

УДК 004.62

КЛАСИФІКАЦІЯ СІЛЬСЬКОГОСПОДАРСЬКИХ ПОСІВІВ З ВИКОРИСТАННЯМ ЧАСОВИХ РЯДІВ СУПУТНИКОВИХ ДАНИХ

С.В. Скакун¹, А.Ю. Шелестов², Б.Я. Яйлимов¹, В.А. Остапенко¹,
М.С. Лавренюк¹, А.В. Вікулов³

¹ Інститут космічних досліджень НАН та ДКА України,

² Національний університет біоресурсів і природокористування України,

³ ЗОШ №286 м. Києва

*serhiy.skakun@gmail.com, andrii.shelestov@gmail.com, yailyimov@gmail.com,
ostapenkovadym@gmail.com, nick_93@ukr.net, andrewvikulov@gmail.com*

Запропоновано метод класифікації сільськогосподарських посівів з використанням часових рядів супутникових даних. Метод ґрунтується на послідовному використанні двох нейромережових парадигм: карт самоорганізації Кохонена та ансамблю багатошарових перцептронів. Точність класифікації підтверджується тестуванням на незалежній вибірці, а також шляхом порівняння з офіційною статистикою.

Ключові слова: класифікація посівів, відновлення даних, Landsat-8, нейронні мережі, ансамбль нейронних мереж, Україна

A method for crop classification using a time series of satellite data is presented. The method is based on the subsequent applications of two neural networks paradigms: self-organising Kohonen maps and an ensemble of multi-layer perceptrons. The accuracy of classification is verified using independent testing data set and compared with official statistics.

Keywords: crop classification; missing data; Landsat-8; neural networks; ensemble; Ukraine

Предложен метод классификации сельскохозяйственных посевов с использованием временных рядов спутниковых данных. Метод основан на последовательном использовании двух нейросетевых парадигм: карт самоорганизации Кохонена и ансамбля многослойных перцептронов. Точность классификации подтверждается тестированием на независимой выборке, а также путем сравнения с официальной статистикой.

Ключевые слова: классификация посевов, восстановление данных, Landsat-8, нейронные сети, ансамбль нейронных сетей, Украина

Вступ

Важливою прикладною задачею супутникового моніторингу є картографування сільськогосподарських посівів в масштабах окремих адміністративних одиниць, наприклад областей України. Карти посівів можуть використовуватись в якості динамічної маски для прогнозування врожайності [1], оцінки ризиків посух [2, 3, 4], тощо. Враховуючи значне різноманіття сільгоспкультур на території України, неможливо здійснити класифікацію посівів за одним супутниковим знімком. Тому виникає задача класифікації за часовими рядами супутникових знімків, отриманих на різних стадіях періоду вегетації. На сьогодні найбільш прийнятним джерелом даних, оптимальним за критерієм ціни, часового і просторового розрізнення, є супутник Landsat-8. Однак розв'язання задачі класифікації за часовими рядами ускладнюється зашумленістю знімків (наявністю хмар, тіней, аерозолів та конденсаційних

слідів). Для її розв’язання пропонується використати індуктивний підхід, заснований на послідовному використанні кількох нейромережових парадигм: карт Кохонена (self-organising map - SOM) для відновлення даних [5, 6, 7, 8] та ансамблю багат шарових перцептронів (MLP – multi-layer perceptron) для класифікації.

1. Метод розв’язання задачі

Запропонований підхід полягає в послідовному застосуванні нейронних мереж різного типу для відновлення невизначеностей (хмар на знімках) та класифікації сільськогосподарських культур. На першому етапі на кожному знімку виділяються пошкоджені дані (відсутні пікселі) і виконується їх відновлення з використанням карт Кохонена, як запропоновано в [6, 7, 8]. Після відновлення даних для класифікації даних за часовими рядами супутникових знімків застосовується ансамбль багат шарових нейронних мереж перцептронного типу.

Багат шаровий перцептрон забезпечує високу точність класифікації окремих супутникових зображень [9, 10]. І хоча навчання перцептрона потребує значних комп’ютерних ресурсів, цей недолік в наші дні легко долається шляхом використання розподілених [11, 12, 14, 15] та високопродуктивних обчислень [13, 16, 17]. Водночас, для підвищення ефективності класифікації доцільно застосовувати не один окремих класифікатор, а їх ансамбль. При цьому можливі дві реалізації (Рис. 1). В першому випадку злиття даних відбувається на рівні прийняття рішень (Рис. 1а), а в другому – також і на рівні окремих пікселів (Рис. 1б).

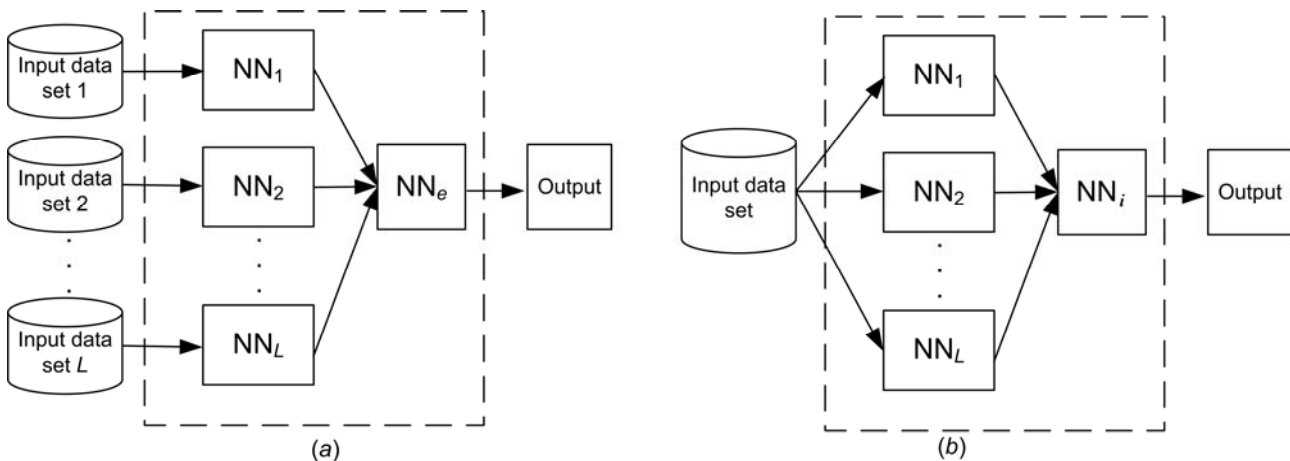


Рис. 1. Дві реалізації класифікації: злиття даних відбувається на рівні прийняття рішень (а) та на рівні окремих пікселів (б)

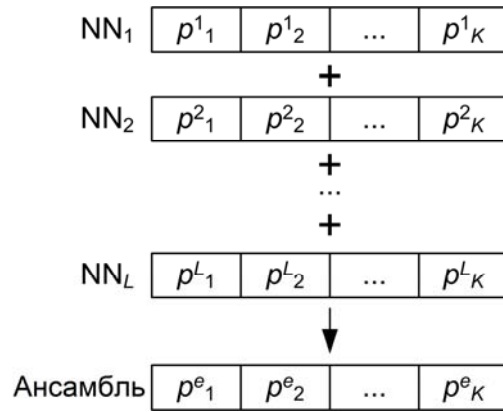


Рис. 2. Злиття даних на рівні прийняття рішень: визначення результатів ансамблю класифікаторів

На виході кожного класифікатора отримуємо апостеріорну імовірність належності вхідного образу до певного класу. Виходи окремих класифікаторів об'єднуються простим сумуванням, після чого обчислюється середня імовірність належності вхідного образу конкретному класу (Рис. 2). Вхідний образ відносять до класу з максимальною апостеріорною імовірністю:

$$k^* = \arg \max_{k=1, K} p_k^e, \quad p_i^e = \frac{1}{L} \sum_{l=1}^L p_i^l, \quad (1)$$

де k^* — це клас, до якого відносить вхідний образ ансамбль класифікаторів, p_i^e — апостеріорна імовірність належності класу для ансамблю, p_i^l — апостеріорна імовірність належності класу для кожного класифікатора MLP, L — кількість класифікаторів в ансамблі, K — кількість класів.

2. Територія досліджень, використані дані та результати експерименту

Запропонована методологія застосована для класифікація посівів на тестовому полігоні ЖЕСАМ, який включає Київську область [10, 18–20, 21] (Рис. 3).



Рис. 3. Київська область як полігон ЖЕСАМ

Площа полігону складає 28,100 км². Основними культурами є кукурудза (25.1% всіх посівних площ в 2013 р.), озима пшениця (16.1%), соя (12.6%), овочі (10.3%), соняшник (9.3%), ярий ячмінь (6.8%), озимий ріпак (4.0%), цукровий буряк (1.3%).

Для формування навчальної та незалежної тестової вибірки було виконано маршрутні дослідження (Рис. 4). Для коректної ідентифікації посівів на полях, що знаходяться на певній відстані від дороги, було використано квадрокоптер (Рис. 5). В цілому навчальна і тестова вибірки містять 386 полігонів, та включають основні класи земного покриття відповідно до стандартизованої номенклатури класів European Land Use and Cover Area frame Survey (LUCAS) (Табл. 1). Загальна площа навчальних даних становить 22,700 га.

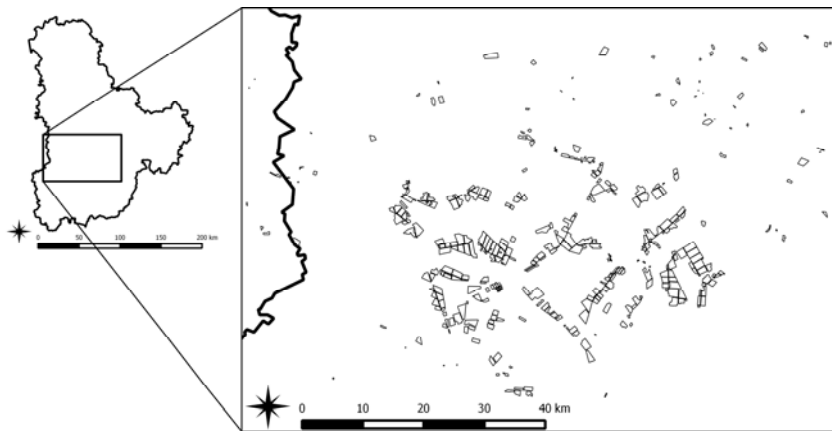


Рис. 4. Маршрутні дослідження на полігоні JESAM



Рис. 5. Квадрокоптер, який використано для наземних досліджень

Таблиця 1

Навчальна вибірка відповідно до номенклатури класів LUCAS

№	Код	Клас	Полігон		Площа	
			№.	%	га	%
1	Axx	Штучні об'єкти	6	1.6	23.0	0.1
2	B11	Озима пшениця	51	13.2	3960.8	17.4
3	B32	Озимий ріпак	12	3.1	937.3	4.1
4	B12, B14	Ярі культури	9	2.3	455.9	2.0
5	B16	Кукурудза	87	22.5	7253.3	31.9
6	B22	Буряк	8	2.1	632.5	2.8
7	B31	Соняшник	30	7.8	2549.0	11.2
8	B33	Соя	60	15.5	3252.3	14.3
9	B19, B39, B40	Інші культури	32	8.3	1364.0	6.0
10	C10, B60	Ліси	17	4.4	1014.3	4.5
11	E01, E02	Луги	48	12.4	747.5	3.3
12	F00	Порожня земля	10	2.6	67.2	0.3
13	G01, G02	Водні об'єкти	16	4.1	448.3	2.0
		Разом	386	100	22705.3	100

Для розв'язання поставленої задачі даних використано супутникові знімки сенсора Operational Land Imager (OLI) супутника Landsat-8. В якості вхідних ознак класифікатора використовувались 6 спектральних каналів (канали 2–7) з просторовим розрізненням 30 м. Тестовий полігон покривається трьома сценами з координатами 181/24, 181/25 181/26 (Табл. 2). За вегетаційний період 2013 р. отримано по 6 знімків кожної сцени з різним ступенем захмареності.

Таблиця 2

Перелік отриманих даних Landsat-8/OLI

Дата (2013 р.)	Коорд. path/row	Захмареність знімку, %	Дата (2013)	Коорд. path/row	Захмареність знімку, %
16 квітня	181/24	4.77	19 червня	181/24	62.58
	181/25	0.60		181/25	33.37
	181/26	0.02		181/26	26.43
2 травня	181/24	0.01	5 липня	181/24	35.30
	181/25	0.77		181/25	21.06
	181/26	4.38		181/26	15.69
18 травня	181/24	9.06	6 серпня	181/24	24.35
	181/25	14.93		181/25	12.69
	181/26	14.32		181/26	40.86

Для отримання каліброваних даних здійснено наступну попередню обробку зображень.

- перетворення цифрових значень до коефіцієнта відбиття верхнього шару атмосфери (top-of-atmosphere — TOA);
- подальше перетворення значень до відбиття земної поверхні (surface reflectance — SR) на основі моделі атмосферної корекції SMAC [22];
- детектування хмар і тіней з подальшим відновленням відсутніх даних [7, 8].

Оскільки в якості вхідних ознак використовується 6 спектральних каналів із 6 знімків, то загальна розмірність вхідного шару класифікатора становить 36. Всі змінні нормуються до діапазону з середнім значенням 0 та середньоквадратичним відхиленням 1. На виході класифікатора маємо 13 класів у відповідності до номенклатури класів LUCAS (див. табл. 1).

Наземні дані випадковим чином поділяються на навчальну (50%) та тестову (50%) вибірки. Точність класифікації і матриці невідповідностей обчислюються для незалежних тестових даних.

Для визначення ефективності використання ансамблю класифікаторів було порівняно 3 різних ансамблевих архітектури. В першій з них використано 1 класифікатор з різною кількістю (від 20 до 80) прихованих нейронів, в другій — ансамбль із 5 класифікаторів, що були навчені на різних даних, в третій — ансамбль із 7 класифікаторів, які були навчені на спільних даних. Результати експерименту представлені в табл. 3. Для порівняння використовувалися наступні метрики: загальна точність класифікації, індекс Карра, точність «виробника» (ТВ) та точність «користувача» (ТК). Індекс Карра є статистичною мірою узгодженості між отриманими класами (по супутниковим даним) та класами, що спостерігаються (наземними дослідженнями). Точність «виробника» показує, на скільки добре результат класифікації для даного класу співпадає з тестовими даними. Точність «користувача», в свою чергу, показує, на скільки імовірно, що даний клас співпадає з результатами класифікації.

Таблиця 3

Результати класифікації (ТВ — точність «виробника»;
ТК — точність «користувача»)

		Схема 1: Крайній одиничний класифікатор (MLP)		Схема 2: Ансамбль із 5 класифікаторів (MLPs)		Схема 3: Ансамбль із 7 класифікаторів (MLPs)	
Загальна точність, %		84.60		85.11		85.32	
Індекс Карра		0.8144		0.8211		0.8235	
№		ТВ, %	ТК, %	ТВ, %	ТК, %	ТВ, %	ТК, %
1	Штучні об'єкти	74.5	93.2	100.0	97.9	100.0	97.9

Продовження таблиці 3

2	Озима пшениця	95.6	90.6	96.0	91.9	95.7	91.8
3	Озимий ріпак	94.5	96.1	93.3	99.2	93.5	99.4
4	Ярі культури	12.1	15.3	46.2	38.8	40.6	34.6
5	Кукурудза	92.6	86.6	90.3	86.8	90.5	86.8
6	Буряк	83.0	93.7	94.4	88.0	94.9	89.6
7	Соняшник	86.1	82.1	83.6	84.2	84.1	85.4
8	Соя	66.6	77.1	68.8	76.6	69.7	77.1
9	Інші культури	71.8	76.9	70.2	78.1	70.9	78.0
10	Ліси	96.7	91.9	96.9	91.9	96.9	92.9
11	Луги	84.2	88.9	90.7	88.0	91.0	89.0
12	Порожня земля	86.7	88.8	86.7	98.5	86.7	99.0
13	Водні об'єкти	99.3	98.1	100.0	98.0	100.0	98.1

Як видно із табл. 3 найвища точність отримана для ансамблю із 7 класифікаторів, які навчались на спільних даних. Матрицю невідповідностей для цього випадку наведено в табл. 4.

Таблиця 4

Матриця невідповідностей для тестової вибірки при застосуванні архітектури 3 (ТВ — точність «виробника»; ТК — точність «користувача»; ПР — помилка першого роду; ДР — помилка другого роду)

Клас	1	2	3	4	5	6	7	8	9	10	11	12	13	ТК,%	ДР,%
1	184	0	0	0	3	0	0	0	0	0	0	1	0	97.9	2.1
2	0	21453	33	51	26	0	29	11	1728	0	2	30	0	91.8	8.2
3	0	1	3788	1	14	0	5	1	2	0	0	0	0	99.4	0.6
4	0	723	0	982	3	0	0	3	1101	0	24	0	0	34.6	65.4
5	0	0	0	177	33972	2	390	4544	34	0	41	0	0	86.8	13.2
6	0	0	0	11	222	3511	118	52	2	0	2	0	0	89.6	10.4
7	0	1	0	10	389	7	13241	1672	91	0	92	0	0	85.4	14.6
8	0	56	0	351	2660	181	1027	14668	29	3	52	0	0	77.1	22.9
9	0	74	230	828	96	0	832	78	7677	0	24	0	0	78.0	22
10	0	27	0	0	78	0	36	11	6	4198	165	0	0	92.9	7.1
11	0	71	1	6	63	0	26	15	159	131	4089	31	0	89.0	11
12	0	0	0	0	0	0	0	0	0	0	4	403	0	99.0	1.0
13	0	0	0	0	0	0	49	0	0	0	0	0	2524	98.1	1.9
ТВ,%	100.0	95.7	93.5	40.6	90.5	94.9	84.1	69.7	70.9	96.9	91.0	86.7	100.0		
ПР,%	0.0	4.3	6.5	59.4	9.5	5.1	15.9	30.3	29.1	3.1	9.0	13.3	0.0		

Для перевірки достовірності отримані результати було порівняно з офіційною статистикою (Табл. 5).

Таблиця 5

Порівняння результатів класифікації з офіційною статистикою
для Київської області

Но- мер класу	Назва класу	Посівні площі: офіційна статис- тика, x 1000, га	Посівні площі: результати класифікації, x 1000, га	Відносна похибка, %
2	Озима пшениця	187.3	184.5	-1.5
3	Озимий ріпак	46.7	59.9	28.3
5	Кукурудза	291.7	342.4	17.4
6	Цукровий буряк	15.5	11.2	-27.9
7	Соняшник	108.2	117.6	8.7
8	Соя	145.9	168.5	15.5

3. Висновки

В статті запропоновано індуктивний підхід до класифікації посівів за часовими рядами супутникових даних. Ідея полягає в послідовному застосуванні двох нейромережових парадигм: нейронних мереж Кохонена для відновлення зашумлених даних та ансамблю мереж перцептронного типу для класифікації. В цілому, отримані результати класифікації можна вважати задовільними. Найбільші помилки класифікації отримані для ярих культур, соняшника, сої і кукурудзи. Вони спричинені тим, що вегетаційний період цих культур співпадає з періодом вегетації інших сільськогосподарських культур, тому для їх розділення необхідно застосовувати додаткові джерела даних (наприклад, радарні дані). Такі дослідження планується проводити в подальшому. Важливо також відзначити, що карти класифікації на основі супутникових даних в операційному режимі можна отримувати в кінці вегетаційного сезону, за півроку до отримання офіційної статистики.

Література

1. Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., Lavrenyuk, A., 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *International Journal of Applied Earth Observation and Geoinformation* 23, 192–203. doi:10.1016/j.jag.2013.01.002.
2. Куссуль Н.М., Скакун С.В., Шелестов А.Ю., 2014. Геопросторовий аналіз ризиків стихійних лих. К.: «Наукова думка» — 2014. — 258 с.
3. S. Skakun, N. Kussul, A. Shelestov, O. Kussul., 2014. Flood Hazard and Flood Risk Assessment Using a Time Series of Satellite Images: A Case Study

- in Namibia. *Risk Analysis*. — 2014. — Vol. 34, No. 8. — P. 1521–1537. doi: 10.1111/risa.12156.
4. N. Kussul, S. Skakun, A. Shelestov, O. Kussul and B. Yailymov., 2014. Resilience Aspects in the Sensor Web Infrastructure for Natural Disaster Monitoring and Risk Assessment Based on Earth Observation Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. — 2014.
5. Kohonen, T., 1995. Self-organizing maps. Series in information sciences 30, Springer, Heidelberg, Germany.
6. Skakun, S., Kussul, N., Basarab, R., 2014. Restoration of Missing Data due to Clouds on Optical Satellite Imagery Using Neural Networks. In: ESA SENTINEL-2 for Science Workshop, ESA-ESRIN, Frascati, Rome, Italy, 20-22 May 2014. Available at: http://seom.esa.int/S2forScience2014/files/05_S2forScience-MethodsI_KUSSUL.pdf.
7. Skakun Sergii; Kussul Nataliia; Basarab Ruslan., 2014. Dealing with missing data in remote sensing images within land and crop classification. 40th COSPAR Scientific Assembly. Held 2-10 August 2014, in Moscow, Russia, Abstract A3.1-3-14.
8. Скакун С.В., Р.М. Басараб. Восстановление отсутствующих данных во временных рядах оптических спутниковых снимков с использованием самоорганизующихся карт Кохонена. *Проблеми управління та інформатики* — 2014., №6. —С. 88 – 94.
9. Gallego, J., Kravchenko, A.N., Kussul, N.N., Skakun, S.V., Shelestov, A.Y., Grypych Y.A., 2012. Efficiency assessment of different approaches to crop classification based on satellite and ground observations. *Journal of Automation and Information Sciences* 44 (5), 67–80. doi:10.1615/JAutomatInfScien.v44.i5.70.
10. Gallego, F.J., Kussul, N., Skakun, S., Kravchenko, O., Shelestov, A., Kussul, O., 2014. Efficiency assessment of using satellite data for crop area estimation in Ukraine. *International Journal of Applied Earth Observation and Geoinformation* 29, 22–30. doi:10.1016/j.jag.2013.12.013.
11. N. Kussul, A. Shelestov, S. Skakun, G. Li, O. Kussul, J. Xie., 2014. Service-oriented infrastructure for flood mapping using optical and SAR satellite data. *International Journal of Digital Earth* . — 2014. — Vol. 7, No. 10 — P. 829–845.
12. Kussul N., Shelestov A., Skakun S., 2011. Technologies for Satellite Data Processing and Management Within International Disaster Monitoring Projects. In . *Grid and Cloud Database Management Grid* — Fiore, S.; Aloisio, G. (Eds.). — 2011, Springer — P. 279–306.

13. Kravchenko, A.N., Kussul, N.N., Lupian, E.A., Savorsky, V.P., Hluchy, L., Shelestov, A.Yu., 2008. Water resource quality monitoring using heterogeneous data and high-performance computations. *Cybernetics and System Analysis* 44 (4), 616–624. doi:10.1007/s10559-008-9032-x.
14. Kussul, N., Shelestov, A., Skakun, S., 2009. Grid and sensor web technologies for environmental monitoring. *Earth Science Informatics* 2 (1-2), 37–51. doi:10.1007/s12145-009-0024-9.
15. Kussul, N., Shelestov, A., Skakun, S., Kravchenko, O., Gripich, Y., Hluchy, L., Kopp, P., Lupian, E., 2010. The Data Fusion Grid Infrastructure: Project Objectives and Achievements. *Computing and Informatics* 29 (2), 319–334.
16. Kussul, N.N., Shelestov, A.Y., Skakun, S.V., Li, G., Kussul, O.M., 2012. The wide area grid testbed for flood monitoring using earth observation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (6), 1746–1751. doi:10.1109/JSTARS.2012.2201447.
17. Shelestov, A., Kussul, N., Skakun, S., 2006. Grid Technologies in Monitoring Systems Based on Satellite Data. *Journal of Automation and Information Sciences* 38 (3), 69–80. doi:10.1615/J Automat Inf Scien.v38.i3.50.
18. Kussul N., Skakun S., Kravchenko O., Shelestov A., Gallego F. J., Kussul O., 2013. Application of satellite optical and SAR images for crop mapping and area estimation in Ukraine. *International Journal "Information Technologies & Knowledge"*. — 2013. — Vol. 7, No. 3 — P. 203–210.
19. Shelestov, A.Y., Kravchenko, A.N., Skakun, S.V., Voloshin, S.V., Kussul, N.N., 2013. Geospatial information system for agricultural monitoring. *Cybernetics and System Analysis* 49 (1), 124–132. doi:10.1007/s10559-013-9492-5.
20. Skakun Sergii; Kussul Nataliia; Kravchenko Oleksiy; Basarab Ruslan; Ostapenko Vadym; Yailymov Bohdan; Shelestov Andrii; Kolotii Andrii; Mironov Andrii, Validation of Global EO Biophysical Products at JECAM Test Site in Ukraine //40th COSPAR Scientific Assembly. 2-10 August 2014, in Moscow, Russia, Abstract A3.1-59-14.
21. A. Shelestov, S. Skakun, R. Basarab , T. Baranova. JECAM Activities in Ukraine. *EARTH Bioresources and Life Quality*. — 2013. — Vol.4. P. 1–13.
Available at: <http://gchera-ejournal.nubip.edu.ua/index.php/ebql/article/view/160/124>.
22. Rahman, H., Dedieu, G., 1994. SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing* 15 (1), 123–143. doi:10.1080/0143116940895405.
— 2014. — Vol. 7, No. 9 — P. 1–7.