

## **ПРИКЛАДНІ АСПЕКТИ ГЕНЕРАЦІЇ ГІДРОЛОГІЧНО-КОРЕКТНИХ ТА ЕКОЛОГІЧНО-ВІДПОВІДНИХ ЦИФРОВИХ МОДЕЛЕЙ МІСЦЕВОСТІ**

Розглянуто ряд прикладних аспектів генерації цифрових моделей місцевості з використанням кросс-валідаційних процедур, застосування яких мінімізує помилки інтерполяції в автоматичному режимі. Показано приклад застосування такого підходу з використанням вільної ГІС GRASS.

**Ключові слова:** цифрова модель місцевості, кросс-валідація, напружені регуляризовані сплайни, ГІС GRASS.

Рассмотрены прикладные аспекты создания цифровых моделей местности с использованием кросс-валидационных процедур, которые минимизируют ошибки интерполяции в автоматическом режиме. Показан пример применения такого подхода с использованием свободной ГИС GRASS.

**Ключевые слова:** цифровая модель местности, кросс-валидация, напряженные регуляризованные сплайны, ГИС GRASS.

Practical aspects were considered when creating a digital terrain models using cross-validation procedures to minimize the interpolation error in the automatic mode. There was an example of such an approach using free GIS GRASS.

**Keywords:** digital terrain model, cross-validation, regularized splines with tension, GIS GRASS.

**Постановка проблеми.** Використання сучасних ГІС-технологій з метою моделювання, прогнозування розвитку різноманітних ситуацій, їх аналізу і запобігання ризикам повинне базуватися на даних з беззаперечною достовірністю. Основою такого підходу виступають цифрові моделі місцевості (ЦММ). Питання їх точності, детальності та репрезентативності є актуальними і важливими в силу різноманітності напрямків практичного застосування, зокрема прогнозування ерозійних явищ, повеней, управління земельними ресурсами, аналізу та розрахунку екологічних показників, різних видів моніторингу тощо.

Залежно від сфери застосування та характеру наступних дій є певні застороги щодо формування ЦМР. Так, просторова неоднорідність земної поверхні зумовлює виникнення ряду похибок при наземній зйомці навіть за використання сучасних приладів [17]: результуюча модель поряд із своїми, буде акумулювати в собі й похибки зйомки. Попри високий рівень розвитку методів ДЗЗ та їх високу роздільну здатність, часто мають місце систематичні погрешності в контексті інтерпретації деяких особливостей поверхні [2, 18].

Серед широкого класу цифрових моделей місцевості найбільш відомими є растрові (на основі регулярної мережі висотних відміток – моделі GRID) та векторні (на базі нерегулярної мережі трикутників – моделі TIN). До загальновідомих проблем останніх належить так званий ефект «терас», що унеможливорює їх використання як основи для моделей вищого рівня (ерозійних, екологічних тощо), які вимагають точного опису перерозподілу потоків речовини та енергії. Виходячи з останньої вимоги, можливість побудови гідрологічно-коректної а, отже, й екологічно-відповідної цифрової моделі місцевості можуть надати растрові моделі. При цьому важливо і необхідно враховувати кілька речей, що найчастіше залишаються поза увагою дослідників, а саме: питання вибору правильної роздільної здатності моделі та алгоритму її генерації. Зокрема, хибний

вибір кроку дискретизації моделі зумовлює втрату значної частини інформації про досліджуваний об'єкт і, відповідно, подальші аналіз та моделювання в будь-яких прикладних додатках, зокрема екологічних, буде супроводжуватись значними систематичними похибками [6].

Серед існуючих методів генерації ЦММ, які використовуються в комплексі наук про Землю, найбільш популярними згідно [13] є 42, поділених на 3 групи: не-геостатистичні, геостатистичні та комбіновані. За такого різноманіття методів, які можуть бути використані при цифровому моделюванні рельєфу, не всі надають можливість отримати гідрологічно-коректну й екологічно-відповідну ЦММ. Причинами цього є як недоліки самих інтерполяційних алгоритмів, так і методичні помилки дослідників, зумовлені, зокрема, недостатнім досвідом та низьким рівнем математичної підготовки [7]. Із розширенням доступності геоінформаційних методів і підходів для широкого кола наукових досліджень вплив останніх факторів зростає. Водночас інтерполяційні алгоритми контролюються значними наборами параметрів, які дозволяють оптимізувати математичні функції для отримання відповідної щодо поставлених критеріїв поверхні. Ці параметри не завжди є добре документовані, а тому отримані ЦММ із використанням значень параметрів за замовчуванням часто є неякісними [7]. Отже, оптимізація та автоматизація процесу вибору значень параметризації моделі – актуальне і необхідне завдання.

**Мета роботи** – дослідження можливості автоматизованого підбору параметрів інтерполяційних функцій для отримання гідрологічно-коректних та екологічно-відповідних ЦММ.

**Об'єкт досліджень** – процес отримання цифрової моделі місцевості із заданими характеристиками; **предмет дослідження** – розгляд алгоритму отримання ЦММ в частині автоматизованого підбору параметрів інтерполяційних функцій та перевірка якості отриманої моделі на основі статистичних показників.

**Методика досліджень.** Дослідження проводилися з використанням інструментальних засобів ГІС GRASS 6.4 [3,12] у середовищі Debian GNU Linux 7.0 [10] із дотриманням умов Загальної громадської ліцензії GNU GPL [11] щодо вживання цих програмних засобів. Для тестової ділянки м. Чернівці (1x1 км) складної геоморфологічної будови в якості топографічної основи була обрана частина карти горизонталей М 1:2000 з перерізом 1 м. Для оцифрування картографічних даних застосували векторизатор Easy Trace 7.99 [5], згладжування горизонталей провели модулем GRASS v.generalize (алгоритм chaiken, число ітерацій n=3). Для генерації ЦММ обрали регуляризовані напружені сплайни [14] реалізовані у модулі GRASS v.surf.rst [1, 16].

**Виклад основного матеріалу.** Із численних методів створення ЦММ, представлених у ГІС GRASS, ми використали метод регуляризованих напружених сплайнів (Regularized Spline with Tension - RST), який обчислює значення висот у вузлах сітки за допомогою функції, що моделює тонку гнучку пластину, яка проходить через/або близько до точок вихідних даних [8,9,14]. Наразі це найточніший метод у ГІС GRASS, хоча підбір параметрів для отримання найкращого результату на різнорідних наборах даних є окремим інженерним завданням [16].

До переваг RST-інтерполяції належить її гнучкість в рамках однієї радіальної базисної функції, на відміну від часто суб'єктивного вибору відповідної варіограми у крігінгу. Для належного врахування мінливості модельованого явища (у т. ч. рельєфу), функція може бути налаштована рядом параметрів, з яких до числа найбільш впливових належать напруженість та згладженість. Їх підбір емпіричним шляхом вимагає великого досвіду та знань модельованого явища. Проте існують методи їх автоматичного вибору шляхом пошуку мінімальної помилки, який здійснюється процедурою кросс-валідації – перехресної перевірки [8,15]. Її суть полягає у видаленні однієї точки початкових даних та виконанні апроксимації для її місцезонашування, користуючись даними, що залишилися. При цьому обчислюється різниця між фактичним і апроксимованим значенням. Потім

видалена точка повертається і вилучається наступна і це повторюється для кожної точки вихідних даних. Надалі ефективність роботи інтерполятора обчислюється як середньоквадратична помилка RMSE (Root Mean Squared Error):

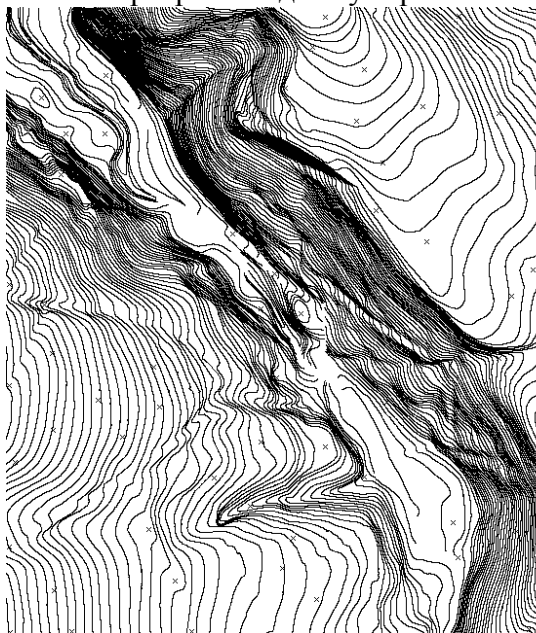
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i^* - a_i)^2}, \text{ де} \quad (1)$$

$a_i^*$  є розрахованим значенням в заданій точці  $i$ , а  $a_i$  є дійсним значенням висоти. Низькі значення RMSE свідчать про більш надійні оцінки в області між точками даних. Отже, кросс-валідація може бути використана для пошуку оптимальних параметрів інтерполяції шляхом мінімізації RMSE. Виходячи з великого досвіду, запропонований для вирішення такого роду задач скрипт на bash [16], розробником якого є J.Hofierka, на відмінних від модельних наборах даних не працює. Аналіз його структури засвідчив, що він призначений для проведення кросс-валідації лише на невеликих наборах даних, обмежених кількістю точок  $n=700$ . Тому нами було дещо видозмінено секцію, яка відповідає власне за розрахунки:

до: `v.surf.rst -c input=$INMAP cvdev=data_cv_$TNSSMTH zcolumn=$ZCOL tension="$i" smooth=0."$j" segmax=700 --o`

після: `v.surf.rst -c input=$INMAP cvdev=data_cv_chvr_$TNSSMTH zcolumn=$ZCOL tension="$i" smooth=0."$j" dmin=70 dmax=100 npmin=120 segmax=25 --o`

з метою можливості проведення кросс-валідації на матеріалі довільного обсягу. У нашому дослідженні для апробації методики з усього пулу відвекторизованих горизонталей (рис. 1а) згідно видозмінених параметрів скрипта було вибрано 5693 базові точки (рис. 1б), які з високою достовірністю репрезентували масив даних та надали можливість проводити розрахунки в розумні терміни (процедура перебору всіх можливих варіантів (всього 150 можливих комбінацій) тривала і збільшення числа точок в арифметичній прогресії подовжує тривалість розрахунків в геометричній).



а) карта горизонталей тестової ділянки

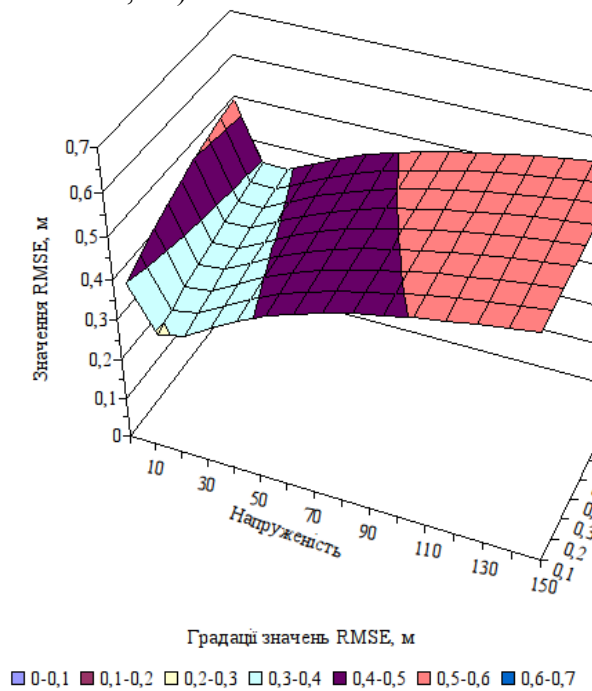


б) розташування апробаційних точок

Рис. 1. Вихідна векторна основа для процедури кросс-валідації

Залежність величини RMSE від різних значень параметрів напруженості сплайну  $\varphi$  та його згладженості  $\omega$  ( $\varphi \in \langle 10, 150 \rangle$ ,  $\omega \in \langle 0.1, 1.0 \rangle$  з кроками дискретизації 10 та 0,1 відповідно) представлена на рис. 2. Отримані результати показують, що найменша середньоквадратична помилка для дослідного набору даних притаманна парі значень

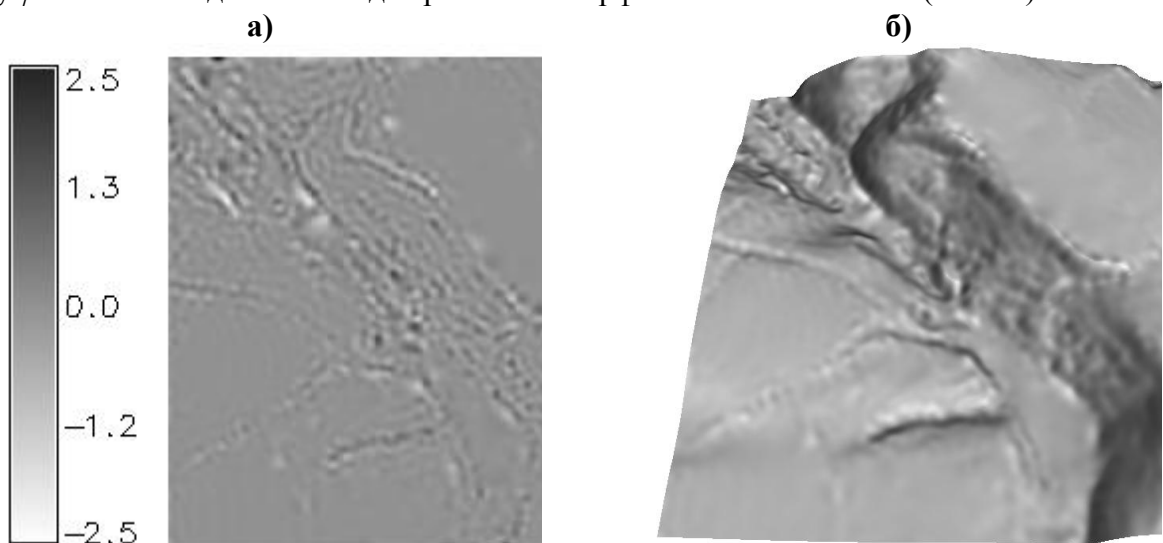
$\varphi=20$  та  $\omega=0,1$  і становить 0,292 м, що є цілком задовільним для ЦММ заданої точності і є меншою, ніж при встановленні параметрів за замовчуванням, при яких її значення становить 0,346 м (різниця склала 18,5%).



**Рис. 2. Залежність величини RMSE від градацій факторів напруженості та згладженості сплайну**

Аналіз просторового розподілу стандартного відхилення в отриманій моделі показує, що найбільші різниці притаманні областям із високою варіабельністю z-координати. Це вказує на те, що для таких ділянок, які чітко візуалізуються як у 2d, так і 3d форматах (рис. 3), доцільним є вибір дещо інших параметрів інтерполювання. Інструментальні засоби ГІС GRASS дозволяють проводити такий аналіз із хорологічними дефініціями щодо кожного параметру окремо: це окреслює напрямки подальших досліджень такого роду.

Порівняння статистичних показників варіантів із найнижчою та найвищою RMSE показує, що навіть за використання оптимальних значень параметризації сплайну є точки окремих локальних екстремумів, обчислені висоти в яких значно відрізняються від реальних: це лише підтверджує необхідність вибору різних величин напруженості сплайну  $\varphi$  та його згладженості  $\omega$  для різних геоморфологічних елементів (табл. 1).



**а) розподіл у 2d; б) приуроченість екстремумів девіації до форм рельєфу у 3d**

**Рис. 3. Просторовий розподіл стандартного відхилення після кросс-валідації**

Таблиця 1

**Порівняння статистичних характеристик векторних карт відхилень модельної ділянки із найкращими та найгіршими значеннями параметрів  $\phi$  та  $\omega$**

Показник	$\phi=20; \omega=0,1$	$\phi=150; \omega=0,9$
число точок	n=5693	
Min	-3.58178	-4.01959
Max	5.96057	6.31467
Діапазон	9.54235	10.3343
стандартне відхилення сукупності	0.514251	0.592672
Дисперсія	0.264454	0.35126
коефіцієнт варіації	219.702	254.86
Екссес	11.9827	10.7414
Асиметрія	0.741917	0.747703
RMSE	0,292	0,346

**Висновки.** Отже, підбір параметрів моделювання з метою генерації гідрологічно-коректних та екологічно-відповідних ЦММ засобами ГІС, зокрема GRASS, є можливим і доступним процесом для широкого кола фахівців. Процедура їх автоматичного вибору шляхом пошуку мінімальної квадратичної помилки при використанні алгоритму кросс-валідації (перехресної перевірки) дозволяє поліпшити якість ЦММ і тим самим сприяє отриманню достовірніших результатів при моделюванні інших видів, зокрема гідрологічного, ерозійного, екологічного тощо. Регуляризовані напружені сплайни забезпечують створення фізичної моделі рельєфу шляхом зміни еластичних властивостей інтерполяційної функції. Відповідно, їх використання для відтворення фізичних явищ, які є наслідком процесів мінімізації енергії, на прикладі ландшафту з його балансом між силою гравітації, взаємодією ґрунтів та клімату є цілком виправданим та рекомендованим у сфері наук про Землю.

### Література

1. Дмитрук Ю.М. Окремі аспекти організації території на основі виокремлення потоково-орієнтованих структур із використанням ГІС GRASS / Ю.М.Дмитрук, В.Р.Черлінка // Землеустрій і кадастр. – 2011. – №3. – С. 57-64.
2. Дмитрук Ю.М. Передискретизація даних Aster GDEM 2 для потреб картування ґрунтів / Ю.М.Дмитрук, В.Р.Черлінка // Біологічні системи. – 2012. – №4. – Вип. 2. – С. 157-161.
3. Черлінка В.Р. Особливості та актуальність використання системи підтримки аналізу географічних ресурсів (GRASS) / В.Р.Черлінка, Ю.М.Дмитрук // Ученые записки Таврійського національного університету ім. В.И.Вернадського. – Серія: Географія. – 2011. – Т.24 (63). – №1. – С. 3-7.
4. Debian GNU Linux – the universal operating system [Електронний ресурс]. – Режим доступу: <http://www.debian.org/index.en.html>
5. Easy Trace 7.99 [Електронний ресурс]. – Режим доступу: [http://www.easytrace.com/site2/program/et799\\_ru](http://www.easytrace.com/site2/program/et799_ru)
6. Hengl T. Mathematical and Digital Models of the Land Surface / T.Hengl, I.S.Evans // Geomorphometry: Geomorphometry: Concepts, Software, Applications. Developments in Soil Science (ed T.Hengl, H.I.Reuter). – Vol. 33. – Elsevier, 2009. – pp. 31-64.
7. Hengl T. Finding the right pixel size / T.Hengl // Computers & Geosciences. – 2006. – №32. – pp. 1283-1298.
8. Hofierka J. Multivariate interpolation of precipitation using regularized spline with tension / J.Hofierka, J.Parajka, H.Mitasova, L.Mitas // Transactions in GIS. – 2002. – №6. – pp. 135-150.

9. Hofierka J. Interpolation of radioactivity data using regularized spline with tension / J.Hofierka // Applied GIS. – 2005. – Vol. 1. – №2. – pp. 16/01-16/13.
10. Hofierka J. Optimisation of Interpolation Parameters Using Cross-validation / J.Hofierka, T.Cebecauer, M.Šúr // Digital Terrain Modelling. Lecture Notes in Geoinformation and Cartography (ed. R.J.Peckham, J.Gyozo). –Berlin: Springer-Verlag Heidelberg, 2007. – pp. 67-82.
11. GNU GENERAL PUBLIC LICENSE [Электронный ресурс]. – Режим доступа: <http://www.gnu.org/licenses/gpl.html>
12. Geographic Resources Analysis Support System [Электронный ресурс]. – Режим доступа: <http://grass.fbk.eu/>
13. Li J. A Review of Spatial Interpolation Methods for Environmental Scientists / J.Li, A.D.Heap. – Canberra: Geoscience Australia, 2008. - Record 23. – 137 pp.
14. Mitašova H. Interpolation by Regularized Spline with Tension: I. Theory and Implementation / H.Mitašova, L.Mitaš // Mathematical Geology. – 1993. – Vol. 25. – №.6. – pp. 641-655.
15. Mitasova H. Modeling spatially and temporally distributed phenomena: New methods and tools for GRASS GIS / H.Mitasova, L.Mitas, W.M.Brown, D.P.Gerdes, I.Kosinovsky // International Journal of GIS.- 2001 – Vol. 9. – pp. 443-446.
16. Neteler M. Open Source GIS: a GRASS GIS approach (3rd edition) / M.Neteler, H.Mitasova. – New York: Springer, 2008. – 406 p.
17. Oksanen J. Digital elevation model error / Oksanen J. – Helsinki: Helsinki University Press, 2006. – 59 p.
18. Remote Sensing of Urban and Suburban Areas / Ed. Tarek Rashed, Carsten Jürgens // Series: Remote sensing and digital image processing. – Vol. 10. – London: Springer Dordrecht Heidelberg, 2010. – 352 p.

*Поступила в редакцию 8 травня 2012 р.*