

Оценка и визуализация на качеството на цифровите модели на терена (DTM) при картографиране на заплахата и риска от наводнения

Антон Попов, Стелиян Димитров
СУ „Св. Климент Охридски“, гр. София, 1504, бул. „Цар освободител“ №15
[e-mail: popovanton@gmail.com](mailto:popovanton@gmail.com)

Абстракт: Настоящата статия има за цел да представи приложените от авторите подходи и методи за оценка и визуализация на качеството на данните в цифровите модели на терена (DTM), използвани при изготвянето на картите за заплахата и риска от наводнения на примера на Западнобеломорския район за управление на водите. Подробно са разгледани математико-статистическите и пространствено-аналитичните техники, чрез които са оценени неопределеността и точността на използваните геопространствени данни, използвани при оценката на заплахата и риска от наводнения.

Ключови думи: цифров модел на терена, DTM, ГИС, картографиране, риск от наводнения

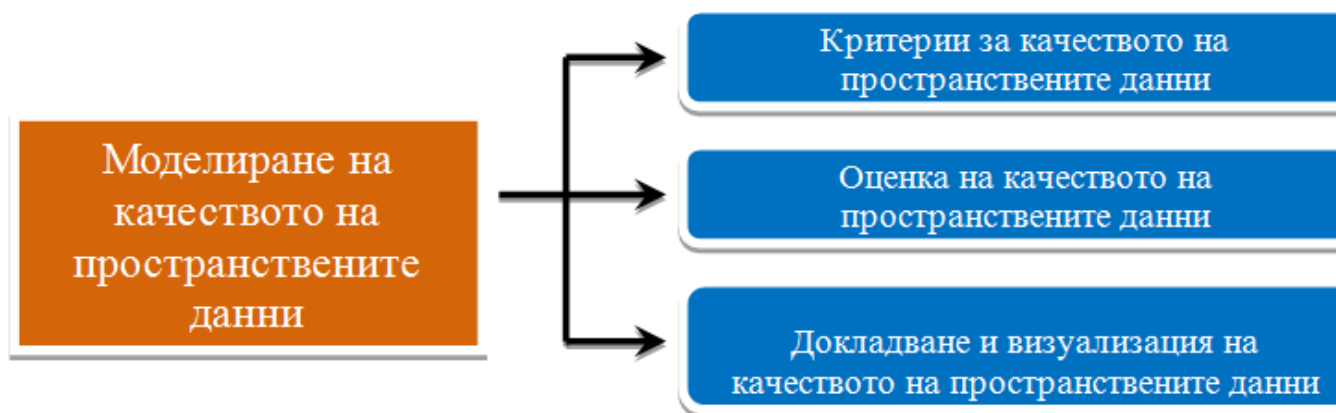
1. Цел и контекст на изследването

В днешно време определянето на зоните с опасност и риск от наводнение става с помощта на ГИС в комбинация с хидроложко и хидравлично моделиране. Съществен момент в такъв тип изследвания е осигуряването на качествени данни за цифровия модел на терена (DTM).

Целта на публикацията е да представи приложените от авторите подходи и методи за оценка и визуализация на качеството на данните в цифровите модели на терена (DTM), използвани при изготвянето на картите за заплахата и риска от наводнения на примера на Западнобеломорския район за управление на водите.

Качеството на данните в DTM има фундаментално значение за целия процес на изготвяне на картите за опасността (запахата) и риска от наводнения. Точността на пространствените данни (позиционна, тематична, темпорална) и логическата им съвместимост са най-важните компоненти на качеството на данните, тъй като точността гарантира годността им за използване за конкретни изследователски и практически цели [1, 2]. От друга страна, големината и пространственото разпределение на грешките са елементи на несигурността при извличане на информацията от данните, като в определени случаи грешките в пространствените данни могат да компрометират изпълнението на конкретния проект [3, 4]. В идеалния случай качеството на пространствените данни трябва да съответства и на определени национални и международни стандарти за оперативна съвместимост на географската информация при обмена на данни в националните и международни инфраструктури на пространствените данни (по ISO TC211, INSPIRE и др.) [5, 6]. Обикновено генерираните набори от пространствени данни се придружават от метаданни, включващи описание на: изходния материал, от който са събирани данните, точността на теренните изследвания и прецизността на използваното оборудване, методите за дигитализиране и процедурите по контрол на качеството при създаването на DTM.

Процедурите по контрола на качеството включват три главни компонента [6] – дефиниране на критерии за качеството на данните, извършване на оценка на качеството на данните (съобразно тези изисвания), и докладване и визуализация на качеството на пространствените данни (фиг. 1).



Фиг. 1. Компоненти на процедурите по контрола на качеството на данните в DTM

В настоящата публикацията се обсъждат преди всичко въпросите, свързани с използваните критерии и методи за оценка на точността и грешките в DTM, като същевременно е предложена и апробирана методология за визуализиране на резултатите от тази оценка.

2. Качество на изходните данни за генериране на DTM

Поради липса на данни от прецизно лазерно сканиране (LiDAR) за района на изследване, бяха използвани два основни източника на изходни данни – едромасщабни топографски карти (ЕТК) и ортофотоизображения. Това наложи да бъдат оценени грешките в изходните материали, допринасящи за несигурността на генерирания DEM и техният кумулативен ефект върху качеството на модела, задаващ геометрията на изследваните речни участъци. Установи се, че ортофотоизображенията са с ниска вертикална точност, което ги прави неподходящи за генериране на цифровия модел на терена. Основната причина за това е, че по същество те не са цифрови модели на терена (DTM), а цифрови модели на повърхността (DSM), тъй като не са „почистени“ дърветата, горските масиви, сградите и съоръженията. Установи се, че за създаване на DTM по-подходящи са изходни данни за височините от ЕТК в М 1:5000 със сечения на хоризонталите 1 m, като допустимата неточност при тях зависи от зададени прагови стойности на наклоните, зависещи от своя страна от хоризонталните разстояния между изохипсите и от това дали конкретните територии са залесени или обезлесени. По подразбиране точността на височините, изобразени с изолинии в ЕТК, не може да бъде по-голяма от 1/2 от сечението между хоризонталите, т.е. 0.5 m. При наклони близки до 0° съществено влияние оказва голямото хоризонтално разстояние при плоски терени между изохипсите, допринасящо за по-голяма несигурност в геометрията на границите на полигоните на заливаемите територии. Друг съществен проблем при използваните ЕТК е актуалността на съдържащата се в тях географска информация. От времето на тяхното изработване и отпечатване до днес са настъпили значителни промени както в природните, така и в антропогенните обекти, изобразени на картата, вкл. корекции на речни легла или настъпили естествени промени в конфигурацията на речните течения.

3. Оценка на качеството на данните в генерирания DTM

Цифровият височинен модел е основен източник на информация за разработване на други модели, изцяло или частично зависещи от топографията. Грешките в растерните матрици на височините (гридове) са от атрибутивен тип, което предполага неправилно присвояване на стойности за височината по оста Z за дадена клетка в грида [7]. На практика от цифровия височинен модел зависят както пространственият обхват, така и дълбочината на заливаемите територии в различните речни участъци. Ефектът от неточностите в DTM върху обхвата на моделираните заливаеми територии е в пъти по-голям в сравнение с неточностите в хидравличния модел. Поради това вниманието ни беше насочено към определянето на вида (систематични и случайни грешки) и тяхната големина във височинния модел, определящи вертикалната точност и качеството на DTM. Тези грешки пораждаат несигурност, тъй като пряко влияят върху генерираната от хидравличния модел геометрия на речните корита и прилежащите им заливаеми територии.

Методите за откриване на аномални стойности на височините в DTM се базират на обективни статистически критерии в комбинация с визуализация на пространственото разпространение на грешките в изследваните райони [8]. В публикацията демонстрираме получените резултати от изследването за район Земен (BG4_APSFR_ST_09) по течението на р. Струма.

3.1 Критерии за определяне на грешките и вертикалната точност на DTM

За да бъдат изчислени грешките и вертикалната точност на DTM, е необходимо да бъдат използвани за сравнение два набора от данни:

- данни със стойностите на височините (z) на центроидите на клетките в грида;
- данни за напречните профили на речните течения със стойностите на височините на същите точки, определени инструментално на терена с GNSS¹, за които се приема, че поради по-високата им точност, ще бъдат референтни при определяне на грешките и вертикалната точност на DTM.²

Сравняваните височини са на точки по картираните на терена напречни профили на речното корито, в прилежащите до него заливаеми територии и за съществуващи по реката съоръжения (мостове, диги, защитни стени и др.).

Използваните статистически критерии за изчисляване на грешките и вертикалната точност на генерирания DTM са следните:

а) средна грешка (ME, m)

$$\bar{x} = \frac{1}{(n)} \sum_{i=1}^n x_i$$

Където:
 x_i е грешка по оста Z в i -та точка,
 n е броят на тестваните точки,
 i е цяло число от 1 до n .

¹ Забел. При статично позициониране може да се постигне позиционна точност от ± 2 cm, докато при кинематично позициониране, изискващо по-кратко време за наблюдения, може да се постигне точност от ± 5 cm. По правило, както при оценката на хоризонталната точност, вертикалната точност на измерванията с GNSS трябва да е поне три пъти по-висока от изискуемата (или определената като приемлива) вертикална точност на тестваните точки от DEM.

² Забел. Хоризонталните координати на тестваните точки (центроиди на клетките) от цифровия модел трябва пространствено напълно да съответстват на хоризонталните координати на същите точки, определени при GNSS измерванията. По същия начин могат да бъдат тествани и височините на възлите в TIN модел с данни от GNSS измервания.

б) стандартно отклонение (SD, m)

$$s_x = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Където:
 x_i е грешка по оста Z в i^{ma} точка,
 \bar{x} е средната грешка (ME) по оста Z ,
 n е броят на тестваните точки,
 $(n - 1)$ са степените на свобода,
 i е цяло число от 1 до n .

в) стандартна грешка (SE, m)

$$SE = \frac{SD}{\sqrt{n}}$$

Където:
SD е стандартното отклонение (m)
 n е броят на тестваните точки

г) средна квадратична грешка (RMSE, m)

$$RMSE_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{i(DEM)} - x_{i(GNSS)})^2}$$

Където:
 $x_{i(DEM)}$ е височината на i^{ma} тествана точка (пиксел) в DEM,
 $x_{i(GNSS)}$ е височината на i^{ma} тествана точка, определена с GNSS,
 n е броят на тестваните точки,
 i е цяло число от 1 до n .

Статистически, в незалесени територии и на други места, където грешките следват нормално (Гаусово) разпределение:

- 68.27% от грешките са в рамките на едно стандартно отклонение (SD) от средната грешка (ME);
- 94.45% от грешките са в рамките на две стандартни отклонения ($2 * SD$) от средната грешка (ME);
- 99.73% от грешките са в рамките на три стандартни отклонения ($3 * SD$) от средната грешка (ME).

На практика за апроксимиране на максималната възможна грешка от двете страни на средната стойност на грешката най-често се използват граничните стойности на интервала (\pm долна и горна граница) на стандартното отклонение при ниво на доверие $P = 95\%$ вероятност (т.е. 95-я процентил) [9]. Процентилът е интерполирана абсолютна стойност в набора от данни за грешките, разделяща разпределението на индивидуалните грешки на сто групи с еднаква честота. При 95-я процентил 95% от грешките в данните имат абсолютни стойности с равна или по-малка стойност, а 5% от грешките имат по-голяма стойност. Чрез този метод точността е приравнена към 95-процентил, което означава, 95% от грешките имат абсолютни стойности, които са равни или по-малки от всички грешки в набора от данни.

При същото ниво на доверие (95-я процентил) съществува шанс 1 към 20 (т.е. при ниво на значимост $\alpha = 0.05$) измерваната стойност да лежи извън доверителния интервал (Confidence Interval, CI). По същия начин, при 94% шансът е 1 към 17, при 96% – шансът е 1 към 25, а при 99% – шансът е 1 към 100. Интервалът, съответстващ на специфично ниво на доверие, се определя чрез коефициента k (Coverage factor). При ниво на доверие $P = 95\%$ (или $1 - 0.05$) и при степени свобода $(n - 1) > 100$ той е равен на 1.960. Обичайните стойности на k в зависимост от доверителния интервал са показани в табл. 1.

Табл. 1 Обичайни стойности на коефициента k

Стойности на коефициента k	на α	Доверителен интервал (1 - α)
1.00		0.68
1.96		0.95
2.00		0.95
2.58		0.99
3.00		1.00

По този начин, за да определим доверителния интервал (CI) на средната грешка, като изходни данни могат да бъдат използвани стойностите на средната грешка (ME), стандартното отклонение (SD), броят на точките (N), нивото на доверие (95-я процентил) и корекционният коефициент ($k = 1.96$). Изчисленията със статистическия пакет връщат като резултат стойности за доверителния интервал: долна граница на интервала (m) < средна грешка (m) < горна граница на интервала (m).

Средната квадратична грешка ($RMSE_z$) също може да съдържа грешка (E_{RMSE_z}), която може да бъде оценена по следния начин:

$$E_{RMSE_z} = \frac{RMSE}{\sqrt{n}}$$

Критерият $RMSE_z$, като най-добър измерител за вертикална точност, е определен за задължителен при оценка на качеството на височинните данни в документа „INSPIRE Data Specification on Elevation – Draft Technical Guidelines” (2014) на ЕС.³ По-точното определяне на вертикалната точност изисква интензивни статистически изчисления на базата на стандартното отклонение, средната грешка, коефициентите на асиметрия (Skew), ексцес (Kurtosis) и средната квадратична грешка. Такива статистически изчисления могат да бъдат извършвани с помощта на стандартни статистически пакети. Върху вертикалната точност влияят и много други фактори, като трансформирането на елипсоидалните височини в ортометрични, трансформациите между различни датуми, проекционните преобразувания, методите за интерполация и др., на които тук няма да се спираме.

3.2 Резултати от оценката на грешките и вертикалната точност на DTM за района на Земен

Наборите от данни за височините в район Земен включват стойности за надморските височини за 264 точки по напречните профили. За всяка точка са анализирани данни от измерванията на терена с GNSS, както и данни за височините на същите точки (центроидите на пикселите), извлечени от генерирания DTM. Разликите между стойностите от DTM и от GNSS измерванията дава представа за отклоненията (или грешките) на височините на точките в DTM спрямо приетите за по-точни стойности на височините, измерени на терена (табл. 2).

Табл. 2 Изходни данни за оценката на вертикалната точност на DTM, район Земен

Point ID	P	DEM	GNSS	Разлика (грешка)
	(1)	z (m)	z (m)	Δz (m)
1	1	592.20	592.63	-0.435
2	1	587.11	588.67	-1.567
3	6	584.31	586.53	-2.223
4	9	587.40	587.52	-0.113
5	6	590.66	588.57	2.091
...
64	0	587.92	586.66	1.257

С помощта на стандартен статистически софтуерен пакет са определени базисните статистически характеристики на стойностите в колони 2 и 3 в горната таблица, като полученият резултат е следният (табл. 3):

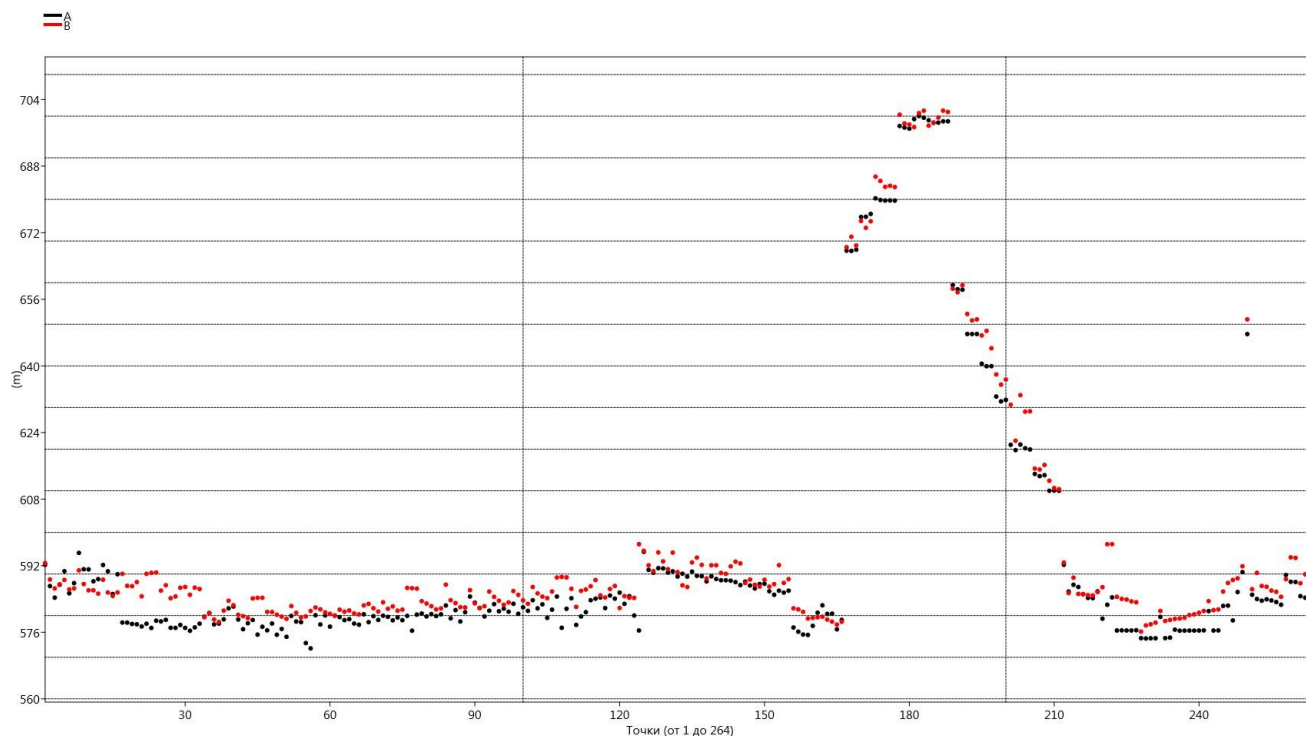
Табл. 3 Базисни статистически параметри на двата набора от данни (височините в DTM и измерените на терена с GNSS)

Basic Statistics		
	DEM _z	GNSS
N	264	264
Min	572.14	576.16
	2	0

³ В.ж. на адрес: http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EL_v3.0rc3.pdf

Max		700.03	701.31
	4		9
Sum		15715	15796
	1.800		7.500
Mean		595.27	598.36
	2		2
Std. error		1.972	1.947
Varian		1026.8	1000.5
nce	33		03
Stand. Dev		32.044	31.631
Media		582.73	586.32
n	4		7
25 prcentil		578.65	582.12
	4		8
75 prcentil		589.70	592.03
	4		1
Skewn		2.217	2.199
ess			
Kurto		3.717	3.587
sis			
Geom. mean		594.47	597.58
	4		7
Coeff. var		5.383	5.286

За да се получи визуална представа за сравняваните стойности, е полезно първо да се генерира графика (Scatter plot), показваща едновременно двата набора от стойности на височините (от DTM и от GNSS) за всяка от 264-те точки (фиг. 2):

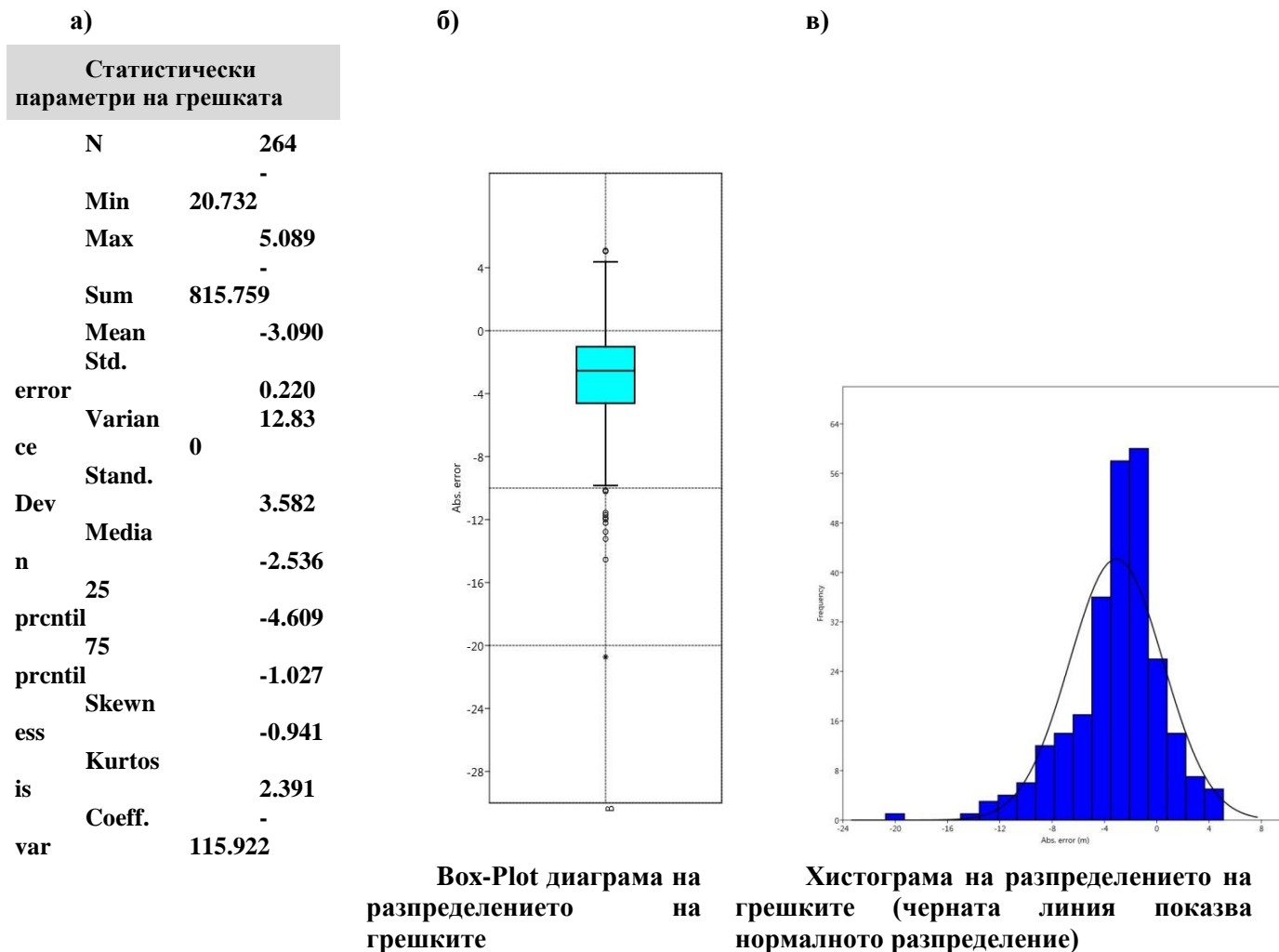


Фиг. 2 Височини на точките по профилните линии, район Земен: по хоризонталната ос – всяка точка от Point_ID_1 до Point_ID_264; по вертикалната ос – н.в. (m); A (черните точки) – височините на точките в DTM (m), B (червените точки) – височините на същите точки, измерени с GNSS (m)

От графиката веднага се забелязва, че в преобладаващата си част черните точки (DTM) са с по-ниски стойности на височините в сравнение с червените точки (GNSS). Това е признак за систематичен компонент в отклоненията на височините в DTM от височините, измерени с GNSS, като при това знакът на тези отклонения е отрицателен.

Възможното обяснение за това, е че височините в DTM са апроксимирани с пиксели, които са площи, а не точкови обекти, т.е. пространствената резолюция на растерната мрежа, както и методът за интерполиране на стойностите в грида, са оказали влияние върху вертикалната точност на DTM.

След това са определени базисните статистически характеристики на стойностите в колона 4 разлики (грешки) в табл. 2, като полученият резултат е показан на фиг. 3, а, б, в.



Фиг. 2 Статистически параметри на грешката, район Земен

Направеното по-горе предположение за систематично по-ниски стойности на височините в DTM се потвърждава от табличните данни, Box-Plot диаграмата и хистограмата на фиг. 2.

Изчисленията показват, че при ниво на доверие $P=95\%$ ($\alpha=0.05$) доверителният интервал (CI) на средната грешка ($ME=3.090$ m) на височините в DTM за района на Земен е в диапазона: -3.522 m < -3.090 m < -2.658 m.

Получената стойност на $(RMSE_z)$ е 4.725 m, а стойността на нейната грешка (E_{RMSEz}) е 0.206 m. При ниво на доверие 95% по формулата $1.9600(RMSE_z) = 1.9600 * (4.725$ m) полученият резултат за вертикалната точност е ± 9.261 m.

От всичките 264 точки, извън рамките на този интервал, попадат 14 точки, или 5.3% от всички точки не отговарят на този стандарт за вертикална точност на DTM. В редица национални и международни стандарти за оценка на точността на пространствените данни е прието процентът на броя на грешките да не надхвърля 10% от всички тествани точки, т.е. полученият процент на тези грешки за района на Земен (5,3%) е в рамките на допустимото. Идентифицираните 14 точки са следните (табл. 4):

Табл. 4 Списък на точките с техните стойности, неотговарящи на стандарта на вертикална точност, район Земен

Точка	Poi nt_ID	Стойност на вертикалната грешка (m)
1	17	-11.703
2	20	-10.137

.	3	22	-11.984
.	4	23	-13.238
.	5	24	-11.568
.	6	30	-9.836
.	7	32	-9.556
.	8	77	-10.205
.	9	10	-12.216
.	8		
0.	1	12	-20.732
0.	4		
1.	1	20	-9.621
1.	1		
2.	1	20	-11.867
2.	3		
3.	1	22	-14.452
3.	1		
4.	1	22	-12.789
4.	2		

Всичките 14 идентифицирани точки са с отрицателен знак на грешките, което илюстрира систематичния компонент (или *bias*) в отклоненията на височините в DTM от приетите стойности за по-точни (или „верни“), измерени на терена с GNSS.

4. Визуализация на получените резултати от оценката грешките и вертикалната точност на генерирания DTM

Практически интерес представлява пространственото разпространение на установените грешки за всяка от тестваните точки. То разкрива локалните географски фактори, влияещи върху несигурността на моделираните височини в дадения район. Освен локалните фактори обаче, върху несигурността на стойностите височините в DTM влияние оказват също: качеството на изходните материали (ЕТК или ортофотоизображенията), използваната интерполация по Хътчинсон, осигуряваща съгласуваност на хидрографската мрежа с релефа, както и билинейната интерполация, приложена за увеличаване на изходната хоризонтална резолюция на растрната мрежа.

За разлика от методите за оценка на грешките и вертикалната точност, няма налични общоприети стандарти за пространствена визуализация на резултатите от тази оценка. Затова авторите пристъпиха към разработване на собствена методология, която да отговаря на две важни изисквания:

- визуализацията трябва да отразява пространственото разпространение на основните компоненти на грешките (вида на грешките, тяхната големина и техния знак);
- тя трябва да е подходяща за комуникация, т.е. по максимално достъпен за възприемане начин да информира за качеството на генерирания DTM.

Съгласно разработената методология грешките за всяка от всичките 264 точки за район Земен са визуализирани с помощта на градириани (пропорционални) точкови символи с помощта на ГИС, по начина, показан на фиг. 3.



Фиг. 3 Пример за визуализиране на пространственото разпространение на грешките в DTM в район Земен

Точковите символи са представени като отделен слой в базата данни, който може да се комбинира с различни по тип пространствени основи. Авторите са предпочели за пространствена основа ортофокартата, тъй като тя най-добре показва конкретната географска обстановка, в която е локализирана всяка тествана точка.

Стойностите на грешките за всяка всяка точка са изобразени с точкови символи в два цвята – с червен цвят за грешките с положителен знак и със син цвят за грешките с отрицателен знак. Диаметърът на точковите символи е определен в зависимост от големината (магнитуда) на грешката. За целта стойностите на грешките са групирани в три интервални класа (голяма, средна и малка стойност на грешката), съответно за положителните и за отрицателните отклонения на стойностите. Групирането на грешките в интервални класове е извършено по метода за класифициране с естествени прекъсвания (Natural Breaks).

В горния ъгъл на изображението, визуализиращо пространственото разпространение на грешките, с помощта на врезка е показано ситуирането на изобразеното пространство и местоположението на напречните профили. В долния ъгъл е поместена легендата, която представя семантиката на точковите символи със споменатите два цвята и диапазоните на грешките от двата вида.

Заклучение

Неотменна част от процедурите при изготвяне на картите на заплахата и риска от наводнения е оценката на качеството на данните. Затова е важно качеството на тази информация да се оцени още преди да бъде използвана. В настоящото изследване вертикалната точност и грешките в DTM са оценени с помощта обективни статистически критерии – средна, стандартна и средна квадратична грешка, стандартно отклонение при ниво на доверие 95-я процентил. Оценката е извършена чрез сравнение на височините между два набора от данни – от генерирания DTM и измерените с GNSS на терена данни по напречните профили на речните течения, като последните са приети за референтни, поради високата им точност. Резултатите показаха, че от всички тествани точки по профилите, 5,3% не отговарят на възприетия стандарт за оценка на вертикална точност, което е в рамките на допустимото. Тези точки са идентифицирани и локализирани в пространството, като отклоненията им от стандарта за вертикална точност са взети предвид при изготвянето на метаданните и калибрирането на модела. Разработената от авторите методология за визуализация на грешките в генерирания DTM позволява ефективно представяне на пространственото им разпространение в изследвания район, което улеснява откриването на точките с „проблемна” вертикална точност. Приложените подходи и методи в изследването могат да бъдат използвани при решаване на аналогични проблеми и в други райони от страната.



Литература:

1. De Bruin, S., Bregt, A., Van de Ven, M. (2001). Assessing fitness for use: the expected value of spatial data sets. *Int. J. Geograph. Inf. Sci.*, 15, 5, 457–475.
2. Frank, A., Grum, E., Vasseur, B. (2004). Procedure to select the best dataset for a task. In: *Proceedings of the Third International Conference on Geographic Information Science (GIScience 2004)*, Adelphi, USA, pp. 81–93.
3. Krok, A., Frank, A. (1999). Optimization of quality of geoinformation products. In: *Proc., 11th Annual Colloquium of the Spatial Information, Research Centre, SIRC'99*, Univ. of Otago, Dunedin, New Zealand, 151–159.
4. Veregin, H. (1999). Data quality parameters. In: *Geographical Information Systems*, P. Longley, M. Goodchild, D. Maguire, D. Rhind (Eds), Wiley, New York, pp. 177–189.
5. Church, R., Curtin, K., Fohl, P., Funk, C., Goodchild, M., Kyriakidis, P., Noronha, V. (1998). Positional distortion in geographic data sets as a barrier to interoperation. In: *Technical Papers, ACSM (American Congress on Surveying and Mapping)*, Bethesda, Md.
6. Devillers, R., Jeansoulin, R. (Eds), (2006). *Fundamentals of Spatial Data Quality*, ISTE, London.
7. Попов, А. (2011). Географски информационни системи: основи на геoinформационното моделиране, ИК „Анубис“.
8. Li, Z. (1991). Effects of check points on the reliability of DTM accuracy estimates obtained from experimental test. In: *Photogramm. Eng. Remote Sens.*, 57, 10, 1333–1340.
9. ICSM (International Community of Surveying & Mapping), *Guidelines for Digital Elevation Data*, August 12, 2008 (www.icsm.gov.au/publications/index.html)