ISSN 2078-6441. Вісник Львівського університету. Серія географічна. 2018. Випуск 52. С. 201–209.
Visnyk of the Lviv University. Series Geography. 2018. Issue 52. P. 201–209.
http://dx.doi.org/10.30970/vgg.2018.52.10185

УДК 519.226:528.28

ОЦІНКА РІВНЯ НАДІЙНОСТІ ЛІНІЙНОГО ЗВ'ЯЗКУ МІЖ ПРОСТОРОВО РОЗПОДІЛЕНИМИ АВТОКОРЕЛЬОВАНИМИ ЗМІННИМИ

Олександр Мкртчян

Львівський національний університет імені Івана Франка,
вул. П. Дорошенка, 41, 79007, м. Львів, Україна,
e-mail: alemkrt@gmail.com

Розглянуто питання оцінювання надійності зв'язків між змінними, які демонструють просторову автокореляцію (подібність значень характеристики між сусідніми місцеположеннями). Зв'язки між змінними описують двома відмінними, хоч і пов'язаними між собою, характеристиками: силою та надійністю (значимістю). Зокрема, коефіцієнт кореляції (у різних формах) описує силу залежності, проте сам по собі не може характеризувати її значимість. Натомість значимість характеризує *p*-рівень, який виражає ймовірність того, що зв'язок між змінними, виявлений у вибірці, є лише випадковою особливістю цієї вибірки. Цей показник залежить як від сили зв'язку, так і від кількості незалежних один від одного пунктів (об'єктів) спостережень. Проте геопросторовим даним, зазвичай, властива автокореляція, тому такі дані загалом не задовольняють вимогу незалежності, і кількість ступенів вільності в них є суттєво меншою від кількості пунктів спостережень. Особливо це стосується аналізу залежностей між змінними, зображеними у формі растрових шарів.

Для коректного аналізу таких залежностей ми створили функцію (програму) мовою R, яка ґрунтується на методі, запропонованому Кліффордом та ін. [6]. Функція як вхідні дані використовує растрові шари та на підставі аналізу автокореляційної структури растрових шарів обчислює ефективний розмір вибірки *M*, на основі якого відбувається розрахунок дисперсії коефіцієнта кореляції *r*, його *p*-рівень (рівень надійності) та довірчі інтервали. Роботу цієї функції протестовано на наявних геопросторових даних для території в Українських Карпатах: обчислено зв'язки між морфометричними характеристиками рельєфу – абсолютними перевищеннями й похилами поверхні, з одного боку, та індексом NDVI, – з іншого.

Ключові слова: автокореляція, статистична значимість, *R*, просторові зв'язки.

Дослідження в географії та екології часто передбачають виявлення й аналіз зв'язків між різними процесами та явищами. Методи математичної статистики мають важливе значення в дослідженнях природних і природно-антропогенних систем як інструмент кількісного формалізованого опису зв'язків між їхніми характеристиками та сформованих цими зв'язками структур. Одним із завдань, які вирішують ці методи, є перевірка гіпотез про наявність зв'язків між окремими показниками (характеристиками). З огляду на тип залежних і незалежних змінних (кількісні, номінальні, порядкові) та їхній розподіл для такої перевірки використовують певні параметричні або непараметричні критерії.

Недостатня обізнаність дослідників у галузях природничих наук з принципами та методиками статистичного аналізу даних часто не дає змоги використовувати їх належно. Найчастіше застосовують найпростіші в обчисленні показники, або такі,

обчислення яких вбудоване в поширені офісні й геоінформаційні програмні продукти. Однак такі показники не завжди адекватні меті, яку ставить дослідник. Будь-яка залежність між змінними має дві різні, хоч і пов'язані між собою, властивості: силу (величину) та надійність (значимість). Такий часто використовуваний для характеристики залежності між змінними показник, як коефіцієнт кореляції (у різних формах) описує силу залежності, тобто відображає, якою мірою варіабельність однієї змінної визначає (пояснює) варіабельність іншої, якщо між цими змінними існує об'єктивний зв'язок. Однак часто першим завданням дослідження є з'ясування питання про саму наявність зв'язку між змінними. Лише після того, як отримано позитивну відповідь, можна переходити до характеристики сили та форми цього зв'язку. Проте коефіцієнт кореляції сам по собі не дає змоги визначити надійність зв'язку; для цього використовують *p*-рівень, який виражає ймовірність того, що зв'язок між змінними, виявлений у вибірці, є лише випадковою особливістю цієї вибірки. Чим менше значення *p*-рівня, тим більша надійність залежності. Традиційно в природничо-наукових дослідженнях прийнятною межею статистичної значимості вважають 0,05; якщо значення *p*-рівня є меншим, то ймовірність помилкового виявлення залежності менша, ніж 5 %.

Наприклад, якщо йдеться про найпростіший випадок двох кількісних змінних, пов'язаних лінійною залежністю, то для визначення *p*-рівня можна використати статистику

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \qquad (1)$$

де *r* – коефіцієнт кореляції Пірсона; *n* – кількість незалежних пар спостережень; *n* – 2 – кількість ступенів вільності (2 – це кількість незалежних параметрів моделі лінійної регресії) [11]. Ця статистика підпорядкована розподілу Стьюдента, що дає змогу використати її для обчислення рівнів надійності зв'язку (автоматично або порівнюючи з табличними значеннями для відповідної кількості ступенів вільності).

Як бачимо, рівень надійності зв'язку залежить не лише від коефіцієнта кореляції, а й від кількості ступенів вільності, що насамперед визначене кількістю даних, на підставі яких досліджують залежність. Одне й те ж значення коефіцієнта кореляції може бути й достатнім, і недостатнім для того, щоб засвідчити зв'язок, залежно від того, за якою кількістю даних його обчислено. І чим слабшим є зв'язок між змінними, тим більший обсяг даних потрібний, щоб його надійно засвідчити.

Обчислення рівня надійності зв'язку за наведеною формулою ґрунтується на низці припущень, чи не найважливішим з яких є припущення про незалежність пар даних між собою. Адже кожна така пара є окремим джерелом інформації, яку беруть до уваги. Якщо йдеться про контрольований експеримент, то цю незалежність забезпечують, відповідно підбираючи піддослідні об'єкти (організми, особи) та/або унеможливлюючи взаємодію між ними протягом експерименту. Якщо ж ідеться про аналіз даних, зібраних з натури (спостереження, вимірювання природних об'єктів, аналіз відібраних зразків), то забезпечити таку незалежність значно важче. У разі, коли властивості одних об'єктів спостережень залежать від властивостей інших об'єктів цієї ж вибірки, реальна кількість ступенів вільності виявляється меншою, а отже, обчислене значення рівня надійності є завищеним. У цьому випадку зростає ризик "виявлення" зв'язку між змінними, який насправді є випадковим збігом обставин.

Ця проблема особливо актуальна для геопросторових (просторово розподілених) даних, які збирають в окремих пунктах спостережень (збору зразків). Поширеною властивістю геопросторових даних є наявність у них просторової автокореляції, яку визначають як кореляцію значень деякої величини між парами точок, відокремлених між собою деякою відстанню [7]. На відміну від звичайної кореляції, яку обчислюють між двома (або більше) змінними, автокореляція може бути притаманна просторовому розподілу однієї змінної. Принцип автокореляції, по суті, є наслідком дії першого закону географії. Цей закон, сформульований 1970 р. американським географом Вальдо Тоблером, стверджує, що все взаємопов'язане, проте ближчі речі пов'язані між собою сильніше, ніж віддаленіші [12]. Значення і потенціал цього закону та його висновків для формування міцного теоретичного фундаменту географічної науки досі належно не розкриті, проте цей закон є в основі низки методів формалізованого географічного аналізу. Наприклад, на цьому законі ґрунтується теорія регіоналізованої змінної, яка характеризує континуальне просторове поле як суму трьох компонентів: детермінованого структурного компонента (тренда), стохастичного автокорельованого компонента та випадкового некорельованого компонента (шуму). Практичним застосуванням цієї теорії є методика геостатистичної інтерполяції (крігінгу), яку ефективно застосовують у картуванні різноманітних континуально розподілених характеристик геосистем за значеннями, виміряними в окремих пунктах земної поверхні. Зокрема, у наших дослідженнях описано можливості цієї методики для складання карт полів розподілу кліматичних характеристик за їхніми значеннями, обчисленими для окремих метеостанцій [1–3].

Завдяки явищу автокореляції можна створювати точні й надійні карти розподілу кількісних характеристик по земній поверхні. Проте поряд з позитивним це явище має для дослідників і негативний момент. Автокорельовані дані загалом не задовольняють вимоги незалежності, і кількість ступенів вільності в них суттєво зменшується. Особливо це стосується аналізу залежностей між змінними у формі растрових шарів. Кількість пікселів у растрових шарах може становити мільйони, створюючи ілюзію великої кількості даних та, відповідно, ступенів вільності (n у формулі (1)). Оскільки ж сусідні піксели, зазвичай, містять подібні значення характеристики, то реальна кількість просторової інформації в растровому шарі є суттєво меншою, ніж кількість пікселів. Автокореляція в растровому шарі залежить від типу шару: для цифрових моделей рельєфу (ЦМР) та їхніх похідних шарів вона, здебільшого, є вищою, ніж для дистанційних зображень, представлених значеннями відбивної здатності земної поверхні в різних спектральних діапазонах. Також вона залежить від характеру території та її просторової структури: чим складніший рельєф і чим більша мозаїчність наземного покриву, тим менша автокореляція у відповідних растрових шарах.

За умов використання хибного значення кількості ступенів вільності, виникає великий ризик зробити неправильний висновок про наявність зв'язку між змінними, коли об'єктивно такого зв'язку нема. У термінах математичної статистики це означає великий ризик помилки першого роду в ході перевірки гіпотези щодо зв'язку між автокорельованими характеристиками [7].

В дослідженні [8] з'ясовано, що під час аналізу зв'язків між двома змінними зазначений ефект простежується в разі, коли автокореляція властива двом змінним. Якщо ж хоча б в одній зі змінних автокореляції нема, то проблема зникає: автокореляція в другій змінній не впливає на розподіл коефіцієнта кореляції і класичні тести на значимість кореляції є коректними [8]. У разі множинної регресії, коли досліджують сукупний

вплив кількох чинників на сумарну характеристику, фактор автокореляції в даних не лише впливає на оцінку надійності статистичного зв'язку, а й може призвести до хибної оцінки відносної ролі і значимості чинників. Значимість чинників, що мають високу просторову автокореляцію, у такому разі виявиться завищеною порівняно з тими, яким властива висока просторова варіабельність. Дж. Леннон назвав це явище червоним зміщенням у географічній екології за аналогією з відомим явищем червоного зміщення в астрономії [8].

Для врахування просторової автокореляції в тестуванні статистичної значимості оцінок параметрів є низка вирішень. Наприклад, метод, запропонований Кліффордом та ін. (1989), виявив добрі результати у випадку позитивної автокореляції у великих вибірках (саме такі властивості часто мають растрові шари) [6]. Розглянемо коротко зміст цього методу.

На початку всі можливі комбінації пар місцеположень розбивають на шари (лаги) так, щоб значення коваріації в межах кожного шару можна було вважати сталими. Для стаціонарних процесів у праці [6] як оцінку автоковаріації однієї змінної *X* для деякого лагу запропоновано використовувати вираз

$$\hat{C}_x(k) = \sum_{S_k}(X_\alpha - \overline{X})(X_\beta - \overline{X}) \,/\, N_k \,,$$

де S_k – шар, що відповідає лагу *k*, N_k – кількість пар місцеположень у шарі S_k. Цю ж формулу використовують для оцінки автоковаріації $\hat{C}_y(k)$ у змінній *Y*. Умовна дисперсія вибіркової коваріації тоді становить $N^{-2}\sum_k N_k \hat{C}_x(k)\hat{C}_y(k)$, де *N* – загальна кількість точок спостережень, з чого випливає такий вираз для оцінки дисперсії коефіцієнта кореляції *r*:

$$\hat{\sigma}_r^2 = \frac{\sum_k N_k \hat{C}_x(k)\hat{C}_y(k)}{N^2 S_x^2 S_y^2},$$

де S_x^2, S_y^2 – дисперсія змінних *X* та *Y*. Звідси ефективний розмір вибірки $M = 1 + \frac{1}{\hat{\sigma}_r^2}$. Значення M, обчислене за цією формулою, підставляють замість *n* у разі оцінки статистичної значимості та довірчих інтервалів для коефіцієнта кореляції.

Растрові шари, зазвичай, складаються з мільйонів пікселів, і в разі використання стандартних методів та алгоритмів статистичного аналізу кожний з них інтерпретується як незалежна одиниця (точка) даних. Проте такі дані майже завжди демонструють певний ступінь автокореляції, яка особливо сильна у випадку гладких поверхонь, з поступовою зміною значень характеристики від піксела до піксела (як-от топографічна поверхня).

Більшість програмних продуктів ГІС, що містять функції аналізу растрових даних, дають змогу серед іншого обчислювати статистичні зв'язки між растровими шарами. Наприклад, в ArcMap 10 відповідна функція міститься у групі інструментів просторового аналізу (Spatial Analyst Tools – Multivariate – Band Collection Statistics). Якщо задати як вхід набір растрових шарів, то на виході отримаємо обчислені коваріаційну та кореляційну матриці. Кореляційна матриця містить коефіцієнти кореляції між окремими растровими шарами. Проте ця функція не дає змоги обчислити рівень надійності (*p*-рівень) зв'язку, а отже, – визначити, наскільки зв'язок між відображеними характеристиками є статистично обґрунтованим. Вільне програмне забезпечення SAGA дає змогу виконувати лінійний регресійний аналіз растрових шарів завдяки інструменту Multiple Regression Analysis (Grid and Predictor Grids). Цей інструмент, окрім коефіцієнта кореляції та параметрів регресійної моделі, також дає змогу обчислити рівень надійності

(*p*-рівень) зв'язку. Однак рівень надійності обчислюють на підставі припущення про незалежність значень в окремих пікселах растру (нема автокореляції). Оскільки це майже ніколи не відповідає реальній ситуації, то визначені цим інструментом рівні значимості є суттєво завищеними, тобто в багатьох випадках можуть спонукати дослідника зробити висновок про наявність зв'язку між змінними в ситуації, коли реально цього зв'язку нема. Отже, ця можливість у застосуванні до географічного аналізу швидше шкодить, уводячи в оману, ніж дає користь. Статистичний аналіз растрових даних можливий, якщо експортувати растри в табличні формати з подальшим аналізом у відповідних програмних продуктах (наприклад, у програмі Statistica або інструментом Аналіз даних програми Excel). Проте в цих випадках маємо аналогічну ситуацію: обчислення рівнів значимості зв'язку, що не бере до уваги просторову автокореляцію в даних, зазвичай, продукує некоректний результат.

Наша мета – розробка методики, яка б визначала рівень надійності лінійного зв'язку між просторово розподіленими змінними з урахуванням наявності автокореляції в їхньому розподілі.

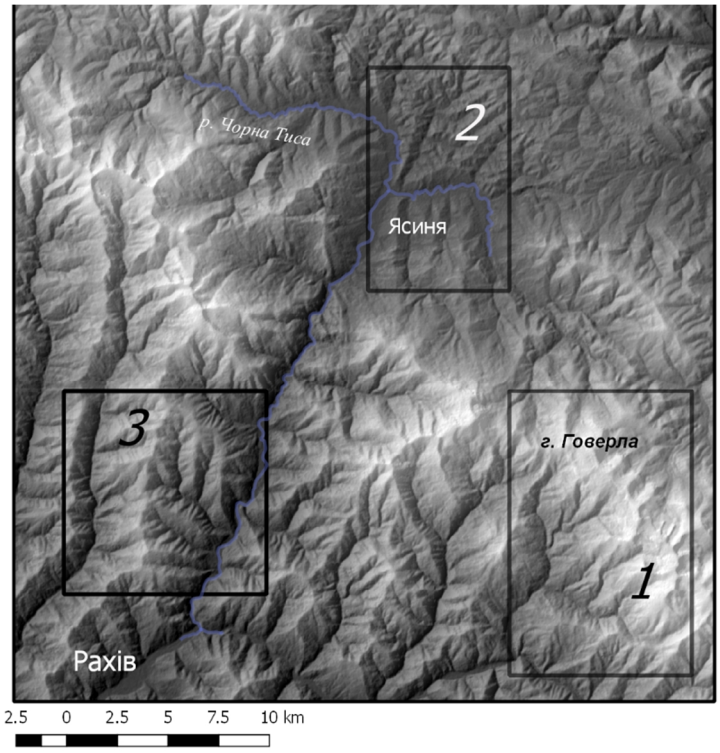
Ми створили спеціалізовану функцію мовою R – поширеною мовою програмування для статистичних обчислень та аналізу даних, яка є стандартом у міжнародній спільноті фахівців у галузі статистики і яку широко використовують у розробках статистичних програм та аналізі даних. Створена функція (алгоритм) ґрунтується на методі, запропонованому в праці [6]. Як вхідні дані використано растрові шари у форматі RasterLayer, які можна створити конвертацією інших растрових форматів засобами R (пакет raster). Функція на підставі аналізу автокореляційної структури растрових шарів обчислює ефективний розмір вибірки M, на основі якого відбувається розрахунок дисперсії коефіцієнта кореляції *r,* його *p*-рівня (рівня надійності) та довірчих інтервалів. Довірчі інтервали для коефіцієнта кореляції можна оцінити, використавши трансформацію Фішера: $F(r) = \frac{1}{2} ln \frac{1+r}{1-r} = arctanh\,(r)$, що приблизно задовольняє нормальний розподіл зі стандартним відхиленням $SE(F) = \frac{1}{\sqrt{n-3}}$.

Отже, на виході функції отримуємо значення коефіцієнта кореляції Пірсона, параметрів моделі лінійної регресії, рівня надійності (*p*-рівня) зв'язку між змінними та довірчих інтервалів. Довірчі інтервали обчислюють для двох рівнів значимості: p = 0,05 (традиційний граничний рівень значимості, який застосовують у природничих науках) та p = 0,001 (жорсткіший рівень, що дає змогу робити більш надійні висновки про зв'язки між змінними).

Роботу цієї функції протестовано на основі наявних геопросторових даних для території в Українських Карпатах (басейн р. Чорна Тиса та прилеглі до нього площі, Закарпатська обл.) Ці дані охоплювали цифрову модель рельєфу (ЦМР) SRTM з просторовою роздільністю 1 кутова секунда ~ 30 м; похідний від неї растровий шар похилів поверхні (з такою ж просторовою роздільністю); оброблений космознімок Landsat 5 TM, що безпосередньо відображає відбивну здатність земної поверхні [4] (знятий сенсором Landsat 5 ETM+, сцена 15 липня 2009 р.), на основі якого обчислено індекс NDVI (різниця значень відбивної здатності в інфрачервоному та червоному діапазонах спектра, поділена на суму цих значень у тих же діапазонах). Цей індекс характеризує інтенсивність фотосинтезу, біологічну продуктивність екосистем, ступінь розвинутості й стан

наземної фітомаси [3]. Отже, об'єктом аналізу були три растрові шари (перевищень, похилів поверхні та значень індексу NDVI) з просторовою роздільністю 30 м, прив'язані до спільної системи координат UTM (зона 35).

Обчислення виконували окремо для цілого басейну (загальний розмір растрів – 1152×1397 пікселів) та для трьох менших за розмірами ділянок у його межах. Перша ділянка (300×467 пікселів) розташована в найвищій частині басейну й охоплює субальпійську зону з полонинними ландшафтами, друга (233×367 пікселів) – у щільно заселеній, освоєній та антропогенно перетвореній ділянці Ясинської улоговини, третя (233×333 пікселів) – у низькогірній південно-західній частині території, переважно зайнятій схилами, що вкриті листяними та мішаними лісами (див. рисунок).



Розміщення території досліджень та окремих ділянок у її межах
Location of the territory of research and particular areas within its boundaries

Предмет аналізу – зв'язки між морфометричними характеристиками рельєфу: абсолютними перевищеннями й похилами поверхні, з одного боку, та індексом NDVI, – з іншого. У таблиці наведено основні результати аналізу, виконаного за допомогою описаної вище функції.

Як бачимо з таблиці, статистично значимий додатний зв'язок між перевищеннями та індексом NDVI зафіксовано лише в межах ділянки 2, де людські поселення, посіви й елементи інфраструктури (що мають низькі значення NDVI) природно тяжіють до розташованих нижче місцеположень (долинно-терасових комплексів річок), тоді як

розташовані вище схили вкриті лісами. Щодо зв'язків між похилами та NDVI, то статистично значимий зв'язок виявився характерним для цілої території, а також (окремо) для ділянок 2 і 3. Цікаво, що значення ефективного розміру вибірки виявились вищими тоді, коли однією зі змінних були похили поверхні, ніж тоді, коли нею були значення перевищень. Пояснити це можна тим, що поверхня абсолютних висот є гладшою (має сильнішу просторову автокореляцію), ніж похили поверхні, які можуть сильно варіювати на менших просторових масштабах.

Результати аналізу зв'язків між просторово розподіленими автокорельованими змінними
Results of the analysis of relationship between spatially distributed autocorrelated variables

Показник	**Ціла територія**	**Ділянки**		
		1	**2**	**3**
Перевищення~NDVI				
r	-0.0678	-0.126	0.4294	-0.0571
M	99.7	54.18	38.34	43.98
p	0.748	0.818	0.0034	0.3565
Похили~NDVI				
r	0.1599	-0.02	0.3168	0.3520
M	213.53	121.04	93.21	102.45
p	0.0097	0.587	0.00098	0.00014

П р и м і т к а : *r* – коефіцієнт кореляції, *M* – ефективний розмір вибірки, *p* – рівень надійності зв'язку.

Отже, сучасні комп'ютерні технології, методи й алгоритми дають змогу ефективно досліджувати зв'язки між різними характеристиками геосистем. Наявність просторової автокореляції в геопросторових даних створює додаткові проблеми, зокрема, у разі визначення надійності зв'язків між змінними. Вирішення цієї проблеми у випадку дослідження лінійного зв'язку між двома змінними можливе завдяки розробленому нам алгоритму. Перспективи подальших досліджень полягають у вдосконаленні цього алгоритму для аналізування зв'язків між більшою кількістю змінних (множинна регресія), інкорпорації складніших методів статистичного аналізу (дисперсійний і дискримінантний аналіз, логістична регресія тощо).

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Мкртчян О., Шубер П.* Інтерполяція даних метеоспостережень кількостей опадів та інших кліматичних змінних методом регресійного криґінґу // Вісник Львів. ун-ту. Сер. геогр. 2013. Вип. 42. С. 258–264.
2. *Мкртчян О. С.* Картування полів кліматичних характеристик із застосуванням відкритих даних та геостатистичної інтерполяції // Національне картографування: стан, проблеми та перспективи розвитку: Збірник матеріалів VII Всеукр. наук.-практ. конференції "Картографічні твори у пізнанні та розвитку регіонів" / Відп. за вип. О. Ю. Дегтяр. Київ : ДНВП "Картографія", 2016. Вип. 7. С. 71–74.
3. *Мкртчян О. С.* Цифровий аналіз мультиспектрального космознімка Landsat 7 ETM+ ділянки центральної частини Українських Карпат // Вісник Львів. ун-ту. Сер. геогр. 2013. Вип. 41. С. 205–212.

4. *Мкртчян О. С., Свідзінська Д. В.* Дослідження змін наземного покриву басейну Чорної Тиси та його факторів на основі аналізу космознімків // Вісник Одес. ун-ту. Сер.: геогр. та геол. науки. 2015. Т. 20, вип. 4. С. 60–73.
5. *Bahrenberg G., Giese E., Nipper J.* Statistische Methoden in der Geographie. 2. Multivariate Statistik. Stuttgart, 1992. 415 p.
6. *Clifford P., Richardson S., Hemon D.* Assessing the significance of the correlation between two spatial processes // Biometrics. 1989. N 45. P. 123–134.
7. *Legendre P.* Spatial autocorrelation: trouble or new paradigm? // Ecology. 1993. N 74. P. 1659–1673.
8. *Lennon J. J.* Red-shifts and red herrings in geographical ecology // Ecography. 2000. N 23. P. 101–113.
9. *Masek J. G., Vermote E. F., Saleous N., Wolfe R., Hall F. G., Huemmrich F., Gao F., Kutler J., Lim T. K.* A Landsat surface reflectance data set for North America, 1990–2000 // IEEE Geoscience and Remote Sensing Letters. 2006. Is. 3. P. 68–72.
10. *Mkrtchian A.* Annual precipitation data processing and interpolation for the weather stations of Western Ukraine // Proceedings of GeoMLA, Geostatistics and Machine Learning, Application in Climate and Environmental Sciences (21–24 June 2016), Belgrade, Serbia. Belgrade : Faculty of Civil Engineering, University, 2016. P. 61–66. URL: http://geomla.grf.bg.ac.rs/site_media/static/Proceedings%20of%20GeoMla%202016%20Conference.pdf
11. StatSoft, Inc. Электронный учебник по статистике. Москва : StatSoft, 2012. URL: http://www.statsoft.ru/home/textbook/default.htm.
12. *Tobler W.* A computer movie simulating urban growth in the Detroit region // Economic Geography. 1970. Vol. 46(2). P. 234–240.

REFERENCES

1. Mkrtchian, O., & Shuber, P. (2013). Interpolation of meteodata on precipitation and other climatic variables by regression-kriging. *Visnyk of the Lviv University. Series Geography, 42,* 258–264 (in Ukrainian).
2. Mkrtchian, O. (2016). Kartuvannia poliv klimatychnykh kharakterystyk iz zastosuvanniam vidkrytykh danykh ta heostatystychnoi interpoliatsii. *Natsionalne kartohrafuvannia: stan, problemy ta perspektyvy rozvytku, 7.* Kyiv: DNVP "Kartohrafiia", 71–74 (in Ukrainian).
3. Mkrtchian, O. (2013). Digital analysis of multispectral space imagery Landsat 7 ETM+ of the area in central part of Ukrainian Carpathians. *Visnyk of the Lviv University. Series Geography, 41,* 205–212 (in Ukrainian).
4. Mkrtchian, O., & Svidzinska, D. (2015). The study of land cover changes in Black Tisza River basin and its factors based on the analysis of spatial imagery. *Visnyk of the Odesa University. Series Geographical and Geological Sciences, 20*(4), 60–73 (in Ukrainian).
5. Bahrenberg, G., Giese, E., & Nipper, J. (1992). *Statistische Methoden in der Geographie, 2. Multivariate Statistik.* Stuttgart, 415 p.
6. Clifford, P., Richardson, S., & Hemon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics, 45,* 123–134.
7. Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology, 74,* 1659–1673.
8. Lennon, J. J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography, 23,* 101–113.

9. Masek, J. G., Vermote, E. F., Saleous, N., Wolfe, R., Hall, F. G., Huemmrich, F., Gao, F., Kutler, J., & Lim, T. K. (2006). A Landsat surface reflectance data set for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters, 3,* 68–72.
10. Mkrtchian, A. (2016). Annual precipitation data processing and interpolation for the weather stations of Western Ukraine. Proceedings from *GeoMLA, Geostatistics and Machine Learning, Application in Climate and Environmental Sciences* (21–24 June 2016), Belgrade, Serbia. Belgrade: Faculty of Civil Engineering, University, 61–66. Retrieved from http://geomla.grf.bg.ac.rs/site_media/static/Proceedings%20of%20GeoMla%202016%20Conference.pdf
11. StatSoft, Inc. (2012). *Jelektronnyj uchebnik po statistike.* Moscow: StatSoft. Retrieved from http://www.statsoft.ru/home/textbook/default.htm
12. Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(2), 234–240.

Стаття: надійшла до редакції 22.05.2017
доопрацьована 17.01.2018
прийнята до друку 06.02.2018

ASSESSMENT OF RELIABILITY OF LINEAR RELATIONSHIP BETWEEN SPATIALLY DISTRIBUTED AUTOCORRELATED VARIABLES

Olexandr Mkrtchian

Ivan Franko National University of Lviv,
P. Doroshenko St., 41, UA – 79007 Lviv, Ukraine,
e-mail: alemkrt@gmail.com

The paper deals with the problem of the assessment of the reliability of the relationship between variables when these variables show a level of spatial autocorrelation (similarity of values between near locations). Relationships between characteristics are being characterized by two different, while similar properties: the strength and the reliability. The commonly used correlation coefficient measures the relationship's strength and cannot determine by its own whether the relationship is statistically significant. In return, statistical significance is measured with p-level, which shows the probability that the relationship revealed in the sample is non-existent in population.

Significance level depends not only on the strength of the relationship but also on the degrees of freedom. Its calculation is based on the assumption of the independent samples. Yet it is often not the case when spatial data are being analyzed. These data are usually characterized by spatial autocorrelation, which leads to highly inflated degrees of freedom and overestimated significance levels.

The method to deal with this problem is proposed, based on (Clifford, Richardson, Hemon, 1989). It has been implemented with R function that takes as an input raster layers and by analyzing their autocorrelation structure calculates the effective sample size. On the basis of the latter the variance of correlation coefficient r, its p-level and confidence intervals are calculated as outputs of the function.

The created function has been tested by analyzing relationships between terrain parameters (elevation and slope) and land cover characteristic (NDVI index) with data gathered for a test area in the Ukrainian Carpathians and three its subareas.

Key words: autocorrelation, statistical significance, *R*, spatial relationships.