

КОНЦЕПТУАЛЬНЕ МОДЕЛЮВАННЯ СПАЛАХІВ ЛІСОВИХ ПОЖЕЖ НА ОСНОВІ ОНТОЛОГІЧНОГО ПІДХОДУ DATAMINING. ЧАСТИНА 2

**М. РАДОВАНОВІЧ, Я.І. ВИКЛЮК, М. МІЛЕНКОВІЧ, А. ЙОВАНОВІЧ,
Д. ВУКОВІЧ, М. СТЕВАНЧЕВІЧ, Н.О. МАЦЮК, Т.Б. ЛЕКО**

Протестовано геліоцентричну гіпотезу причин спалахів лісових пожеж. Знайдено докази кореляції між раптовим надходженням заряджених частинок з боку сонця і виникненням лісових пожеж із затримкою від одного до чотирьох днів. Проведено порівняльний аналіз методів ANFIS та нейронних мереж в задачі пошуку функціональної залежності між виникненням лісових пожеж і факторами, що характеризують сонячну активність. Для цієї мети використано декілька методів аналізу (методи усунення сезонності, *R/S* аналіз, DataMining) для встановлення потенційних зв'язків між коливаннями певних параметрів, що характеризують сонячну активність, і виникненням лісових пожеж з врахуванням затримки в часі. Знайдено наявність взаємозв'язку і розроблено прогностичний сценарій, заснований на ANFIS та нейромережових технологіях. Ці методи, в деяких випадках, дозволяють досягнути точності прогнозування до 93%.

НЕЙРОННІ МЕРЕЖІ

У роботі було досліджено два типи багат шарових нейронних мереж:

- нейронні мережі з прямим поширенням помилки;
- нейронні мережі із зворотнім поширенням помилки.

Для визначення необхідної кількості нейронів були використані емпіричні формули [1]:

$$\frac{mN}{1 + \log_2 N} \leq L_w \leq m \left(\frac{m}{N} + 1 \right) (n + m + 1), \quad (1)$$

$$L = \frac{L_w}{n + m}, \quad (2)$$

$$\frac{N}{10} - n - m \leq L \leq \frac{N}{2} - n - m, \quad (3)$$

$$2(L + n + m) \leq N \leq 10(L + n + m), \quad (4)$$

де N — число елементів навчальної вибірки, m — розмірність вихідного сигналу, n — розмірність вхідного сигналу, L_w — необхідне число синаптичних ваг, L — кількість елементів масиву.

У нашому випадку навчальна вибірка містить 7 входів ($n = 7$) і 1 вихід ($m = 1$). Число елементів навчальної вибірки залежить від лагу, при збільшенні якого зменшується кількість елементів навчальної вибірки (табл. 1).

Таблиця 1. Залежність числа елементів навчальної вибірки від лагу

lag	0	1	2	3	4	5
<i>N</i>	710	706	702	698	694	690

У результаті розрахунків було отримано: $63 \leq L \leq 347$. Отже, сумарна кількість нейронів має бути більшою за 63 та меншою за 347. Тому в розрахунках було використано нейронні мережі, що містять 50×50 та 60×60 нейронів у прихованих прошарках для двох типів вищезазначених нейронних мереж (рис. 1).

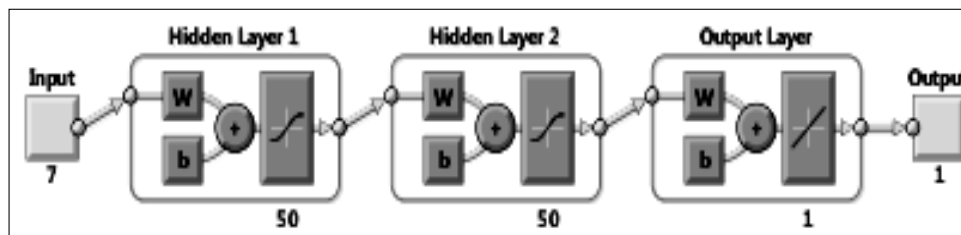


Рис. 1. Схема нейронної мережі з 7 входами та 1 виходом розмірністю 50×50

ГІБРИДНІ НЕЙРОННІ МЕРЕЖІ (ANFIS)

ANFIS — нейронна мережа, що базується на основі нечіткої системи виводу Такагі-Сугено. Ця система виведення містить набір нечітких *If-Then* правил, які отримують під час навчання на великих базах даних на основі нелінійних функцій [2–3]. Ці методи добре зарекомендували себе в моделюванні складних соціальних систем у наших попередніх роботах [4].

Для побудови нечіткої моделі всі вхідні параметри представлено як лінгвістичні змінні. Як було зазначено, у 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 X_i (21 терм для кожної моделі). У випадку 2-х термів моделі не були адекватними. Якщо ж цих термів більше, ніж 3, кількість емпіричних параметрів перевищує обсяг навчальної вибірки, що унеможливує процес 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 результаті було отримано продуктивні бази знань, що містять 6561 нечітке правило.

РЕЗУЛЬТАТИ РОЗРАХУНКІВ

Отже, у результаті навчання для кожного з шести лагів було отримано по 4 моделі на основі нейронних мереж (всього 24) та по 1 (всього 6) на основі

ANFIS для великих і малих пожеж відповідно (разом: 48 — нейронні мережі, 12 — ANFIS). Як відомо, результат навчання нейромережі залежить від конфігурації, методу навчання, стохастичних параметрів. Результат навчання ANFIS характеризується більшою стійкістю під час навчання. Тому для аналізу були взяті усереднені дані для 4-х нейронних мереж. Як було показано в роботі [5], такий підхід дозволяє відсіяти випадкові флуктуації у функціонуванні нейронних мереж, а, отже, досягти кращих результатів. Для перевірки точності моделей було проведено кореляційний аналіз між реальними значеннями кількості пожеж $\tilde{F}^{\text{small}(\text{large})}$ і прогнозованими за допомогою моделей $M_L^{\text{small}(\text{large})}$ для кожного лагу окремо. Це дало змогу встановити інтервал часу між початком лісових пожеж та сонячною активністю (рис. 2).

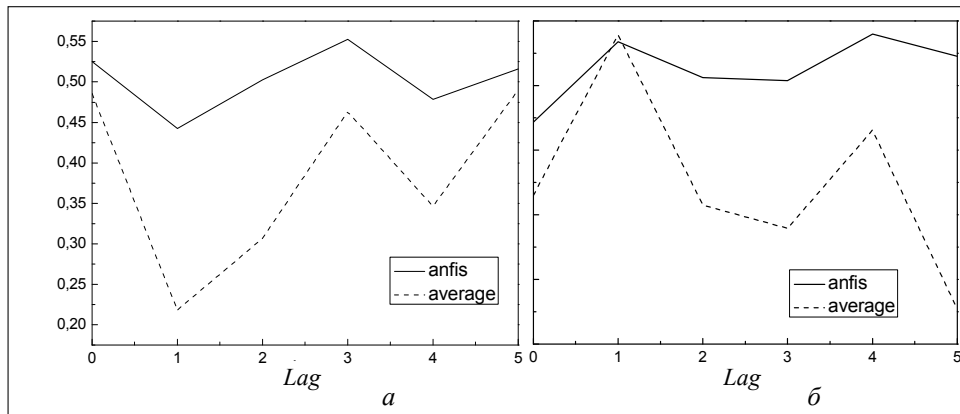


Рис. 2. Залежність коефіцієнта кореляції від лагу L : a — $\tilde{F}^{\text{small}(\text{large})}$, b — $M_L^{\text{small}(\text{large})}$

Як видно з рис. 2, коефіцієнти кореляції для моделі на основі ANFIS є більшими за нейронні мережі. Крім того видно, що на графіках спостерігаються схожі тенденції, зокрема у випадку великих пожеж наявні незначні піки для $lag = 1, 4$. Отже, можна дійти висновку, що існує затримка в 1 або 4 доби від початку сонячної активності і настанням великих лісових пожеж, що спричинені нею. Аналогічна ситуація спостерігається для невеликих пожеж. Однак максимальна кореляція спостерігається при $lag = 0, 3$. Як видно з графіку ANFIS, різниця між коефіцієнтами кореляції є незначною. Натомість нейромережі демонструють більшу «чутливість» до лагу, незважаючи на те, що абсолютні значення коефіцієнта кореляції є меншими. Для перевірки отриманого висновку про залежність від часового лагу було проведено порівняльний аналіз збігів числа малих і великих лісових пожеж між реальними даними та моделями (рис. 3). Також було проаналізовано помилкові піки та відмінність амплітуди піків.

За даними з рис. 3 всі моделі дають можливість пояснити основний вплив сонячної активності на малі та великі лісові пожежі. Модельні піки за положенням та амплітудою є близькими до реальних. Це вказує на достовір-

ність цих моделей. Однак для точного аналізу необхідно кількісно підрахувати вищезазначені показники.

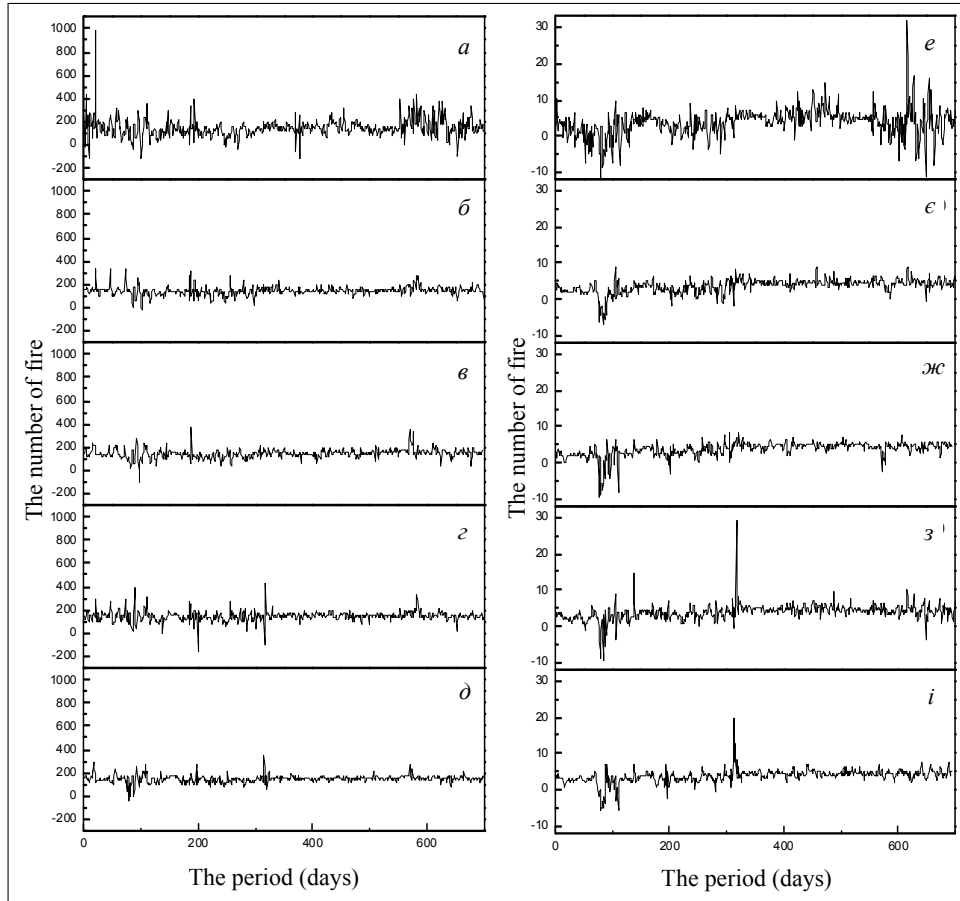


Рис. 3. Порівняння результатів моделювання з реальними даними. Малі пожежі: *a* — реальні дані, *б* — ANFIS ($lag = 0$), *в* — ANFIS ($lag = 3$), *г* — нейронні мережі ($lag = 0$), *д* — нейронні мережі ($lag = 3$). Великі пожежі: *e* — реальні дані, *е* — ANFIS ($lag = 1$), *ж* — ANFIS ($lag = 4$), *з* — нейронні мережі ($lag = 1$), *и* — нейронні мережі ($lag = 4$)

Для перевірки точності результатів, було проведено порівняльний аналіз між числом, положенням, амплітудою реальних спалахів пожеж (піки на рис. 3, (*a, e*)) і спалахами пожеж, що прогнозуються моделями (піки на рис. 3. (*б, в, г, д, е, ж, з, и*)). Розглянемо два випадки:

- порівняльний аналіз прогнозованих спалахів пожеж з точністю в межах доби (тобто прогноз вважається правильним, якщо пожежа настає у той самий день, що й згідно з прогнозом);
- порівняльний аналіз прогнозованих спалахів пожеж з точністю в межах однієї однієї доби (тобто, якщо модель передбачає спалах пожежі наприклад у середу, а реальна пожежа сталась у проміжку з вівторка до четверга, то прогноз вважається точним).

Результати розрахунків наведено в таблиці 2 та 3.

Таблиця 2. Аналіз точності пожеж, проведених методом ANFIS

Lag	Реальна кількість пожеж	Прогнозовані спалахи пожеж	Пояснені моделлю спалахи пожеж з точністю 1 доба		Середня різниця в амплітуді	Помилкові піки з точністю 1 доба		Не пояснені моделлю спалахи пожеж	Пояснені моделлю спалахи пожеж з точністю ±1 доба		Середня різниця в амплітуді	Помилкові піки з точністю ±1 доба		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Малі пожежі														
0	207	189	73	35%	-4,6%	116	61%	48	23%	169	82%	-4,4%	20	11%
1	206	187	59	29%	-7,7%	128	68%	53	26%	170	83%	-3,4%	17	9%
2	204	197	78	38%	-5,1%	119	60%	44	22%	178	87%	-3,4%	19	10%
3	202	185	78	39%	-5,1%	107	58%	41	20%	170	84%	-2,1%	15	8%
4	202	180	65	32%	-1,2%	115	64%	42	21%	162	80%	-2,9%	18	10%
5	201	182	76	38%	6,7%	106	58%	38	19%	159	79%	-4,1%	23	13%
Великі пожежі														
0	229	191	71	31%	11,6%	120	63%	55	24%	186	81%	-6,2%	5	3%
1	229	210	82	36%	2,1%	128	61%	60	26%	210	93%	2,3%	0	0%
2	226	194	75	33%	-1,9%	119	61%	52	23%	194	86%	2,8%	0	0%
3	225	189	66	29%	-2,4%	123	65%	58	26%	188	88%	13,7%	1	1%
4	223	193	69	31%	33,1%	124	64%	56	25%	177	79%	3,2%	16	8%
5	222	197	71	32%	13,3%	126	64%	56	25%	189	85%	22,3%	8	4%

Таблиця 3. Аналіз точності пожеж, проведених методом нейронних мереж

Lag	Реальна кількість пожеж	Прогнозовані спалахи пожеж	Пояснені моделлю спалахи пожеж з точністю 1 доба		Середня різниця в амплітуді	Помилкові піки з точністю 1 доба		Не пояснені моделлю спалахи пожеж	Пояснені моделлю спалахи пожеж з точністю ±1 доба		Середня різниця в амплітуді	Помилкові піки з точністю ±1 доба		
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	
Малі пожежі														
0	207	204	88	43%	-7,4%	116	57%	39	19%	185	89%	0,6%	19	9%
1	206	187	70	34%	-7,4%	117	63%	54	26%	164	80%	6,6%	23	12%
2	204	196	75	37%	-3,4%	121	62%	40	20%	171	84%	-1,4%	25	13%
3	202	207	85	42%	-5,3%	122	59%	41	20%	185	92%	1,5%	22	11%
4	202	197	74	37%	-8,4%	123	62%	47	23%	175	87%	3,3%	22	11%
5	201	204	75	37%	7,7%	129	63%	49	24%	182	91%	-1,0%	22	11%
Великі пожежі														
0	229	176	57	25%	39,03%	119	68%	59	26%	175	76%	-7,63%	1	1%
1	229	201	82	36%	-2,92%	119	59%	54	24%	197	86%	11,56%	4	2%
2	226	198	68	30%	0,27%	130	66%	63	28%	196	87%	8,75%	2	1%
3	225	179	73	32%	22,74%	106	59%	48	21%	169	75%	6,52%	10	6%
4	223	193	68	30%	-9,68%	125	65%	60	27%	180	81%	2,75%	13	7%
5	222	186	54	24%	7,87%	132	71%	63	28%	178	80%	-2,58%	8	4%

Як видно з табл. 2–3, розроблені моделі характеризуються високою точністю прогнозу в наближенні однієї доби (стовпець 4). Найбільша точність прогнозу невеликих пожеж спостерігається при $lag = 0$ та $lag = 3$ (також високою є точність для $lag = 5$). У випадку великих пожеж найбільш точними виявились моделі з $lag = 1$. Моделі на основі ANFIS показують також високу точність для лагу 4 та 5. ANFIS моделі можуть передбачити до 39% малих пожеж і 36% великих пожеж з точністю прогнозу в одну добу. Моделі на основі нейронних мереж показують більшу точність для прогнозу малих пожеж — 43%. У випадку великих пожеж точність залишається такою самою. Отже, ці результати підтверджують попередні висновки кореляційного аналізу (рис. 2) в розрізі залежності від часової затримки. Тобто затримка між спалахом на сонці та спалахом малих пожеж становить 0 або 3 доби, що свідчить про наявність декількох механізмів, які призводять до спалаху лісових пожеж. Стосовно великих пожеж — затримка становить одну добу (підтверджено 3-ма моделями) 4–5 діб (підтверджено 2-ма моделями).

Якщо ж розглядати точність у межах ± 1 доби, то результати прогнозу стають надзвичайно оптимістичними (стовпець 8): за допомогою ANFIS можна отримати прогноз до 87 % малих пожеж ($lag = 2$) і 93 % великих пожеж ($lag = 1$). Нейронні мережі показали знову більшу точність для малих пожеж: до 89% ($lag = 0$) та 92% ($lag = 3$). У випадку великих пожеж прогноз дещо гірший: 86–87% ($lag = 1$ та 2). Отже, у цій точності прогнозу менш ніж 21% (100% — колонка 8) спалахів лісових пожеж не залежить від активності сонця.

Варто зазначити, що якщо розглядати точність прогнозу в межах однієї доби, то в середньому до 57–65% прогнозованих моделлю спалахів пожеж (для встановлених «точних» лагів) на практиці не спостерігаються (колонка 6 табл. 2–3). Ці помилкові прогнози властиві як великим, так і малим пожежам. Однак на практиці важливіше знати, 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 7, табл. 2–3). За розрахунками, тільки 19–26% реальних спалахів малих пожеж не можуть бути передбачені розробленими моделями. Для великих пожеж це число складає приблизно 23–27%.

Однак, якщо точність передбачення складає ± 1 добу, то кількість помилкових піків є меншою за 13 % для всіх розрахунків (10 стовпець). Також відсутні спалахи реальних пожеж, що неможливо передбачити.

Цікавою є інформація щодо прогнозованої амплітуди піків у порівнянні з реальними піками на рис. 3. Тобто, як співвідноситься кількість прогнозованих спалахів пожеж у конкретний день з реальною кількістю пожеж, зареєстрованими в цей самий день. Як показано в табл. 2–3 (5 колонка), у випадку невеликих пожеж амплітуда, зазвичай, є меншою в середньому на 5%, ніж фактичне число спалахів для ANFIS моделей. Нейронні мережі показують дещо гірший результат — (–7%). Якщо прогноз зроблено в наближенні ± 1 доба, то похибка інтенсивності є меншою: (–4) – (–2) % для ANFIS та в межах (–1%)–(–3%) для нейронних мереж (9-й стовпець).

Для великих пожеж — інша ситуація. У випадку «точних» лагів ($lag = 1$), спостерігається найменша похибка по амплітуді: від -2% до 2% (5 колонка). Для $lag = 4$ та 5 спостерігається сильніше відхилення по амплітуді аж до 33% для ANFIS та -9% – 8% для нейронних мереж. У разі точності прогнозу в наближенні ± 1 доба, похибка амплітуди складе від -7% до $+22\%$. Однак для «точних» lag ця помилка складе: $2-3\%$ для ANFIS та $3-11\%$ для нейронних мереж.

Незважаючи на точність прогнозування, як по часу так і по амплітуді, ці моделі не дозволяють передбачати географічне положення джерел пожеж. Причина полягає у відсутності геопросторової інформації в навчальній вибірці. Цей недолік може бути усунено, якщо долучити цю інформацію до бази даних.

Для визначення ступеня залежності кількості спалахів пожеж від зміни вхідних параметрів, було проведено аналіз чутливості. Оскільки результати моделювання методами ANFIS та нейронних мереж показали схожі результати, аналіз чутливості було проведено саме для ANFIS моделей для «точних» лагів. Для цього, значення всіх вхідних факторів усереднюються і досліджується залежність кількості спалахів пожеж, що прогнозує модель, від послідовних змін кожного фактора. Результати цього аналізу представлено на рис. 4.

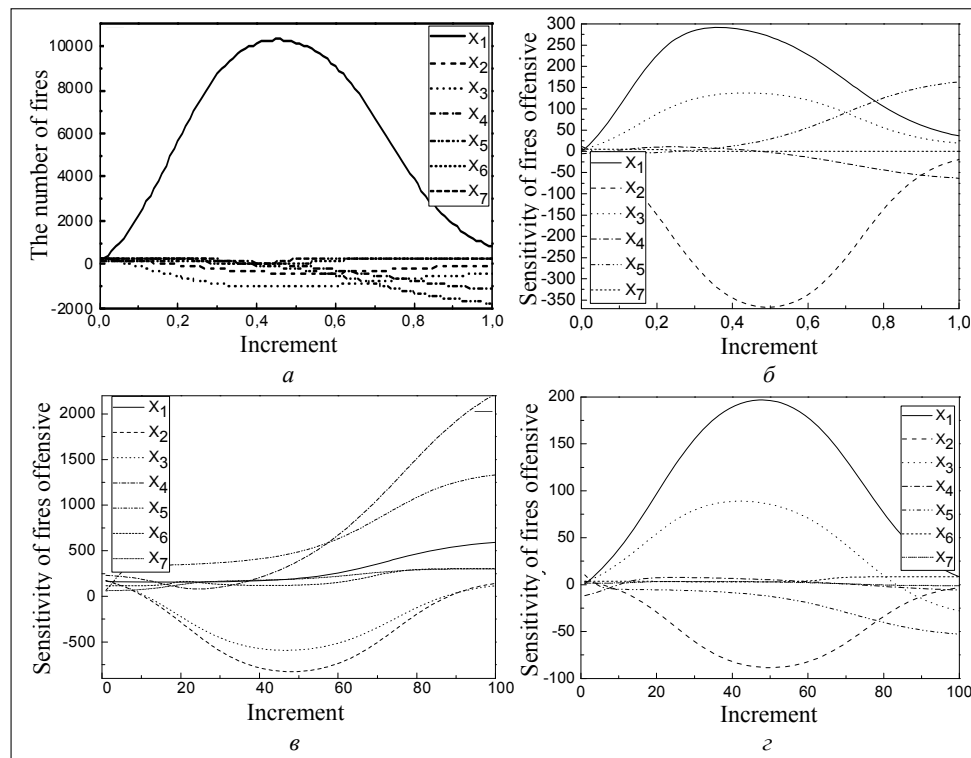


Рис. 4. Залежність кількості пожеж від почергової зміни факторів моделей X_i . Малі пожежі: a — $lag = 0$, $в$ — $lag = 3$, великі пожежі: $б$ — $lag = 1$, $г$ — $lag = 4$

На рис. 4 продемонстровано, що залежність кількості спалахів лісових пожеж від вхідних параметрів є нелінійною. Зокрема, невеликі пожежі є найбільш чутливими до X_1 ($lag = 0$). Залежність від останнього парамет-

ру має квадратичну форму. Коли активність X_1 збільшується від середнього значення 0,008 до 0,5, прогнозована кількість пожеж стрімко зростає. Збільшення цього фактора від 0,5 до 1 призводить до їх зменшення. Це можна пояснити тим, що такого великого збільшення зазначеного параметру без зміни інших факторів ніколи раніше не спостерігалось. Тому діапазон 0,5–1 є неможливим у природі і висновки моделі по ньому можуть не братись до уваги. Параметри $X_2 - X_7$ практично не впливають на кількість невеликих пожеж. Зовсім інша ситуація спостерігається для $lag = 3$ (рис. 4. в). Найбільш чутливим є параметр X_5 : у діапазоні від 0 до 0,1. Від 0,1 до 0,5 цей фактор не впливає на пожежі. Після 0,5, збільшення цього фактора знову призводить до різкого збільшення пожеж. Однак після 0,6 найбільш впливовим параметром стає X_4 .

Інша ситуація спостерігається для великих лісових пожеж. Залежності для $lag = 1$, $lag = 4$ є схожими. Як видно з рис. 4 б, з, найбільш вагомими параметрами є X_1 та X_3 , залежність кількості великих пожеж від їх почергової зміни є аналогічною для X_1 у випадку невеликих пожеж. Залежність від X_5 має експоненційну форму. Як видно з рис. 4 б, кількість великих пожеж стрімко зростає, коли X_5 стає більшою за 0,5 (це характерно лише для $lag = 1$).

Така поведінка потребує подальших експериментальних та теоретичних досліджень для підтвердження чи спростування отриманих висновків.

ВИСНОВОК

У роботі знайдено докази щодо наявності нелінійних зв'язків між настанням лісових пожеж і сонячною активністю. Це дало змогу використати нелінійні методи SoftComputing також для встановлення аналізу функціональних залежностей між ними. Було показано, що між збільшенням сонячної активності і спалахами лісових пожеж існують періоди затримки, які становлять від однієї до декількох діб у випадку великих пожеж. Це дає можливість передбачити настання пожеж і вжити заходів щодо їх запобігання. На основі розроблених нейромережових моделей є можливість передбачити до 89% займань невеликих лісових пожеж у день спалаху на сонці й до 92% із затримкою в 3 доби. У той самий час точність прогнозу для великих пожеж із затримкою в один день становить 93% (для ANFIS). Це справедливо у випадку прогнозування з точністю ± 1 доба. У випадку прогнозування з точністю 1 доби, точність прогнозу спадає до 36–43% і відповідно збільшується кількість помилкових піків. Незважаючи на це, лише 22% випадків пожеж сучасні методи SoftComputing передбачити не можуть. Також варто зазначити, що у всіх випадках відносна похибка під час прогнозування амплітуди кількості пожеж становить менше 5%.

На противагу цьому, варто зазначити, що в постановці задачі існують певні недоліки, які відобразились у результатах. Наприклад, якщо супутник вимірює збільшення потоку будь-якого параметра сонячної активності, це

не означає, що заряджені частинки вступають у контакт з рослинною масою. Навіть якщо довести в лабораторії, що вони можуть викликати початкову фазу займання, це не обов'язково означатиме, що будь-який раптовий потік частинок потрапить на територію США. Натомість база даних містить цілодобову інформацію про спалахи на сонці і лише інформацію про спалахи пожеж в США. Тобто для збільшення адекватності та точності моделі треба враховувати спалахи пожеж на всій території Земної кулі або брати до уваги тільки ті спалахи, які зумовили вплив на територію США. Крім того, коли проходження атмосфери високоенергетичними частинками відбувається в умовах підвищеної вологості і/або хмарності, вони, як правило, не досягають поверхні Землі за рахунок поглинання [6].

За розрахунками, від моменту реєстрації потоку заряджених частинок до спалаху лісової пожежі проходить певний час. Лише для невеликих пожеж лаг становить 0 днів. Повторна активізація спалахів малих пожеж відбувається із затримкою 2–3 доби, а великих — 1 доба. Така особливість потребує подальших досліджень.

З іншого боку, виникають проблеми реєстрації кількості невеликих пожеж в США. Це пов'язано з тим, що велика частина території цієї країни є ненаселеною. З іншого боку, великі пожежі можуть утворюватися зі злиття декількох малих. Про це теж відсутня інформація.

Як видно з рис. 3 і табл. 2 та 3, результати досліджень слабо залежать від часових затримок (лагу). Різниця між результатами кореляції коливається в межах 10% – 30% залежно від лагу. Це може бути зумовлено такими факторами:

- У роботі досліджуються лісові пожежі, що зареєстровані на території Сполучених Штатів. Різниця в кліматі та атмосферних умовах через велику територію країни, а також у рослинному покриві, призводить до різних інерційних процесів займання. Це, у свою чергу, «ослаблює» залежність від часової затримки.

- Спалахи пожеж також залежать від інших чинників, а не лише від сонячної активності, однак вони не були включені до моделей.

- Спалахи пожеж можуть бути пов'язані з коливаннями клімату, погоди та інших стохастичних факторів, які не враховані в моделях.

Однак, незважаючи на відносну точність прогнозування за допомогою розроблених моделей на основі *ANFIS* та *нейронних мереж*, для остаточного підтвердження запропонованої гіпотези, необхідно провести експериментальні лабораторні дослідження. У наступних теоретичних дослідженнях планується провести прогнозування місця і часу займання лісової пожежі. Практична цінність результатів полягає в тому, що з'являється можливість передбачити та повідомити про потенційну небезпеку і відповідно ужити запобіжні заходи для її усунення.

Подяка: Ці результати є частиною проекту Ш47007, заснованого Міністерством освіти і науки Республіки Сербія. Изд-во СОЛОН-пресс,

ЛІТЕРАТУРА

1. Дьяконов В.П. MatLab 6.5 SP1 (7.0). Simulink 5/6. Основы применения, Изд-во СОЛОН-пресс, 2005. — 800 с.

2. *Abraham A.* Adaptation of Fuzzy Inference System Using Neural Learning. In *Fuzzy Systems Engineering: Theory and Practice* (eds. Nedjah, N. & de Macedo Mourelle, L.). Springer Verlag, 2005. — P. 53–83.
3. *Mohandes M., Rehman S.* Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS) // *Appl. Energy*. — 2011. — 88(11). — P. 4024–4032.
4. *Vyklyuk Y.* Simulation of spatial form of urban systems by diffusion methods // *Journal of the Geographical Institute «Jovan Cvijic» SASA*. — 2013. — **63**, № 1. — P. 89–100. — № 2. — P. 67–77.
5. *Yaroslav Vyklyuk, Darko Vuković, Ana Jovanović.* Forex prediction with neural network: usd/eur currency pair // *Actual Problems of Economics*. — 2013. — № 10. — P. 261–273.
6. *Labitzke K.* The global signal of the 11-year sunspot cycle in the atmosphere: When do we need the QBQ? *Meteorolog. Zeitschrift*. — 2003. — 12(4). — P. 209–216.

Надійшла 02.06.2014