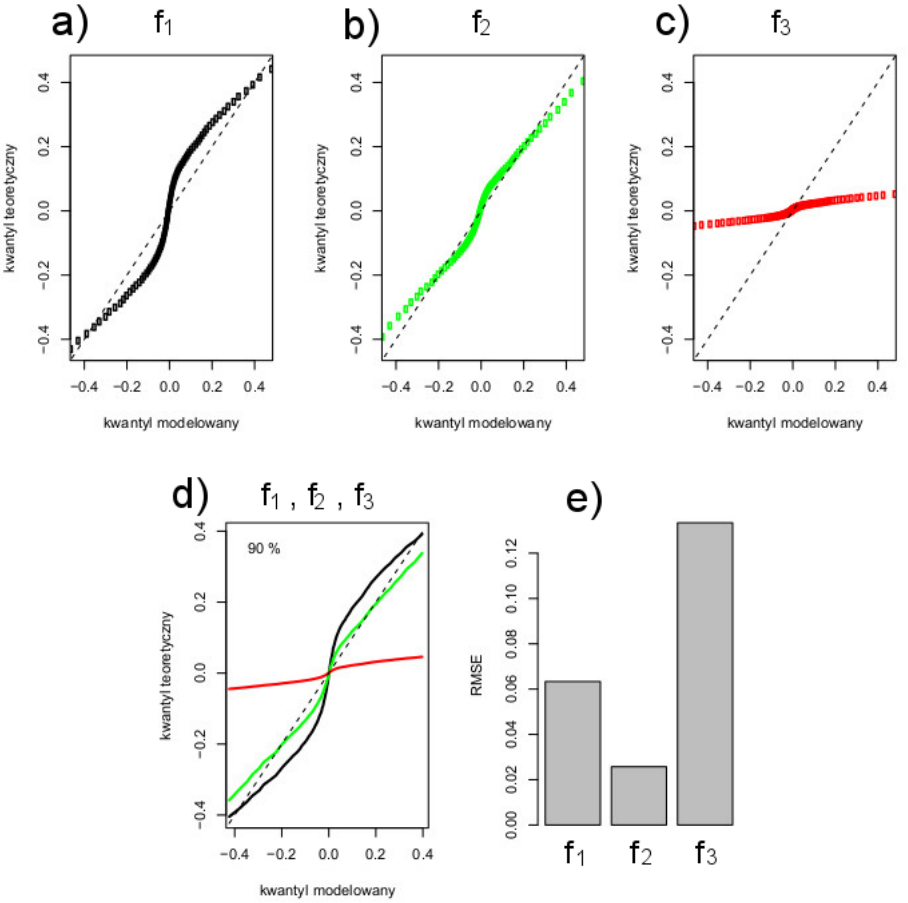


d)

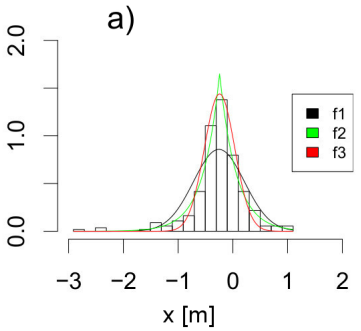
Liczba punktów	4515
Liczba wartości odstających	0
Min	1,498 m
Max	-1,426 m
μ	0,006 m
σ	0,250 m
1,64 σ	0,410 m
1,96 σ	0,490 m
RMSE	0,250 m
1,96*RMSE	0,490 m
k	4,778
s	1,968
m	0,002
P 0,683	0,176 m
P 0,95	0,572 m
NMAD	0,119 m
b	0,158 m
2^0,5*b	0,223 m

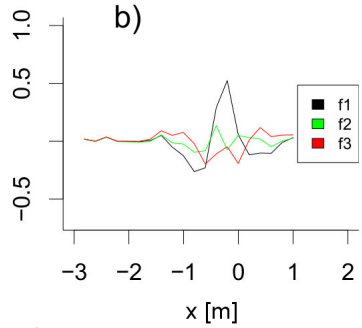
Rysunek 39: Histogram różnic wysokości pomiędzy NMT LPIS i chmurą punktów wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

rów. Rozkład przestrzenny odchyłek względnych i bezwzględnych można przeanalizować na rysunku 43. Tego rodzaju mapy zwierają więcej informacji niż pojedyncze wskaźniki ilościowe niepewności.



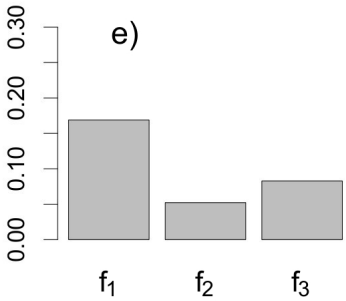
Rysunek 40: Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i chmurą punktów





c)

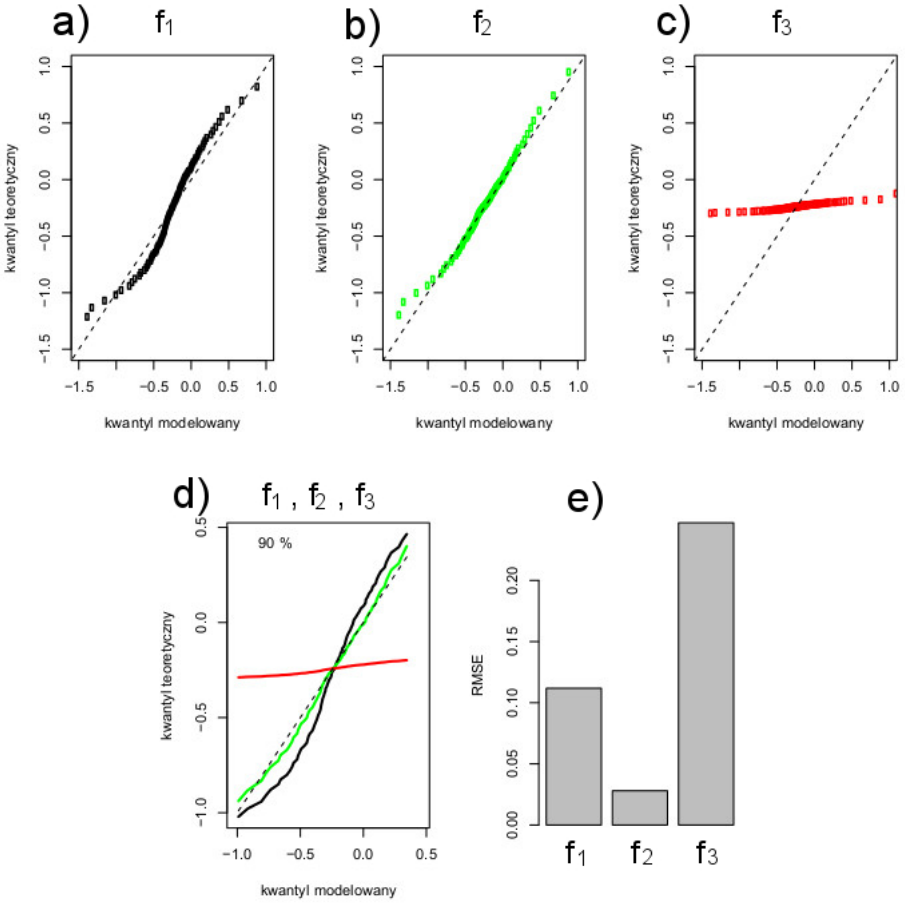
$F_1^{-1}(0.95)$	$F_2^{-1}(0.95)$	$F_3^{-1}(0.95)$
0,657 m	0,591 m	0,303 m



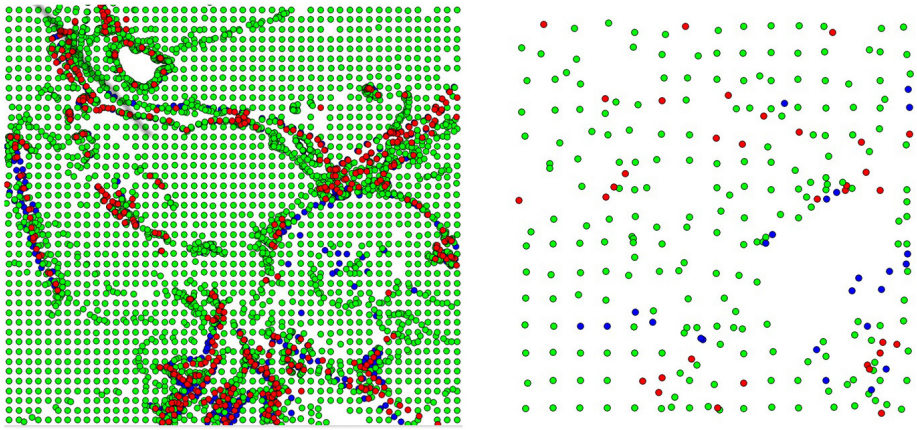
d)

Liczba punktów	275
Liczba wartości odstających	0
Min	-2,912 m
Max	1,102 m
μ	-0,254 m
σ	0,465 m
1,64 σ	0,763 m
1,96 σ	0,911 m
RMSE	0,530 m
1,96*RMSE	1,039 m
k	7,190
s	-1,423
m	-0,242
P 0,683	0,389 m
P 0,95	1,028 m
NMAD	0,278 m
b	0,302 m
2^0,5*b	0,427 m

Rysunek 41: Histogram różnic wysokości pomiędzy NMT LPIS i RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)



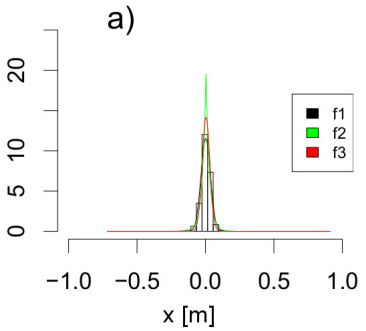
Rysunek 42: Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i RTK

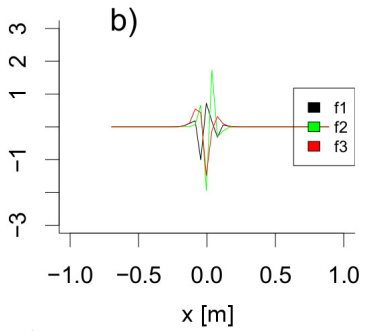


Rysunek 43: Odchyłki pomiędzy NMT LPIS i "chmurą punktów" w zakresie jednego odchylenia standardowego, (z lewej, kolor zielony (-0,250 m; 0,250 m), kolor niebieski poniżej - 0,250 m, kolor czerwony powyżej 0,250 m) oraz punktami RTK (z prawej, kolor zielony (-0,456 m ; 0,456 m), kolor niebieski poniżej - 0,456 m, kolor czerwony powyżej 0,456 m)

5.3. Analiza jakości NMT ISOK

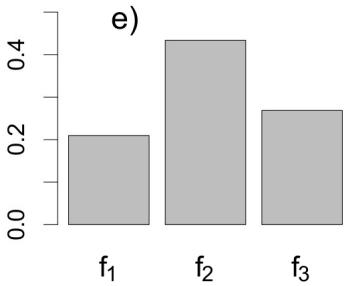
W ostatniej części zajmiemy się wynikami analiz modelu NMT z ostatnio realizowanego projektu ISOK.





c)

$F_1^{-1}(0.95)$	$F_2^{-1}(0.95)$	$F_3^{-1}(0.95)$
0,072 m	0,079 m	0,059 m

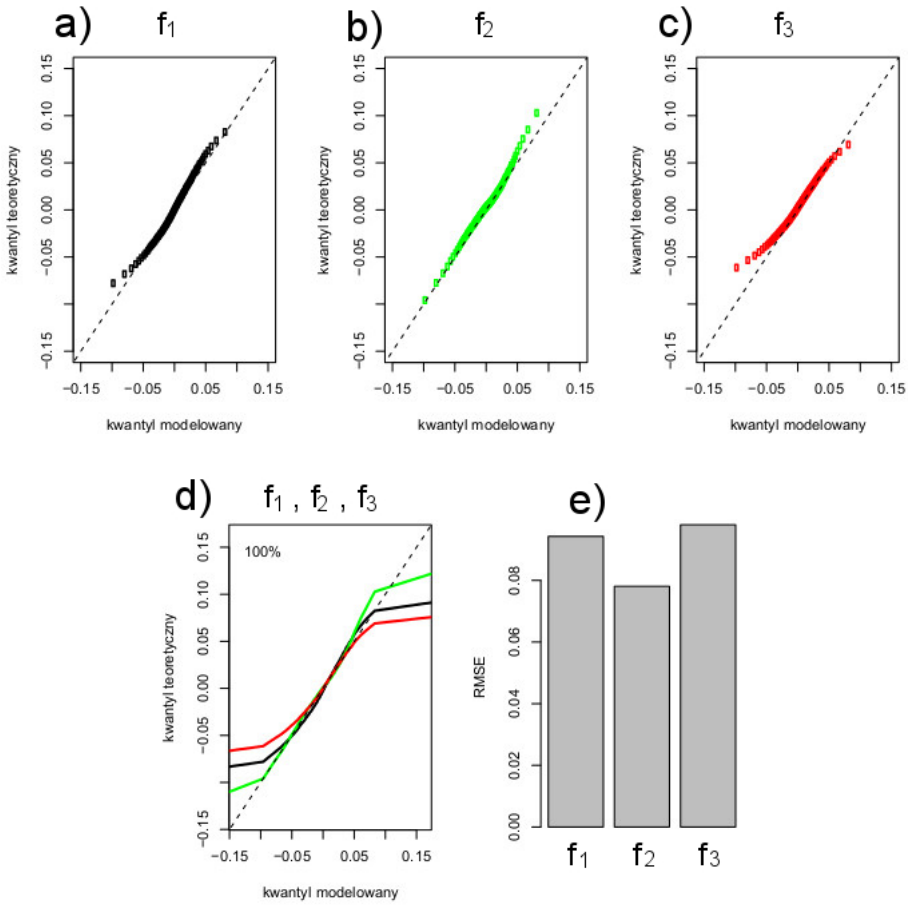


d)

Liczba punktów	1106438
Liczba wartości odstających	0
Min	-0,717 m
Max	0,911 m
μ	0,004 m
σ	0,035 m
1,64 σ	0,057 m
1,96 σ	0,069 m
RMSE	0,035 m
1,96*RMSE	0,069 m
m	0,004
k	6,129
s	-1,490
P 0,683	0,030 m
P 0,95	0,068 m
NMAD	0,028 m
b	0,025 m
2^0,5*b	0,035 m

Rysunek 44: Histogram różnic wysokości pomiędzy NMT ISOK i chmurą punktów wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

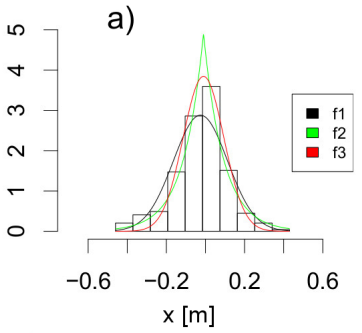
W związku z brakiem systematycznego przesunięcia pomiędzy NMT a chmurą punktów wartości: $1,96\ \sigma$, $1,96\ RMSE$ i odwrotność dystrybuanty rozkładu Gaussa są bliskie siebie i wynoszą około 0,070 m. W tym

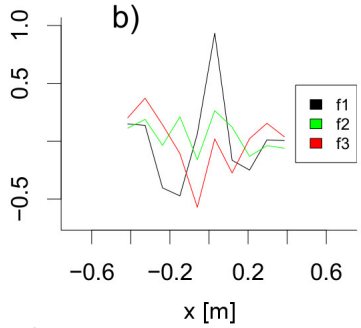


Rysunek 45: Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT ISOK i chmurą punktów

przypadku, pomimo różnic w stopniu dopasowania poszczególnych funkcji gęstości do histogramu empirycznego oraz do wykresów Q-Q, uzyskano bardzo zbliżone wskaźniki dokładności, niezależnie od sposobu obliczeń (Rys. 44 c). Na ich podstawie można określić dokładność wewnętrzną na poziome 7 cm. Odchylenia większe można zaobserwować wzdłuż krawędzi dróg, oraz budynków (Rys.48).

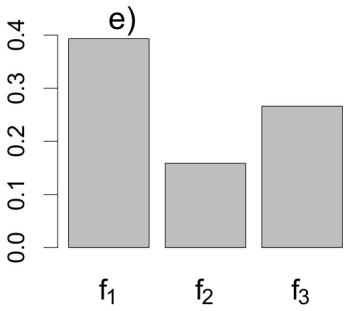
W dalszej kolejność przeprowadzono analizę niepewności bezwzględniej modelu NMT ISOK. Wykorzystano do tego celu wyniki pomiarów RTK, które zostały wykonane tak jak w przypadku modelu LPIS, czyli w siatce o wymiarach około 60 x 60 m (Kulesza 2013).





c)

$F_1^{-1}(0.975)$	$F_2^{-1}(0.975)$	$F_3^{-1}(0.975)$
0.244 m	0,296 m	0,192 m

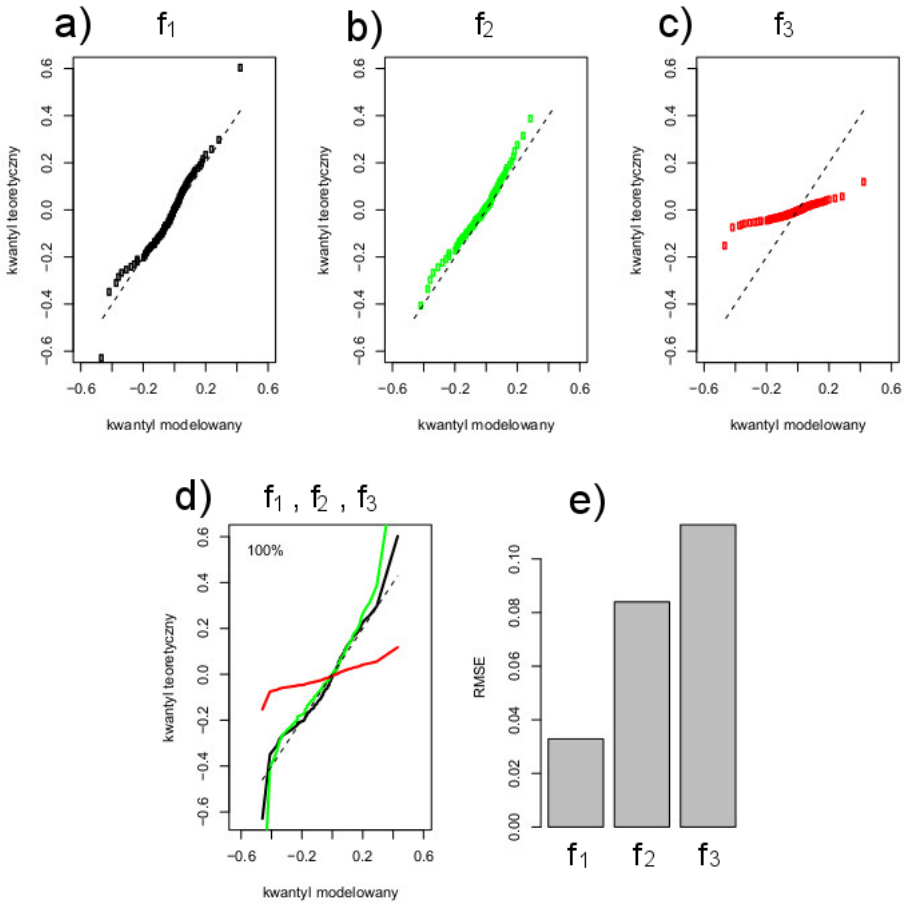


d)

Liczba punktów	275
Liczba wartości odstających	0
Min	-0,460 m
Max	0,430 m
μ	-0,026 m
σ	0,138 m
1,64 σ	0,226 m
1,96 σ	0,270 m
RMSE	0,141 m
1,96*RMSE	0,276 m
m	-0,010
k	1,896
s	1,474
P 0,683	0,120 m
P 0,95	0.330 m
NMAD	0,103 m
b	0,102 m
2^0,5*b	0,144 m

Rysunek 46: Histogram różnic wysokości pomiędzy NMT ISOK i RTK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

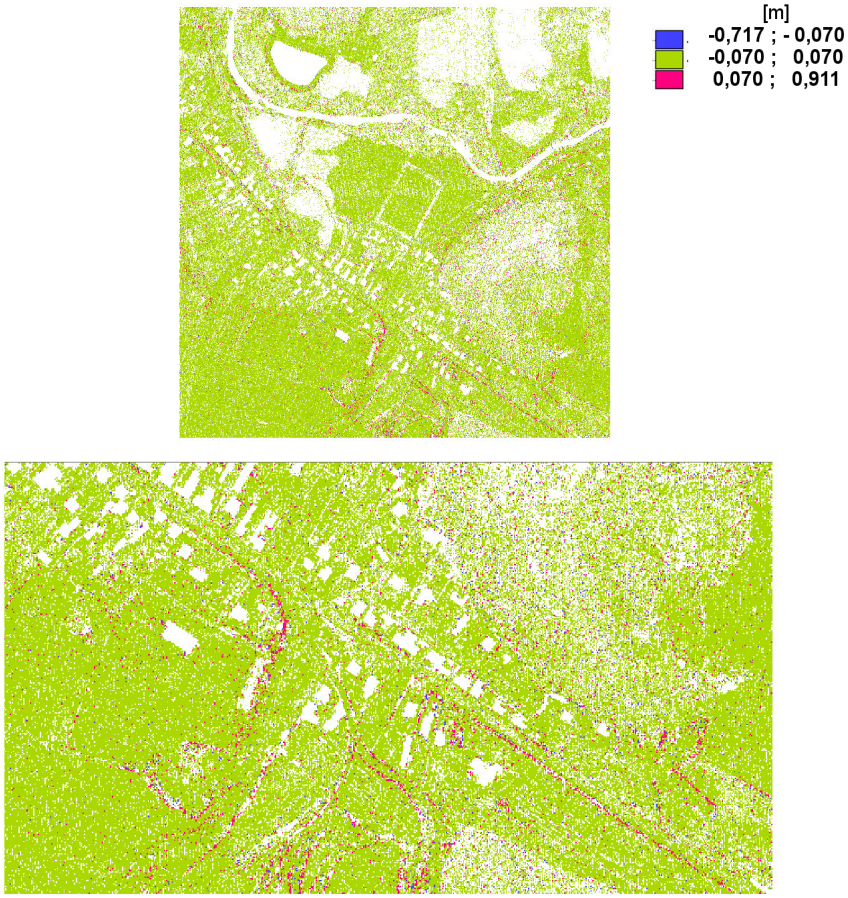
W tym przypadku, jak pokazują rysunki (Rys. 47 i Ry.48) uzyskano z analizowanych funkcji większą rozbieżność wyników. Zaobserwo-



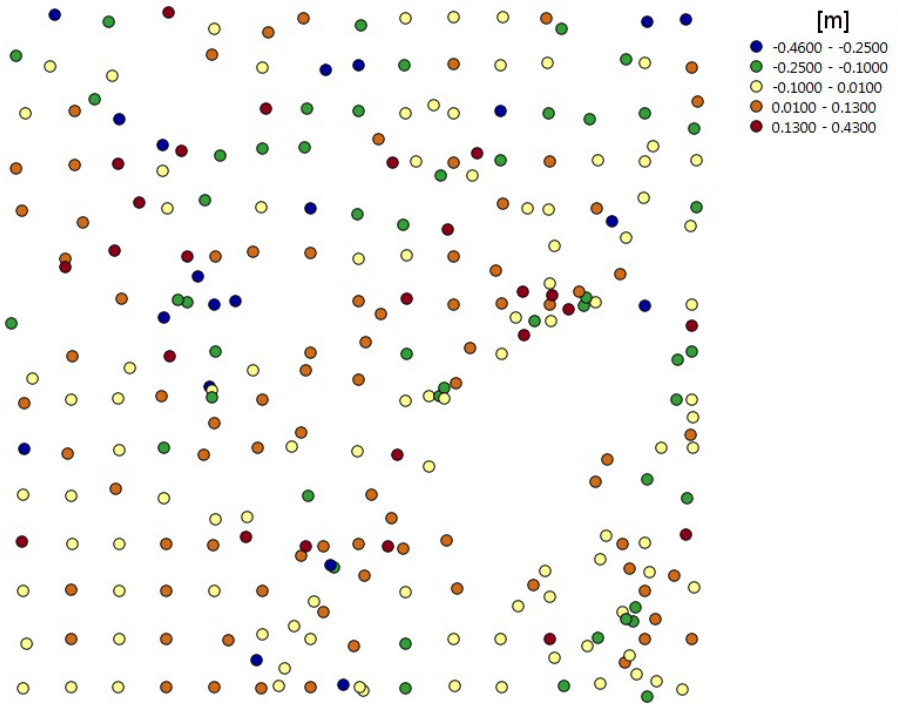
Rysunek 47: Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT i RTK

wano nieznaczne przesunięcie w górę NMT w stosunku do pomiaru RTK (Rys. 44 d), $\mu = -0,026\ m$ i $m = -0,01\ m$). Wielkości odwrotności dystrybuanty wynoszą dla rozkładu Gaussa 0,244 m, dla rozkładu Laplace'a 0,296 m i dla metody odpornej 0,192 m (Rys. 46 c). Patrząc na dopasowanie funkcji (Rys. 46 e) w zestawieniu z wykresami Q-Q (Rys. 47 e), wydaje się, że najbardziej wiarygodnym wskaźnikiem niepewności

jest wartość odwrotności dystrybuanty rozkładu Gaussa równa 0,244 m. Przy czym wartość ta niewiele różni się od wartości dla rozkładu Laplace'a równej 0,296 m i od rozszerzenia odchylenia standardowego na poziomie 95%: 0,270 m czy RMSE 0,276 m. Najmniej wiarygodną wartością w tym przypadku jest odwrotność dystrybuanty dla metody odpornej równa 0,192 m. Rozkład przestrzenny odchyłek względnych znajduje się odpowiednio na rysunkach 47 i 48.



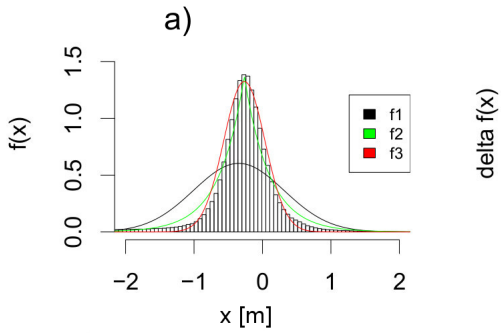
Rysunek 48: Rozkład przestrzenny odchyłek pomiędzy NMT ISOK a chmurą punktów z progiem rozszerzenia niepewności równym 0,07m



Rysunek 49: Rozkład przestrzenny odchyłek bezwzględnych NMT ISOK RTK

5.4. Porównanie NMT LPIS i ISOK

Ostatnim poruszanym zagadnieniem jest porównanie obu omówionych powyżej modeli NMT LPIS i NMT ISOK. Obliczone odchyłki wysokościowe zostały poddane analizie zgodnie ze stosowaną we wszystkich przypadkach metodyką.

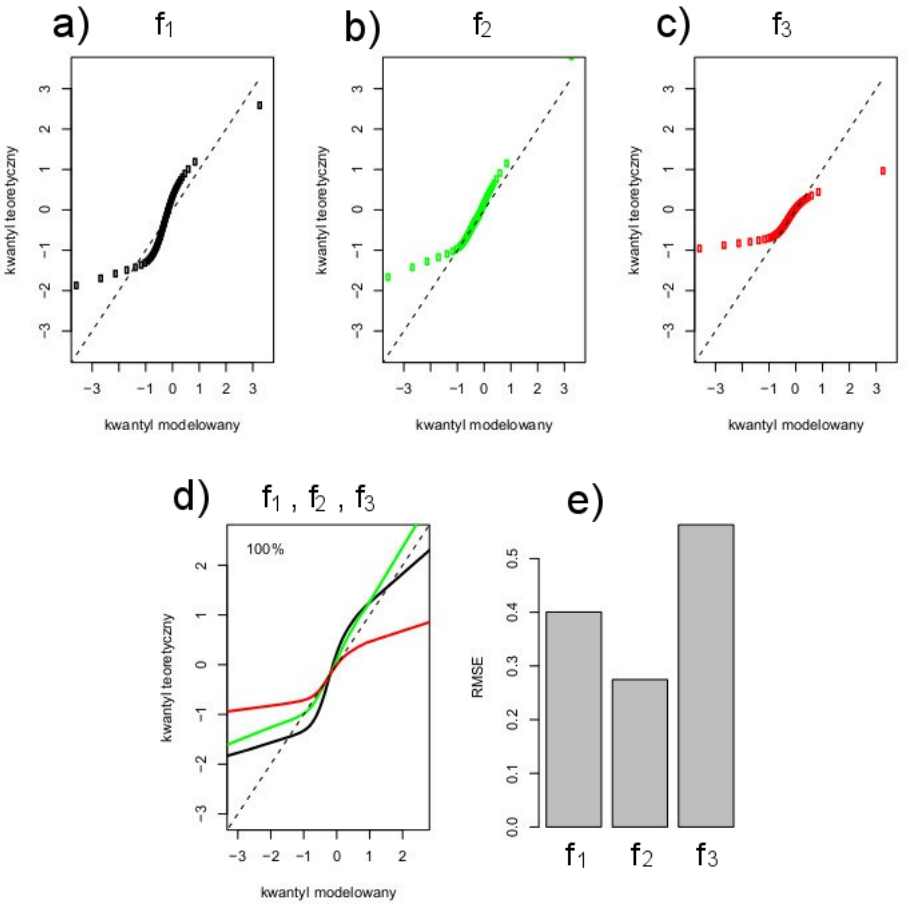


$F_1^{-1}(0.95)$	$F_2^{-1}(0.95)$	$F_3^{-1}(0.95)$
0,946 m	0,825 m	0,325 m

Liczba punktów	904401
Liczba wartości odstających	0
Min	-5,422 m
Max	3,308 m
μ	-0,347 m
σ	0,660 m
1,64 σ	1,082 m
1,96 σ	1,294 m
RMSE	0,746 m
1,96*RMSE	1,462 m
k	13,764
s	-2,848
m	-0,265
P 0,683	0,440 m
P 0,95	1,418 m
NMAD	0,301 m
b	0,364 m
2^0,5*b	0,515 m

Rysunek 50: Histogram różnic wysokości pomiędzy NMT LPIS i NMT ISOK wraz z funkcjami gęstości prawdopodobieństwa (f_1, f_2, f_3)

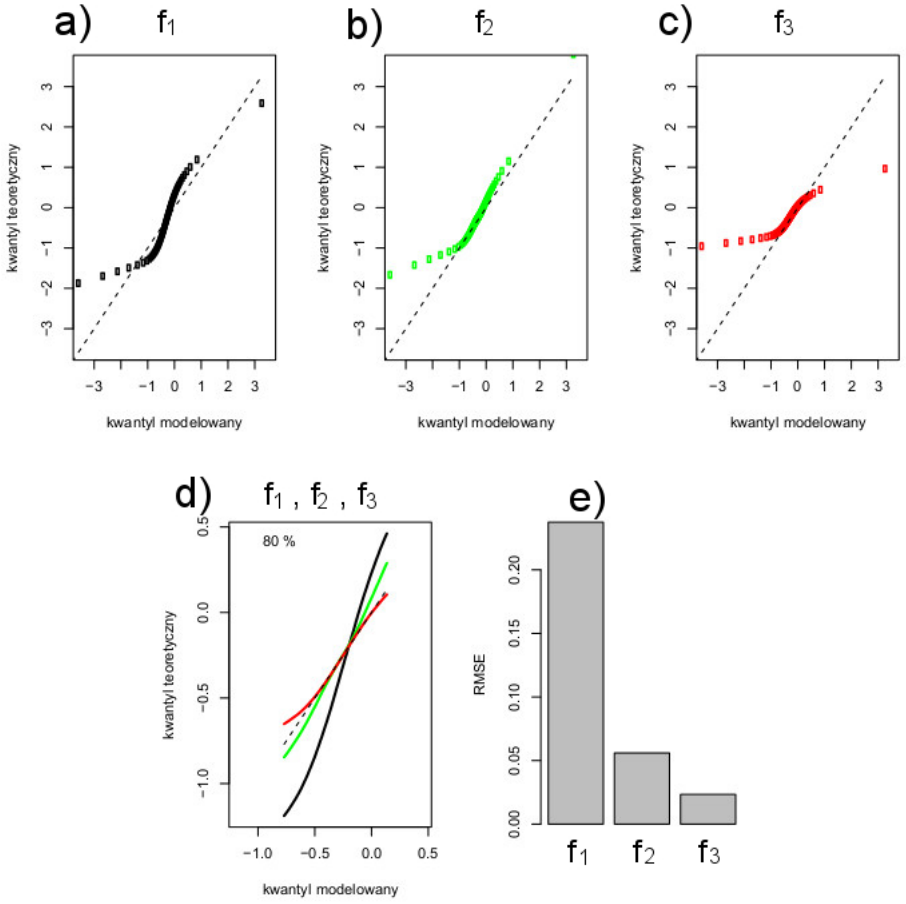
Na rysunku (Rys.50) widoczne jest dobre dopasowanie funkcji od-



Rysunek 51: Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i NMT ISOK - wszystkie punkty

pornej i rozkładu Laplace'a do danych empirycznych, podczas gdy funkcja gęstości rozkładu Gaussa, również w tym przypadku, znacznie odbiega od histogramu empirycznego. Widoczna jest ponadto systematyczna odchyłka równa -0,347 m, oznaczająca, że NMT LPIS jest średnio o taką wartość powyżej NMT ISOK. W związku z tym najbezpieczniejsze byłoby przyjęcie jako parametru niepewności wartości 1,96 RMSE lub percentyla

P(0,95), czyli wartości około 1,5 m. W tej sytuacji wartości odwrotności dystrybuant nie stanowią wiarygodnych parametrów oceny niepewności modelu LPIS (zakładając, że model NMT ISOK jest w stosunku do niego referencyjny). Po usunięciu stałego przesunięcia obu modeli, odwrotności dystrybuant wynoszą odpowiednio: 1,294 m, 1,090 m, 0,590 m.

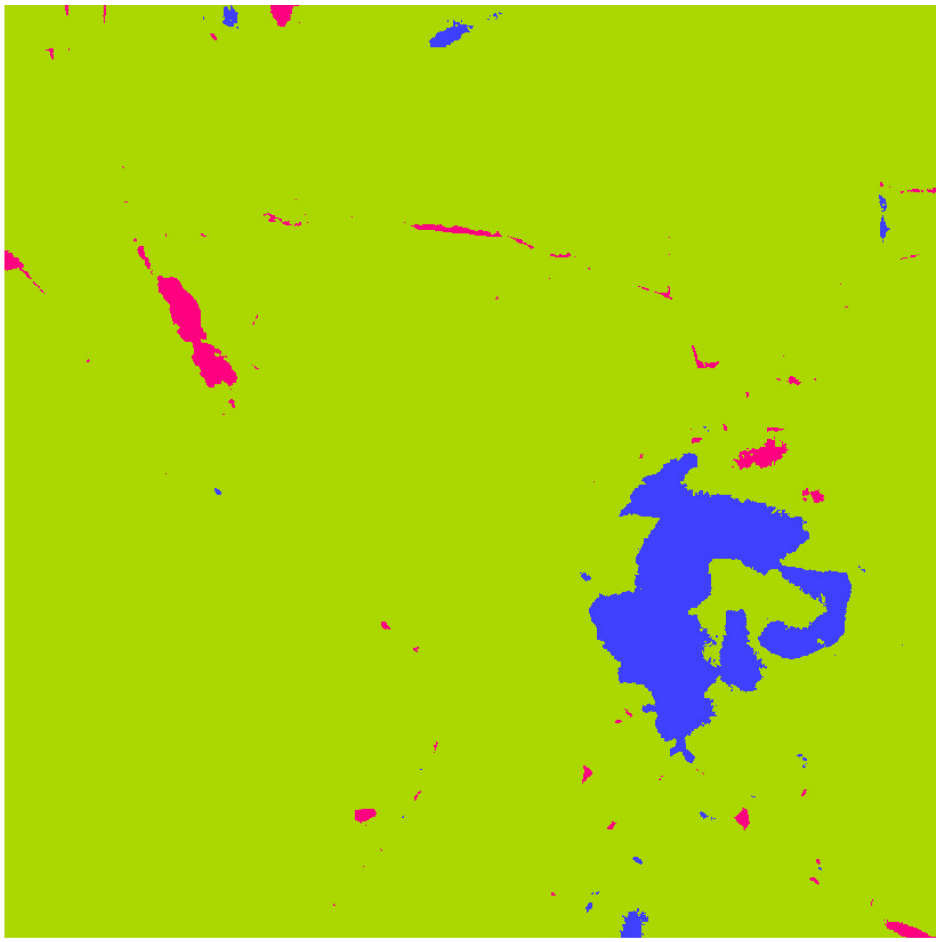


Rysunek 52: Wykres Q-Q wraz z błędami wpasowania, analiza różnic wysokości pomiędzy NMT LPIS i NMT ISOK - 80% punktów

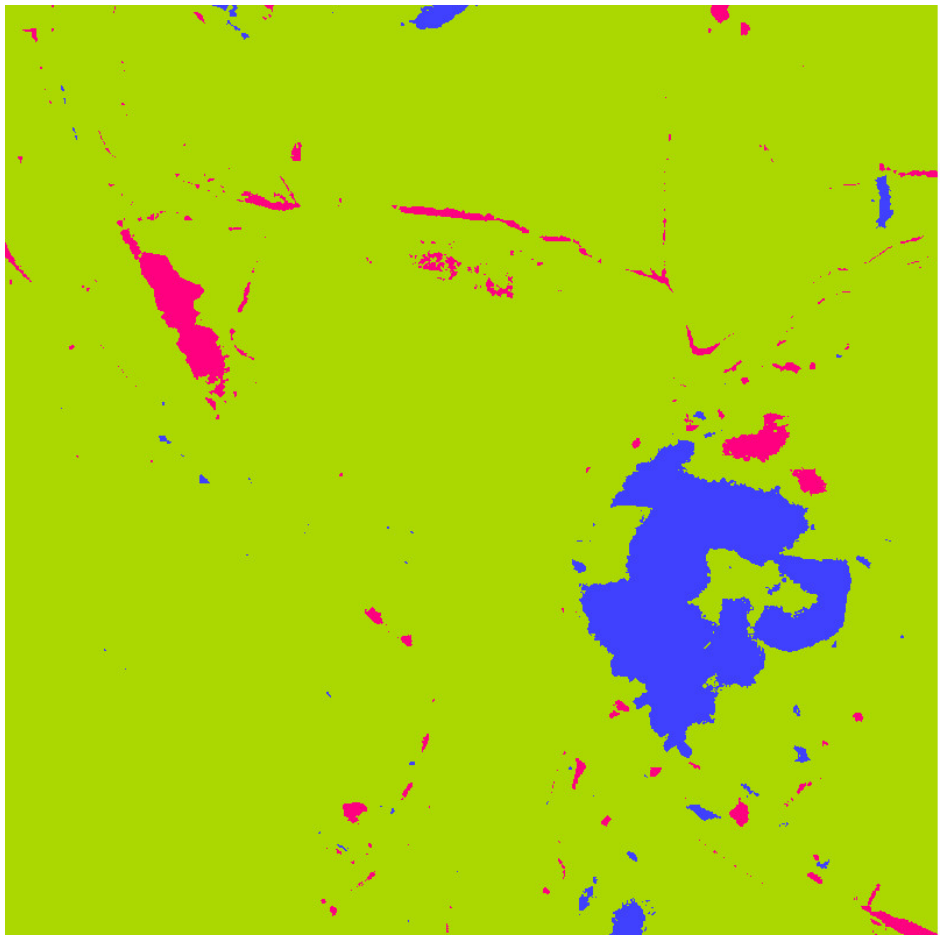
Przed wyciągnięciem ostatecznych wniosków przeanalizujmy jesz-

cze stopień dopasowania wykresów Q-Q. W przypadku, gdy wszystkie wartości odchyłek weźmiemy pod uwagę błąd dopasowania jest najmniejszy dla funkcji Laplace'a (Rys.51). Natomiast jeśli odrzucimy, jako odstające 20% danych, wtedy najlepsze dopasowanie uzyskamy dla metody odpornej (Rys.52). Wziąwszy te wszystkie fakty pod uwagę, najbardziej wiarygodną wartością niepewności wydaje się być odwrotność dystrybuanty funkcji rozkładu gęstości Laplace'a, który w zaokrągleniu wynosi około 1 m i znajduje się pomiędzy wartościami dla rozkładu Gaussa i metody odpornej.

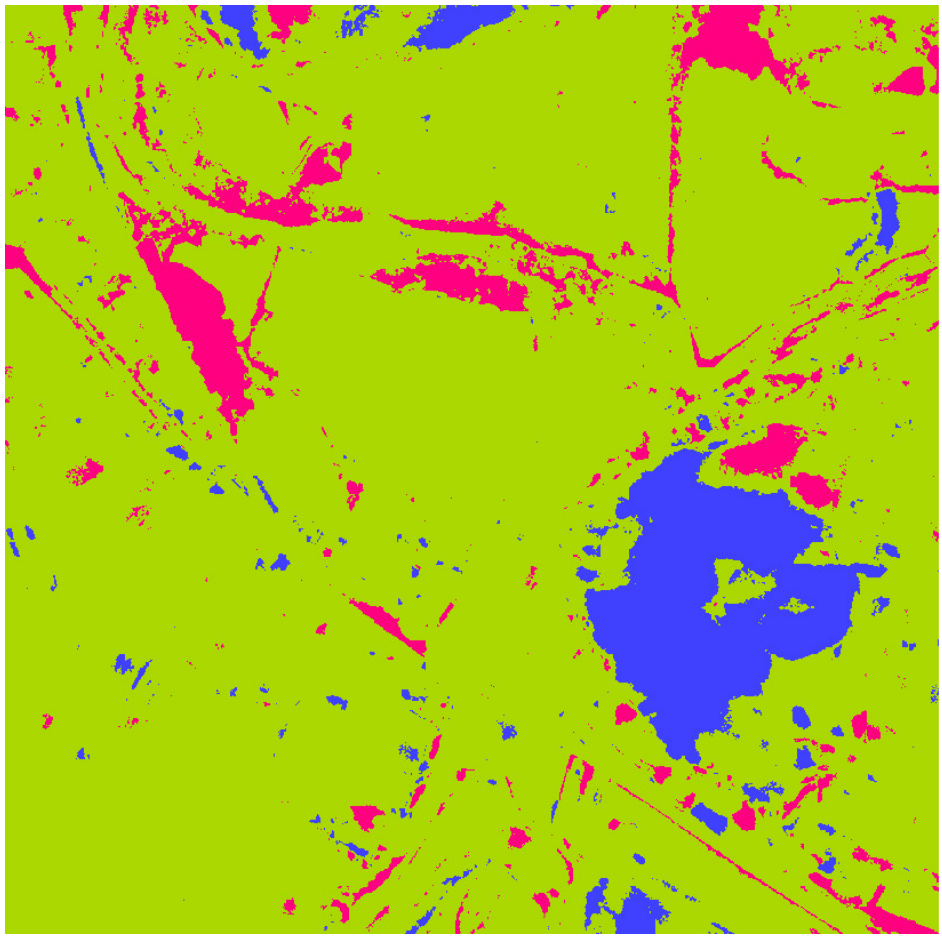
Pomocną ilustracją mogą być mapy rozkładu przestrzennego odchyłek między oboma modelami, przeklasyfikowane zgodnie z obliczoną wartością odwrotności dystrybuant (Rys.53, Rys.54 i Rys.55). Kolorem zielonym przedstawione są obszary na poziomie ufności 95% dla odpowiednich funkcji gęstości prawdopodobieństwa, kolorem niebieskim obszar NMT LPIS znajdujący się najbardziej powyżej powierzchni modelu NMT ISOK, a w kolorze czerwonym najbardziej poniżej NMT ISOK. W kolorze niebieskim i czerwonym przedstawiono odpowiednio obszary znajdujące się poniżej i powyżej przedziału przedziału ufności na poziomie 95%.



Rysunek 53: Rozkład przestrzenny różnic pomiędzy NMT LPIS i NMT ISOK - rozkład Gaussa; kolorem zielonym przedstawione są obszary na poziomie ufności 95% (-1,294 m ; 1,294 m), kolorem niebieskim obszary poniżej -1,294 m, a w kolorze czerwonym powyżej 1,294 m



Rysunek 54: Rozkład przestrzenny różnic pomiędzy NMT LPIS i NMT ISOK - rozkład Laplace'a; kolorem zielonym przedstawione są obszary na poziomie ufności 95% (-1,090 m ; 1,090 m), kolorem niebieskim obszary poniżej -1,090 m, a w kolorze czerwonym powyżej 1,090 m



Rysunek 55: Rozkład przestrzenny różnic pomiędzy NMT LPIS i NMT ISOK - metoda odporna; kolorem zielonym przedstawione są obszary na poziomie ufności 95% (-0,590 m ; 0,590 m), kolorem niebieskim obszary poniżej -0,590 m, a w kolorze czerwonym powyżej 0,590 m

Rozdział 6
Podsumowanie i wnioski końcowe

Podsumowując wyniki zamieszczone i omówione w niniejszej monografii, można stwierdzić co następuje:

1. Podejście do określania jakości danych przestrzennych można uzgodnić z podejściem proponowanym w Przewodniku, który zawiera zalecenia dotyczące szerokiej dziedziny jaką jest metrologia i do której należą pomiary geodezyjne.
2. Dla potrzeb analizy jakości danych przestrzennych można w pełni zaadaptować proponowane w Przewodniku pojęcia: błąd i niepewność, poszerzając jedynie definicję warunków ich stosowania.
   W Przewodniku rozróżnienie pomiędzy błędem i niepewnością wprowadzone zostało ze względu na rzadki przypadek idealny, kiedy mamy do czynienia z w pełni znanymi i opisywalnymi warunkami pomiaru oraz znamy wartość referencyjną. Możemy wtedy mówić o błędzie. We wszystkich innych przypadkach powinniśmy raczej używać pojęcia niepewność. W Przewodniku szczególną uwagę zwraca się na możliwość opisu warunków pomiaru, które mają znaczący wpływ na budżet niepewności oraz na dostępność wartości referencyjnej. W związku z dziedziną, której dotyczy monografia, a która obejmuje zagadnienia pomiaru, wprowadzania i analizowania danych przestrzennych do systemów GIS, podjęto próbę połączenia

podejścia metrologicznego z propozycjami z zakresu analizy danych przestrzennych. W publikacjach Fishera zwraca się uwagę na nieco inny aspekt pomiarowy, mianowicie na trudność w definicji obiektu. Rzeczywiście przedmiot pomiarów geodezyjnych nierzadko nie jest dobrze zdefiniowany, np. kiedy mierzymy jakiś obiekt naturalny, którego położenie nie jest jednoznacznie sygnalizowane. Równocześnie warunki pomiarów w terenie daleko odbiegają od warunków laboratoryjnych. W wyniku syntezy tych dwóch podejść (Fischer 1999, 2005 i Przewodnik) zaproponowano model koncepcyjny określania jakości danych przestrzennych. Pojęcie błąd mogłoby być zarezerwowane dla idealnych warunków pomiaru i idealnego obiektu mierzonego. W warunkach naturalnych, kiedy można zdefiniować obiekt i warunki pomiaru, przy czym nie mamy takiej pewności ich określenia jak w warunkach laboratoryjnych, moglibyśmy mówić o niepewności. Jeśli obiekt i/lub warunki pomiaru są trudne do zdefiniowania mówimy zwykle o nieokreśloności i stosujemy zbiory rozmyte. Ostatnim przypadkiem jest sytuacja kiedy obiekt i/lub warunki pomiaru są niejednoznacznie zdefiniowane. Na uwagę zasługuje propozycja linearnego charakteru klasyfikacji, w porównaniu z podejściem hierarchicznym zaproponowanym przez Fischera.

3. Poddano pod dyskusję zasadność opisywania niepewności modelu NMT i powierzchni działek z wykorzystaniem odchylenia standardowego i błędu RMSE bez analizy rozkładu odchyłek pomiędzy wielkościami mierzonymi a referencyjnymi. W wyniku przeprowadzonych badań potwierdzono fakt znany z literatury, dotyczący

przeszacowywania błędu/niepewności modelu NMT utworzonego z danych ze skaninigu laserowego, jeśli wykorzystamy do jego estymacji odchylenie standardowe rozkładu normalnego. Analiz histogramu błędów pomiarów powierzchni nie znaleziono w literaturze. Zapewne wynika to z faktu, że w naszej dziedzinie do tej pory rzadko była dostępna wystarczająca liczba pomiarów referencyjnych, dla wykonywania analizy histogramu.

4. Zaprezentowano zasadność szacowania jakości modelu NMT poprzez analizy dokładności wewnętrznej i zewnętrznej, zamiast jedynie zewnętrznej, jak to ma miejsce do tej pory.
5. Jeśli histogram odchyłek ma wysmukły kształt, bardzo skoncentrowany wokół wartości średniej, jednocześnie z tzw. "długimi ogonami", to nie należy, w celu określania błędu/niepewności, wykorzystywać parametrów rozkładu normalnego. Należy natomiast przeanalizować kształt histogramu i dobrać najlepiej pasującą funkcję gęstości prawdopodobieństwa. W omawianych przypadkach odpowiedni wydaje się rozkład Laplace'a, choć być może lepiej można by dopasować jakiś inny rozkład.
6. Daje się zauważyć następującą prawidłowość: im obiekt jest lepiej zdefiniowany (działki wybrane do walidacji) lub stosujemy lepszą metodykę pomiaru (ALS ISOK, precyzyjny pomiar RTK), tym bardziej odchyłki mają charakter rozkładu normalnego. Przeciwnie jeśli obiekt lub/i warunki pomiaru są gorzej zdefiniowane, mamy do czynienia z widoczną rozbieżnością pomiędzy rozkładem normalnym a empirycznym.

7. W przypadku analiz powierzchni należy zwracać uwagę, czy przedmiotem analizy są odchyłki bezwzględne, czy względne, gdyż w obu przypadkach możemy mieć do czynienia z różnym charakterem rozkładu. W przypadku odchyłek względnych rozkład był bardzo odbiegający od rozkładu normalnego. Ma to szczególne znaczenie dla procedur kontroli systemu LPIS. Wykorzystanie w tym przypadku odchylenia standardowego z rozkładu normalnego przeszacowuje niepewność tych danych około dwukrotnie, powodując niewiarygodne przekroczenie dopuszczalnych limitów. Najbardziej wiarygodne okazało się w tym przypadku zastosowanie rozkładu Laplace'a.
8. Pomiary "idealnych" obiektów wykonywanych w warunkach kontrolowanych, takie jak: pomiary referencyjne dla potrzeb walidacji metod pomiaru powierzchni działek rolnych w systemie IACS, pomiary walidacyjne GNSS oraz z wykorzystaniem ortofotomapy można charakteryzować poprzez rozszerzenie odchylenia standardowego rozkładu Gaussa. Analizy histogramów i stopnia dopasowania funkcji rozkładu prawdopodobieństwa wskazują na normalny charakter rozkładu odchyłek.
9. Jakość modelu NMT z lotniczego skaningu laserowego była zróżnicowana. Pomiary wykonane w latach 2005-2006 (NMT z JRC i Błoń Krakowskich) charakteryzowały się rozkładem innym niż normalny. Funkcja gęstości rozkładu Gaussa przeszacowywała w tym przypadku niepewność danych, podczas gdy analizowana funkcja odporna ją niedoszacowywała, jak się wydaje. Z analiz histogramów

i błędów wpasowania funkcji gęstości w histogramy empiryczne wynika, że niepewność lepiej oddają parametry rozkładu Laplace'a.

10. W systemach GIS znajdują się dane źródłowe oraz dane modelowe, które służą z definicji do wspomagania decyzji. Wspomaganie decyzji polega na dostarczaniu informacji pomocnych w rozwiązywaniu różnych problemów. Ponadto systemy GIS są wykorzystywane do badań naukowych, np. dotyczących testowania różnych algorytmów obliczeniowych. Jedną z podstawowych procedur jest analiza map różnicowych. Przykładem może być testowanie różnych algorytmów tworzenia modelu NMT z danych ze skaningu ALS i porównywanie wyników modelowania. Często też porównuje się modele NMT pochodzące z wielu źródeł. Również w tym przypadku tworzy się i analizuje mapy różnicowe. W monografii przeanalizowano dwie mapy różnicowe i w obu przypadkach odchyłki nie miały charakteru rozkładu normalnego. Tak jak we wszystkich innych przypadkach, tym razem również funkcja gęstości rozkładu Laplace'a wydaje się być najbardziej odpowiednia. Funkcja gęstości rozkładu Gaussa powodowała przeszacowanie niepewności a metoda odporna jej niedoszacowanie.

Bibliografia

1. Axelson P. 2000 - DEM Generation from Laser Scanner Data Using Adaptive TIN Models. International Archives of Photogrammetry and Remote Sensing, Amsterdam, Netherlands 32, B4/1, s. 110 - 117
2. Aguilar F.,J., Mills J.P., Delgado J., Manuel A., Aguilar M.A., Negreiros J.G., Pérez J. L. – 2010, Modelling vertical error in LiDAR-derived digital elevation models, ISPRS Journal of Photogrammetry and Remote Sensing 65 s. 103 - 110
3. Atkinson Gordo A. D. J., Ariza López F.J., José L., García-Balboa J. L. 2005 - Positional accuracy control using robust estimator, Proceedings of the 21st International Cartographic Conference, 09 - 16 lipiec, Acoruna, Spain
4. Bielawa A. 2011 - Postrzeganie i rozumienie jakości - przegląd definicji jakości, Studia i prace Wydziału Nauk Ekonomicznych i Zarządzania nr 21
5. Darnel A. R., Nicholas J., Tate N. J., Chris Brunsdon C. 2008 – Improving user assessment of error implications in digital elevation models, Computers, Environment and Urban Systems 32 s. 268 - 277
6. Doroszewski W. (red.) 2013 - Słownik języka polskiego, http://doroszewski.pwn.pl/, (dostęp lipiec 2013)

7. Devos W., Milenov P., Wojda P., Tarko A., Franielczyk R. 2012 - The first year of implementation of the LPIS quality assessment in the frame of Comm. Reg. Nr 1122/2009 art. 6.2, Luxembourg: Publications Office of the European Union, JRC 68193, EUR 25205 EN, ISBN 978-92-79-23067-7, ISSN 1831-9424, ftp://mars.jrc.ec.europa.eu/LPIS/Documents/LBNA25205ENN.pdf, (dostęp lipiec 2013)
8. Fisher P.F 1999 - Models of uncertainty in spatial data, w Longley P.A., Goodchild M.F, Maguire D.J., Rhind D.W.(red.) - Geographical Information Systems: Principles, Techniques, Management, and Applications, John Wiley and Sons: New York, NY, s. 191–205
9. Fisher P.F 2005 - Models of uncertainty in spatial data, w Longley P.A., Goodchild M.F, Maguire D.J., Rhind D.W. (red.)- Geographical Information Systems, Abridged edition - New Developments in Geographical Information Systems: Principles, Techniques, Management and Applications
10. Fisher P.F., Tate N.J. 2006 - Causes and consequences of error in digital elevation models, Progress in Physical Geography 30, 4, s. 467-489
11. Gel Y. R. 2010 - Test of fit for a Laplace distribution against heavier tailed alternatives, Computational Statistics and Data Analysis 54 s. 958 - 965
12. Goodchild M. F. 1992 - Geographical information science – Int. J. Geographical Inforrmation Systems, 1992, vol. 6, Nr 1, s. 31 - 45
13. Goodchild M.F. 1997 - What is Geographic Information Science?,

NCGIA Core Curriculum in GIScience, http://www.ncgia.ucsb.edu/giscc/units/u002/u002.html,(dostęp lipiec 2013)

14. Höhle J., Höhle M., 2009 – Accuracy assessment of digital elevation models by means of robust statistical methods, ISPRS Journal of Photogrammetry and Remote Sensing 64 s. 398 - 406
15. Höhle J., 2011 - The EUROSDR project Automated checking and improving of digital terrain models http://www.asprs.org/publications/proceedings/tampa2007/0027.pdf, (dostęp lipiec 2013)
16. Hejmanowska B., 2005 – Wpływ jakości danych na ryzyko procesów decyzyjnych wspieranych analizami GIS, Rozprawy, monografie AGH nr 141, ISSN 0867 - 6631,
17. Hejmanowska B., Borowiec N. Badurska M. 2008 a. – Processing airborne data to Digital Surface Model and Digital Terrain Model. Final report http://home.agh.edu.pl/ galia/research/Processing%20LIDAR%20%202007%20final%20report_5_03_2008.pdf, (dostęp lipiec 2013)
18. Hejmanowska B., Drzewiecki W., Kulesza Ł. 2008 b. - Zagadnienie jakości numerycznych modeli terenu, Archiwum Fotogrametrii, Kartografii i Teledetekcji, ISSN 2083-2214, vol. 18a s. 163–175
19. Hejmanowska B. Kay S. 2011 - A new approach to DTM error estimation basing on Laplacian probability distribution function, Archives of Photogrammetry, Cartography and Remote Sensing, ISSN 2083-2214. 1 vol. 22 s. 201–213

20. Hejmanowska B., Loudjani P., Luckau C., Ganisheva K., Fasbender D. 2012 - Maussane study on GNSS measurements: preliminary results, JRC Monograph, Geomatics in support of the Common Agricultural Policy, Luxembourg: Office for Official Publications of the European Communities, ISBN 978-92-79-26644-7 (pdf)
21. Jahn M., Grabarse S., Ewen M., Linden S., Dammann H.P., Ehlers M., Behncke K., Klonus S. 2012 - Investigation of tolerance for on-the-spot-checks results of the German study 2011, JRC Monograph, Geomatics in support of the Common Agricultural Policy, Luxembourg: Office for Official Publications of the European Communities, ISBN 978-92-79-26644-7 (pdf)
22. Jędrychowski I. 2007 - Lotnicze skanowanie laserowe Kraków, Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol. 17a, 2007, ISBN 978-83-920594-9-2
23. Kay S, Sima A. 2006 - Area measurement validation scheme, Luxembourg: Office for Official Publications of the European Communities - ISBN 978-92-79-08714-1
24. Klir G J, Yuan B 1995 - Fuzzy sets and fuzzy logic: theory and applications, Englewood Cliffs, Prentice-Hall
25. Kramarczyk P., Hejmanowska B., Dąbrowski J. 2012 - Walidacja odbiorników GNSS dla potrzeb kontroli wielkości powierzchni działek rolnych w systemie dopłat bezpośrednich dla rolnictwa IACS: wytyczne, aplikacja, problemy, Wydawnictwo Państwowej Wyższej Szkoły Techniczno-Ekonomicznej, Jarosław, ISBN 978-83-88139-52-9

26. Kramarczyk P, Warchoł A., 2011 - GPS validation ProXT Beacon, niepublikowane sprawozdanie z walidacji Trimble w Hanowerze
27. Kramarczyk P., 2011 - GPS validation GeoExplorerXT6000 Series + EGNOS, niepublikowane sprawozdanie z walidacji Trimble w Krakowie
28. Kraus K., Karel W., Briese Ch., Mandlburger G., 2006 - Local Accuracy Measures for Digital Terrain Model, Photogrammetric Record 21 (116), s. 342-354
29. Kulesza Ł., 2013 - Porównanie metod estymacji przestrzennego rozkładu błędu numerycznych modeli terenu, niepublikowana praca doktorska
30. Milenov P., Galabova K., 2009 - Use of VHR PAN image for the CwRS Results of validation test of EROS B and Ikonos imagery, 2009 CAP Control Methods Workshop - Casa Don Guanella (Ispra), Italy, 06 - 08 kwiecień 2009, http://marswiki.jrc.ec.europa.eu/wikicap/images/4/47/Use_of_VHR_PAN_image_for_the_CwRS_Results_of_validation_test_of_EROS_B_and_Ikonos_imagery.pdf, (dostęp lipiec 2013)
31. Milenov P., Wojda P., DeSantis S., 2009 - Executive Test Suite explained: Belgium - Flanders experience, LPIS Workshop: Applications and Quality, 6-8 październik, 2009, Tallinn, http://mars.jrc.ec.europa.eu/mars/News-Events/LPIS-2009-Workshop/Agenda-and-Presentations/S3_ETSexperience_

Milenov_Roggemans_JRC_v2.pdf (dostęp lipiec 2013)

32. Oksansen J. 2006 - Digitale elevation model error in terrain analysis, PhD Thesis http://ethesis.helsinki.fi/julkaisut/mat/maant/vk/oksanen/digitale.pdf, (dostęp lipiec 2013)
33. Schunn C. D., Wallach D., 2005 - Evaluating goodness-of-fit in comparison of models to data, w Tack W. (wyd.), Psychologie der Kognition: Reden and Vorträge anlässlich der Emeritierung von Werner Tack (s. 115-154), Saarbrueken, Germany: University of Saarland Press
34. Wechsler S.P., 2001 - DEM Uncertainty: Evaluation and Effect on Topographic Parameters, Dissertation State University of New York, College of Environmental Science and Forestry, Syracuse, New York
35. 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), s. 103–130
36. ASPRS Lidar Committee 2004 - ASPRS Lidar Committee Guidelines Vertical Accuracy Reporting for Lidar Data, http://www.asprs.org/a/society/committees/lidar/Downloads/Vertical_Accuracy_Reporting_for_Lidar_Data.pdf, (dostęp lipiec 2013)
37. Główny Urząd Miar 1999 - Wyrażanie niepewności pomiaru - Przewodnik
38. GUGIK 2000 - Zasady wykonywania ortofotomap w skali 1:10 000, wytyczne techniczne, Główny Urząd Geodezji i Kartografii, War-

szawa

39. PN-EN ISO 9001:2009 Systemy zarządzania jakością
40. Norma ISO 19114:2003 - Informacja Geograficzna - Procedury Oceny Jakości
41. Norma PN-EN-ISO 19113 Informacja geograficzna – Podstawy opisu jakości
42. JRC MARS 2013 a. - Area measurements http://marswiki.jrc.ec.europa.eu/wikicap/index.php/Category:Area_measurement, (dostęp lipiec 2013)
43. JRC MARS 2013 b. - Buffer tolerance validation method http://marswiki.jrc.ec.europa.eu/wikicap/index.php/Buffer_tolerance_validation_method, (dostęp lipiec 2013)
44. JRC MARS 2013 c. - LPIS Quality Assessement http://marswiki.jrc.ec.europa.eu/wikicap/index.php/Category:LPISQA, (dostęp lipiec 2013)
45. JRC MARS 2013 d. - LPISQA Tresholds http://marswiki.jrc.ec.europa.eu/wikicap/index.php/LPISQA_3.e.iii_Thresholds, (dostęp lipiec 2013)
46. Rozporządzenie MSWiA 2011 - Rozporządzenie ministra spraw wewnętrznych i administracji z dnia 3 listopada 2011 r. w sprawie baz danych dotyczących zobrazowań lotniczych i satelitarnych oraz ortofotomapy i numerycznego modelu terenu, Dz.U. 2011 nr 263 poz. 1571
47. Rozporządzenie Komisji (WE) NR 1122/2009 z dnia 30 listopada 2009 r. ustanawiające szczegółowe zasady wykonania rozporządze-

nia Rady (WE) nr 73/2009 odnośnie do zasady wzajemnej zgodności, modulacji oraz zintegrowanego systemu zarządzania i kontroli

48. Komisja Europejska 2007 - Managing the agriculture budget wisely, http://ec.europa.eu/agriculture/fin/clearance/factsheet_en.pdf (dostęp lipiec 2013)
49. Słownik języka polskiego - Wydawnictwo naukowe PWN, http://sjp.pwn.pl/ (dostęp lipiec 2013)