

## ПРОГНОЗУВАННЯ СОНЯЧНОЇ АКТИВНОСТІ АЛЬТЕРНАТИВНИМИ МЕТОДАМИ

П.І. БІДЮК, І.В. КАРАЮЗ, В.С. ВАРАВА, О.Л. ЖИРОВ

**Анотація.** Розглянуто актуальну задачу прогнозування нестационарних процесів сонячної активності альтернативними методами, яку розв'язують дослідники багатьох країн світу. Досліджувані процеси належать до класу нелінійних та нестационарних, що потребує вибору спеціальних методів для їх моделювання і прогнозування. Запропоновано підхід до прогнозування на основі трьох фільтрів: адаптивного фільтра Калмана, оптимального фільтра Калмана з оцінюванням параметрів за методом максимальної правдоподібності і ймовірнісного гранулярного фільтра. Вибір цих фільтрів зумовлено тим, що вони дають змогу враховувати у моделі випадкові зовнішні збурення і похибки вимірювань. Отримані результати обчислювальних експериментів свідчать про можливість застосування запропонованого підходу до розв'язання поставленої задачі. Більш точні результати короткострокового прогнозування експоненціально згладжених даних отримано за допомогою адаптивного фільтра. Аналіз якості результатів виконано за допомогою відомих статистичних характеристик якості, зокрема середньої абсолютної похибки у відсотках.

**Ключові слова:** адаптивний фільтр Калмана, оптимальний фільтр Калмана, метод максимальної правдоподібності, ймовірнісний гранулярний фільтр, сонячна активність, короткострокове прогнозування.

### ВСТУП

Сонячна активність — сукупність явищ, які періодично виникають у сонячній атмосфері. Прояв сонячної активності тісно пов'язаний з магнітними властивостями сонячної плазми. Виникнення активної ділянки починається з поступового збільшення магнітного потоку в деякій зоні фотосфери, унаслідок чого у відповідних місцях хромосфери спостерігається збільшення яскравості. Усі рухи в атмосфері або атмосферну циркуляцію зумовлює безперервний потік сонячної радіації, а саму атмосферу можна уподібнити до гігантської теплової машини. Її нагрівником є тропіки, а холодильником — полярні області. Атмосфера перетворює енергію випромінювання, що надходить від Сонця, у кінетичну енергію руху повітря (вітру). Її коефіцієнт корисної дії прямо пропорційний різниці температур між нагрівником (екватором) і холодильником (полюсами) і обернено пропорційний температурі нагрівника. Прогнозування особливостей сонячного циклу допомагає прогнозувати появу сонячних плям і виникнення сонячної бурі, а також мінімізувати проблеми, що виникають унаслідок їх впливу на нашу планету.

Існує також космогонічний аспект проблеми, що має важливе методологічне і світоглядне значення. Під час всебічного ознайомлення з великою різноманітністю явищ, що відбуваються у Сонячній системі в цілому, створюється враження, що нинішній стан сонячної активності (і сонячно-

планетних зв'язків) є не що інше, як прояв поточного етапу еволюції нашої планетної системи. Інакше кажучи, взаємодія в системі Сонце – міжпланетне середовище – планети призводить до змін орбіт, фізичних характеристик планет і, меншою мірою, самого Сонця, тобто Сонячна система продовжує еволюціонувати, хоча і досягла певного рівня еволюційної зрілості і стійкості.

Про значущість та необхідність розв'язання задач прогнозування сонячної активності, зокрема кількості сонячних плям, підкреслюється у праці [1]. Ці задачі є важливими для наукових і практичних цілей, оскільки з ними пов'язано багато питань подальшого розвитку природних процесів та змін клімату. Особливо важливим є питання про можливий вплив сонячної активності на період обертання Землі і процеси, що перебігають у земних глибинах. Ідеться про нерівномірність добового обертання Землі і землетруси. Питання сталості довжини доби почало цікавити вчених ще із середини XVIII ст. Але лише в 40-х роках минулого століття було отримано переконливу відповідь: довжина доби мінлива. Так, за період 1892–1974 рр. середньорічне значення довжини доби  $T$  систематично збільшувалося, причому середньорічні значення зміни  $T$  коливалися в межах  $\Delta T = (8 - 482) \cdot 10^{-5}$  с. Виняток становили лише 1896, 1927, 1935 і 1936 рр., коли середньорічні значення  $\Delta T$  були від'ємними, а також 1928 р., коли  $\Delta T = 0$ . Ці дані вказують на систематичне («вікове») уповільнення обертання Землі в сучасну епоху.

Задачу прогнозування радіофлюксів на довжині хвилі 10,7 см (2,8 ГГц, скорочено  $F_{10,7}$ ) з використанням сучасних методів аналізу глобального сонячного магнітного поля розглянуто у праці [2]. Для побудови прогнозної моделі запропоновано додаткові індекси, обчислені на основі абсолютних значень сонячного магнітного поля. Взята за основу модель перенесення потоку випромінювання дала можливість описати спостереження значень потоку  $F_{10,7}$  з високою адекватністю.

У праці [3] розв'язується задача прогнозування сонячних спалахів, що виникають унаслідок вивільнення енергії, яка нагромаджується у магнітному полі активних регіонів Сонця. Автори запропонували модель, що ґрунтується на методі глибокого навчання. Дані, необхідні для навчання і тестування моделі, сформовані з видимих магнітограм сонячних регіонів. Працю [4] присвячено аналізу сонячного випромінювання на довжині хвилі 10,7 см ( $F_{10,7}$ ), інтенсифікація якого приводить до збільшення кількості сонячних плям. Автори вказують, що для вимірювання сонячної активності у світі використовують множину різних величин. У дослідженні порівнюється інтенсивність магнітного поля, виміряного для повного сонячного циклу з індексом ефективних циклів за 1986–2015 рр. Аналізуються (прогнозуються) обидві вказані величини окремо, а також їх взаємодія.

Оскільки індекс  $F_{10,7}$  корелює з багатьма сонячними показниками, то Теппінг отримав таке співвідношення між потоком на хвилі 10,7 см і кількістю сонячних плям:  $SN = 1,147 \cdot F_{10,7} - 73,21$  [5]. Площа сонячних плям: Денісе (1948) виявив 76%-ву кореляцію між потоком на хвилі 10,7 см і пло-

щею сонячних плям [6]. Іонізація Е-шару: Денісе і Кунду (1957) виявили суттєву кореляцію між індексом іонізації Е-шару Землі і потоками на хвилі 10,7 см [7]. Виявлено залежність між потоком на хвилі 10,7 см та даними, отриманими за допомогою двох бортових магнітографів: MDI на SOHO (Solar & Heliospheric Observatory), яка обертається навколо точки Лагранжа L1 (тут сила земного тяжіння дорівнює силі сонячного тяжіння) на відстані приблизно 1,5 млн км від Землі, та HMI на SDO (Solar Dynamics Observatory), яка перебуває на геосинхронній орбіті Землі на висоті 36000 км [8].

У праці [9] досліджувалась залежність між сонячним випромінюванням на п'яти різних довжинах хвилі (30; 15; 10,7; 8 та 3,2 см). Компроміс між складністю моделі та кореляцією виявився за комбінації двох довжин хвиль. Задовільний результат отримано з використанням на хвилях потоків 10,7 см і 30 см. Побудовано деякі алгоритми передбачення сонячної активності (для індексу  $F_{10,7}$ ) за зображеннями Сонця. Але всі ці результати отримано для величезних вибірок, довжиною не меншою ніж один сонячний цикл (11 років) [9]. Для таких вибірок часто використовують не щоденні значення активності Сонця, а середньомісячні. У таких масштабах дійсно спостерігається залежність між різними сонячними показниками. Але за прогнозування сонячної активності на 1 день ці залежності неістотні, а тому часто внесення їх до моделі погіршують результати. Загалом процеси сонячної активності є нелінійними та нестационарними, що робить задачу їх прогнозування ще більше цікавою та складною. У роботі розглядається задача короткострокового прогнозування індексу  $F_{10,7}$  з використанням цифрових та оптимальних фільтрів, зокрема, фільтра Калмана і ймовірнісного байєсівського (particle) фільтра.

## ПОСТАНОВКА ЗАДАЧІ

На основі сучасних методів цифрової та оптимальної адаптивної фільтрації даних розробити ефективний метод прогнозування індексу  $F_{10,7}$  (потік випромінювання Сонця на довжині хвилі 10,7 см) на один період дискретизації даних (1 день) наперед. Виконати необхідні обчислювальні експерименти з метою побудови математичних моделей, обчислення оцінок прогнозів та порівняння результатів і прогнозування на основі наявних статистичних даних про сонячну активність.

## ФУНКЦІОНАЛЬНА СХЕМА ПРОЦЕДУР ФІЛЬТРАЦІЇ ТА КОРТКОСТРОКОВОГО ПРОГНОЗУВАННЯ

Функціональну схему процедур фільтрації та короткострокового прогнозування даних у формі часового ряду за допомогою запропонованої комплексної моделі зображено на рис. 1. Дані надходять на вхід трьох фільтрів. Перший з них — адаптивний фільтр Калмана, який не потребує початкових оцінок параметрів, оскільки оцінки обчислюються в процесі оброблення даних. Перед запуском другого фільтра у робочий режим визначаються всі початкові параметри за методом максимальної правдоподібності, необхідні

для роботи фільтра Калмана. Третій фільтр — це ймовірнісний гранулярний фільтр (particle filter) байєсівського типу, який ідеологічно доповнює два попередні.

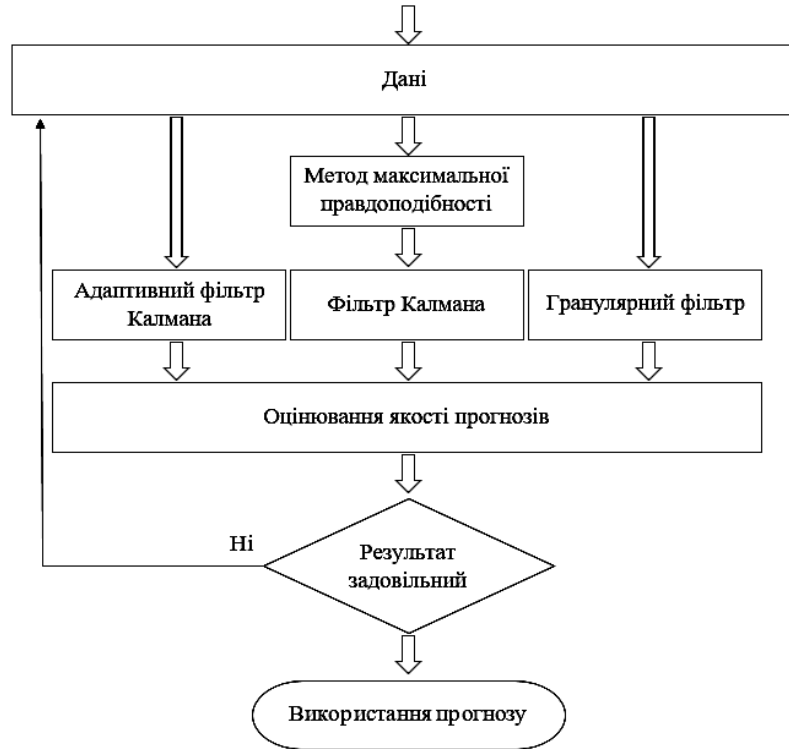


Рис. 1. Функціональна схема процесу фільтрації та короткострокового прогнозування на основі фільтрів трьох типів

Адекватна модель визначається за допомогою статистичних параметрів адекватності моделі ( $R^2$ ,  $DW$ ) і суми квадратів похибок, а кращий прогноз вибирається за допомогою середньої абсолютної похибки у відсотках (або  $MARPE$ ). Запропонована функціональна схема забезпечує отримання точкових і ймовірнісних оцінок прогнозів, а також комбінування оцінок прогнозів, отриманих за допомогою трьох моделей. Якщо отриманий результат прогнозування незадовільний, то відбувається повернення на перший етап обчислень, тобто отримуються додаткові дані, виконується підготовка даних до моделювання (попереднє оброблення) і всі три фільтри використовуються для повторного обчислення оцінок прогнозів.

## ТЕОРЕТИЧНІ ОСНОВИ МЕТОДІВ ЗГЛАДЖУВАННЯ І ПРОГНОЗУВАННЯ

### Згладжувач Калмана

Згладжування даних може суттєво полегшити подальший аналіз даних. Ефективним для отримання остаточного результату методом згладжування даних є згладжувач Калмана [10]. Нехай побудовано модель процесу у просторі станів:

$$X_{t+1} = AX_t + W_t, W_t = N(0, Q),$$

$$Y_t = CX_t + V_t, V_t = N(0, R).$$

Алгоритм згладжувача Калмана складається з таких кроків:

1) реалізується фільтрація даних для моментів часу  $t = 0 \dots T$  :

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t},$$

$$P_{t+1|t} = AP_{t|t}A^T + Q,$$

$$K_{t+1} = P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1},$$

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - C\hat{x}_{t+1|t}),$$

$$P_{t+1|t+1} = P_{t+1|t} - K_{t+1}CP_{t+1|t};$$

2) запускається зворотне проходження для всіх моментів часу  $t = 0 \dots T$  :

$$L_t = P_{t|t}A^T P_{t+1|t}^{-1},$$

$$\hat{x}_{t|T} = \hat{x}_{t+1|T} + L_t(\hat{x}_{t+1|T} - \hat{x}_{t+1|t}),$$

$$P_{t|T} = P_{t+1|T} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T.$$

### Експоненційне згладжування

Експоненційне згладжування також є одним з найпоширеніших прийомів, що використовуються для згладжування часових рядів і короткострокового прогнозування. В основу процедури згладжування покладено розрахунок експоненційних ковзних середніх згладжуваного ряду, тобто це простий цифровий фільтр. Нехай  $X = \{X_0, \dots, X_T\}$  — часовий ряд. Процедура простого експоненційного згладжування здійснюється за формулами:

$$S_1 = X_0,$$

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}, \quad (1)$$

де  $X_t$  — фактичне спостереження в момент часу  $t$ ;  $S_t$  — значення експоненційного середнього в момент часу  $t$ ;  $\alpha$  — параметр згладжування,  $\alpha = \text{const}$ ,  $\alpha \in (0; 1]$ .

Експоненційне середнє в момент часу  $t$  виражено як зважена сума поточного спостереження і експоненційного середнього минулого спостереження з вагами  $\alpha$  і  $1 - \alpha$  відповідно. Якщо послідовно використовувати рекурентне співвідношення (1), то значення  $S_t$  можна виразити через значення часового ряду  $X$ :

$$S_t = \alpha \cdot \sum_{i=0}^{\infty} (1 - \alpha)^i X_{t-i}.$$

Таким чином, величина  $S_t$  виявляється зваженою сумою всіх членів ряду. Причому значення ваг зменшуються експоненційно і залежать від віддаленості спостереження від моменту  $t$ .

### Оптимальний адаптивний фільтр Калмана

Для застосування оптимального фільтра скористаємось простою моделлю випадкового кроку, яку подамо у просторі станів:

$$X_{i+1} = X_i + W_i,$$

$$z_{i+1} = X_i + V_i,$$

де  $X_i$  — значення часового ряду  $z_i$  у момент часу  $i$ ;  $W_i$  — випадковий приріст рівня на інтервалі часу від  $i$  до  $i+1$  — це некорельована послідовність значень з невідомим середнім  $E[W_i] = q^*$  і дисперсією  $E[(W_i - q^*)^2] = (\sigma_W^*)^2$ ;  $V_i$  — випадкова послідовність з нульовим середнім і невідомою дисперсією  $(\sigma^*)^2$ . Послідовність нев'язок цього фільтра має вигляд

$$v_i^{(1)} = z_i - z_{i-1}.$$

Рівняння вимірювань середнього значення  $q^*$  запишемо як

$$v_i^{(1)} = \check{H} + \check{V}_i, \quad i > 1,$$

де  $\check{H} = 1$ ,  $\check{V}_i = V_i - V_{i-1} + (W_{i-1} - q^*)$ . Оцінка середнього значення  $q^*$  за припущення про його сталість на інтервалі розраховується за формулою [10]:

$$\hat{q}^{(i)} = \hat{q}^{(i-1)} + \frac{1}{i+1}(v_i^{(1)} - \hat{q}^{(i-1)}), \quad i = 2, 3, \dots$$

$$q^{(1)} = 0.$$

Нев'язки фільтра, обчислені за трьома спостереженнями, мають вигляд

$$v_i^{(2)} = z_i - \frac{1}{2}z_{i-1} - \frac{1}{2}z_{i-2}.$$

Середні значення нев'язок:

$$E[v_i^{(1)}] = q^*, \quad E[v_i^{(2)}] = \frac{3}{2}q^* \quad \text{та} \quad E[(v_i^{(1)} - \frac{3}{2}q^*)(v_{i-1}^{(1)} - q^*)] = \frac{1}{2}(\sigma_W^*)^2.$$

Таким чином, послідовність значень дисперсії  $(\sigma_W^*)^2$  визначається так:

$$\tilde{\sigma}_{W_i}^2 = 2 \cdot \left[ \left( v_i^{(2)} - \frac{3}{2}\hat{q}^{(i)} \right) (v_{i-1}^{(1)} - \hat{q}^{(i)}) \right], \quad (2)$$

а оцінка постійної дисперсії  $(\sigma_W^*)^2$  розраховується за формулами:

$$(\hat{\sigma}_W^{(i)})^2 = (\hat{\sigma}_W^{(i-1)})^2 + \frac{1}{i-2}(\tilde{\sigma}_{W_i}^2 - (\hat{\sigma}_W^{(i-1)})^2), \quad i = 3, 4, \dots; \quad (\hat{\sigma}_W^{(2)})^2 = 0; \quad (3)$$

$$E[(v_i^{(1)} - q^*)^2] = 2(\sigma^*)^2 + (\sigma_W^*)^2.$$

Тому послідовність

$$\tilde{\sigma}_i^2 = \frac{1}{2}[(v_i^{(1)} - \hat{q}^{(i)})^2 - (\hat{\sigma}_W^{(i)})^2] \quad (4)$$

можна розглядати як послідовність вимірювань дисперсії  $(\sigma^*)^2$ , оцінка якої за умови припущення про її сталість розраховується за виразом:

$$(\hat{\sigma}^{(i)})^2 = (\hat{\sigma}^{(i-1)})^2 + \frac{1}{i-2}(\tilde{\sigma}_i^2 - (\hat{\sigma}^{(i-1)})^2), i = 3, 4, \dots; (\hat{\sigma}^{(2)})^2 = 0. \quad (5)$$

Прогнозоване значення рівня ряду на один крок  $\hat{X}_{i+1,i}$  обчислюється відповідно до алгоритму фільтра Калмана:

$$\hat{X}_{i+1,i} = \hat{X}_{i,i} + \hat{q}^{(i)}, i = 3, 4, \dots; \quad (6)$$

де  $\hat{X}_{i,i} = \hat{X}_{i,i-1} + \hat{K}_i(z_i - \hat{X}_{i,i-1})$ ,  $\hat{X}_{2,2} = z_2$ .

Коефіцієнт підсилення фільтра

$$\hat{K}_i = \frac{\hat{\rho}_{i,i-1}}{\hat{\rho}_{i,i-1} + (\hat{\sigma}^{(i)})^2},$$

де  $\hat{\rho}_{i,i-1} = \hat{\rho}_{i-1,i-1} + (\hat{\sigma}_W^{(i)})^2$ ;  $\hat{\rho}_{i,i} = (1 - \hat{K}_i)\hat{\rho}_{i,i-1}$ ;  $\hat{\rho}_{2,2} = (\hat{\sigma}^{(3)})^2$ .

Припустімо також, що рівень ряду  $q^*$  не зміщується, тому оцінку середнього значення  $\hat{q}^{(i)}$  передбачимо рівною нулю. Прогнозування рівня ряду виконується відповідно до формул (2) – (6), якщо  $i \geq 3$  і  $q^{(i)} = 0$ .

### Адаптивний фільтр Калмана з оцінюванням параметрів за методом максимальної правдоподібності

Алгоритм оцінювання адаптивного фільтра такий [10]:

1. Покладемо початкову умову  $\theta^{(0)}$ ,  $i = 0$ .
2. За допомогою фільтра Калмана обчислюємо величини  $\tilde{Y}_k(\theta^{(i)})$  і  $S_{k|k-1}(\theta^{(i)})$  для  $k = 1, 2, \dots, N$ , а також функцію правдоподібності.
3. Отриману функцію максимізуємо за вектором параметрів  $(\theta)$ .
4. У результаті отримуємо новий вектор параметрів  $\theta^{(i+1)}$ ,  $i = i + 1$ .
5. Після обчислення нового вектора параметрів повторюємо кроки 2–4 до збіжності алгоритму.

### Гранулярний фільтр

Метод гранулярної (particle) фільтрації — це метод, що ґрунтується на застосуванні модифікованого методу Монте-Карло для розв'язання задачі оцінювання стану динамічної системи. Гранулярний фільтр також відомий як бутстреп (bootstrap) фільтр, алгоритм конденсації, наближення взаємодійних частинок та виживання найбільш придатних [11, 12]. Ключовою ідеєю фільтра є подання необхідної функції апостеріорної щільності множиною випадкових частинок (значень) з відповідними вагами та обчислення оцінок на

основі цих частинок і відповідних вагових коефіцієнтів. Коли кількість частинок стає дуже великою, ця характеристика методу Монте-Карло стає еквівалентним поданням функції апіорної імовірності і розв'язання наближається до оптимальної байссівської оцінки.

Гранулярний фільтр на основі послідовного методу Монте-Карло — це рекурсивний алгоритм для числового розв'язання задач оцінювання (фільтрації, згладжування), особливо для нелінійних та негаусівських випадків. У цьому полягає універсальність такої процедури. Порівняно з розширеними фільтрами Калмана гранулярні фільтри не залежать від методів лінеаризації або апроксимації. Звичайний розширений фільтр Калмана незадовільно функціонує в істотно нелінійних процесах, а також малоприматний у випадку шумів системи і вимірювань, що дуже відрізняються від гаусівських. Варто зазначити, що, у свою чергу, гранулярні фільтри більш вимогливі до обчислювальних ресурсів.

Гранулярний фільтр призначений для оцінювання послідовності прихованих змінних  $x_n$  для  $n=1,2,\dots$  на підставі спостережень  $y_n$ , якщо  $n=1,2,\dots$ . Будемо вважати, що розглядається динамічна система, а  $x_n$  та  $y_n$  — дійсні вектори стану і вимірювань відповідно.

Стохастичне рівняння стану системи має такий загальний вигляд:

$$x_k = f_k(x_{k-1}, v_k),$$

де  $f_k$  — функція зміни стану системи;  $v_k$  — випадкова величина — збуджувальний вплив (збурення). Рівняння вимірювань:

$$y_k = h_k(x_k, w_k),$$

де  $h_k$  — функція (матриця) вимірювань;  $w_k$  — випадкова величина, шум вимірювань.

Функції  $f_k$  і  $h_k$  у загальному випадку нелінійні, а статистичні характеристики шуму системи ( $v_k$ ) і вимірювань ( $w_k$ ) вважаються відомими. Завданням фільтрації є отримання оцінки  $\hat{x}_k$  на підставі відомих до моменту  $k$  результатів вимірювань  $y_{1:k}$ . Розглянемо дискретний марковський процес  $\{X_n\}_{n \geq 1}$  з такими розподілами ймовірностей:

$$X_1 \sim \mu(x_1) \text{ і } X_n | (X_{n-1} = x_{n-1}) \sim f(x_n | x_{n-1}), \quad (7)$$

де  $\mu(x_1)$  — щільність ймовірності;  $f(x_n | x_{n-1})$  — умовна щільність ймовірності.

Позначення  $X | Y \sim f(\dots)$  означає, що  $X$  за умови  $Y$  має розподіл  $f(\dots)$ . Реалізації процесу  $\{X_n\}$  (приховані змінні  $x_n$ ) спостерігаються за допомогою іншого випадкового процесу  $\{Y_n\}_{n \geq 1}$  — процесу вимірювань з маргінальною щільністю:

$$Y_n | (X_n = x_n) \sim h(y_n | x_n), \quad (8)$$

де  $h(y_n | x_n)$  — умовна щільність імовірності, вимірювання вважаються статистично незалежними. Визначена таким чином модель системи і вимірювань відома як прихована марковська модель.



Рівняння (7) визначає апіорний розподіл для процесу  $\{X_n\}$ :

$$\rho(x_{1:n}) = \mu(x_1) \prod_{k=2}^n f(x_k | x_{k-1}).$$

Аналогічно рівняння (8) задає функцію правдоподібності:

$$\rho(y_{1:n}) = \prod_{k=2}^n h(y_k | x_k).$$

Нотація  $x_{k:l}$  для  $k \leq l$  означає, що  $(x_k, \dots, x_l)$ . Отже, байєсівський висновок для  $\{X_{1:n}\}$  за відомих реалізацій вимірювань  $\{Y_{1:n}\}$ , позначених як  $\{x_{1:n}\}$  та  $\{y_{1:n}\}$ , буде спиратися на апостеріорний розподіл

$$\rho(x_{1:n} | y_{1:n}) = \frac{\rho(x_{1:n})\rho(y_{1:n} | x_{1:n})}{\rho(y_{1:n})},$$

де  $\rho(y_{1:n}) = \int \rho(x_{1:n})\rho(y_{1:n} | x_{1:n})dx_{1:n}$ .

Метод Монте-Карло дозволяє оцінювати властивості досить складних розподілів ймовірностей, наприклад, обчисленням середніх і дисперсії у вигляді інтеграла:

$$\bar{\theta} = \int \theta(x)\rho(x)dx,$$

де  $\theta(x)$  — функція для оцінювання. Наприклад, для середнього можна покласти  $\theta(x) = x$ . У разі неможливості аналітичного знаходження розв'язку завдання може бути розв'язане чисельно генеруванням випадкових вибірок зі щільністю  $\rho(x)$ , позначимо їх як  $x_{1 \leq i \leq N}^{(i)}$ , і отриманням середнього арифметичного по точках вибірки:

$$\bar{\theta} \approx \frac{1}{N} \sum_{i=1}^N \theta(x^{(i)}).$$

У більш загальному випадку, коли вибірка з  $\rho$  ускладнена, застосовується інший розподіл  $q$ , а для збереження незміщеності оцінки вводяться вагові коефіцієнти  $w_i$  на основі відношення  $r(x^{(i)}) = \rho(x^{(i)})/q(x^{(i)})$ :

$$w_i = \frac{r(x^{(i)})}{\sum_{j=1}^N r(x^{(j)})},$$

після чого обраховується зважене середнє:

$$\bar{\theta} = \int \theta(x)r(x)q(x)dx \approx \frac{1}{N} \sum_{i=1}^N w_i \theta(x^{(i)}).$$

Хоча допоміжний розподіл використовується здебільшого для спрощення вибірки з основного розподілу  $\rho$ , часто застосовується процедура вибірки і повторної вибірки за значущістю. Ця процедура складається з двох етапів: власне вибірки за значущістю з обчисленням ваг  $w_i$  і додаткової ви-

бірки точок, які враховують ці ваги. Повторна вибірка особливо потрібна для послідовних фільтрів.

Методи гранулярної фільтрації і згладжування є найбільш відомими прикладами алгоритмів послідовного методу Монте-Карло, на основі яких будуються алгоритми байєсівського програмування. Проте послідовний метод Монте-Карло включає в себе більш широкий клас алгоритмів, які можна застосовувати для опису більш складних наближених методів фільтрації і згладжування.

Якщо записати щільність у формі

$$f(x_{1:n}) = \frac{\varphi_n(x_{1:n})}{Z_n},$$

де  $\varphi_n : X^n \rightarrow \mathbb{R}^{\neq}$  відомо поточково;  $Z_n = \int \varphi_n(x_{1:n}) dx_{1:n}$  — нормуюча, можливо невідома, константа, то алгоритм послідовного методу Монте-Карло буде знаходити наближення  $f_k(x_{1:k})$  та оцінки  $Z_k$  для  $k=1,2,\dots$ . Наприклад, для випадку фільтрації можна покласти

$$\varphi_n(x_{1:n}) = \rho(x_{1:n})\rho(y_{1:n} | x_{1:n}) \text{ і } Z_n = \rho(y_{1:n}),$$

звідки

$$f_n(x_{1:n}) = \frac{\rho(x_{1:n})\rho(y_{1:n} | x_{1:n})}{\rho(y_{1:n})} = \rho(x_{1:n} | y_{1:n}).$$

Схему предиктор–коректор можна подати у вигляді:

$$\rho(x_{1:n} | y_{1:n-1}) = \rho(x_{1:n-1} | y_{1:n-1})f(x_n | x_{n-1}) \text{ — предиктор;}$$

$$\rho(x_{1:n} | y_{1:n}) = \frac{h(y_n | x_n)\rho(x_{1:n} | y_{1:n-1})}{\rho(y_n | y_{1:n-1})} \text{ — коректор.}$$

Множник  $\rho(y_n | y_{1:n-1})^{-1}$  — нормуюча константа, якої не потребує звичайний алгоритм послідовного методу Монте-Карло. Точність гранулярного фільтра залежить від кількості частинок  $N$ ; більшій кількості частинок відповідає точніша апостеріорна щільність.

Алгоритм функціонування гранулярного фільтра:

1. *Ініціалізація.* Генеруються частинки з початкової функції щільності ймовірності  $x^{i,(0)} \sim \rho(x^{(0)})$ ; устанавлюється номер часового кроку  $k=1$ .

2. *Прогнозування.* Обраховуються  $N$  нових частинок з моделі переходу  $x^{i,(k)} \sim \rho(x^{(k)} | x^{i,(k-1)})$ .

3. *Оновлення.* Обчислюються ваги частинок на основі моделі вимірювання  $w^{i,(k)} = w^{i,(k-1)}\rho(y^{(k)} | x^{i,(k)})$ .

4. *Нормалізація.* Ваги частинок нормалізуються — їх сума має дорівнювати 1.

5. *Повторна вибірка.* Обраховуються  $N$  нових частинок на основі апостеріорної функції щільності ймовірності, отриманої на кроках 2–4.

6. *Кінець ітерації.* Обраховується оцінюване значення вектора стану  $\hat{x}^{(k)}$ ; збільшується часовий крок  $k = k + 1$ , Перехід на крок 2.

Загальна проблема гранулярного фільтра — це явище виродження, коли після декількох кроків, усі, крім однієї частинки, матимуть незначні ваги [12]. Це виродження означає, що значні обчислювальні витрати спрямовані на оновлення частинок, внесок яких у наближення функції апостеріорної щільності майже дорівнює нулю. Цю проблему можна подолати шляхом збільшення кількості частинок. Крім того, рекомендується використовувати метод повторної вибірки, щоб уникнути виродження частинок.

Повторна вибірка передбачає відображення випадкової величини  $\{x_k^{(i)}, w_k^{(i)}\}$  у випадкову величину  $\{x_k^{(i)*}, N^{-1}\}$  з рівномірними вагами [12]. Це можна виконувати, якщо кількість ефективних частинок з великими вагами зменшується нижче від певного порогового числа. Як альтернативу повторну вибірку також можна застосовувати несистемно на кожному моменті  $t_k$ . Розглянемо два методи генерування повторної вибірки.

### Мультиномальна повторна вибірка

Мультиномальна повторна вибірка, відома також як проста випадкова повторна вибірка, була запропонована разом з першим гранулярним фільтром у праці [12]. Вона складається з генерування  $N$  чисел з рівномірного розподілу

$$u^i \sim U[0,1), \quad i = 1, \dots, N$$

та вибору частинки  $x^j$  для копіювання такої, що

$$u^i \in \left[ \sum_{p=1}^{j-1} w^p, \sum_{p=1}^j w^p \right). \quad (9)$$

Тут можна розрізнити дві реалізації:

- сортування за зростанням згенерованих чисел  $u$  для отримання впорядкованого набору  $u_0$  та порівняння з послідовними діапазонами ваг;
- створення додаткового набору чисел  $Q = [Q_1, \dots, Q_N]$  на основі виразу

$$Q_j = \sum_{p=1}^j w^p = Q_{j-1} + w^j,$$

з використанням бінарного пошуку та вибір для копіювання частинки  $x^j$  такої, що  $u^i \in [Q_{j-1}, Q_j)$ .

### Стратифікована повторна вибірка

У цьому алгоритмі передбачається, що виконується поділ на страти (шари). У кожному шарі повторна вибірка може виконуватися одночасно. Однак і в цьому випадку з'явилися варіації методу.

Підхід, який можна легко застосувати, передбачає, що діапазон  $[0,1)$  поділяється на  $N$  однакових частин, і генерування відбувається у кожному такому шарі:

$$u^i \sim \left[ \frac{i-1}{N}, \frac{i}{N} \right).$$

Частинки  $x^j$  вибираються для реплікації таким чином, щоб виконувався вираз (9).

Інший метод полягає в розщепленні частинок на  $n_s$ -страти; у  $j$ -й страті є  $N_j$  частинок із загальною вагою  $W_j$ . Це більш загальний підхід, проте слід зазначити, що якщо умова

$$\frac{W_i}{N_i} = \frac{W_j}{N_j}$$

не задовольняється для будь-яких  $i$  та  $j$ , ваги частинок після повторної вибірки відрізняються. Це означає, що на кроці 3 наведеного алгоритму потрібно враховувати ваги з попереднього часового кроку. Цей метод важливо вибирати тоді, коли можна виконати паралельне обчислення. Розподіл на шари може бути виконаний відповідно до кількості частинок (однак у першому шарі ймовірно опиняться частинки тільки з нульовими або майже нульовими вагами) або відповідно до ваг шарів  $W_j$ .

## РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

Результати прогнозування сонячної активності за допомогою оптимального фільтра за припущення, що зміщення рівня ряду  $q^*$  немає, зображено на рис. 2. Для цього прогнозу  $MAPE = 2,5447$  і коефіцієнт детермінації  $R^2 = 0,7261$ .

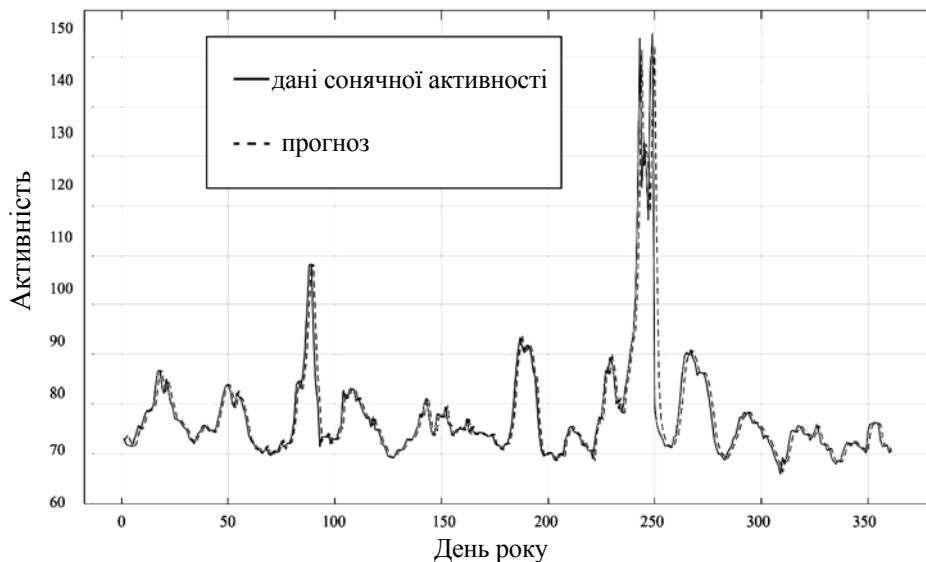


Рис. 2. Короткостроковий прогноз сонячної активності, якщо  $q^* = 0$ : по осі абсцис — кількість днів; по осі ординат — значення індексу  $F_{10,7}$

Результати однокрокового прогнозування сонячної активності за допомогою гранулярного фільтра графічно зображено на рис. 3. Використаний тип повторної згенерованої вибірки — пропорційний. Мінімальне відношення ефективних частинок до всіх згенерованих становило 0,5. Використаний метод повторної вибірки — залишковий; кількість згенерованих частинок — 1000.

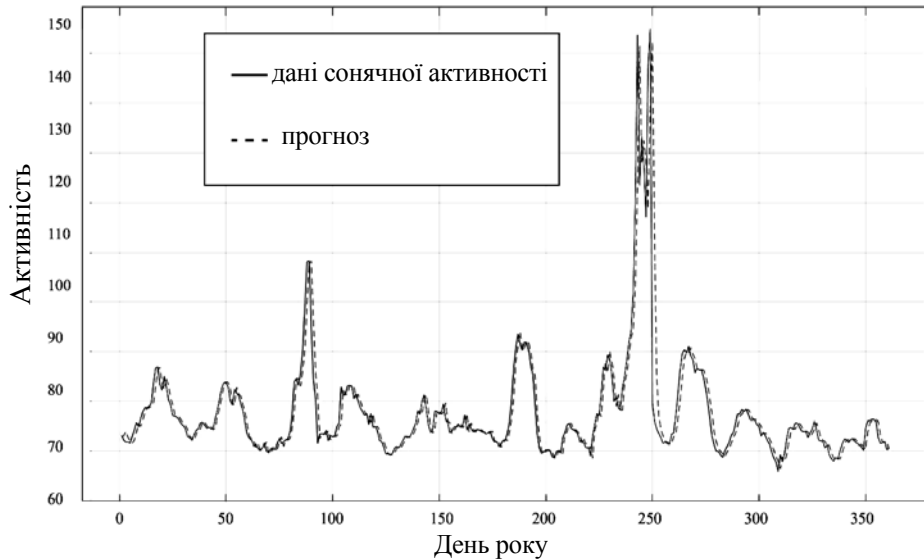


Рис. 3. Короткостроковий прогноз сонячної активності: по осі абсцис — кількість днів; по осі ординат — значення індексу  $F_{10,7}$

Значення  $MAPE$  та статистики  $R^2$  для оцінок прогнозів сонячної активності за допомогою оптимального адаптивного фільтра Калмана наведено в табл. 1, у якій вказано припущення про зміщення або незміщення  $q^*$ , а також тип згладжування даних, якщо такий є.

**Таблиця 1.** Результати прогнозування за допомогою оптимального адаптивного фільтра Калмана

Номер експерименту	Дані експерименту	$MAPE$	$R^2$
1	$q \neq 0$	2,5942	0,7215
2	$q \neq 0$ , експоненційне згладжування	0,987	0,9645
3	$q \neq 0$ , калманівське згладжування	1,9175	0,8499
4	$q = 0$	2,5447	0,7261
5	$q = 0$ , експоненційне згладжування	0,9453	0,966
6	$q = 0$ , калманівське згладжування	1,8486	0,8538

Результати прогнозування сонячної активності графічно зображено на рис. 1. Для згладжених за допомогою експоненційного згладжування даних отримано мінімальне значення  $MAPE$  та максимальне значення статистики  $R^2$ . Максимальне значення  $MAPE$  та мінімальне значення статистики  $R^2$  отримано для вихідного ряду.

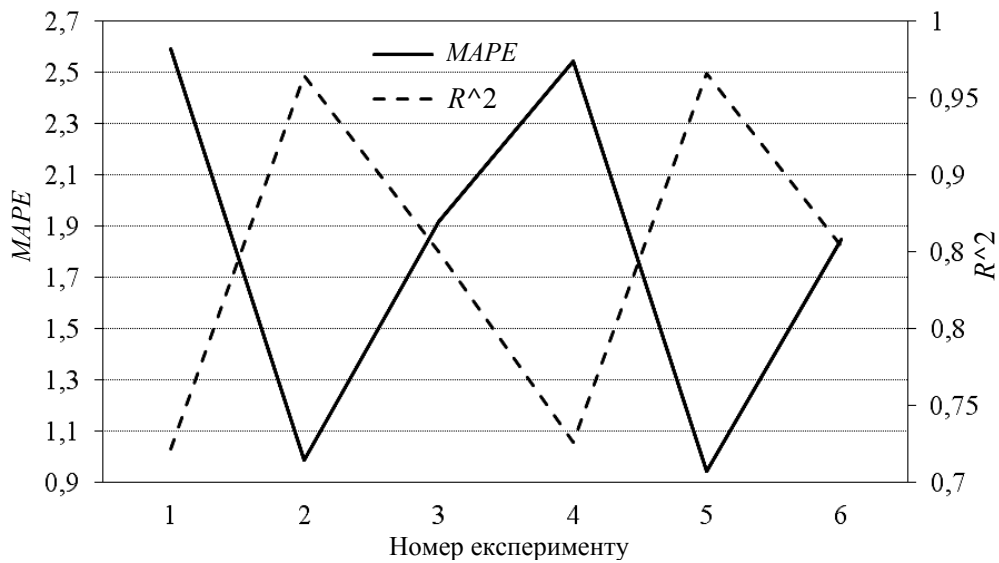


Рис. 4. Результати прогнозування за допомогою оптимального адаптивного фільтра Калмана: по осі абсцис — номер експерименту; по осях ординат — значення статистик

Найменше значення  $MAPE$  та найбільше значення статистики  $R^2$  отримано для адаптивного фільтра за припущення незміщення рівня ряду  $q^*$ . Значення середньої абсолютної відсоткової похибки  $MAPE$  та статистики  $R^2$  для прогнозів сонячної активності за допомогою адаптивного фільтра Калмана наведено в табл. 2, у якій вказано довжину вибірки, яка використана у методі максимальної правдоподібності для визначення параметрів фільтра (2015 означає, що для визначення параметрів фільтра використовувалася вибірка за 2015 р.), і тип згладжування даних, якщо такий є.

Таблиця 2. Результати прогнозування за допомогою адаптивного фільтра Калмана з використанням методу максимальної правдоподібності (ММП)

Номер експерименту	Дані експерименту	MAPE	R <sup>2</sup>
1	Довжина вибірки: 20	2,7281	0,7359
2	Довжина вибірки: 100	2,3977	0,747
3	Довжина вибірки: 200	2,406	0,7464
4	Уся вибірка	2,5271	0,7445
5	20–2015	3,0988	0,7295
6	100–2015	2,5189	0,7454
7	200–2015	3,7064	0,6961
8	Уся вибірка–2015	3,2916	0,7176
8	20–експоненційне згладжування	1,2465	0,9546
10	100–експоненційне згладжування	1,1065	0,9591
11	200–експоненційне згладжування	1,1059	0,9589

Результати однокрокового прогнозування сонячної активності графічно зображено на рис. 5. Для експерименту, у якому навчання виконувалось на всій вихідній вибірці з використанням експоненційного згладжування даних, отримано мінімальне значення  $MAPE$  та максимальне значення статистики  $R^2$ . Максимальне значення  $MAPE$  та мінімальне значення статистики  $R^2$  отримано для експерименту, у якому навчання проводилось на перших 200 значеннях ряду за 2015 р., без згладжування вихідного ряду. Для експоненційно згладженого ряду отримано максимальне значення  $MAPE$  та мінімальне значення статистики  $R^2$ .

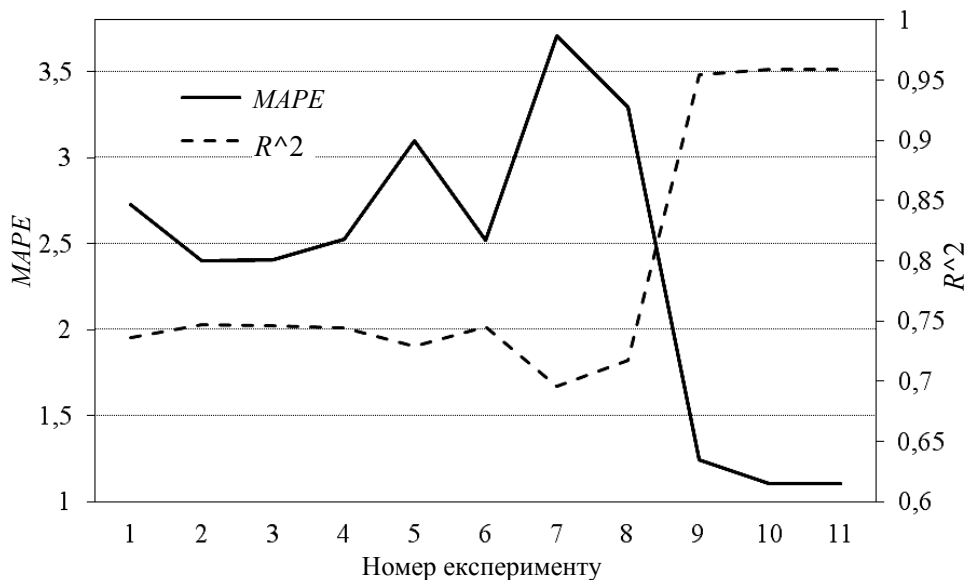


Рис. 5. Результати прогнозування за допомогою адаптивного фільтра Калмана з використанням ММП: по осі абсцис — номер експерименту; по осях ординат — значення статистик

Значення середньої абсолютної відсоткової похибки  $MAPE$  та статистики  $R^2$  для прогнозів сонячної активності за допомогою гранулярного фільтра наведено в табл. 3. Для *real*,  $MAPE$  і  $R^2$  дослідження виконувалось на вихідній вибірці, а для *exp\_sm*,  $MAPE$ ,  $R^2$  — на експоненційно згладжених даних; для *Kalman\_sm*,  $MAPE$ ,  $R^2$  — на згладжених за Калманом даних. Виявилось, що тип повторної вибірки несуттєво впливає на якість прогнозу, а тому в таблиці наведено лише результати прогнозування з найбільшим значенням статистики  $R^2$ . У табл. 3 вказано також мінімальне відношення ефективних частинок до всіх або довжину інтервалу для інтервального типу повторної вибірки та кількість частинок.

Результати прогнозування експоненційно згладженої сонячної активності за допомогою гранулярного фільтра графічно зображено на рис. 6. Для експериментів, у яких не реалізовувалась повторна вибірка, отримано найнижчу якість прогнозів (максимальне  $MAPE$  та мінімальне значення статистики  $R^2$ ). Мінімальне значення  $MAPE$  та максимальне значення статистики  $R^2$  отримано для інтервальної повторної вибірки з довжиною інтервалу 1 та 100000 частинками.

Таблиця 3. Результати прогнозування за допомогою гранулярного фільтра

Номер експерименту	<i>real</i>	<i>MAPE</i>	$R^2$	<i>exp_sm</i>	<i>MAPE</i>	$R^2$	<i>Kalman_sm</i>	<i>MAPE</i>	$R^2$
1	0,5 100	3,358	0,5667	0,5 100	1,7942	0,8186	0,5 100	3,045	0,6072
2	0,5 1000	3,083	0,6057	0,5 1000	1,6985	0,8510	0,5 1000	2,8106	0,6786
3	0,5 100000	4,1949	0,3655	0,5 100000	1,5571	0,8949	0,5 100000	2,6473	0,718
4	1–100	3,3659	0,5691	1–100	1,8134	0,8159	1–100	3,101	0,6157
5	1–1000	4,1053	0,4098	1–1000	1,7057	0,8498	1–1000	2,8595	0,6696
6	1–100000	4,156	0,3705	1–100000	1,5472	0,8976	1–100000	2,677	0,7135
7	30–100	4,8714	0,466	30–100	3,3447	0,5853	30–100	5,0505	0,4483
8	30–1000	4,631	0,4262	30–1000	2,82	0,6772	30–1000	4,011	0,5683
9	30–100000	3,9897	0,5404	30–100000	2,2434	0,753	30–100000	3,5777	0,5941

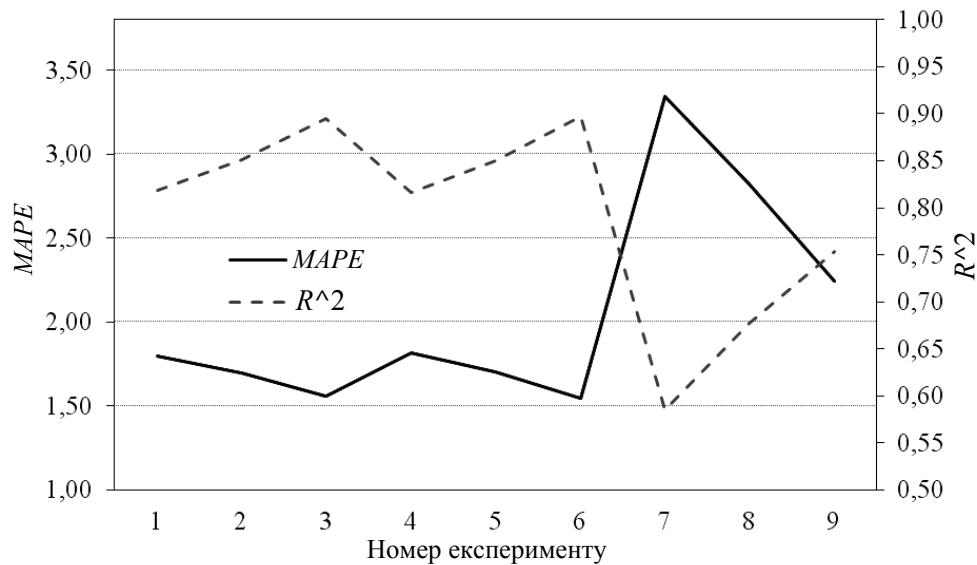


Рис. 6. Результати прогнозування експоненційно згладжених даних за допомогою гранулярного фільтра: по осі абсцис — номер експерименту; по осях ординат — значення статистик

Результати прогнозування згладженої за Калманом сонячної активності за допомогою гранулярного фільтра графічно зображено на рис. 7. Для експериментів, у яких не реалізовувалась повторна вибірка, отримано найнижчу якість прогнозів (максимальне *MAPE* та мінімальне значення статистики  $R^2$ ). Мінімальне значення *MAPE* та максимальне значення статистики  $R^2$  отримано для пропорційної повторної вибірки з 100000 частинками. Отже, прийнятний прогноз (мінімальне значення *MAPE* та максимальне значення статистики  $R^2$ ) отримано для оптимального адаптивного фільтра Калмана за припущення про незміщення рівня ряду  $q^*$  та експоненційно згладжених даних.



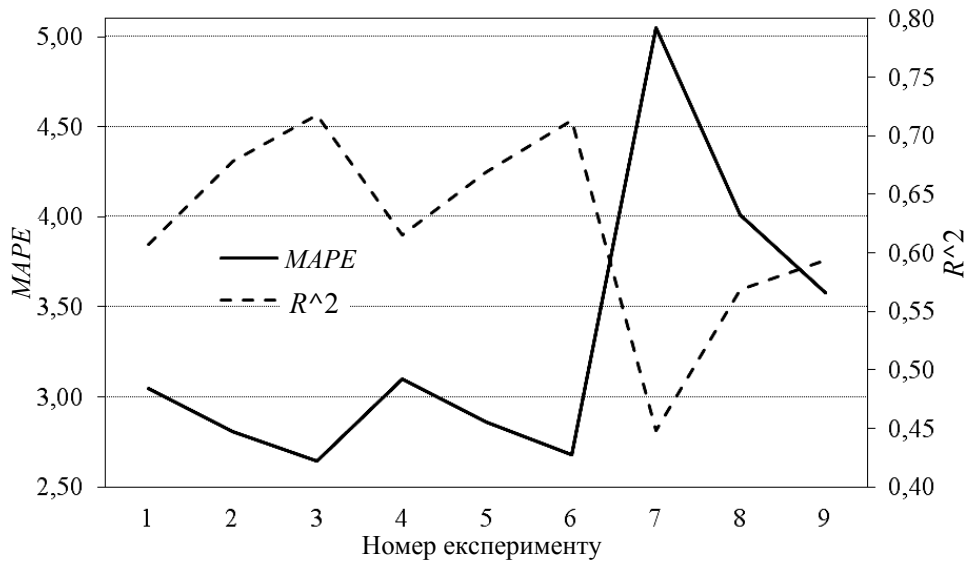


Рис. 7. Результати прогнозування експоненційно згладжених даних за допомогою гранулярного фільтра: по осі абсцис — номер експерименту; по осях ординат — значення статистик

Для вихідного ряду прийнятний прогноз (мінімальне значення  $MAPE$  та максимальне значення статистики  $R^2$ ) отримано для адаптивного фільтра Калмана з використанням методу максимальної правдоподібності для навчання на 100 перших значеннях вихідної вибірки. Для згладженого за Калманом ряду мінімальну  $MAPE$  отримано для оптимального адаптивного фільтра Калмана за припущення про незміщення рівня ряду  $q^*$ . Для згладженого за Калманом ряду максимальне значення статистики  $R^2$  отримано для адаптивного фільтра Калмана з використанням ММП для навчання на 100 перших значеннях вихідної вибірки.

## ВИСНОВКИ

У всіх випадках найкращий (мінімальне значення  $MAPE$  та максимальне значення статистики  $R^2$ ) прогноз отримано на експоненційно згладженій вибірці. Для вихідної вибірки отримано найнижчу якість прогнозу (максимальне значення  $MAPE$  та мінімальне значення статистики  $R^2$ ). Це цілком передбачувані результати, адже згладжування даних має підвищувати якість прогнозу. Виявилось, що оптимальний адаптивний фільтр Калмана за припущення про незміщення рівня ряду  $q^*$  дає дещо кращі результати, ніж фільтр за припущення про його зміщення.

Для адаптивного фільтра Калмана з використанням ММП кращі результати (мінімальне значення  $MAPE$  та максимальне значення статистики  $R^2$ ) отримано у випадку, коли навчання параметрів за ММП проводилось на вихідній вибірці.

Під час дослідження гранулярного фільтра виявилось, що тип повторної вибірки несуттєво впливає на якість прогнозу для процесу сонячної активності. Зі збільшенням кількості частинок, збільшувався час роботи програми. Тому необхідно знаходити компроміс між кількістю частинок та

якістю прогнозу у випадках, коли збільшення кількості частинок покращує прогноз. Гранулярний фільтр давав найкращі результати для пропорційної повторної вибірки з частотою 0,5 та для інтервальної повторної вибірки з довжиною інтервалу 1. Також виявилось, що проводити повторну вибірку в гранулярному фільтрі дійсно важливо, адже якщо її не проводити, якість прогнозу суттєво знижується.

У подальших дослідженнях для моделювання і прогнозування нелінійних нестационарних процесів, у тому числі індексу  $F_{10,7}$ , доцільно розширити номенклатуру моделей як методами класичної статистики та регресійного аналізу, так і сучасними методами інтелектуального аналізу даних (байєсівські мережі, дерева рішень, нейронечіткі структури). Створення спеціалізованої системи підтримання прийняття рішень на основі таких моделей надасть можливість далі підвищувати якість короткострокових прогнозів сонячної активності.

## ЛІТЕРАТУРА

1. *Clette F.* Revisiting the sunspot number / F. Clette, L. Svalgaard, J.M. Vaquero, E.W. Cliver. — Brussels: World Data Center Silso, 2016. — 80 p.
2. *Henney C.J.* Forecasting F10.7 with solar magnetic flux transport modeling / C.J. Henney, W.A. Toussaint, S.M. White, C.N. Arge // *Space Weather*. — 2012. — Vol. 10. — P. 1–9.
3. *Huang X.* Deep learning based solar flare forecasting / X. Huang, H. Wang, L. Xu et al. // *The Astrophysical Journal*. — 2018. — Vol. 856. — N 7. — P. 1–11.
4. *Tharsini A.D.* The Study of solar activity in relation with high frequency variations of solar radio flux / A.D. Tharsini, G. Shanti // *American Journal of Astronomy and Astrophysics*. — 2015. — Vol. 3. — N. 6. — P. 87–92.
5. *Tapping K.F.* Recent solar radio astronomy at centimeter wavelengths: The temporal variability of the 10.7-cm flux / K.F. Tapping // *Journal of Geophysical Research: Atmospheres*, 1987. — Vol. 92. — P. 829–838.
6. *Schmahl E.J.* Synoptic radio observations / E.J. Schmahl, M.R. Kundu // *Synoptic Solar Physics*. — 1998. — P. 387–400.
7. *Svalgaard L.* F10.7 Microwave Flux Matches the Total Disk Unsigned Magnetic Flux from MDI and HMI / L. Svalgaard, X. Sun // Technical Report, Hansen Experimental Physics Laboratory. — Stanford, CA 94305: Stanford University, 2016. — Available at: <http://hmi.stanford.edu/hminuggets/?p=1510>
8. *Dudok de Wit T.* Synoptic radio observations as proxies for upper atmosphere modeling / T. Dudok de Wit, S. Bruinsma, K. Shibasaki // *Journal of Space Weather and Space Climate*. — 2014. — P. 13–26.
9. *Schonfeld S.J.* Coronal Sources of the Solar F10.7 Radio Flux / S.J. Schonfeld et al. // *The Astrophysical Journal*. — 2015. — P. 29–39.
10. *Згуровский М.З.* Аналитические методы калмановской фильтрации для систем с априорной неопределенностью / М.З. Згуровский, В.Н. Подладчиков. — К.: Наук. думка, 1995. — 278 с.
11. *Arulampalam S.* A Tutorial on Particle Filters for on-line Non-linear/Non-Gaussian Bayesian Tracking / S. Arulampalam, S. Maskell, N. Gordon, T. Clapp // *IEEE Trans. Signal Processing*. — 2001. — Vol. 50. — P. 174–188.
12. *Kozierski P.* Resampling in particle filtering – comparison / P. Kozierski, M. Lis, J. Zietkiewicz // *Studia z automatyki informatyki*. — 2013. — Vol. 38. — P. 35–64.
13. *Довгий С.О.* Системи підтримки прийняття рішень на основі статистично-ймовірнісних методів / С.О. Довгий, О.М. Трофимчук, П.І. Бідюк. — К.: Логос, 2014. — 419 с.
14. *De Gooijer J.* Elements of nonlinear time series analysis and forecasting / J. De Gooijer. — Berlin: Springer, 2017. — 620 p.

Надійшла 23.10.2018