

DARIUSZ GOŚCIEWSKI*

ZASTOSOWANIE KOMPRESJI RLE DO REDUKCJI WIELKOŚCI ZBIORÓW TYPU GRID

APPLICATION OF RLE COMPRESSION FOR SIZE REDUCTION OF GRID TYPE FILES

Streszczenie

Współczesne systemy informacji o terenie (GIS) mogą przechowywać dane o numerycznym modelu terenu w zbiorach wykorzystujących struktury TIN lub GRID. Zastosowanie struktury GRID stwarza dodatkowe możliwości uporządkowania zapisu informacji. Wykorzystywany w tej strukturze zapis danych w postaci regularnej siatki kwadratów może zostać poddany kompresji, co stwarza możliwości redukcji wielkości plików potrzebnych do przechowywania żądanej informacji. W artykule przeanalizowano możliwości zastosowania kompresji typu RLE do redukcji wielkości zbiorów struktury typu GRID przechowujących informacje o numerycznym modelu terenu. Przeanalizowano również, w jaki sposób tego typu kompresja wpływa na dokładność numerycznego modelu terenu.

Słowa kluczowe: numeryczny model terenu, DTM, kompresja danych, struktura GRID

Abstract

Contemporary land information systems (GIS) can store information on digital terrain models in sets using TIN or GRID structure. GRID structure application creates additional possibilities of organizing the information recording. The data recording applied in that structure in the form of a regular grid of squares can be subjected to compression, which allows reducing the size of files needed to store the required information. The paper presents the analysis of RLE type compression application possibilities for reducing GRID type structure sets storing information on the digital terrain model. The influence of that compression type on accuracy of the digital terrain model was analyzed.

Keywords: digital terrain model, DTM, data compression, GRID structure

* Dr inż. Dariusz Gościewski, Instytut Geodezji, Wydział Geodezji i Gospodarki Przestrzennej, Uniwersytet Warmińsko-Mazurski w Olsztynie.

1. Wstęp

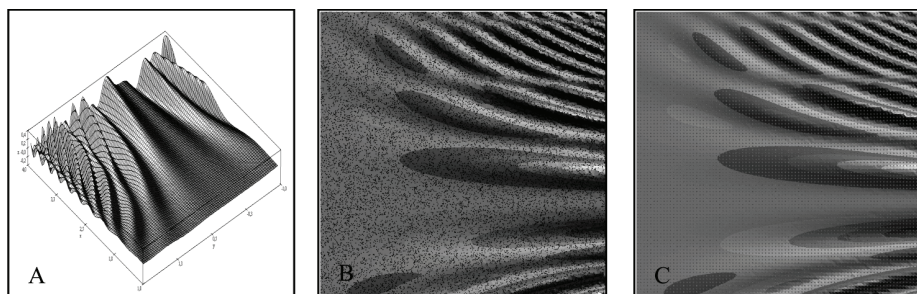
Numeryczny model terenu (DTM) stanowi jedną z podstawowych warstw informacyjnych we współczesnych systemach informacji geograficznej (GIS) [2, 5, 6]. Dane pozwalające na konstrukcję DTM pozyskiwane są coraz częściej dzięki automatycznym pomiarom masowym. Aktualnie, zintegrowane systemy pomiarowe, takie jak skaniny lotniczy (LIDAR), laserowe stacje pomiarowe czy echosondy wielowiązkowe, umożliwiają pozyskiwanie bardzo dużej ilości danych pomiarowych w stosunkowo krótkim czasie [1, 2, 4, 9]. Informacji tego typu z reguły nie można bezpośrednio wykorzystać do budowy DTM, ponieważ ze względu na ich ilość oraz organizację zbiorów pomiarowych wymagają one uprzedniego przetworzenia. Jednym ze sposobów uporządkowania informacji opisującej powierzchnię jest przedstawienie jej w postaci struktury regularnej siatki węzłów typu GRID [3, 7, 10]. Większość współczesnych systemów informacji przestrzennej umożliwia stworzenie i wizualizację numerycznego modelu terenu na bazie regularnej siatki kwadratów. Powstaje ona najczęściej przez zastosowanie interpolacji wartości w punktach węzłowych na podstawie otaczających je punktów pomiarowych [3, 5, 6, 10]. Pozwala to na znaczną redukcję ilości danych i ogranicza ich redundancję. Ponadto zastosowanie odpowiedniej organizacji zapisu informacji w zbiorach typu GRID umożliwia dokonanie kompresji danych. Dzięki takim zabiegom można znacznie ograniczyć wielkość przestrzeni dyskowej potrzebnej do przechowywania informacji o numerycznym modelu terenu [4, 8]. W zależności od stopnia kompresji należy się spodziewać różnej dokładności modelu powierzchni. Warto zatem przeanalizować, w jaki sposób zmniejszenie objętości zbiorów przechowujących informację o DTM wpływa na zmniejszenie dokładności i pogorszenie jakości numerycznego modelu terenu.

2. Tworzenie modelu testowego

Do przeprowadzenia analiz wykorzystano oprogramowanie autorskie. Za pomocą funkcji dwóch zmiennych (1) (rys. 1A) utworzony został teoretyczny model powierzchni wzorcowej. Opierając się na funkcji (1), w podanym przedziale argumentów x i y , wygenerowano w sposób losowy 20 000 punktów pseudopomiarowych (pp), będących odpowiednikami punktów pomiarowych. Następnie cały obszar wzorcowy został przeskalowany i przesunięty tak, że w rezultacie utworzył kwadrat o boku 450 m, a zbiór punktów pp, tworzących powierzchnię wzorcową, zawierał wyłącznie dodatnie wartości współrzędnych x , y , h (rys. 1B). W celu utworzenia struktury typu GRID na wybranym obszarze, opierając się na punktach pp, dokonano interpolacji 8100 węzłów siatki kwadratów o boku bazowym 5 m. Interpolację przeprowadzono, wykorzystując algorytm płaszczyzny i prostej, gdzie wartość w węźle jest wyznaczana jako wynik rozwiązania układu równań: płaszczyzny przechodzącej przez trzy najbliższe punkty pomiarowe otaczające węzeł i prostej prostopadłej do płaszczyzny xy przechodzącej przez ten węzeł [3, 6]. Tak wygenerowany GRID pozwolił na utworzenie modelu powierzchni przedstawionego na rys. 1C.

$$f(x, y) = \sin(x^2 y^2) \sin(5y) x^2 \quad (1)$$

$$x \in <1; 4>, \quad y \in <-1; 2>$$



Rys. 1. Tworzenie numerycznego modelu powierzchni
Fig. 1. Creating the digital terrain model

Na podstawie funkcji (1) zostały również wygenerowane teoretyczne wartości w poszczególnych węzłach siatki kwadratów struktury GRID. Obliczenie różnic pomiędzy wartością praktyczną w węźle (interpolacyjną) a wartością teoretyczną (z funkcji (1)) pozwoliło na obliczenie błędów prawdziwych w poszczególnych punktach, co umożliwiło następnie sporządzenie analiz dokładności. W celu dokonania porównań analizę przeprowadzono również dla 100 000 pp i siatki kwadratów o boku 2 m.

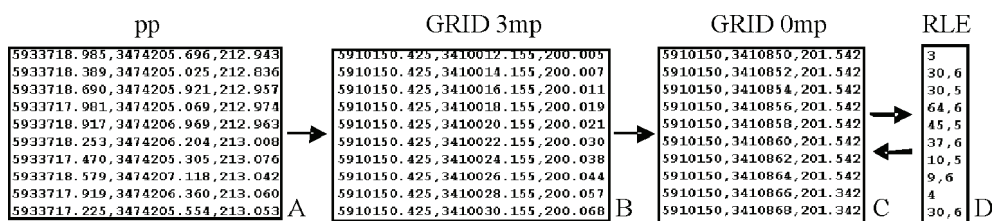
3. Redukcja wielkości zbiorów

Sposób organizacji zapisu danych ma kluczowe znaczenie podczas redukcji wielkości zbiorów. Przygotowanie danych do kompresji wymaga kilku etapów pośrednich. Wszystkich przekształceń danych oraz właściwej kompresji i dekompresji zbiorów dokonano, wykorzystując oprogramowanie autorskie. Zapisanie danych pomiarowych wymaga podania współrzędnych x , y , h z ustaloną dokładnością rejestracji, która jest określona przez urządzenie pomiarowe (rys. 2A). Przekształcenie tego typu danych za pomocą interpolacji do postaci GRID (rys. 2B) pozwala na znaczne ograniczenie ich ilości dzięki odpowiedniemu zorganizowaniu struktury zapisu. Wykres przedstawiony na rys. 3A pokazuje stopień redukcji wielkości pliku (w MB) podczas takiej operacji dochodzący do 24% wartości początkowej. Mając dane w postaci regularnej siatki kwadratów, dodatkową redukcję zbioru można osiągnąć, zmieniając sposób zapisu informacji o położeniu każdego węzła. W tym celu należy znać współrzędne węzła początkowego, rozmiar siatki bazowej struktury GRID oraz sklasyfikować wartość wysokości w węźle do odpowiedniego przedziału (rys. 2C).

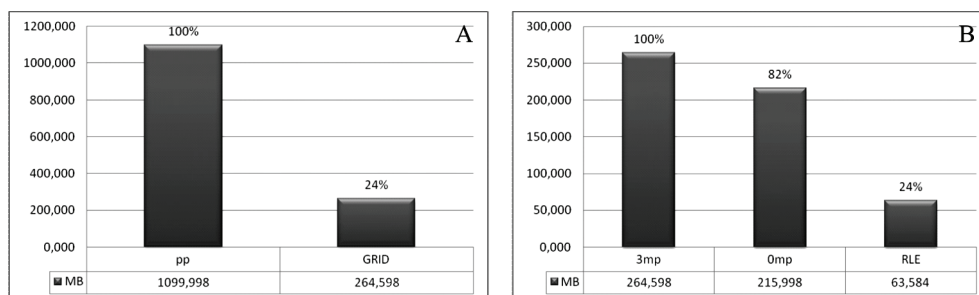
Kompresja typu RLE (*Run Length Encoding*) zakłada zamianę ciągów liczb złożonych z tych samych wartości w pary typu (liczba wystąpień, wartość) [5, 7, 8]. Dokładności pozyskiwania punktów pomiarowych przez urządzenia do skaningu powierzchniowego zależą od techniki pomiarowej i wahają się od $\pm 0,01$ m (laserowe stacje pomiarowe) do $\pm 2,0$ m (echosondy wielowiązkowe). W zależności od ustalonej dokładności pomiaru całość spektrum pomiarowego można podzielić na przedziały odpowiadające tej dokładności.

Po zakwalifikowaniu wartości pomiarowej do odpowiedniego przedziału otrzymuje ona wartość równą wartości środka przedziału. Właściwą kompresję poprzedza również uporządkowanie węzłów (sortowanie) wzdłuż osi X lub Y . Takie działanie umożliwia zapis

danych o punktach węzłowych w postaci numeru wiersza lub kolumny oraz pary liczb (liczba wystąpień, wartość przedziału). Daje to z kolei znaczne ograniczenie objętości zbioru (rys. 2D i 3B – wielkość zbiorów w MB). Dane o wartościach przedziałów przechowywane są w oddzielnym pliku. Odwrotną operacją jest dekompresja, która wykończuje zbiory RLE, informacje o współrzędnych węzła początkowego, rozmiar siatki bazowej struktury GRID i zbiór z danymi o wartościach przedziałów. Po takiej operacji ze zbioru typu RLE (rys. 2D) odtwarzany jest w sposób bezstratny zbiór x, y, h (rys. 2C).

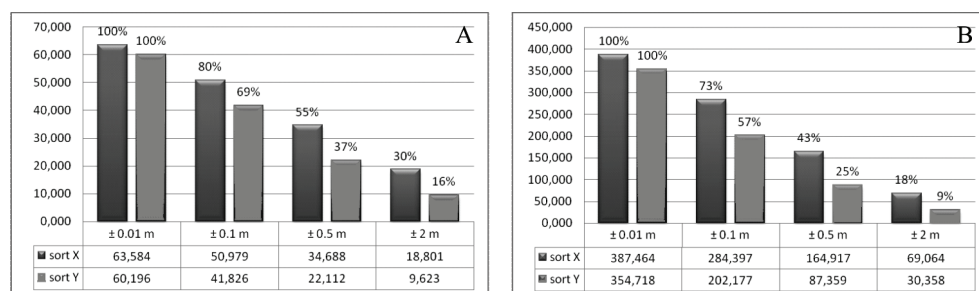


Rys. 2. Redukcja ilości informacji w poszczególnych zbiorach
Fig. 2. Reduction of the volume of information in individual sets



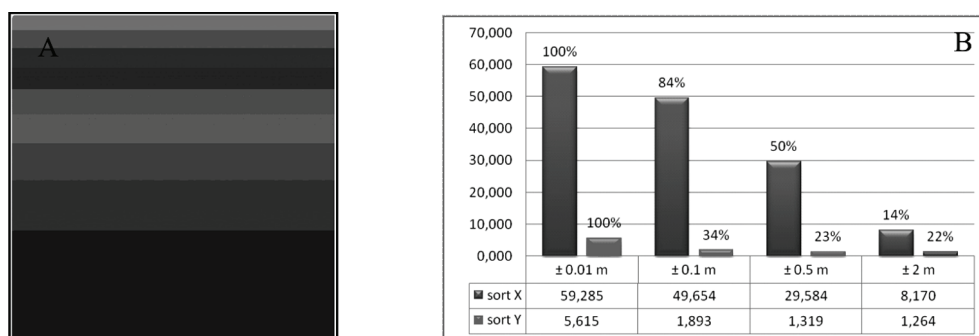
Rys. 3. Redukcja wielkości zbiorów
Fig. 3. Reduction of sets' size

W zależności od ustalonej rozpiętości przedziału można uzyskać odpowiadającą mu dokładność danych oraz założony stopień kompresji i odpowiednią redukcję wielkości zbioru. Na wykresach przedstawionych na rys. 4 porównano wielkości (w MB) plików typu GRID po zastosowaniu różnych stopni kompresji.



Rys. 4. Porównanie wielkości plików typu GRID dla różnych stopni kompresji
Fig. 4. Comparison of GRID type files size for different degrees of compression

W celu porównania w zależności od określonej dokładności pomiaru ustalone zostały 4 przedziały: $\pm 0,01$ m, $\pm 0,1$ m, $\pm 0,5$ m i ± 2 m, dla których dokonano kompresji. Nad wykresami pokazano procent wielkości zbioru po kompresji w stosunku do zbioru bazowego. Wykresy zamieszczone na rys. 4A przedstawiają sytuację dla siatki kwadratów o boku wynoszącym 5 m wygenerowanej na podstawie 20 000 pp, a na rys. 4B dla siatki o boku 2 m wygenerowanej na podstawie 100 000 pp. Na prezentowanych wykresach pokazano także różnice w kompresji dla różnego sposobu organizacji punktów węzłowych w zbiorach (sortowanie wzdłuż X lub Y).



Rys. 5. Porównanie wielkości plików po kompresji dla różnej organizacji punktów węzłowych
Fig. 5. Comparison of files size after compression for different organizations of nodal points

We wszystkich przypadkach zwiększenie rozpiętości przedziału (zmniejszenie założonej dokładności pomiaru) skutkuje zwiększeniem stopnia kompresji, a co za tym idzie – zmniejszeniem wielkości zbioru z danymi. Z porównania wielkości przedstawionych na wykresach A i B (rys. 4) wynika, że większy stopień kompresji uzyskano w przypadku większego zagęszczenia punktów węzłowych (rys. 4B). Ponadto można zauważyć, że stopień kompresji ma związek z ukształtowaniem powierzchni wzorcowej. Większy stopień kompresji występuje dla sortowania względem poziomej osi Y , ponieważ na przekrojach dokonanych wzdłuż tej osi morfologia terenu jest mniej zróżnicowana niż wzdłuż osi X (rys. 1C). Zatem odpowiedni dobór sortowania przed kompresją ma wpływ na jej jakość i pozwala na jeszcze większe ograniczenie wielkości pliku przy zachowaniu tej samej informacji. Skrajny przypadek pokazano na rys. 5, gdzie ukształtowanie powierzchni terenu jest jednostajne, a różnice wysokości zmieniają się tylko w kierunku osi X (rys. 5A). Wykresy zaprezentowane na rys. 5B pokazują bardzo wyraźną różnicę pomiędzy wynikami kompresji poprzedzonej sortowaniem wzdłuż osi X i Y niezależnie od przyjętego przedziału dokładności.

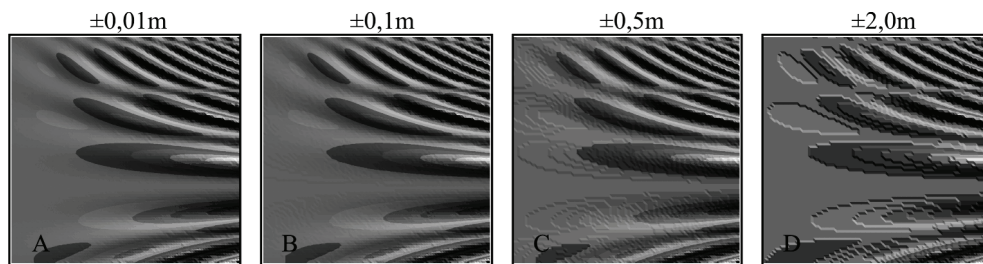
4. Analiza dokładności modelu

Zastosowanie większego stopnia kompresji przez ustalenie większych przedziałów dla analizowanych wartości powoduje zmniejszenie dokładności modelu terenu. Dokładność dopasowania modelu powierzchni po kompresji do modelu teoretycznego może zostać wyrażona liczbowo za pomocą współczynnika RMS

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i, y_i) - z_i)^2} \quad (2)$$

gdzie:

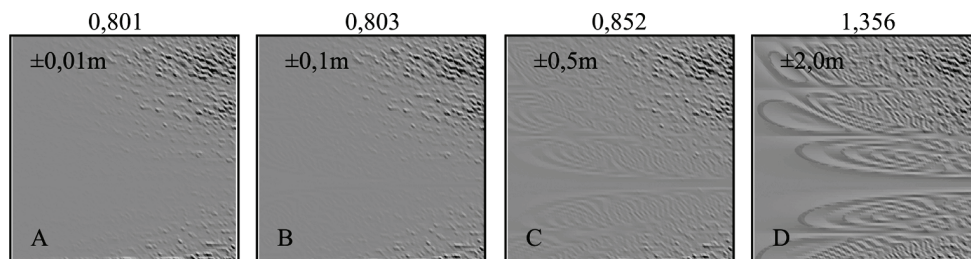
- $f(x, y)$ – wartość funkcji (1) w teoretycznym punkcie węzłowym o współrzędnych x, y ,
- z – wartość obliczona przez dany algorytm interpolacyjny na podstawie punktów pomiarowych w punkcie węzłowym o współrzędnych x, y ,
- n – liczba punktów węzłowych.



Rys. 6. Modele powierzchni dla różnych stopni kompresji
Fig. 6. Surface models for different compression levels

Na rysunku 6 przedstawiono modele badanej powierzchni utworzone z danych, które zostały poddane kompresji. W przypadku kompresji w przedziale $\pm 0,01$ m i $\pm 0,1$ m (rys. 6A i B) jakość modelu nie odbiega od modelu wzorcowego (rys. 1C).

W pierwszym przypadku (rys. 6A) mamy do czynienia z kompresją bezstratną. Dla trzech pozostałych przypadków widoczne są zniekształcenia modelu wynikające z przyjętych przedziałów dokładności (zwłaszcza w przypadku modeli na rys. 6C i D). Należy pamiętać, że pogorszenie jakości modelu nie zależy od kompresji, a wynika głównie z dokładności pomiaru, ponieważ przyjęty stopień kompresji zależy głównie od dokładności pomiaru.

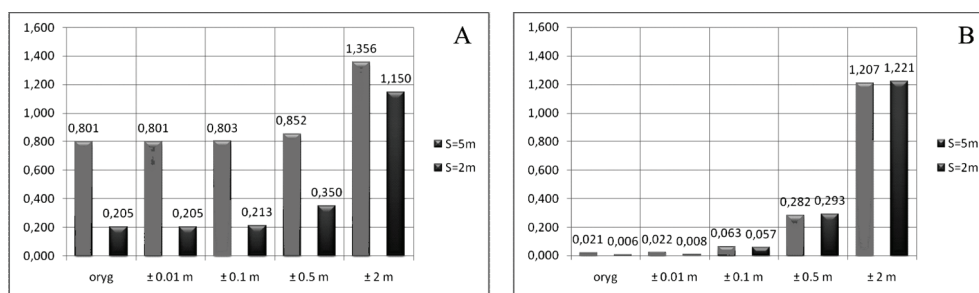


Rys. 7. Rozmieszczenie błędów dla różnych stopni kompresji
Fig. 7. Distribution of errors for different compression levels

Na rysunku 7 za pomocą diagramów różnicowych przedstawiono rozmieszczenie błędów prawdziwych na powierzchni modeli terenu po kompresji. Największe zniekształcenia występują na obszarach najbardziej zróżnicowanych morfologicznie. Nad poszczególnymi diagramami pokazano wartość współczynnika RMS. Również ta analiza pozwala stwierdzić, że różnica pomiędzy modelami A i B (rys. 7) jest niewielka. W przy-

padku modelu B większy współczynnik RMS wskazuje na gorsze dopasowanie powierzchni. Znacznie większe różnice występują dla modeli C i D (rys. 7). Różnice te wyraźnie zobrazowane są przez pasy zniekształceń zlokalizowane na obszarach o większych błędach. W tych przypadkach również współczynniki RMS są większe.

Na rysunku 8A pokazano zbiorcze zestawienie wartości współczynnika RMS dla badanej powierzchni wzorcowej przy różnym zagęszczeniu punktów węzłowych. Na rysunku 8B przedstawiono sytuację dla skrajnego przypadku ukształtowania terenu (zaprezentowanego na rys. 5).



Rys. 8. Porównanie wielkości współczynnika RMS dla różnych stopni kompresji
Fig. 8. Comparison of RMS coefficient values for different compression levels

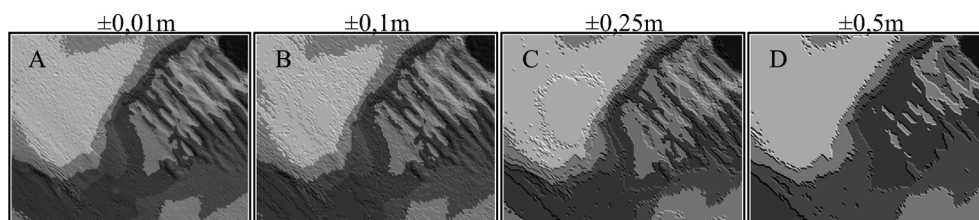
W obu przypadkach porównania dokonano w stosunku do zbioru bez kompresji RLE (oryginalnego) uzyskanego bezpośrednio z interpolacji. Różnice dokładności pomiędzy oryginalnym modelem powierzchni (bez kompresji) a modelem po kompresji dla przedziału $\pm 0,01$ m są niezauważalne (taki sam współczynnik RMS) zarówno dla siatki o boku 5 m, jak i 2 m. Mniejsze wartości współczynników RMS występujące w przypadku siatki o boku 2 m wynikają z lepszego dopasowania struktury GRID o większym zagęszczeniu (rys. 8A). W jednym i w drugim przypadku współczynniki RMS dla przedziałów $\pm 0,01$ m i $\pm 0,1$ m są porównywalne i można stwierdzić, że zastosowanie kompresji nie pogorszyło jakości modeli powierzchni przy jednoczesnym znacznym zredukowaniu wielkości zbiorów przechowujących dane.

5. Zastosowanie kompresji RLE w numerycznym modelu terenu

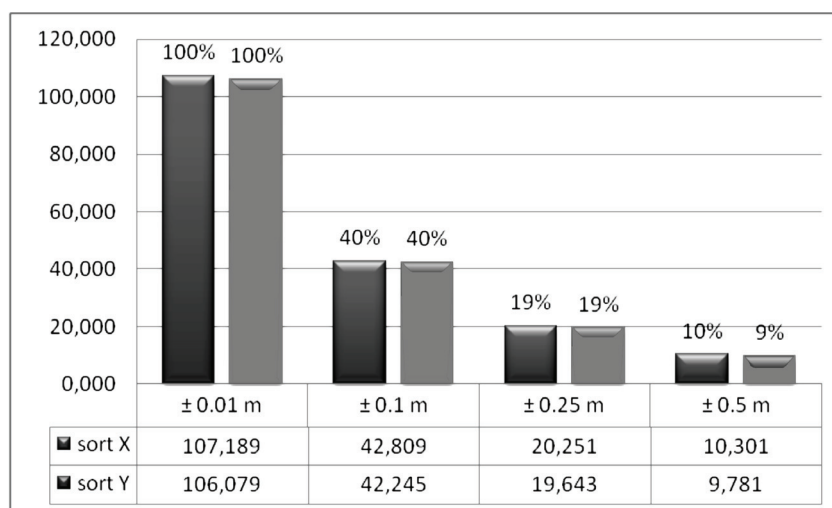
W celu sprawdzenia wyników badań w praktyce poddano kompresji różnego stopnia dane zawierające informacje o ukształtowaniu rzeczywistego terenu. Na rysunku 9 przedstawiono modele terenu po dokonaniu kompresji dla różnie przyjętej dokładności pomiaru (od $\pm 0,01$ m do $\pm 0,5$ m).

Wykresy przedstawione na rys. 10 ilustrują wielkości plików dla kompresji różnych stopni. Przy założeniu mniejszego poziomu dokładności ($\pm 0,1$ m) wielkości zbiorów zredukowane zostały do 40% w stosunku do zbiorów przechowujących dane z dokładnością $\pm 0,01$ m bez wyraźnego pogorszenia jakości modelu (rys. 10A i B).

Należy również zauważyć, że poziom redukcji zbiorów nie zależy w tym przypadku od sposobu organizacji węzłów i jest jednakowy dla obu przypadków sortowania. Wynika to z podobnego ukształtowania terenu wzdłuż profili równoległych do osi X i Y.



Rys. 9. Numeryczny model terenu dla różnych stopni kompresji
Fig. 9. Digital terrain model for different compression levels



Rys. 10. Porównanie wielkości plików DTM dla różnych stopni kompresji
Fig. 10. Comparison of DTM files size for different compression levels

6. Wnioski

Zastosowanie kompresji RLE do organizacji zapisu danych struktury typu GRID pozwoliło na znaczne zmniejszenie wielkości zbiorów przechowujących dane o numerycznym modelu terenu. Przed przystąpieniem do kompresji należy ustalić pożądany poziom dokładności modelu powierzchni i w zależności od potrzeb można wykorzystać kompresję bezstratną lub stratną. Kompresja bezstratna pozwoliła na zmniejszenie wielkości zbioru do 24% wielkości początkowej. W przypadku ustalenia przedziału dokładności dla wartości wysokości na poziomie $\pm 0,01$ m nie spowodowało to pogorszenia dokładności modelu powierzchni, a współczynnik RMS pozostał na takim samym poziomie jak dla zbioru przed kompresją. Zastosowanie kompresji stratnej dodatkowo redukuje wielkość zbiorów, jednak pogarsza jednocześnie dokładność modelu. W tym przypadku należy określić poziom żądanej dokładności modelu, a następnie ustalić stopień kompresji. Przy korzystaniu z algorytmu kompresji RLE należy brać pod uwagę również sposób organizacji zbiorów typu GRID i dostosować porządek sortowania punktów węzłowych do morfologii terenu. Poz-

wala to na dodatkowe zwiększenie stopnia redukcji zbiorów wynikowych. Prawidłowy dobór kompresji umożliwia zachowanie wymaganej dokładności modelu powierzchni przy jednoczesnej znacznej redukcji zbiorów. Daje to wymierne korzyści podczas archiwizacji dużej ilości danych oraz dodatkowe możliwości podczas ich dynamicznego przetwarzania w systemach informacji przestrzennej.

Literatura

- [1] Axelsson P., *DEM generation from laser scanner data using adaptive TIN models*, International Archives of Photogrammetry and Remote Sensing, Amsterdam 2000.
- [2] Burrough P., McDonnell R., *Principles of Geographical Information Systems – Spatial Information Systems and Geostatistic*, Oxford University Press, Oxford, UK 1998.
- [3] Gościewski D., *Tworzenie numerycznego modelu terenu typu GRID z wykorzystaniem kombinacji algorytmów interpolacyjnych*, Prace Naukowe Instytutu Górniczego Politechniki Wrocławskiej, Nr 114 „Geoinformacja dla wszystkich”, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2005.
- [4] Gościewski D., *Influence of measurement points positioning on accuracy of interpolation surfaces*, Research Group of Engineering Geodesy, Vienna University of Technology, Baden, Austria 2006.
- [5] Li Z., Zhu Q., Gold C., *Digital Terrain Modeling – Principles and Methodology*, CRC Press, Boca Raton, FL 2005.
- [6] Maune D., *Digital Elevation Model Technologies and Applications*, The DEM Users Manual, American Society for Photogrammetry and Remote Sensing, Bethesda, MD 2001.
- [7] Oksanen J., Jaakkola O., *Interpolation and Accuracy of Contour-based Raster DEM's*, Reports of the Finnish Geodetic Institute 2000, Finnish Geodetic Institute, Kirkkonummi, Finland 2000.
- [8] Schabenberger O., Gotway C., *Statistical Methods for Spatial Data Analysis*, Chapman and Hall, CRC Boca Raton, FL 2005.
- [9] Wack R., Wimmer A., *Digital terrain models from airborne laser scanner data – a GRID based approach*, International Archives of Photogrammetry and Remote Sensing, Graz 2002.
- [10] Wilson J., Gallant J., *Digital terrain analysis*, Terrain Analysis, Principles and Applications, John Wiley and Sons, New York 2000.