

ВСТАНОВЛЕННЯ ФУНКЦІОНАЛЬНОЇ ЗАЛЕЖНОСТІ МІЖ ВИНИКНЕННЯМ ЛІСОВИХ ПОЖЕЖ І ХАРАКТЕРИСТИКАМИ СОНЯЧНОЇ АКТИВНОСТІ НА ОСНОВІ DATA MINING

*Географічний інститут "ЈованСвјііс", Сербська академія наук та мистецтв – SASA, Белград, Сербія

**Буковинський університет, Чернівці, Україна

Анотація. Робота присвячена перевірці «геліоцентричної» гіпотези, згідно з якою значна частина лісових пожеж спричинена спалахами на сонці. В результаті дослідження знайдено найбільш вагомими фактори сонячної активності (з урахуванням часової затримки), що спричиняють лісові пожежі. Була побудована функціональна залежність між характеристиками сонячної активності та числом лісових пожеж на території США на основі фрактального аналізу та гібридних нейронних мереж. Розроблені моделі дали змогу пояснити до 47% лісових пожеж.

Ключові слова: сонячна активність, лісові пожежі, «геліоцентрична гіпотеза», гібридні нейронні мережі ANFIS, коефіцієнт кореляції, R/S аналіз.

Аннотация. Работа посвящена проверке «гелиоцентрической» гипотезы, согласно которой значительная часть лесных пожаров вызвана вспышками на солнце. В результате исследования найдены наиболее значимые факторы солнечной активности (с учетом временной задержки), которые вызывают лесные пожары. Была построена функциональная зависимость между характеристиками солнечной активности и числом лесных пожаров на территории США на основе фрактального анализа и гибридных нейронных сетей. Разработанные модели позволили объяснить до 47% лесных пожаров.

Ключевые слова: солнечная активность, лесные пожары, «гелиоцентрическая гипотеза», гибридные нейронные сети ANFIS, коэффициент корреляции, R/S анализ.

Abstract. The work is dedicated to check of "heliocentric" hypothesis, according to which a significant part of forest fires caused by solar flares. The most important factors of the solar activity (including time delay) that cause forest fires were found in the study. The functional relationship between the characteristics of solar activity and the number of forest fires in the United States using fractal analysis and hybrid neural networks was develop. The obtained models allowed explain up to 47% of forest fires.

Keywords: solar activity, forest fires, "heliocentric hypothesis", ANFIS hybrid neural networks, correlation coefficient, R/S analysis.

1. Вступ

У наш час лісові пожежі становлять глобальну світову проблему. Пожежі масштабних розмірів здатні знищити цілі ліси, поля, житлові об'єкти тощо. Крім того, продукти горіння викликають великі екологічні проблеми. Для ефективного подолання цієї проблеми необхідно впроваджувати відповідні превентивні заходи, що базуються на науково-обґрунтованих математичних методах прогнозування. Розробка останніх є досі не вирішеною задачею, оскільки існує велика кількість лісових пожеж з невстановленою причиною займання.

Аналіз наукових джерел дав змогу встановити, що частка лісових пожеж з непоясненими причинами займання становить від 38% до 95%. Зокрема, в країнах Балканського півострова за період з 1988 по 2004 рік відсоток таких пожеж становив 37,9% [1]. В той самий час Болгарія має найбільший відсоток виникнення лісових пожеж з невідомих причин (67,9%). У Португалії за 2011 рік сталася 25 221 пожежа, при чому 40% зафіксованих випадків – пожежі невідомого походження, а у Німеччині з 888 пожеж в тому ж році – 48% [2]. У багатонаселеній Індії в 1995 році із зареєстрованих 208793 лісових пожеж кількість непояснених становить 50,78% [3]. Впродовж 2013 року на території США сталася

47579 лісових пожеж, серед яких 39,4% – причини виникнення яких залишаються таємницею [4]. Найбільший відсоток таких пожеж спостерігається в Угорщині (95%) [2].

Вважається, що основними причинами лісових пожеж є людський фактор, спалахи блискавок та пекельна спека. Проте, як було показано вище, вагомий відсоток лісових пожеж спричинений невідомими факторами. У своїх працях професор Мілан Радованович запропонував «геліоцентричну гіпотезу», згідно з якою лісові пожежі можуть бути спричинені спалахами на Сонці. Джерелом пожежі є займання рослинної маси під дією заряджених частинок сонячного вітру, що дійшли до поверхні Землі [5–8].

У нашій попередній роботі була перевірена ця гіпотеза на невеликих навчальних вибірках [5]. У дослідженні використовувались дані за період з травня по жовтень 2004 – 2007 рр. Як математичні методи використовувалися R/S аналіз, нейронні мережі та нечітка логіка. В ході розрахунків було встановлено наявність нелінійного зв'язку між раптовим надходженням заряджених частинок сонячного вітру і виникненням лісових пожеж із затримкою від одного до чотирьох днів. Також було встановлено, що найбільш адекватним методом дослідження є гібридні нейронні мережі ANFIS.

Отже, метою роботи є встановлення та дослідження функціональної залежності між кількістю лісових пожеж та характеристиками сонячної активності на основі навчальної вибірки за більший часовий інтервал з використанням гібридних нейронних мереж, що дасть змогу отримати результати з вищим ступенем точності.

2. Попередня обробка даних

У дослідженні використовувались дані за період з травня по жовтень 1999 – 2013 рр. Інформаційна база даних налічувала 2760 елементів. Вхідні параметри (показники сонячної активності) були обрані таким чином: X_1 – потік протонів > 1 MeV, X_2 – потік протонів > 10 MeV, X_3 – потік протонів > 100 MeV, X_4 – потік електронів > 0,6 MeV, X_5 – потік електронів > 2 MeV, X_6 – індекс F10.7 (міра рівня шуму, генерованого сонцем на довжині хвилі 10,7 см на орбіті Землі), X_7 – швидкість сонячного вітру, X_8 – густина частинок сонячного вітру, X_9 – температура сонячного вітру.

Як вихідне поле виступала загальна кількість пожеж, зареєстрована в певний день на території США – F . Такий вибір обґрунтовується наявністю великого обсягу статистичних даних про пожежі на відносно великій площі і на щоденній основі.

Для встановлення функціональної залежності між параметрами сонячної активності та кількістю лісових пожеж було проведено комплексне дослідження, що складалось з наступних етапів.

Заповнення пропусків даних. Як показав попередній аналіз бази даних, у ній міститься велика кількість пропущених даних («дірок»). У більшості випадків «дірки» в інтервалі від 1 до 3 днів. Як правило, часовий період «дірок» для різних вхідних параметрів не співпадає один із одним, що пов'язано з різними джерелами отримання вхідних параметрів. Видалення записів, які містять «дірки», створить часовий розрив даних, що в подальшому унеможливить перевірку лагової (часової) залежності. Оскільки кількість «дірок» складає невеликий відсоток даних (табл. 1), було прийнято рішення для апроксимації останніх використати метод максимальної вірогідності.

Таблиця 1. Кількість «дірок» у базі даних

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	F
К-ть «дірок»	48	48	48	13	13	0	6	6	276	48
%	1,74	1,74	1,74	0,47	0,47	0,00	0,22	0,22	10,00	3,48

Якісний аналіз часових рядів та усунення сезонності і циклічності. Результати попереднього етапу представлені на рис. 1. Як видно з рисунку, графіки, що характеризують енергію протонів ($X_1 - X_3$), мають серію різких піків, пов'язаних із спалахами на сонці. Водночас на графіку спалахів пожеж (F) не спостерігається відповідних їм різких піків. Це може означати, що ці три вхідні фактори не впливають на лісові пожежі. На графіках $X_4 - X_6$ та F присутня велика стохастичність в усьому досліджуваному діапазоні. Водночас на графіку X_6 спостерігається мультиплікативна залежність від активності сонця. Збільшення активності сонця (11-літній цикл) призводить до збільшення флуктуацій цього вхідного параметра. І навпаки, при зменшенні сонячної активності коливання цього параметра теж зменшуються. Якщо якісно порівняти цей графік з графіком лісових пожеж (F), то стає очевидно, що 11-літня циклічність сонця не впливає на пожежі. У протилежному випадку, в роки найбільшої активності сонця, спостерігалось би збільшення лісових пожеж. Відповідно графік лісових пожеж описувався б синусоїдальною залежністю з накладеними на неї стохастичними коливаннями аналогічно графіку X_6 . Натомість графік F містить тільки яскраво виражену річну сезонність та не містить циклічну складову сонячної активності. Тому для встановлення впливу активності сонця на лісові пожежі необхідно усунути циклічну залежність фактора X_6 та сезонність вихідного фактора.

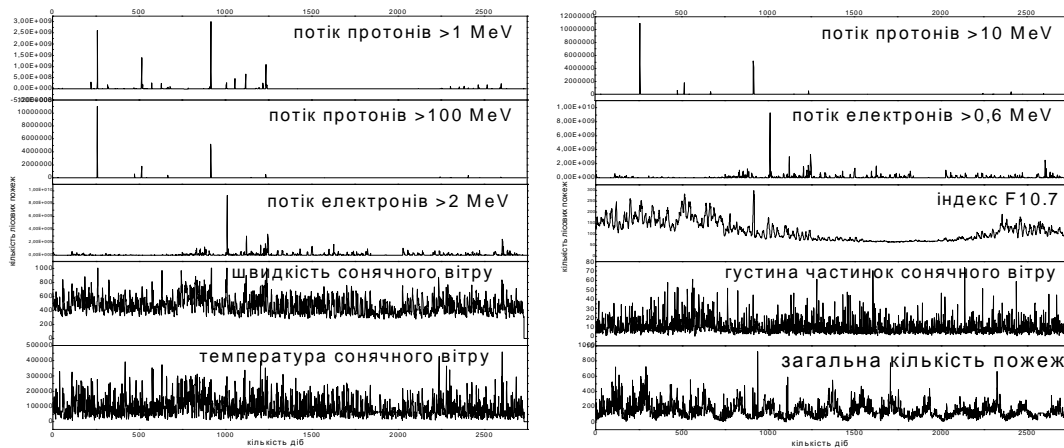


Рис. 1. Часові ряди вхідних та вихідного полів після усунення пропущених даних

Позбутись сезонної компоненти рядів даних можна шляхом розкладання часових рядів X_6 і F на компоненти з використанням адитивної моделі (мультиплікативна модель, як показали розрахунки, виявилась неадекватною). Адитивна часова модель, у нашому випадку для F (для X_6 аналогічно), має вигляд [9]:

$$F = T + S + \tilde{F},$$

де $T = \{t_j\}_{j=1,n}$ – трендова компонента кількості пожеж;

n – кількість спостережень (у нашому випадку $n=2760$ – днів у період з травня по жовтень 1999 – 2013 рр.);

$S = \{s_j\}_{j=1,n}$ – сезонна компонента – кількість пожеж, пов'язаних з підвищенням (зниженням) температури протягом року або з впливом туристів на появу лісових пожеж;

$\tilde{F} = \{\tilde{f}_j\}_{j=1,n}$ – флуктуаційна компонента, пов'язана з впливом неврахованих моделлю факторів, як, наприклад, сонячна активність.

Методика видалення сезонної складової проводилась згідно з класичним методом індексів сезонності [10]. Видалення трендової компоненти з \hat{F} проводилось методом найменших квадратів [11]. У часовому ряду пожеж було знайдено лінійний тренд: $y = -0,024x + 180,9$. Оскільки X_6 демонструє синусоїдальний характер зміни, розрахований тренд задається рівнянням $y = 117,874 + 53,835 \cdot \cos(0,002x - 0,556)$.

Таким чином, завдяки наведеним перетворенням було отримано часові ряди \tilde{F} та \tilde{X}_6 , очищені від сезонної складової та тренду (рис. 2, 3), що в подальшому використовувались у розрахунках. Слід зазначити, що фактор X_6 кожної доби вимірюється з поверхні Землі, тому в його часовому ряду, крім циклічної компоненти активності Сонця, спостерігається сезонність, пов'язана саме із зміною сезонів на планеті Земля.

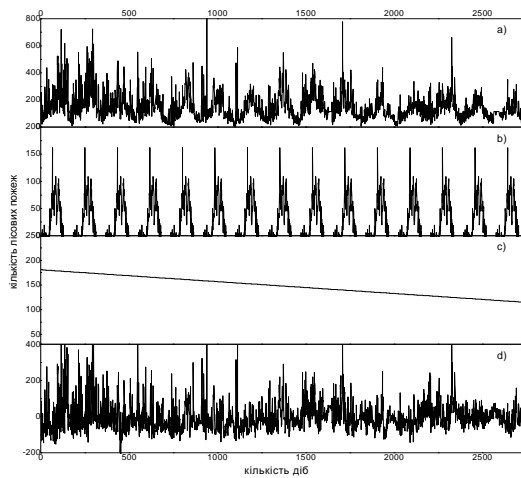


Рис. 2. Часовий ряд серій F :

а) вхідний ряд, б) сезонна складова, в) тренд, г) ряд, очищений від сезонної компоненти і тренду

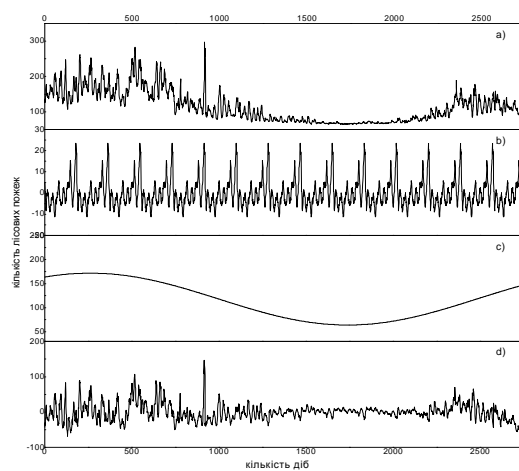


Рис. 3. Часовий ряд серій X_6 :

Кореляційний аналіз. Для перевірки гіпотези наявності функціональної залежності між компонентами сонячної активності та спалахами лісових пожеж був проведений кореляційний аналіз між параметрами X_i та кількістю пожеж \tilde{F} з урахуванням часу затримки (лагу) між настанням пожеж і сонячною активністю. Результати цього аналізу приведені в табл. 2.

Таблиця 2. Коефіцієнти кореляції між вхідними (X_i , $i = \overline{1,9}$) і вихідними (\tilde{F}_L) параметрами залежно від лага $L = \overline{0,5}$

	X_1	X_2	X_3	X_4	X_5	\tilde{X}_6	X_7	X_8	X_9
\tilde{F}_0	-0,018	-0,024	-0,016	-0,046	-0,045	0,060	0,023	-0,011	0,000
\tilde{F}_1	-0,002	-0,011	-0,033	-0,050	-0,045	0,064	0,017	-0,003	-0,017
\tilde{F}_2	0,033	0,053	0,017	-0,034	-0,035	0,058	0,005	-0,018	-0,041
\tilde{F}_3	0,001	0,020	0,065	-0,027	-0,010	0,048	0,008	0,024	-0,025
\tilde{F}_4	0,015	0,025	0,019	-0,020	0,015	0,038	0,012	0,015	-0,016
\tilde{F}_5	0,015	0,023	0,036	-0,017	0,011	0,030	0,008	-0,029	-0,018

Як можна побачити, будь-який коефіцієнт кореляції не перевищує 0,01. Це свідчить про повну відсутність лінійних зв'язків між згаданими факторами. Тому необхідно застосовувати методи нелінійного аналізу, щоб перевірити гіпотезу про функціональний взаємозв'язок між виникненням пожеж і сонячною активністю.

3. R/S аналіз

Для визначення ступеня стохастичності часових рядів вхідних і вихідних параметрів був використаний R/S аналіз [12–14]. R/S аналіз дозволяє встановити факт наявності довгострокової пам'яті у часових рядів. Для цього було використано таке співвідношення [15]:

$$R/S = cn^H, \quad (1)$$

де R/S – нормований розмах, тобто відношення часткових сум відхилень часових рядів від його середнього, масштабований за допомогою стандартного відхилення;

c – константа;

H – коефіцієнт Херста.

Це рівняння було розв'язане для кожної із змінних X_i і \tilde{F} . Значення коефіцієнта Херста інтерпретуються таким чином [16]:

- якщо $H = 0,5$, часові ряди є стохастичними (“білий шум”);
- якщо $0,5 < H < 1$, часовий ряд характеризується персистентністю, тобто властивістю тривалої пам'яті (“чорний шум”);
- якщо $0 < H < 0,5$, часові ряди є антиперсистентні, тобто часовий ряд змінюється швидше, ніж у випадку випадкового процесу (“рожевий шум”).

Використання критеріїв персистентності чи антиперсистентності часових рядів дозволяє прогнозувати розвиток досліджуваного часового ряду у відносно простій формі на базі своєї історії.

На основі коефіцієнта Херста був розрахований інший показник – фрактальна розмірність D :

$$D = 2 - H. \quad (2)$$

Фрактальна розмірність є кількісною характеристикою, яка визначає зміну графіка часового ряду залежно від масштабу, тобто ступінь самоподібності. Результати цих обчислень наведені в табл. 3.

Таблиця 3. Результати R/S аналізу для часових рядів

Параметр		Коефіцієнт Херста	Фрактальна розмірність
Потік протонів > 1 MeV	X_1	0,52	1,48
Потік протонів > 10 MeV	X_2	0,43	1,57
Потік протонів > 100 MeV	X_3	0,4	1,6
Потік електронів > 0,6 MeV	X_4	0,94	1,06
Потік електронів > 2 MeV	X_5	0,72	1,28
Індекс F10.7	X_6	0,71	1,29
Швидкість сонячного вітру	X_7	0,84	1,16
Густина частинок сонячного вітру	X_8	0,96	1,04
Температура сонячного вітру	X_9	0,84	1,16
Загальна кількість пожеж зареєстрована в певний день на території США	F	0,74	1,26

Як видно з табл. 3, середнє значення коефіцієнтів Херста для X_1 є близьким до 0,5. Це означає, що цей часовий ряд описують випадкові процеси. Фактори X_2 та X_3 характеризуються антиперсистентністю. На відміну від перших трьох факторів, всі інші є персистентними. При чому коефіцієнт Херста для часового ряду лісових пожеж є найбільш близьким до X_5 та X_6 . Тобто ці три часових ряди мають схожі фрактальні характеристики. Варто зазначити, що коефіцієнти Херста для X_7 та X_9 є однаковими і можуть функціонально залежати один від одного. Отже з високим ступенем імовірності можна стверджувати, що фактори $X_1 - X_3$ не впливатимуть на кількість спалахів пожеж. Однак цей факт потребує подальшої перевірки.

4. ANFIS моделювання

Задача пошуку прихованих залежностей у великих базах даних відноситься до задач DataMining. Тому в роботі було досліджено та проведено порівняльний аналіз моделей на основі гібридних нейронних мереж ANFIS. Цей підхід, як показали наші попередні розрахунки, дозволив отримати кращі результати на відміну від нейронних мереж [17].

ANFIS – нейронна мережа, що базується на основі нечіткої системи виводу Такагі-Сугено. Її система виведення містить набір нечітких If-Then правил, які отримуються при навчанні на великих базах даних на основі нелінійних функцій [18–19].

Для побудови нечіткої моделі всі вхідні параметри представлені як лінгвістичні змінні. Як було показано вище, в досліджуваній системі наявні нелінійні зв'язки, тому кожний терм у всіх лінгвістичних змінних описується нелінійними гаусівськими функціями належності. Як показали тестові розрахунки, найкращий результат отримувався при кількості 3-х термів у кожній лінгвістичній змінній для кожного вхідного параметра. У випадку 2-х термів моделі не були адекватними. Якщо ж цих термів більше, ніж 3, різко зростає час навчання. Як метод виведення нечіткої системи було обрано функцію Сугено нульового порядку. Методом навчання був гібридний спосіб, що об'єднує метод зворотного поширення помилки з методом найменших квадратів.

У загальному випадку задача зводиться до знаходження залежності у вигляді $M : X_1 \times \dots \times X_9 \rightarrow \tilde{F}$. При формуванні навчальної вибірки слід врахувати і часову затримку (лаг) між спалахом на сонці і пожежею, до якої вона призвела. Як було показано в роботі [20], ця затримка може тривати до 3-х діб. Для цього навчальну вибірку було трансформовано з урахуванням присутніх розривів у зимові місяці.

Враховуючи складність фізико-біологічних процесів взаємодії сонячної активності та процесів запалювання, можна висунути гіпотезу, що різні характеристики сонячної активності можуть впливати на процеси запалювання з властивою тільки їм часовою затримкою. Це призводить до різкого збільшення вхідних параметрів моделі. Так, у випадку врахування lag від 0 до 3-х кількість вхідних полів зростає в 4 рази ($9 \cdot 4 = 36$). Для розв'язання задачі з урахуванням часової затримки (lag) була сформована навчальна множина у вигляді кортежів, що містить 36 вхідних полів і 1 вихідне:

$$Tr = \left\{ \left\langle \begin{matrix} -0 & -3 \\ x_{1,j}, \dots, x_{9,j}, \tilde{f}_j \end{matrix} \right\rangle \right\}_{j=1, \dots, n}, \quad (3)$$

де $\bar{x}_{i,j}$ – нормовані компоненти часових рядів X_i ;

$$\left(\bar{x}_{i,j} = \frac{x_{i,j} - \min(X_i)}{\max(X_i) - \min(X_i)} \right), \text{ верхній індекс означає лаг;}$$

n – потужність навчальної множини (після лагової трансформації вона склала 2684).

Необхідність нормалізації всіх вхідних параметрів зумовлена значною різницею між абсолютними значеннями max-min компонентів окремих вхідних векторів, що можуть змінюватися від одного до 11 порядків (наприклад, X_1 , X_5). Також наявною є велика різниця між абсолютними значеннями різних вхідних векторів. Наприклад: $\max(X_4) - \max(X_6) \approx 10^{11}$, $\min(X_4) - \min(X_6) \approx 10^8$ (табл. 4). Комп'ютерний розрахунок без нормалізації цих даних призводить до значних помилок заокруглення, що повністю нівелює адекватність моделі [21–29].

Таблиця 4. Статистичні характеристики вхідних та вихідних параметрів

Значення параметрів	Максимальне значення	Мінімальне значення	Середнє значення	Середнє значення \bar{X}_i
X_1	$3 \cdot 10^9$	0	8547429	0,0078
X_2	10^9	8500	1341435	0,00133
X_3	$1,1 \cdot 10^7$	1500	15604	$1,28 \cdot 10^{-3}$
X_4	$1,8 \cdot 10^{11}$	10^7	12289233061	0,068
X_5	$9,3 \cdot 10^9$	0	104077420	0,0158
X_6	146	-67	-0,1615	0,314
X_7	1005	273	475	0,276
X_8	73,5	0,4	9,7	0,127
X_9	458592	8219	92481	0,187
\tilde{F}	746	-216	0,0499	–

Необхідність створення та навчання ANFIS моделі, що містить 36 вхідних полів і одне вихідне, при потужності навчальної вибірки 2684 записи вимагає величезних комп'ютерних ресурсів. Час навчання для однієї епохи залежно від кількості вхідних параметрів, приведено у табл. 5 (AMD Athlon 64, Core 4*3000 GHz, 8 Gb RAM).

Таблиця 5. Час навчання ANFIS моделі для однієї епохи (AMD Athlon 64, Core 4*3 000 GHz, 8 Gb RAM)

Кількість входів	1	2	3	4	5	6	7
Час (с)	0,26	0,27	0,42	1,24	7,3	116	1718
Час (хв)	0,004	0,004	0,007	0,02	0,121	1,933	28,633

Як видно з табл. 5, час навчання однієї епохи експоненційно зростає. Починаючи з 8 вхідних параметрів, обсягу оперативної пам'яті комп'ютера недостатньо для навчання і необхідно підключати файл підкачки, що сповільнює розрахунок близько в 10 разів. Крім того, для якісного навчання необхідно, як показали розрахунки, близько 100 епох навчання. Отже, навіть у випадку перевірки гіпотези окремо для кожного лага (9 вхідних параметрів) комп'ютерний час навчання вимірюватиметься місяцями. Тому для перевірки гіпотези необхідно або використовувати специфічні методи навчання чи трансформації даних, або скористатись методами системного аналізу для пошуку найбільш вагомих параметрів. Так як навчання гібридним методом ANFIS моделі, що містить 36 вхідних полів, може оцінюватись роками, то розробка нового математичного апарату, який зможе на 3–4 порядки зменшити час навчання, виглядає сумнівним. Однак ця задача є актуальною для подальших досліджень. Тому в роботі ми зосередились на факторному аналізі та методах системного аналізу.

Факторний аналіз. Суть полягає у виділенні з множини характеристик об'єкта нових факторів, які більш адекватно відображають властивості об'єкта. В ході аналізу створюються нові фактори, які є лінійними комбінаціями колишніх і «вбирають» у себе велику частину загальних флуктуацій вхідних факторів. Тому вони містять велику частину інформації, що міститься в первинній навчальній вибірці. В роботі був використаний класичний метод головних компонент [30]. Цей метод зводиться до вибору нової ортогональної системи координат у просторі спостережень. Як першу головну компоненту обирають напрямком, уздовж якого масив даних має найбільший розкид. Вибір кожної наступної головної компоненти відбувається так, щоб розкид даних уздовж неї був максимальним і щоб ця головна компонента була ортогональна іншим головним компонентам, обраним перед цим.

Так як обсяг бази даних не дозволяє аналізувати ANFIS моделі з кількістю вхідних параметрів, що перевищують 7–8, а факторний аналіз не дав змоги зменшити кількість вхідних полів з 36 до 8, було прийнято рішення провести факторний аналіз окремо для кожного лага. Для цього були створені 6 навчальних множин для лагів від 0 до 5:

$$Tr_L = \left\{ \left\langle x_{1,j}^{-L}, \dots, x_{9,j}^{-L}, \tilde{f}_j \right\rangle \right\}_{j=1, \bar{n}}, L = \overline{0,5}. \quad (4)$$

У результаті факторного аналізу вдалось зменшити розмірність з 9 до 6: $Tr_L = \left\{ \left\langle x_{1,j}^{-L}, \dots, x_{9,j}^{-L}, \tilde{f}_j \right\rangle \right\}_{j=1, \bar{n}} \rightarrow \left\{ \left\langle \xi_{1,j}^L, \dots, \xi_{6,j}^L, \tilde{f}_j \right\rangle \right\}_{j=1, \bar{n}} = \Xi_L$. Отримані навчальні множини використовувались при навчанні ANFIS моделей.

У результаті навчання були отримані продукційні бази знань, що містять по 729 нечітких правил кожна. Для перевірки точності моделей був проведений кореляційний аналіз між реальними значеннями кількості пожеж \tilde{F} і прогнозованими за допомогою моделей M_L для кожного лага окремо. Результати аналізу приведено в табл. 6.

Таблиця 6. Залежність коефіцієнта кореляції для ANFIS моделей на основі факторного аналізу

Лаг	0	1	2	3	4	5
Коеф. кореляції	0,22	0,24	0,23	0,22	0,22	0,24

Як видно з табл. 6, всі коефіцієнти кореляції є достатньо низькими і не залежать від лага. Такий результат протирічить нашим попереднім розрахункам [31], в яких було чітко доведено лагову залежність та досягнуто суттєво вищих коефіцієнтів кореляції (в межах 0,5). Це означає те, що ці моделі є неадекватними і не можуть використовуватись у подальшому аналізі. Це пов'язано з тим, що лінійна трансформація вхідних полів у рамках факторного аналізу «обрізає» всі нелінійні залежності, без яких гіпотеза про вплив сонячної активності на настання лісових пожеж не підтверджується. Отже, це є ще одним доказом нелінійності задачі. Тому будь-які лінійні перетворення можуть призвести до втрати прихованих залежностей і недопустимі в задачах такого типу.

Встановлення вагомих факторів. Для встановлення найбільш вагомих факторів був використаний метод повного перебору. Основна ідея полягає в тому, що із усієї множини вхідних параметрів, яка із урахуванням лага складає $n = 36$ полів, слід перевірити всі можливі комбінації ANFIS моделей, що містять $m < n$ вхідних параметрів. Згідно з теорією ймовірності ця кількість моделей визначається як

$$C_n^m = \frac{n!}{m!(n-m)!}. \quad (5)$$

Для визначення оптимального значення m було проведено оцінку приблизного часу навчання для 100 епох (табл. 7).

Таблиця 7. Час навчання ANFIS моделей (100 епох) залежно від кількості вхідних параметрів

Кількість вхідних параметрів m	Тривалість навчання (с) для 1 моделі	Множина вхідних параметрів моделей $n=36$	Тривалість навчання			
			Секунд	Хвилин	Годин	Діб
1	26	36	936	15,6	0,26	0,010833
2	27	630	17010	283,5	4,725	0,196875
3	42	7140	299880	4998	83,3	3,470833
4	124	58905	7304220	121737	2028,95	84,53958
5	730	376992	2,75E+08	4586736	76445,6	3185,233
6	11600	1947792	2,26E+10	3,77E+08	6276219	261509,1
7	171800	8347680	1,43E+12	2,39E+10	3,98E+08	16598743

Як видно з табл. 7, при кількості вхідних параметрів, що дорівнює 3, необхідно побудувати 7140 моделей і на це потрібно затратити близько 3,4 доби. Вже при 4-х вхідних параметрах цей час складає близько 84 діб, тобто трохи менше 3-х місяців. Як відомо із системного аналізу [32], у складних системах є, як правило, не більше 3–4 ключових параметрів. Тому було прийнято рішення провести 3 експерименти:

Експеримент 1 (3 вхідних параметри).

- Провести навчання для ANFIS моделей, що містять 3 вхідних поля, та для кожної з них оцінити точність за допомогою коефіцієнта кореляції.
- Вибрати найбільш вагомні фактори (моделі, для яких дають найвищий коефіцієнт кореляції) та на основі їх побудувати одну ANFIS модель.

Експеримент 2 (4 вхідних параметри).

- Провести навчання для ANFIS моделей, що містять 4 вхідних поля для кожного лага окремо, та для кожної з них оцінити точність за допомогою коефіцієнта кореляції.
- Вибрати найбільш вагомні фактори для кожного лага окремо та сформувану навчальну вибірку з цих факторів.

Експеримент 3 (4 вхідних параметри).

На основі навчальної вибірки з попереднього кроку провести навчання для ANFIS моделей, що містять 4 вхідних поля, для кожної з них оцінити точність за допомогою коефіцієнта кореляції

- Вибрати найбільш вагомні фактори та на основі їх побудувати одну ANFIS модель.

Так як кількість моделей, які необхідно побудувати, сягає декількох тисяч, для реалізації вищезазначених експериментів була розроблена програма, що автоматизує всі кроки експериментів в середовищі Matlab & Simulink. Побудована засобами імітаційного моделювання програма, з одного боку, дозволяє візуалізувати процес аналізу, а з іншого, автоматизувати всі етапи незалежно від початкових даних.

5. Результати розрахунків

Визначення ключових факторів.

Експеримент 1 (3 вхідних параметри).

Як було показано вище, в ході експерименту було побудовано, навчено та протестовано в автоматичному режимі 7140 ANFIS моделей. Це дало змогу встановити найбільш вагомі фактори. Як показали розрахунки, фактори, які характеризують енергію протонів та електронів, не впливають на активність настання лісових пожеж. Це підтверджує попередні висновки. В тестових моделях, що містили по три вхідних фактори, найбільший коефіцієнт кореляції отримувався при комбінаціях таких шести факторів: X_7 (lag=1; 3); X_6 (lag=0; 2) та X_9 (lag=0; 1). До факторів за значимістю можна віднести X_8 (lag=1; 2); X_7 (lag=2); X_6 (lag=1). Тобто на останньому кроці експерименту необхідно дослідити модель, що містить 6 факторів або ще додатково 4. Так як розрахункові потужності не дозволяють дослідити модель, що містить 10 вхідних факторів, була побудована, навчена та досліджена ANFIS модель, що містить 6 зазначених факторів (експеримент 1). Сумарний час розрахунку експерименту становив 86 годин.

Експеримент 2 (4 вхідних параметри).

Як було зазначено вище, в ході першого етапу розрахунку було побудовано, навчено та досліджено для кожного лага, що містить по 9 факторів, по 126 ANFIS моделей, що містять по 4 фактори. Це дало змогу визначити ключові поля для кожного лага окремо. До ключових факторів також не потрапили характеристики високоенергетичних протонів та електронів. Це дало змогу виділити 17 ключових полів та продовжити наступний етап, що складався з автоматичної побудови 2380 ANFIS моделей. В ході цього етапу було виділено такі ключові поля: X_7 (lag=1 та 2); X_6 (lag=0 та 2); X_9 (lag=3) та X_8 (lag=1). До них за важливістю віднесено: X_6 (lag=1 та 3) та X_8 (lag=2). Три ключові поля для обох експериментів співпадають. Аналогічно попередньому випадку була досліджена ANFIS модель, що містить 6 ключових полів (експеримент 2). Отже, в цьому експерименті було побудовано 2885 моделей та затрачено 103 години машинного часу.

Експеримент 3 (4 вхідних параметри).

Як показали попередні експерименти, фактори, які характеризують енергію електронів і протонів, не впливають на спалахи лісових пожеж. Тому в даному експерименті було запропоновано вилучити з початкової навчальної вибірки ці поля. В результаті кількість вхідних факторів з 36 зменшилась до 16 (по 5 факторів для кожного лага були вилучені). Після чого слід провести експеримент, аналогічний першому, але враховувати не 3 а 4 вхідні фактори. В результаті було проаналізовано 1820 моделей в автоматичному режимі. Як виявилось, більшість ключових полів співпадають з полями, отриманими в попередніх експериментах. До них відносяться: X_7 (lag=0, 1 та 3); X_6 (lag=0 та 2); X_9 (lag=3). До них за важливістю віднесено: X_6 (lag=3); X_7 (lag=2) та X_8 (lag=0). Аналогічно попереднім експериментам було отримано 6 ключових полів, побудована і навчена ANFIS модель, що містить їх як вхідні поля (експеримент 3). Всього на цей експеримент було витрачено 66 годин машинного часу.

Аналіз адекватності та точності моделей. Так як в результаті експериментів отримані різні ключові поля, то необхідно провести аналіз адекватності й точності отриманих моделей. Результати моделювання представлені на рис. 4. Як видно з рисунків, графіки мають багато спільних піків, однак їх велика кількість потребує комп'ютерного аналізу.

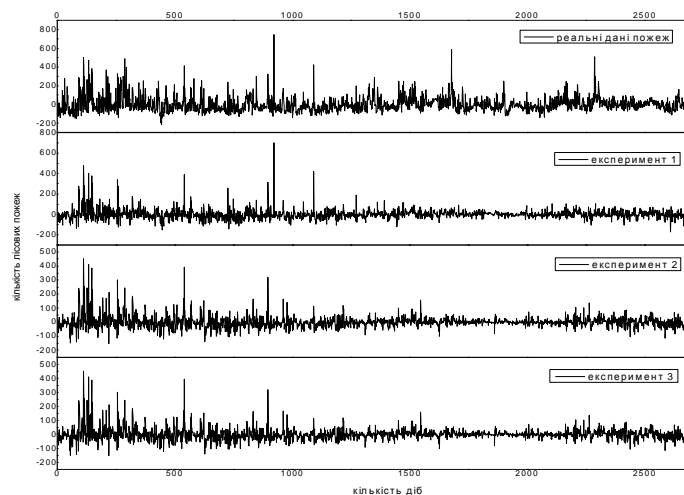


Рис. 4. Порівняння результатів моделювання з реальними даними

Для перевірки отриманого висновку про залежність від часового лага був проведений порівняльний аналіз за допомогою коефіцієнта кореляції та підрахунку збігів числа пожеж між реальними даними та моделями. Також були проаналізовані помилкові піки та відмінність амплітуди піків. Як можна побачити з рисунку, всі моделі дають можливість пояснити основний вплив сонячної активності на лісові пожежі. Модельні піки за положенням та амплітудою є близькими до реальних. Це вказує на достовірність даних моделей. Однак для точного аналізу необхідно кількісно підрахувати вищезазначені показники. Результати комп'ютерних розрахунків наведені в табл. 8.

Таблиця 8. Аналіз точності пожеж, проведених методом *ANFIS*

№ моделі	Коефіцієнт кореляції	Реальні спалахи пожеж	Модельні спалахи пожеж	Піки, пояснені спалахами моделі пожеж		Середнє відхилення інтенсивності	Помилкові піки		Непояснені піки	
				кількість	відсоток		кількість	відсоток	кількість	відсоток
1	0,617258	785	807	367	47%	-4,0%	440	53%	144	18%
2	0,558773	785	848	366	47%	-7,5%	482	53%	151	19%
3	0,588654	785	846	365	46%	-4,5%	481	54%	166	21%

Як видно з таблиць, розроблені моделі характеризуються високою точністю прогнозу, про що свідчить високий коефіцієнт кореляції. Як видно з табл. 2, нам удалось у великій мірі усунути нелінійність та підняти коефіцієнт кореляції з $\sim 0,02$ до $\sim 0,6$. Цей показник є вищим, ніж у нашій попередній роботі [33], не зважаючи на те, що там були враховані 8 вхідних факторів. Це свідчить про адекватність підходу. Крім того, всі показники, наведені в таблиці для даних трьох моделей, є кращими, ніж у попередніх роботах [36]. Як видно з таблиці, всі три моделі можуть пояснити до 47% спалахів пожеж (стовбець 6). Слід зазначити, що в середньому до 53–54% прогнозованих спалахів пожеж виявляються помилковими (колонка 9). Тобто, згідно з прогнозом має настати пожежа, а насправді пожежі не сталось. Однак більш важливою інформацією є те, скільки реальних пожеж розроблені моделі не в змозі передбачити. Щоб перевірити це, було підраховано кількість випадків, коли на графіку реальних пожеж спостерігалися піки, а на модельних графіках значення

було нижче за середнє (колонка 11). Як показали розрахунки, тільки 18–21% реальних спалахів пожеж не можуть бути передбачені розробленими моделями.

Цікавою є інформація про прогнозовану амплітуду піків у порівнянні з реальними піками на рис. 6. Тобто, як співвідноситься кількість прогнозованих спалахів пожеж у конкретний день з реальною кількістю пожеж, зареєстрованими в цей самий день. Як показано в таблиці (7 колонка), прогнозована амплітуда зазвичай є меншою в середньому на 4,0–7,5%, ніж фактичне число спалахів.

Незважаючи на точність прогнозування, як за часом, так і за амплітудою, ці моделі не дозволяють передбачати географічне положення джерел пожеж. Причина полягає у відсутності геопросторової інформації в навчальній вибірці. Даний недолік може бути усунутий, якщо долучити цю інформацію до бази даних.

Як видно з таблиці, точність усіх трьох моделей є близькою. Тому необхідно провести аналіз чутливості, здатний дати відповідь, яка з моделей є більш адекватною.

Для цього значення всіх вхідних факторів усереднюються (табл. 4) і досліджується залежність кількості спалахів пожеж, що прогнозує модель, від послідовних змін кожного фактора. Результати цього аналізу представлені на рис. 5.

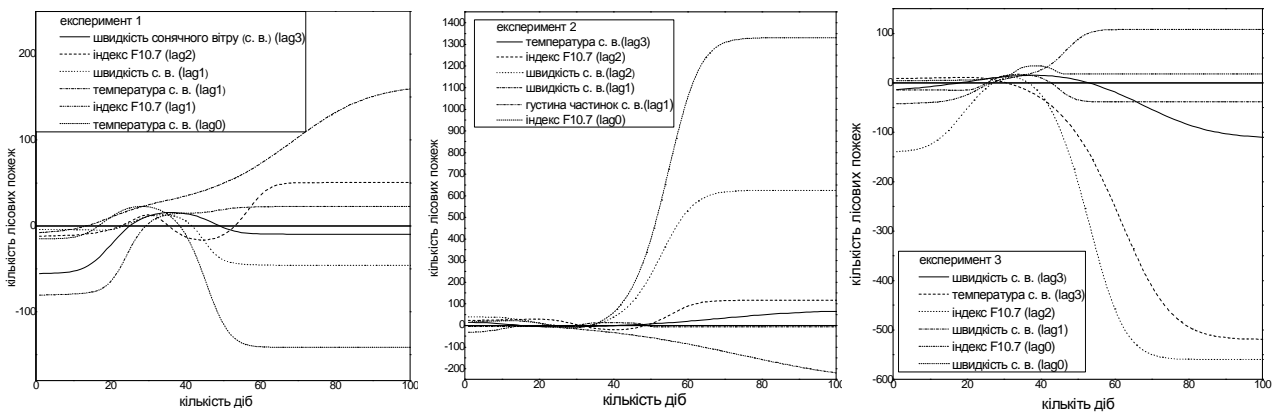


Рис. 5. Чутливість спалахів лісових пожеж від зміни X_i

Для перевірки адекватності треба уточнити, чи відображають графіки очевидну поведінку залежностей.

Очевидно, що при зменшенні інтенсивності параметрів, які характеризують сонячну активність, кількість лісових пожеж зменшується або залишається незмінною. Це яскраво спостерігається на всіх трьох графіках. Навпаки, при збільшенні значень вхідних параметрів кількість пожеж зростає. Найкраще процеси спостерігаються в експерименті 2. Згідно з цією моделлю, найбільш вагомими параметрами є X_6 (lag=0) та X_7 (lag=2). Для першої і третьої моделей спостерігається зовсім інша картина. Зокрема, для експерименту 3 спостерігається зменшення кількості пожеж при зростанні значень X_9 (lag=3) та X_6 (lag=2). Аналогічно спостерігається і для трьох факторів експерименту 1: X_7 (lag=1) та X_9 (lag=0;1). Це аномальна поведінка залежностей.

Отже, доведено, що найбільш адекватною моделлю є експеримент 2. Однак для більш точного розрахунку необхідно провести ANFIS моделювання, яке враховує всі вищезазначені фактори.

6. Висновки

У результаті дослідження були встановлені ключові фактори і часова затримка, які стали основою при побудові трьох ANFIS моделей, що дозволили встановити в явному вигляді

функціональну залежність між характеристиками сонячної активності та настанням лісових пожеж.

У ході дослідження було встановлено, що ця задача відноситься до задач нелінійного програмування і будь-які лінійні перетворення можуть призвести до втрати прихованих залежностей, які недопустимі в задачах такого типу.

У результаті аналізу чутливості було встановлено, що отримані моделі дають змогу пояснити до 47% лісових пожеж, що, у свою чергу, добре корелює з кількістю непояснених лісових пожеж (39,4% в США). Відхилення по амплітуді в межах 5% свідчить про точність моделювання. А непояснені лісові пожежі в рамках цих розрахунків якраз і відносяться до тих, що пов'язані з людським фактором, блискавками чи спекою.

Отже, отримані моделі є адекватними та точними і можуть бути використані при побудові інформаційних систем запобігання кризовим явищам.

СПИСОК ЛІТЕРАТУРИ

1. Nikolov N. Global Forest Resources Assessment 2005 [Електронний ресурс] / N. Nikolov // Report on fires in the Balkan Region. Forestry Department, FAO of the UN, Fire Management Working Papers FM/11/E, Rome. – 2006. – 320 p. – Режим доступу: <http://www.fao.org/docrep/009/j7567e/j7567e00.htm>.
2. Forest Fires in Europe, Middle East and North Africa 2011 [Електронний ресурс] / G. Schmuck, J. San-Miguel-Ayanz, A. Camia [et al.] // Joint Research Centre of the European Commission, 2012. – 109 p. – Режим доступу: http://forest.jrc.ec.europa.eu/media/cms_page_media/9/forest-fires-in-europe-2011.pdf.
3. FAO Forestry country information. Fire Situation in India [Електронний ресурс]. – Режим доступу: <http://www.fao.org/docrep/006/ad653e/ad653e50.htm>.
4. The National Interagency Fire Center (NIFC) [Електронний ресурс] // National Report of Wildland Fires and Acres Burned by State. Figures from the Fire and Aviation Management Web Applications Program. – Режим доступу: http://www.predictiveservices.nifc.gov/intelligence/2013_Statsumm/fires_acres13.pdf.
5. Radovanovic M. Solar Activity and Forest Fires / M. Radovanovic, J.F.P. Gomes. – Nova Science Publishers Inc., 2009. – 109 p.
6. Radovanović M. The Heliocentric Hypothesis on the Origin of Forest Fires (Abstract book) / M. Radovanović, M. Pecelj // International Scientific Conf. “Forestry Science and Practice for the Purpose of Sustainable Development of Forestry”. – Banja Luka, Republic of Srpska/B&H, 2012. – 1-4th November. – 10 p.
7. Radovanović M. The Heliocentric Analysis of the Forest Fire Phenomenon (Abstract book) / M. Radovanović // Fires in Russia on August 2010, Case Study. – 2nd Serbian Geographers' Congress. – Novi Sad, 2010. – 10-11 December. – 30 p.
8. Radovanović M. Heliocentric Electromagnetic Long-Term Weather Forecast and its Applicable Significance / M. Radovanović, V. Lukić, N. Todorović // Journal of the Geographical Institute “Jovan Cvijić” SASA. – 2005. – N 54. – P. 5 – 18.
9. Voxall M. ESS Guidelines on Seasonal Adjustment. Eurostat [Електронний ресурс] / M. Voxall. – 2009. – Режим доступу: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF.
10. Bell W.R. Economic Time Series: Modeling and Seasonality / W.R. Bell, S.H. Holan, T.S. McElroy // Chapman and Hall/CRC. – 2012. – 544 p.
11. Hansen B.E. Econometrics / Hansen B.E. – University of Wisconsin, Department of Economics, 2014. – 378 p.
12. Labitzke K. The global signal of the 11-year sunspot cycle in the atmosphere: When do we need the QBQ? / K. Labitzke // Meteorolog. Zeitschrift. – 2003. – Vol. 12, N 4. – P. 209 – 216.
13. Lenskiy A.A. The analysis of R/S estimation algorithm with applications to WiMAX network traffic / A.A. Lenskiy, S. Seol // International Journal of Multimedia and Ubiquitous Engineering. – 2012. – Vol. 7, N 3. – P. 27 – 34.

14. Spatial Variability of the Hurst Exponent for the Daily Scale Rainfall Series in the State of Zacatecas, Mexico / V. Velásquez, M.A. García, G. Medina [et al.] // Journal of Applied Meteorology and Climatology. – 2013. – Vol. 52, N 12. – P. 2771 – 2780.
15. Özger M. Prediction of ocean wave energy from meteorological variables by fuzzy logic modeling / M. Özger // Expert Syst. Appl. – 2011. – Vol. 38, N 5. – P. 6269 – 6274.
16. Peters E.E. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics / Peters E.E. – London, UK: John Wiley & Sons, 1994. – 167 p.
17. Examination of the correlations between forest fires and solar activity using Hurst index / M. Radovanović, Y. Vykylyuk, A. Jovanović [et al.] // Journal of the Geographical institute «Jovan Cvijic» SASA. – 2013. – Vol. 63, N 3. – P. 23 – 32.
18. Kourtz P.H. Predicting the daily occurrence of lightning-caused forest fires [Электронний ресурс] / P.H. Kourtz, J.B. Todd // Forestry Canada, Petawawa National Forestry Institute. – Chalk River, Ontario, 1991. – Information Report PI-X-112. – Режим доступу: <http://cfs.nrcan.gc.ca/publications/?id=10706>.
19. Hall L.B. Precipitation associated with lightning-ignited wildfires in Arizona and New Mexico / L.B. Hall // Int. J. Wildland Fire. – 2007. – Vol. 16, N 2. – P. 242 – 254.
20. Cumming S.G. Forest type and wildfire in the Alberta boreal mixedwood: What do fires burn? / S.G. Cumming // Ecol. Appl. – 2001. – Vol. 11, N 1. – P. 97 – 110.
21. Wotton M.B. An index for tracking sheltered forest floor moisture within the Canadian Forest Fire Weather Index System / M.B. Wotton, J.B. Stocks, L.D. Martell // Int. J. Wildland Fire. – 2005. – Vol. 14, N 2. – P. 169 – 182.
22. Forest fires caused by lightning as an indicator of connections between atmosphere, lithosphere, and biosphere / S.N. Sannikov, A.I. Zakharov, L.G. Smol'nikova [et al.] // Russ. J. Ecol.+. – 2010. – Vol. 41, N 1. – P. 1 – 6.
23. Viegas D.X. Forest fire propagation / D.X. Viegas // Phil. Trans. R. Soc. London Ser. A. – 1998. – Vol. 356. – P. 2907 – 2928.
24. Predicting Fire Frequency with Chemistry and Climate / P.R. Guyette, C.M. Stambaugh, C.D. Dey // Ecosystems. – 2012. – Vol. 15, N 2. – P. 322 – 335.
25. Gomes J.F.P. Solar activity as a possible cause of large forest fires – a case study: Analysis of the Portuguese forest fires / J.F.P. Gomes, M. Radovanovic // Sci. Total Environ. – 2008. – Vol. 394, N 1. – P. 197 – 205.
26. Ducic V. Contemporary Climate Variability and Forest Fires in Deliblatska pescara / V. Ducic, M. Milenkovic, M. Radovanovic // Journal of the Geographical institute "Jovan Cvijic" SASA. – 2008. – N 58. – P. 59 – 74.
27. Radovanović M. Forest fires in Europe from July 22nd to 25th 2009 / M. Radovanović // Arch. Biol. Sci. – 2010. – Vol. 62, N 2. – P. 419 – 424.
28. Radovanovic M. Solar Activity – Climate Change and Natural Disasters in Mountain Regions / M. Radovanovic // Sustainable Development in Mountain Regions (ed. Zhelezov G.). Springer Science+Business Media B.V. – 2011. – P. 9 – 17.
29. Voxall M. ESS Guidelines on Seasonal Adjustment. Eurostat [Электронний ресурс] / M. Voxall. – 2009. – Режим доступу: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF.
30. Labitzke K. The global signal of the 11-year sunspot cycle in the atmosphere: When do we need the QBQ? / K. Labitzke // Meteorolog. Zeitschrift. – 2003. – Vol. 12, N 4. – P. 209 – 216.
31. Lenskiy A.A. The analysis of R/S estimation algorithm with applications to WiMAX network traffic / A.A. Lenskiy, S.Seol // International Journal of Multimedia and Ubiquitous Engineering. – 2012. – Vol. 7, N 3. – P. 27 – 34.
32. Vykylyuk Y. Simulation of spatial form of urban systems by diffusion methods / Y. Vykylyuk // Journal of the Geographical Institute "JovanCvijic" SASA. – 2013. – Vol. 63, N 1. – P. 89 – 100; 2013. – Vol. 63, N 2. – P. 67 – 77.
33. Vykylyuk Y. Forecast with neural network: usd/eur currency pair / Y. Vykylyuk, D. Vuković, A.Jovanović // Actual Problems of Economics. – 2013. – N 10. – P. 261 – 273.

Стаття надійшла до редакції 09.02.2015