

Путренко В.В.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Пашинська Н.М.

Київський національний університет імені Тараса Шевченка

РЕГІОНАЛЬНИЙ МОНІТОРИНГ СТАНУ РОСЛИННОСТІ З ВИКОРИСТАННЯМ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

У статті викладено методуку дослідження змін вегетаційного покриву за сукупністю показників, які оцінюють вегетацію з використанням методів просторового інтелектуального аналізу даних. Використання результатів дистанційного зондування Землі для розрахунку вегетаційних індексів, які характеризують динаміку зміни стану рослинного покриву, надає дані для сумісного аналізу і детекції трендів у зміні кліматичних умов та стану рослинності з використанням методів просторової кластеризації. Для розрахунків були використані дані європейської програми Copernicus. У результаті було розроблено модель сумісної обробки даних вегетаційних індексів із застосуванням підходу сумісної обробки даних різної природи. Під час обробки даних використовується геоінформаційна технологія зонування, яка дає змогу обрахувати статистичні показники для окремих адміністративних одиниць – районів та міст обласного підпорядкування. Це дає змогу в подальшому використовувати результати аналізу для прийняття управлінських рішень. На наступному кроці було застосовано аналіз кластеризації гарячих-холодних точок, що дало змогу отримати статистично значущі кластери зміни річної динаміки стану вегетаційного покриву. Проведено аналіз і розроблено методуку регіонального моніторингу стану рослинного покриву на основі сумісного використання різних характеристик вегетаційних індексів та методуку просторової кластеризації. Розроблена методика може використовуватися для виявлення просторово-часових трендів у зміні стану вегетації, що може бути враховано в умовах регіональних змін клімату та ведення сільського господарства. Застосування методів просторової кластеризації дає змогу отримати математично обґрунтовані оцінки на регіональному рівні, для яких можуть обраховуватися статистичні значення величини вегетаційних індексів.

Ключові слова: *вегетаційний індекс, рослинний покрив, ГС, просторовий інтелектуальний аналіз даних, кластеризація.*

Постановка проблеми. Сучасні космічні системи розвиваються швидкими темпами в напрямі удосконалення знімального обладнання, зростання його роздільної здатності, урізноманітнення знімальних технологій, розширення спектральних можливостей, часового та просторового охоплення території. Найбільшими постачальниками даних ДЗЗ у вільному доступі є геологічна служба США та Європейське космічне агентство, яке протягом останніх років значно зміцнило групування космічних супутників ДЗЗ та розгорнуло програму Copernicus, в межах якої надаються безплатні дані космічної зйомки та похідні тематичні продукти, що можуть використовуватися під час аналізу стану рослинного покриву, кліматичних змін та забезпеченості водними ресурсами [1–3].

Copernicus є європейською системою для моніторингу Землі [4]. Дані зібрані з різних джерел,

зокрема супутників спостереження Землі та датчиків на місці. Дані обробляються і забезпечують достовірну та актуальну інформацію в шести тематичних областях: земний покрив, морські акваторії, атмосфера, зміна клімату, управління в надзвичайних ситуаціях та безпека.

Глобальний сервіс земного покриття є складовою частиною сервісу земного покриву Copernicus, який забезпечує низку біогеофізичних продуктів про стан і еволюцію земної поверхні у глобальному масштабі із середнім та низьким просторовим розрізненням. Продукція використовується для моніторингу рослинності, водного циклу і енергетичного балансу.

У зв'язку з цим постає актуальне питання використання тематичних продуктів програми Copernicus для вивчення динаміки регіональних кліматичних змін та природних загроз на основі

вивчення динамічної та структурної складової частини оцінок стану рослинного покриву для території України. Ці продукти отримуються на основі використання супутникових даних системи Proba-V. [Copernicus Global Land Service, 2017].

Дослідження довготривалих трендів у зміні стану вегетації дає змогу дійти висновків про регіональні та локальні флуктуації зміни клімату та прогнозні розрахунки щодо адаптації сільського господарства.

Аналіз останніх досліджень і публікацій. Розробка і дослідження вегетаційних індексів як основи для прогнозування врожайності і умов ведення сільського господарства були розпочаті в 70-ті рр. минулого століття. Бурхливий розвиток цієї тематики пов'язаний із розвитком дистанційного зондування Землі та появою великої кількості даних різного ступеня обробки. Дослідженню зв'язків між врожайністю та вегетаційними індексами присвячені роботи Boken et al., 2002, Bolton et al., 2013, проведенню інструментальних наземних досліджень із метою оцінки вегетаційних індексів – роботи Broge et al., 2002; Beckschäfer et al., 2014; Canisius et al., 2010. Визначення трендів із використанням вегетаційних індексів розглянуто в роботах Martinez-Casasnovas et al., 2005, Cicek et al., 2010, Pettorelli et al., 2005. У вітчизняних дослідженнях опрацювання даних вегетаційних індексів щодо сільського господарства розглянуто в таких працях: Кохан, 2011; Лялько та ін., 2006; Мозговой та ін., 2009; Станкевич та ін., 2011. Використання вегетаційних індексів як показника пожежної небезпеки рослинного покриву в умовах кліматичних змін розглянуто у таких роботах: Путренко та ін., 2017, Pashynska et al., 2016. Таким чином, дослідження вказаної проблематики залишається одним із перспективних напрямів.

Постановка завдання. Метою роботи є визначення методів змін вегетаційного покриву за сукупністю показників, що оцінюють вегетацію. Завданнями дослідження є визначення перемінних, які характеризують вегетацію, збір даних для проведення дослідження, розробка методики сумісної оцінки показників вегетації, кластеризація усереднених значень за адміністративними одиницями за методом «гарячих точок».

Виклад основного матеріалу дослідження. Сервіси земного покриву Copernicus дають змогу провести багатостороннє дослідження змін сезонної вегетації із використанням низки вегетаційних індексів. До найбільш значущих індексів можна зарахувати індекси NDVI, LAI, VPI, VCI.

NDVI – нормалізований відносний індекс рослинності – простий кількісний показник кількості фотосинтетично активної біомаси (зазвичай має назву «вегетаційний індекс»). Це один із найпоширеніших і використовуваних індексів для вирішення завдань, що застосовують кількісні оцінки рослинного покриву.

Обчислюється за такою формулою:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

де NIR – відображення в ближній інфрачервоній області спектра, RED – відображення в червоній області спектра.

Відповідно до цієї формули, щільність рослинності (NDVI) в певній точці зображення дорівнює різниці інтенсивностей відбитого світла в червоному і інфрачервоному діапазоні, поділеній на суму їхніх інтенсивностей.

Низькі значення NDVI вказують на поганий стан рослинного покриву, що може бути спричинено посухою і, відповідно, може призвести до підвищеної пожежної небезпеки. У цьому разі важливо зберігати відповідний масштаб та роздільну здатність даних, які мають бути більш узагальненими, ніж у разі звичайного сільськогосподарського моніторингу.

LAI (Leaf Area Index) – Індекс листової площі, який визначається як половина загальної площі зелених елементів купола на одиницю горизонтальної ділянки. Значення, отримане із супутника, відповідає загальному зеленому LAI всіх шарів пологу лісу, в тому числі підліску, які можуть становити дуже важливий внесок для лісів. Практично LAI вимірює товщину рослинного покриву. LAI визнається однією з найважливіших кліматичних змінних (ECV) у рамках Глобальної системи спостережень за кліматом (GCOS).

VPI (Vegetation Productivity Index) – Індекс продуктивності рослинності – оцінює загальний стан рослинності шляхом посилення на поточне значення NDVI порівняно з довгостроковою статистикою за той самий період. VPI є процентиллю ранжування поточного значення NDVI порівняно зі своїм історичним діапазоном мінливості: значення 0%, 50% і 100% відповідно свідчать, що поточне спостереження співвідноситься з історичним мінімумом (найгірший стан рослинності), медіаною (нормальний) або максимумом (найкраща ситуація), що коли-небудь спостерігалися.

VCI (Vegetation Condition Index) – Індекс стану рослинності – порівнює поточний NDVI в діапазоні значень, які спостерігалися в той самий

період у попередні роки. VCI виражається у відсотках і дає уявлення про те, де спостережуване значення знаходиться між крайніми значеннями (мінімум і максимум) у попередні роки. Нижчі і вищі значення вказують, відповідно, на погані і гарні умови вегетаційного стану.

Основними методологічними засадами розробки динамічних моделей є визначення усереднених показників за певний час спостережень, щодо яких відбувається аналіз. Чим більшою буде низка спостережень, тим якіснішими можуть бути отримані результати. З іншої сторони, довший ряд спостережень дає змогу спрогнозувати глобальні та регіональні тренди з високим ступенем вірогідності. Сукупність індексів оцінки вегетації не дає змоги вибрати оптимальний показник, який мав би найбільшу кореляцію з регіональними змінами клімату. Тому вибір оптимальних показників залишається складною науковою задачею. Проте підхід до сумісного аналізу даних різної природи може допомогти у вирішенні цього завдання. З цією метою необхідно вирішити базові завдання зонування, нормалізації, зважування та сумісної обробки отриманих рядів статистичних даних.

Для розрахунку інтегрального індексу за показниками вегетаційних індексів пропонується формула, яка дає змогу поєднати усі різниці міжсезонної зміни індексів для визначання регіональних трендів:

$$I = \frac{\sum_{t=1}^n (NDVI_t - NDVI_{t-1}) / N \sum_{t=1}^n (LAI_t - LAI_{t-1}) / L \sum_{t=1}^n (VCI_t - VCI_{t-1}) \sum_{t=1}^n (VPI_t - VPI_{t-1})}{n}$$

де I – інтегральний індекс; N – кількість продуктів, які беруть участь у розрахунку; N , L , S , P – кількість результатів класифікації знімків за індексами.

Для проведення експериментальних розрахунків були використані дані вегетаційних індексів за червень 2017 – червень 2018 рр. для південно-східної території України, яка водночас є регіоном високо розвинутого землеробства з високим експортним потенціалом, а також значною частиною розташована в зоні ризикованого землеробства.

Загальний процес геоінформаційної обробки даних у середовищі ArcGIS 10.4 включає завантаження відповідних індексів із порталу, маскування території з використанням кордону України, обчислення попіксельної різниці між даними за різний період. На наступному етапі отримуємо зональну статистику для кожної адміністративної одиниці окремо. Дані в такому вигляді простіше нормалізуються та можуть бути використані безпосеред-

ньо у процесі підтримки прийняття рішень місцевими органами влади.

Розрахунок інтегрального індексу демонструє наявність виражених регіональних змін, які прослідковуються навіть для річної динаміки даних. Чітко видно, що показники вегетаційних індексів для сходу країни погіршилися, тоді як для півдня та центральної частини України вони дещо покращилися.

З метою перевірки гіпотези регіоналізації було використано метод гарячих точок, який дає змогу визначити просторові кластери стійких позитивних або негативних змін.

Ключова ідея полягає в тому, що значення в середині нормального розподілу становлять очікуваний результат. Коли абсолютне значення z -оцінки є великим, а ймовірності є маленькими (у хвостах нормального розподілу), спостерігається незвичний розподіл. Для методу «Аналіз гарячих точок» це означає статистично істотну «гарячу» або «холодну» точку (Ord et al., 1995).

Метод «гарячі точки» розраховує статистичний показник для кожної події в наборі даних. Підсумкові p -значення (ймовірності) та z -оцінки (стандартні відхилення) говорять про те, в якій області простору кластеризуються події з високими або низькими значеннями [6]. Метод працює шляхом аналізу кожної події в контексті сусідніх за географією подій. Щоб бути статистично суттєвою гарячою точкою, подія повинна мати високе значення і бути оточена іншими подіями з також високими значеннями. Локальна сума для події та її сусідів порівнюється пропорційно із сумою всіх подій; якщо локальна сума відрізняється від очікуваної локальної суми і ця відмінність є занадто великою, щоб бути результатом випадкового процесу, отримується статистично значуща z -оцінка.

Метод «гарячих точок» використовує формули:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1}}}, \quad \bar{X} = \frac{\sum_{j=1}^n x_j}{n},$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2},$$

де x_j – атрибутивне значення для подій j , $w_{i,j}$ – просторова вага між подіями i та j , n – загальна кількість подій, \bar{X} – середнє арифметичне значень подій, S – дисперсія.

Статистична величина G^* для кожної події в наборі даних є z -оцінкою. Щодо статистично значущих позитивних z -оцінок, чим вищою є z -оцінка,

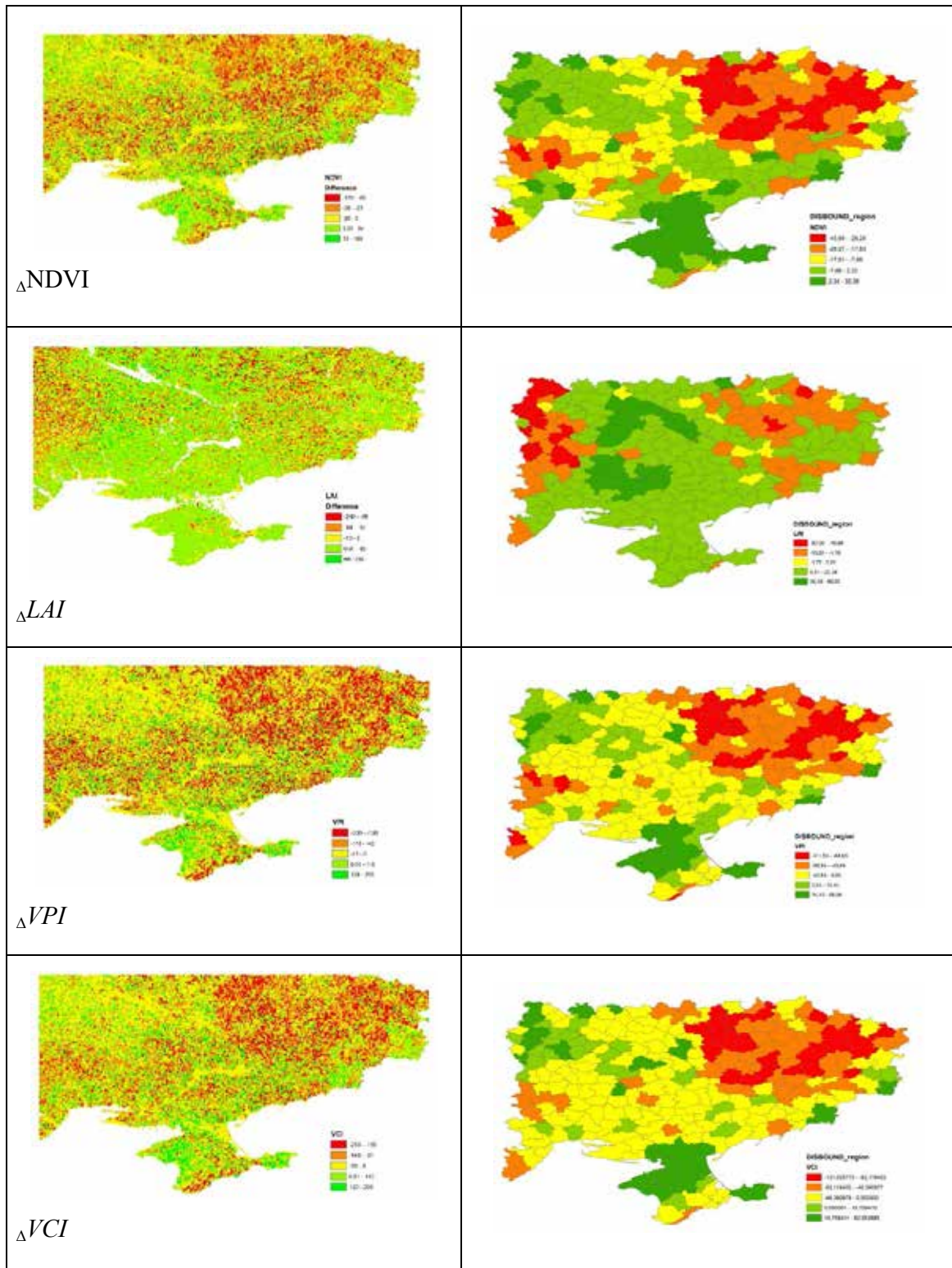


Рис. 1. Розподіл значень різниці вегетаційних індексів та усереднена зональна статистика за адміністративними районами

тим інтенсивнішою є кластеризація високих значень (гаряча точка). Щодо статистично значущих негативних z-оцінок, чим меншою є z-оцінка, тим інтенсивнішою є кластеризація низьких значень (холодна точка). На виході отримуємо новий набір

даних про події із z-оцінкою, p-значенням і рівнем достовірності G_i_Bin для кожної події у вхідному масиві.

Використання методу «гарячих точок» дало змогу отримати статистично значущі просторові

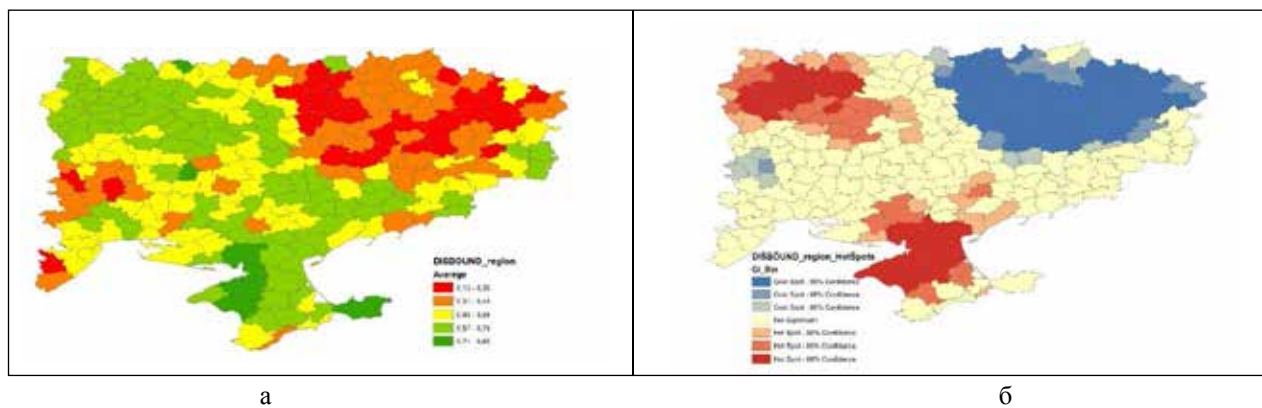


Рис. 2. Розподіл результуючого індексу за результатами а) класифікації, б) кластеризації методом «гарячих точок»

кластери високих та низьких значень у розподілі інтегрального індексу вегетації.

Отримані статистичні кластери підтверджують гіпотезу про існування регіональних трендів погіршення вегетації у 2015 р. порівняно з 2014 р. у північно-східній частині України та певне покращення цих показників для Херсонської області та АР Крим, а також Кіровоградської, Черкаської областей.

Висновки. Система збору, накопичення та обробки супутникових даних Copernicus надає цінну інформацію про стан вегетації на планеті, яка може бути використана в сільському господарстві, дослідженні клімату та землекористуванні. Вивчення динамічних рядів цих даних дає

змогу визначити довгострокові просторові тренди у зміні кліматичних умов та стану рослинного покриву. Сучасний математичний апарат системного аналізу та інструментарій геоінформаційних систем дають змогу отримати інтегральну оцінку цих просторово-часових трендів із використанням методів сумісного аналізу даних різної природи. Застосування цього підходу до даних про стан вегетації в Україні з використанням чотирьох видів індексів дало змогу визначити значні регіональні диспропорції в динаміці рослинного покриву. Подальші дослідження пов'язані зі збільшенням часових рядів досліджень та території спостереження.

Список літератури:

1. Кохан С.С. Застосування вегетаційних індексів на основі серії космічних знімків IRS-1D LISS-III для визначення стану посівів сільськогосподарських культур. *Космічна наука і технологія*. 2011. Т. 17. № 5. С. 58–63.
2. Лялько В.І. Особливості прогнозування врожайності зернових культур забагато спектральними даними ДЗЗ / В.І. Лялько, О.І. Сахацький, Г.М. Жолобак. *Багатоспектральні методи дистанційного зондування Землі в задачах природокористування*. Київ : Наук. думка, 2006. С. 176–191.
3. Мозговой Д.К. Использование многоспектральных снимков для классификации посевов сельхозкультур / Д.К. Мозговой, О.В. Кравец. *Екологія та ноосферологія*. 2009. Т. 20. № 1–2. С. 54–58
4. Путренко В.В. Застосування інструментарію геоінформаційного моделювання для інтелектуального аналізу даних пожежної небезпеки / В.В. Путренко, Н.М. Пашинська. *Вісник НТУ «ХПІ», Серія: Нові рішення в сучасних технологіях*. 2017. № 7 (1229). С. 156–163.
5. Станкевич С. Комплексна обробка багатоспектральних аерокосмічних зображень для виявлення сільськогосподарських угідь / С. Станкевич, А. Васько. *Сучасні досягнення геодезичної науки та виробництва* : збірник наукових праць Західного геодезичного товариства УТГК. Львів : Видавництво Львівської політехніки, 2011. Випуск 2 (22). С. 171–175.
6. Beckschäfer P. Mapping leaf area index in subtropical upland ecosystems using rapideye imagery and the randomforest algorithm / P. Beckschäfer, L. Fehrmann, R.D. Harrison, J. Xu, C. Kleinn. *iForest*. 2014. № 7. P. 1–11.
7. Boken V.K. Improving an operational wheat yield model using phenological phase-based Normalized Difference Vegetation Index / V.K. Boken, C.F. Shaykewich. *International Journal of Remote Sensing*. 2002. № 23. P. 4155–4168.
8. Bolton D.K. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics / D.K. Bolton, M.A. Friedl. *Agric. For. Meteorol.* 2013. № 173. P. 74–84
9. Broge N.H. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density / N.H. Broge, E. Leblanc. *Remote Sens. Environ.* 2001. № 76. P. 156–172.

10. Canisius F. Comparison and evaluation of Medium Resolution Imaging Spectrometer leaf area index products across a range of land use / F. Canisius, R. Fernandes, J. Chen. *Remote Sens. Environ.* 2010. № 114. P. 950–960
11. Cicek H. Using vegetation indices from satellite remote sensing to assess corn and soybean response to controlled tile drainage / H. Cicek, M. Sunohara, G. Wilkes, H. McNairn, F. Pick, E. Topp, D.R. Lapen. *Agric. Water Manage.* 2010. № 98. P. 261–270.
12. Copernicus Global Land Service. URL: <http://land.copernicus.eu/global/products/dmp> (08.10.2017).
13. Getis A. The Analysis of Spatial Association by Use of Distance Statistics / A. Getis, J.K. Ord. *Geographical Analysis.* 1992. № 24(3). P. 189–206
14. Martinez-Casasnovas J.A. Mapping multi-year cropping patterns in small irrigation districts from time-series analysis of Landsat TM images. *Eur. J. Agron.* 2005. № 23. P. 159–169.
15. Ord J.K. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application / J.K. Ord, A. Getis. *Geographical Analysis.* 1995. № 27(4). P. 286–306.
16. Pashynska N. A decision tree in a classification of fire hazard factors / N. Pashynska, V. Snytyuk, V. Putrenko, A. Musienko. *Eastern-European Journal of Enterprise Technologies.* 2016. Vol 5. № 10(83). P. 32–37
17. Pettorelli N. Using the satellite-derived NDVI to assess ecological responses to environmental change / N. Pettorelli, J.O. Vik, A. Mysterud, J.M. Gaillard, C.J. Tucker, N.C. Stenseth. *Trends Ecol. Evol.* 2005. № 20. P. 503–510.

Putrenko V.V., Pashynska N.M. REGIONAL MONITORING OF THE STATE OF VEGETATION BY USING METHODS OF DATA MINING

Determination of the method of changes in vegetation cover by a set of indicators that evaluate the vegetation using spatial data mining methods is studied in paper. Using the results of remote sensing of the Earth to calculate vegetation indices that characterize the dynamics of vegetation condition changes provides data for a coherent analysis and trend detection in changing the climatic conditions and vegetation state using spatial clustering methods. The calculations used data from the European Copernicus program. The model of compatible data processing of vegetation indices was developed in the south-eastern part of Ukraine, using the approach of compatible data processing of different nature. During data processing geographic information technology zoning is used, which allows to calculate statistical indicators for certain administrative units - districts and cities of regional subordination. This allows you to further use the results of the analysis to make managerial decisions. The combined analysis provides a more complete picture of the change in vegetation status. The next step was to analyze the clustering of hot-cold points, which allowed to obtain statistically significant clusters of changes in the annual dynamics of the state of vegetation cover. The analysis and the methodology of regional monitoring of the vegetation state on the basis of the joint use of various characteristics of vegetation indices and spatial clusterization techniques have been developed. The developed method can be used to detect spatial-temporal trends in changing the state of vegetation that can be taken into account in conditions of regional climate change and agriculture. The application of spatial clustering methods allows to obtain mathematically grounded estimates both at the regional level and at the levels of individual administrative units, for which the statistical values of vegetation index values can be calculated separately or together.

Key words: *vegetation index, vegetation cover, GIS, spatial data mining, clusterization.*