

ОЦІНКА ТОЧНОСТІ ПРОГНОЗУВАННЯ ЗМІНИ СТАНІВ КОЛИВНИХ ПРОЦЕСІВ ІЗ НЕКРАТНИМИ ЧАСТОТАМИ

М.І. Горбійчук, М.А. Шуфнарів, О.Т. Лазорів

ІФНТУНГ; 76019, м. Івано-Франківськ, вул. Карпатська, 15, тел. (0342) 504521,
e-mail: gorb@nung.edu.ua

Розглянуто новий метод побудови математичних моделей складних коливних процесів з некрatними частотами, який ґрунтується на ідеях генетичних алгоритмів. В порівнянні з методом групового урахування аргументів даний метод дає можливість синтезувати математичні моделі будь-якої складності без попереднього вибору числа рядів селекції. Крім того, використання генетичних алгоритмів значно зменшить число обчислень у порівнянні з індуктивним методом самоорганізації моделей, який передбачає значний обсяг перебору моделей із заданої множини. Новий метод може бути застосований для прогнозування як фізичних явищ, так і складних технологічних процесів. Як приклад застосування методу здійснено прогнозування рівня р. Дністер та оцінена вірогідність такого прогнозування за $\Delta^2(S)$ -критерієм та шляхом побудови довірчих інтервалів.

Ключові слова: синтез математичних моделей, генетичний алгоритм, хромосома, зовнішній критерій, некрatні частоти, гармонічний ряд, точність прогнозу

Rассмотрен новый метод построения математических моделей сложных колебательных процессов с некрatными частотами, основанный на идеях генетических алгоритмов. По сравнению с методом группового учета аргументов, данный метод дает возможность синтезировать математические модели любой сложности без предварительного выбора числа рядов селекции. Кроме того, использование генетических алгоритмов значительно уменьшит число вычислений по сравнению с индуктивным методом самоорганизации моделей, который предполагает значительный объем перебора моделей из заданного множества. Новый метод может быть применен для прогнозирования, как физических явлений, так и сложных технологических процессов. В качестве примера применения метода осуществлен прогноз уровня р. Днестр и оценена вероятность такого прогноза по $\Delta^2(S)$ -критерию и путем построения доверительных интервалов.

Ключевые слова: синтез математических моделей, генетический алгоритм, хромосома, внешний критерий, некрatные частоты, гармоничный ряд, точность прогноза

The article describes the new method for development of mathematical models of complex oscillating processes with nonmultiple frequencies, based on the ideas of genetic algorithms. When compared with the group method of data handling, this method provides an opportunity to synthesize mathematical models of any complexity without prechoosing of a number of selection rows. Besides, utilization of genetic algorithms will significantly decrease the number of calculations when compared with the inductive method of model self-organisation, which involves considerable model enumeration from the given set. The new method can be applied to forecast both physical phenomena, and complex technological processes. To exemplify this method utilization, the forecast of water level in the Dniester river was conducted and the reliability of such a forecast in accordance with the $\Delta^2(S)$ -criterion and confidence interval construction were evaluated.

Keywords: synthesis of mathematical models, genetic algorithm, chromosome, external criterion, nonmultiple frequencies, harmonic series, forecast accuracy.

Вступ

Методи прогнозування знаходять широке застосування у багатьох сферах людської діяльності особливо у складних технічних, екологічних, економічних, соціальних та інших системах.

Серед процесів, поведінку яких необхідно прогнозувати, слід виділити коливні процеси з некрatними частотами. Так, до них можна віднести процеси, які мають місце в екології, що зумовлено циклічністю цілого ряду природних факторів, наприклад, зміна рівня води у ріках, урожайність сільськогосподарських культур та ін.

Для регіонів Західної України актуальною науковою задачею є прогнозування паводків, які можуть призвести до катастрофічних наслідків, як це мало місце у 2008 році.

Аналіз сучасних публікацій

Процес прогнозування складається із двох етапів [1]: побудова моделі досліджуваного об'єкта; вибір структури моделі, визначення її параметрів і оцінка точності прогнозу за отриманою моделлю.

Процес прогнозування здійснюється з використанням математичних моделей, які можна отримати двома способами – шляхом побудови аналітичних або емпіричних моделей.

Якщо процеси, що протікають у моделі, мають зрозумілу фізичну природу, можна отримати модель, яка адекватно відтворює поведінку такого об'єкта при зроблених припущеннях і за певних умов. У багатьох випадках мають справу з об'єктами, складність яких зумовлена не тільки їх структурою, але й складною взаємодією такого об'єкта зі зовнішнім середовищем. У таких випадках для опису по-

ведінки складних об'єктів звертаються до емпіричного моделювання.

У загальному випадку емпірична модель включає у себе три складові – лінійний тренд $h(t)$, гармонічну складову $G(t)$ і залишкову складову $H(t)$, тобто

$$\hat{H}(t, \bar{x}) = H(\bar{x}) + G(t) + h(t). \quad (1)$$

Найчастіше залишкову складову вибирають у вигляді рівняння регресії, яке є лінійним відносно своїх параметрів

$$H(\bar{x}) = \sum_{k=0}^r a_k f_k(\bar{x}), \quad (2)$$

для визначення яких застосовують добре відомий метод найменших квадратів (МНК), зміст якого полягає у такому. Зміна залишкової складової $H(\bar{x})$ визначається вектором вхідних величин $\bar{x} = (x_1, x_2, \dots, x_n)^T$. У результаті спостережень за вхідними і вихідною величиною $Y = H(\bar{x})$ отримують множину значень – $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(N)}$ та $Y^{(1)}, Y^{(2)}, \dots, Y^{(N)}$, де N – кількість спостережень. Сукупність векторів $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(N)}$ утворюють, так звану, матрицю спостережень, рядки якої $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(N)}$.

У МНК припускають, що структура моделі (2) відома. Задача полягає у тому, щоб на основі спостережень за вхідними і вихідною величинами визначити параметри $a_k, k = \overline{1, r}$ моделі (2), таким чином, щоб якомога точніше наблизити вихід системи до виходу моделі. За критерій такого наближення вибирають суму квадратів відхилень розрахункових $y^{(i)}$ від експериментальних значень $Y^{(i)}, i = \overline{1, N}$.

За відомим $Y^{(i)}$ та обчисленим значенням $y^{(i)}$ згідно з формулою (2) можна отримати критерій наближення (апроксимації) $J(\bar{a}) = \sum_{i=1}^N (Y^{(i)} - y^{(i)})^2$, який у векторній формі набуде такого вигляду:

$$J(\bar{a}) = (\bar{Y} - F\bar{a})^T (\bar{Y} - F\bar{a}), \quad (3)$$

де \bar{Y} – вектор спостережень з компонентами $Y^{(i)}, i = \overline{1, N}$;

F – матриця, рядки якої значення функцій $f_i(\bar{x}^{(j)}), i = \overline{0, r}, j = \overline{1, N}$ у точках спостережень за величинами $\bar{x}^{(j)}$.

Значення параметрів моделі (2) обчислюють із умови, що критерій наближення (3) набуде мінімального значення відносно вектора параметрів \bar{a} . Мінімізація (3) приводить до такого результату:

$$M_F \bar{a} = F^T \bar{Y}, \quad (4)$$

де $M_F = F^T F$ – матриця Фішера.

У тому випадку, коли розмірність вектора \bar{a} невелика і матриця M добре обумовлена, із

рівняння (4) можна безпосередньо визначити $\bar{a} = M_F^{-1} F^T \bar{Y}$.

У переважній більшості випадків структура моделі (2) невідома, що приводить до необхідності вибору як структури самої моделі, так і її параметрів. Критерій наближення (2) є внутрішнім критерієм, [2] і його застосування призводить до помилкового висновку: чим складніша модель, тим вона точніша. Оскільки на вихід системи накладається перешкода (допускають, що вона адитивна), то надмірна точність моделі може значно спотворити об'єктивно існуючу функціональну залежність між виходом системи і її входами.

Тому для вибору структури моделі (2) акад. О. Г. Івахненком був запропонований метод, який дістав назву індуктивний метод самоорганізації моделей [3], ідейну сторону якого визначає теорема Геделя. Ця теорема стверджує, що жодна система аксіом не може бути логічно замкнутою: завжди знайдеться така теорема, для доведення якої необхідне зовнішнє доповнення – розширення початкової системи аксіом. Стосовно задачі синтезу моделі гедельвський підхід означає застосування зовнішнього критерію для однозначного вибору структури моделі із заданого класу моделей. Критерій називають зовнішнім, якщо його обчислення ґрунтується на даних, які не використовувались при розв'язанні задачі (2). Це означає, що всі дані, отримані у результаті експерименту, розбиваються на дві частини – навчальну N_A і перевірну N_B .

Для вибору структури моделі (2) здебільшого використовують критерій регулярності

$$\Delta^2(B) = \frac{\sum_{i=1}^{N_B} (Y^{(i)}(B) - y^{(i)}(B))^2}{\sum_{i=1}^{N_B} Y^{(i)}(B)^2} \quad (5)$$

і зміщення

$$\Delta^2(A, B) = \frac{\sum_{i=1}^N (y^{(i)}(A) - y^{(i)}(B))^2}{\sum_{i=1}^N (Y^{(i)})^2}, \quad (6)$$

де $y^{(i)}(A), y^{(i)}(B)$ – значення виходу моделі, обчислені відповідно на множинах експериментальних значень N_A і N_B .

Якщо вибрано критерій регулярності (5), то вибирають такий розподіл даних експерименту [4]: $N_A = 0,7N$ і $N_B = 0,3N$, а при виборі критерію (6) – розподіл $N_A = 0,5N$ і $N_B = 0,5N$.

Реалізація індуктивного методу самоорганізації моделей здійснюється поетапно: перший етап – генерування моделей-претендентів (у певному порядку підвищення їх складності); другий етап – вибір найкращої моделі за мінімумом критерію селекції (5) або (6).

Розрізняють три способи генерування моделей-претендентів. Перший із них комбіна-

рний [4], за яким вибір моделей-претендентів здійснюється шляхом прирівнювання до нуля деяких коефіцієнтів у виразі (2), що дає змогу отримати сукупність моделей. Вибір найкращої із них здійснюється на основі одного із критеріїв селекції (5) або (6). Другий спосіб відомий як метод групового врахування аргументів (МГУА) [5], у якому генерування моделей-претендентів здійснюється за допомогою багаторядної процедури. У першому ряду селекції комбінують всі можливі пари аргументів (вхідних величин), і для кожної із них утворюють моделі, наприклад, у вигляді повного полінома. Із всіх часткових моделей, утворених у такий спосіб, вибирають K найкращих за одним із критеріїв селекції. Із виходів цих K моделей знову утворюють комбінації всіх можливих пар, які є входами другого ряду селекції. Для кожної із цих пар знову формують часткові моделі, із яких вибирають K найкращих. Процес нарощування рядів селекції відбувається до тих пір, поки вибраний критерій селекції зменшує своє значення. Оцінка параметрів часткових моделей здійснюється за допомогою МНК. Третій метод [3] подібний до другого. Різниця полягає лише у тому, що на кожному ряді селекції часткові моделі утворюють шляхом прирівнювання до нуля певного числа їх коефіцієнтів.

Недоліком комбінаторного методу селекції моделей є необхідність перебору великої кількості часткових моделей. Так, якщо за початкову модель вибрано повний поліном степені m , то загальне число моделей претендентів буде $2^M - 1$, де M – загальне число членів початкового полінома. Навіть сучасні ЕОМ не здатні реалізувати комбінаторний алгоритм селекції моделей при значній кількості вхідних величин і при високій степені початкового полінома. МГУА синтезує моделі, в яких фігурують проміжні змінні, що значно ускладнює процес переходу до вхідних змінних системи. Сказане відноситься і до третього методу, оскільки він, по суті, є модифікацією МГУА.

МГУА має той недолік, що у відповідності до прийнятої процедури синтезу емпіричних моделей вихідні змінні досить складно виразити у явному вигляді через вхідні змінні об'єкта.

Із усіх трьох методів найпривабливішим є комбінаторний метод, оскільки він дає змогу отримати оптимальну модель, де аргументами виступають вхідні змінні системи.

Гармонічна складова $G(t)$ з некрatними частотами відноситься до ширшого класу моделей, які отримали назву сезонних. До сезонних відносять такі явища, які у своєму розвитку мають тенденцію до більш-менш регулярного повторення. Для синтезу сезонних моделей використовують два різних підходи [6]: метод корегуючих коефіцієнтів і подання циклічних змін у вигляді ряду Фур'є. Метод коригуючих коефіцієнтів чутливий до зовнішніх впливів, а точність прогнозу за методом Фур'є залежить від допрогнозного проміжку спостережень, який визначає кількість гармонік ряду. А це

вимагає значного усереднення амплітуди коливань, тобто отримані амплітуди не будуть відповідати прогнозним значенням. Це означає, що використовувати такий метод доцільно лише для короткотермінових прогнозів.

Для прогнозування сезонних змін також використовують модель, яка дістала назву “процес авто – регресії” – проінтегрованого ковзкого середнього (АПККС) [7, 8] або метод Бокса-Дженікса.

У методі Бокса-Денікса не використовуються незалежні змінні. Синтез моделі АПККС ґрунтується на інформації, що міститься у передісторії прогнозованих рядів.

Виклад основного матеріалу

Тренд $h(t)$, який є адитивною складовою рівняння (1), описується лінійним рівнянням регресії (наприклад, першого порядку), а його коефіцієнти визначають за методом найменших квадратів.

Гармонічну складову з некрatними частотами будемо описувати таким виразом:

$$G(t) = A_0 + \sum_{j=1}^m (A_j \sin(t\omega_j) + B_j \cos(t\omega_j)), \quad (7)$$

де t – такти відліку часу, $t = 1, 2, 3, \dots, N$;

A_0, A_j, B_j – параметри гармонічного ряду (7);

$\omega_j = \omega_{j-1} + \Delta\omega_j$ – некрatні частоти, $j = 1, 2, 3, \dots$;

Для того, щоб можна було б оцінити параметри ряду (7), необхідно виконати умову [3] $N \geq 3m + 1$.

Суму декількох гармонік ряду (7), у якому коефіцієнти A_0, A_j, B_j визначені за методом найменших квадратів, а число гармонік і їх частоти вибрані так, щоб отримати мінімум деякого зовнішнього критерію селекції, називають [3] гармонічним трендом оптимальної складності.

Ідентифікація параметрів моделі (7) відбувається у декілька етапів [4].

Перший етап – обчислення коефіцієнтів α_p із умови

$$\sum_{p=0}^{m-1} \alpha_p (g(i+p) + g(i-p)) = g(i+m) + g(i-m),$$

$$i = \overline{m+1, N-m},$$

де $g(t)$ – відліки реалізації процесу у моменти часу, що симетрично розміщені відносно довільної точки i .

Балансові коефіцієнти α_p знаходять [3] із умови мінімізації нев'язки

$$B = \sum_{i=m+1}^{N-m} b_i^2, \quad (8)$$

де

$$\bar{z}_m = \begin{bmatrix} \bar{g}(2m+1) + \bar{g}(1) \\ \bar{g}(2m+2) + \bar{g}(2) \\ \dots \\ \bar{g}(N) + \bar{g}(N-2m) \end{bmatrix},$$

$$F_m = \begin{bmatrix} 2\bar{g}(m+1) & \bar{g}(m+2) + \bar{g}(m) & \bar{g}(m+3) + \bar{g}(m-1) & \dots & \bar{g}(2m) + \bar{g}(2) \\ 2\bar{g}(m+2) & \bar{g}(m+3) + \bar{g}(m+1) & \bar{g}(m+4) + \bar{g}(m) & \dots & \bar{g}(2m+1) + \bar{g}(3) \\ \dots & \dots & \dots & \dots & \dots \\ 2\bar{g}(N-m) & \bar{g}(N-m+1) + \bar{g}(N-m-1) & \bar{g}(N-m+2) + \bar{g}(N-m-2) & \dots & \bar{g}(N-1) + \bar{g}(N-2m+1) \end{bmatrix}$$

Рисунок 1 – Матриці \bar{z}_m, F_m

$$b_i = g(i+m) - \sum_{p=0}^{m-1} \alpha_p (g(i+p) + g(i-p)) + g(i-m),$$

$$i = \overline{m+1, N-m}.$$

Отже, будемо розв'язувати задачу

$$\min_{\bar{\alpha}} : J(\bar{\alpha}) = \sum_{i=m+1}^{N-m} \left(z_{i,m} - \sum_{p=0}^{m-1} \alpha_p g_{i,p} \right)^2, \quad (9)$$

де $\bar{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{m-1})^T$ – вектор вагових коефіцієнтів;

$$z_{i,m} = \bar{g}(i+m) + \bar{g}(i-m);$$

$$g_{i,p} = \bar{g}(i+p) + \bar{g}(i-p);$$

T – символ транспонування матриць.

Задачу (3) запишемо у матрично-векторній формі

$$\min_{\bar{\alpha}} : J(\bar{\alpha}) = (\bar{z}_m - F_m \bar{\alpha})^T (\bar{z}_m - F_m \bar{\alpha}), \quad (10)$$

де матриці \bar{z}_m, F_m зображені на рис. 1.

Мінімізація виразу (10) приводить до нормального рівняння Гауса, яке у матричній формі матиме такий вигляд:

$$F_m^T F_m \bar{\alpha} = F_m^T \bar{z}_m. \quad (11)$$

Із останнього рівняння можна знайти

$$\bar{\alpha} = (F_m^T F_m)^{-1} F_m^T \bar{z}_m. \quad (12)$$

Використовувати формулу (12) можна лише тоді, коли розмірність вектора $\bar{\alpha}$ невелика і матриця $F_m^T F_m$ є добре обумовленою. Якщо така умова не виконується, то для знаходження $\bar{\alpha}$ слід розв'язувати рівняння (11) одним із числових методів, наприклад, методом Гауса зі зворотнім ходом [9].

На другому етапі складається рівняння $\alpha_0 + \sum_{p=1}^{m-1} \alpha_p \cos(p\omega) = \cos(m\omega)$ для довільної частоти ω , яке за допомогою рекурентного співвідношення [3]

$$\cos(p\omega) = 2\cos((p-1)\omega)\cos\omega - \cos((p-2)\omega),$$

$$p = \overline{2, m}$$

приводиться до алгебраїчного рівняння m -ого степеня відносно $\cos\omega$

$$P_m z^m + P_{m-1} z^{m-1} + \dots + P_1 z + P_0 = 0, \quad (13)$$

де $z = \cos\omega$.

Рівняння (13) має m коренів, які однозначно визначають частоти $\omega_j, j = \overline{1, m}$.

Суть третього етапу полягає у тому, що, знаючи вагові коефіцієнти α_p , можемо скласти рівняння (13), розв'язок якого відносно z дає змогу однозначно визначити частоти гармонік $\omega_j, j = \overline{1, m}$. Тепер задача полягає в оптимальному синтезі гармонічного ряду (7).

Відомі два підходи [4] до вирішення поставленої задачі. Перший з них передбачає викреслювання гармонік у різних комбінаціях із повного ряду $\omega_j, j = \overline{1, m}$.

Другий метод ґрунтується на ідеях багаторядних алгоритмів групового урахування аргументів (МГУА). У відповідності з цим методом число гармонік, що включаються у модель, постійно зростає. Це відбувається до тих пір, поки критерій селекції зменшується. Найпростішим є алгоритм з послідовним виділенням найкращої моделі у кожному ряду. Але ефективнішим є алгоритм, коли виділяється кілька гармонік у кожному ряді. Нехай отримана деяка реалізація вихідної величини процесу довжиною N . Деяка частина цих даних, яка вміщує N_R послідовних точок спостережень, виділяється у навчальну послідовність. Інші точки розбиваються на дві частини: перша N_Q – перевірна і друга N_S – екзаменаційна. Всього точок:

$N = N_R + N_Q + N_S$. На першому ряді селекції за всіма заданими точками виділяються всі можливі тренди гармонічного ряду; максимальне число трендів [2] $m_{i \text{ до } 3} \leq \frac{N}{3}$. Із них вибирається

не єдиний тренд, а q трендів, які якнайкраще задовольняють вибраному критерію селекції. Після цього обчислюється q залишків (залишком називають різницю ординат коливного процесу і кожного із трендів першого ряду). На другому ряду селекції із кожного залишку знову виділяється q_1 трендів. Із всієї множини qq_1

отриманих трендів другого ряду за тим же критерієм селекції вибирається q кращих трендів цього ряду і т. д. Величину свободи вибору q рекомендується вибирати на основі ряду проб, а вибір кращих трендів здійснюється за точками окремої перевіркої послідовності. Складність моделі (число рядів селекції) збільшується до тих пір, поки зменшується величина критерію селекції. На останньому ряді селекції вибирається єдиний розв'язок, що відповідає мінімуму критерію селекції.

Недоліком першого підходу до вирішення поставленої задачі є необхідність перебору великої кількості варіантів, визначається як сума

$$S_v = \sum_{i=1}^m C_m^i. \quad \text{Відомо,} \quad \text{що}$$

$$(a + b)^n = \sum_{i=0}^n C_n^i a^{n-i} b^i = a^n + \sum_{i=1}^n C_n^i a^{n-i} b^i. \quad \text{Якщо}$$

$$a = b = 1, \text{ то } \sum_{i=1}^m C_m^i = 2^m - 1. \text{ Отже, } S_v = 2^m - 1.$$

Наприклад, при $m = 20$ необхідно перебрати 1048575 варіантів, що потребує значних затрат машинного часу. Для другого підходу характерним є те, що у результаті реалізації багаторядного алгоритму МГУА неможливо отримати математичну модель у явному вигляді, що є суттєвим недоліком такого методу.

Нами запропоновано інший підхід до побудови математичних моделей коливних процесів, який базується на ідеях генетичних алгоритмів. Суть такого підходу полягає у наступному.

Вся реалізація вихідної величини процесу або явища розбивається на три частини у такій пропорції [2]: $N_R = 0,7N$, $N_Q = 0,2N$ і $N_S = 0,1N$. Для множини даних $N_R + N_Q$ визначаються вагові коефіцієнти α_p як розв'язок лінійного алгебраїчного рівняння (4) за методом виключення Гауса з вибором головного елемента [9]. Розв'язок рівняння (11) відносно змінної z дає можливість знайти частоти ω_j , $j = \overline{1, m}$. За відомими частотами ω_j на множині точок $N_R + N_Q$ необхідно знайти параметри моделі (8) A_0 , A_j і B_j . Поставлену задачу будемо розв'язувати, використовуючи генетичні алгоритми [10]. Утворимо упорядковану структуру довжиною m , в якій на i -тому місці стоятиме нуль або одиниця, що залежить від того,

вилучена частота ω_j із вибраного повного ряду m , чи залишена. У теорії генетичних алгоритмів така упорядкована послідовність має назву хромосоми або особи, а атомарний елемент хромосоми (одиниця або нуль) – це ген. Набір хромосом утворює популяцію. Важливим поняттям у теорії генетичних алгоритмів є функція пристосування, яка визначає ступінь пристосування окремих осіб у популяції. Вона дає змогу із всієї популяції вибрати прийнятно пристосовані особи, тобто такі, які мають найбільше (найменше) значення функції пристосування. У задачі синтезу моделей коливних процесів функцією пристосованості виступає комбінований критерій селекції [5]

$$\rho = \sqrt{n_d^2 + B^2}, \quad (14)$$

де n_d^2 – критерій зміщення, який обчислюється за такою формулою:

$$n_d^2 = \frac{\sum_{i=1}^N (g_i(R) - g_i(S))^2}{\sum_{i=1}^N \bar{g}_i^2}; \quad (15)$$

B – функція нев'язки, що визначається як (8);

$g_i(R)$, $g_i(S)$ – величини, значення яких обчислені відповідно на множині точок N за формулою (7), а коефіцієнти моделі (7) знайдені відповідно на множинах $N_R + N_Q$ і N_S .

Таким чином, задачу синтезу моделі коливного процесу сформуємо наступним чином: із початкової популяції хромосом шляхом еволюційного відбору вибрати таку, хромосому, яка забезпечує найкраще значення функції пристосування (мінімальне значення критерію селекції (14)).

Запропонований алгоритм синтезу моделей коливних процесів оптимальної складності, як і класичний генетичний алгоритм, складається із таких кроків [10, 11].

K1. Формування початкової популяції (ініціалізація). На першому кроці роботи алгоритму випадковим чином формується популяція із I осіб, кожна із яких є хромосомою довжиною m .

K2. Оцінка пристосованості хромосоми у популяції. Для кожної хромосоми обчислюється критерій селекції (14). Здійснюється така процедура наступним чином. У відповідності з моделлю (7) формується матриця F (рис. 2).

$$F = \begin{bmatrix} 1 & \sin \omega_1 & \cos \omega_1 & \sin \omega_2 & \cos \omega_2 & \dots & \sin \omega_m & \cos \omega_m \\ 1 & \sin(2\omega_1) & \cos(2\omega_1) & \sin(2\omega_2) & \cos(2\omega_2) & \dots & \cos(2\omega_m) & \cos(2\omega_m) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \sin(N\omega_1) & \cos(N\omega_1) & \sin(N\omega_2) & \cos(N\omega_2) & \dots & \cos(N\omega_m) & \cos(N\omega_m) \end{bmatrix}$$

Рисунок 2 – Матриця F

У сформованій хромосомі подвоюємо одиниці і нулі. Наприклад, якщо згенерована на першому кроці хромосома була такою: $Ch = [1001011]$, то після виконання операції подвоєння вона набуде такого вигляду: $Ch_d = [11000011001111]$. Оскільки у моделі (7) завжди присутній коефіцієнт A_0 , то до хромосоми Ch_d на першу позицію додаємо одиничний ген, у результаті чого отримаємо $Ch_{do} = [111000011001111]$. Необхідність операції подвоєння пояснюється тим, що кожній частоті ω_j відповідає пара коефіцієнтів A_j, B_j . У відповідності до сформованої хромосоми Ch_{do} із матриці F формуємо нову матрицю F_{new} шляхом вилучення тих стовпців із матриці F , які асоційовані із нулями хромосоми Ch_{do} . Із отриманої матриці утворимо дві матриці F_A і F_B розмірами $(N_R + N_Q) \times m_1$ і $N_S \times m_1$. Матрицю F_R утворюють перші $N_R + N_Q$ стовпці матриці F_{new} , а другу – останні N_S стовпці матриці F_{new} . На множині точок $N_R + N_Q$ і N_S обчислюються ненульові коефіцієнти A_0, A_j, B_j моделі (7) шляхом розв'язання нормального рівняння Гаусса

$$F_R^T F_R \bar{A}_R = F_R^T \bar{g}_R, \quad (16)$$

$$F_S^T F_S \bar{A}_S = F_S^T \bar{g}_S, \quad (17)$$

де

$$\bar{A}_R = (A_0^{(R)}, A_1^{(R)}, B_1^{(R)}, A_2^{(R)}, B_2^{(R)}, \dots, A_{m_1}^{(R)}, B_{m_1}^{(R)})^T,$$

$$\bar{A}_S = (A_0^{(S)}, A_1^{(S)}, B_1^{(S)}, A_2^{(S)}, B_2^{(S)}, \dots, A_{m_1}^{(S)}, B_{m_1}^{(S)})^T -$$

вектори параметрів моделі, яка асоційована з черговою хромосомою із початкової популяції і обчислені за формулою (11);

$$\bar{g}_R = (\bar{g}^{(1)}, \bar{g}^{(2)}, \dots, \bar{g}^{(N_R+N_Q)})^T,$$

$$\bar{g}_S = (\bar{g}^{(1)}, \bar{g}^{(2)}, \dots, \bar{g}^{(N_S)})^T -$$

вектори експериментальних даних на множині точок $N_R + N_Q$ і N_S .

За відомою сукупністю коефіцієнтів \bar{A}_R і \bar{A}_S моделі (7) на множині точок N обчислюють

$$g(R) = F_{new} \bar{A}_R,$$

$$g(S) = F_{new} \bar{A}_S.$$

За формулами (14) і (15) обчислюють критерій селекції, де B знаходять у відповідності з (8). Значення критерію селекції обчислюють для кожної хромосоми, в результаті чого отри-

мують множину значень $\rho_i, i = \overline{1, M}$, де M – кількість хромосом у популяції.

K3. Перевірка умови зупинки алгоритму. Визначають

$$\rho_{\min} = \min_{i \in M} \rho_i. \quad (18)$$

Якщо мінімальне значення (18) критерію селекції (14) не перевищує деякого додатного значення ε , то відбувається зупинка алгоритму. Зупинка алгоритму також може відбутися у випадку, коли його виконання не приводить до покращення функції пристосування або у тому випадку, коли алгоритмом уже виконано задане число ітерацій.

Після виконання однієї із трьох умов із популяції вибирається хромосома ch^* , для якої виконується умова (18). Після операції подвоєння і приєднання одиничного гену до хромосоми Ch_d^* отримуємо Ch_{do}^* . Ця хромосома задає структуру моделі оптимальної складності і формує матрицю F^* таким чином, що із початкової матриці F вилучаються стовпці, які асоційовані з нульовими генами хромосоми Ch_{do}^* . Перерахунок параметрів моделі (7) здійснюється на множині всіх точок початкового масиву даних.

K4. Селекція хромосом. За розрахованими на другому кроці значеннями функції пристосування здійснюється відбір тих хромосом, які будуть брати участь в створенні потомків для наступної популяції. Такий вибір проводиться у відповідності з принципом природного відбору, коли найбільші шанси у створенні нової популяції мають хромосоми з найкращим значенням функції пристосування, тобто такі, що забезпечують мінімальне значення критерію селекції (14). Найбільш поширеними методами селекції [10] є метод рулетки і метод турнірної селекції. У даному алгоритмі використано турнірний метод, який можна використовувати як у задачах максимізації, так і у задачах мінімізації функцій. При турнірній селекції всі хромосоми розбиваються на підгрупи з наступним вибором із кожної утвореної підгрупи хромосоми з найкращою пристосованістю. Підгрупи можуть мати довільний розмір, але найчастіше популяції ділять на підгрупи по 2 – 3 особи у кожній.

K5. Формування нової популяції потомків здійснюється за допомогою двох основних операторів: схрещування і мутації [11]. Слід відмітити, що оператор мутації відіграє другорядну роль у порівнянні з оператором схрещування. Це означає, що у генетичному алгоритмі схрещування проводиться майже завжди, а мутація – досить рідко. Вірогідність схрещування досить велика ($0,5 \leq P_c \leq 1$), тоді як ймовірність мутації вибирається досить малою ($0 \leq P_m \leq 0,1$).

Оператор мутації з ймовірністю P_m змінює значення гена в хромосомі на протилежне, тобто з 1 на 0 чи з 0 на 1. Мутація може здійснюватися як над пулом родичів, так і над пулом потомків.

Після виконання оператора схрещування відбувається перехід до K2.

Після вилучення тренду $h(t)$ і складової $G(t)$ із моделі (1) отримаємо залишкову складову $H(\bar{x}) = \tilde{H}(t, \bar{x}) - G(t) - h(t)$, яка є функцією незалежних змінних $\bar{x} = (x_1, x_2, \dots, x_n)^T$.

Співвідношення (2) будемо шукати у вигляді полінома

$$y = \sum_{i=0}^{M-1} a_i \prod_{j=1}^n x_j^{s_{ji}}, \quad (19)$$

де M – кількість членів полінома;

a_i – коефіцієнти полінома;

s_{ji} – степені аргументів, які повинні задовольняти обмеженню – $\sum_{j=1}^n s_{ji} \leq m$. Число членів

M полінома (19) визначають за такою формулою [12]: $M = \frac{(m+n)!}{m!n!}$.

Вхідні величини $x_j, j = \overline{1, n}$, які є аргументами залежності $y = H(\bar{x})$, у кожному спостереженні $t = \overline{1, N}$ приймають певне значення так, що їх сукупність утворює матрицю X . Реакція деякої системи на сукупність значень X характеризується вектором \bar{Y} .

Допустимо, що нам відомі параметри $a_i, i = \overline{1, M-1}$ моделі (19). Тоді за відомими значеннями величин $x_j^{(t)}$ можна обчислити

$$y_t = \sum_{i=0}^{M-1} a_i \prod_{j=1}^n (x_j^{(t)})^{s_{ji}}, \quad t = \overline{1, N}.$$

Множина значень $y_t, t = \overline{1, N}$ утворює вектор \bar{y} . Знаючи \bar{Y} і \bar{y} із рівняння (4), можна визначити параметри моделі (19).

На практиці, як правило, структура моделі (19) невідома, що вимагає довільного вибору як числа функцій, так і вигляду самих функцій у моделі (19).

Проведений аналіз показав, що з усіх трьох методів найпривабливішим є комбінаторний метод, оскільки він дає можливість отримати модель, де аргументами виступають вхідні величини системи. Його реалізація наштовхується на проблему під назвою «прокляття розмірності»; зі збільшенням числа незалежних змінних і степені полінома (19) катастрофічно зростає кількість можливих варіантів перебору моделей-претендентів.

Для зняття проблеми великої розмірності застосуємо генетичний підхід. Як емпіричну модель будемо розглядати поліном (19) степені m . При комбінаторному методі синтезу моделі із повного полінома (19) отримують емпіричну модель, де частина параметрів набуває значення «нуль». Інші параметри, що залишилися, будуть відмінні від нуля. Утворимо упорядковану структуру довжиною M , в якій на i -тому місці буде стояти одиниця або нуль залежно від того, чи параметр $a_i, i = \overline{1, M}$ моделі (19) відмінний від нуля, чи нульовий.

Таким чином, задачу синтезу емпіричної моделі сформуємо наступним чином: із початкової популяції хромосом шляхом еволюційного відбору вибрати таку, хромосому, яка забезпечує найкраще значення функції пристосування (мінімальне значення критерію селекції (5) або (6)).

Алгоритм розв'язку поставленої задачі [13] аналогічний раніше розробленому для виділення гармонічного тренду з тією лише різницею, що не здійснюється операція подвоєння генів у хромосомах.

Оскільки, спостереження за реальними процесами, як правило, завжди проводиться в умовах різних завад, а протікання самих процесів піддається впливу різноманітних випадкових факторів, тому не можна впевнитись в тому, що прогнозоване майбутнє значення процесу за математичною моделлю буде отримане абсолютно точним. Важливим етапом побудови прогнозу є оцінка точності результатів прогнозування.

Для оцінки виконаного прогнозу використовують різноманітні критерії точності в залежності від вибраного методу моделювання.

Для оцінки якості прогнозу при самоорганізації прогнозуючих моделей з вхідних експериментальних даних крім навчальної N_R і перевірної N_Q вибірок виділяють частину даних на екзаменаційну вибірку N_S . На новій множині даних N_S , що не використовується в побудові математичної моделі, визначають середньоквадратичне відхилення обчисленого за моделлю прогнозованого значення y_i і дійсного значення Y_i [14]:

$$\Delta^2(S) = \frac{\sum_{i=1}^{N_S} (Y_i - y_i)^2}{\sum_{i=1}^{N_S} Y_i^2} \rightarrow \min. \quad (20)$$

Обчислене за формулою (20) значення критерію дає можливість встановити якість прогнозу: за умови $\Delta^2(S) \leq 0,5$ прогноз вважається високої якості, при $\Delta^2(S) \leq 0,8$ – прогноз задовільний. При $\Delta^2(S) > 1$ застосовувати прогнозуючу модель немає змісту, оскільки модель не точна, тобто мають місце значні відхилення від реального досліджуваного процесу.

Інший метод визначення надійності прогнозу – це побудова довірчих інтервалів.

Помилку прогнозу в кожній прогнозованій точці $N+k$ можна визначити у вигляді різниці

$$\delta_{N+k} = y_{N+k} - \tilde{y}_{N+k},$$

де $\tilde{y}_{N+k}, k = \overline{1, N_{pr}}$ – точковий прогноз, зроблений за результатами спостережень за проце-

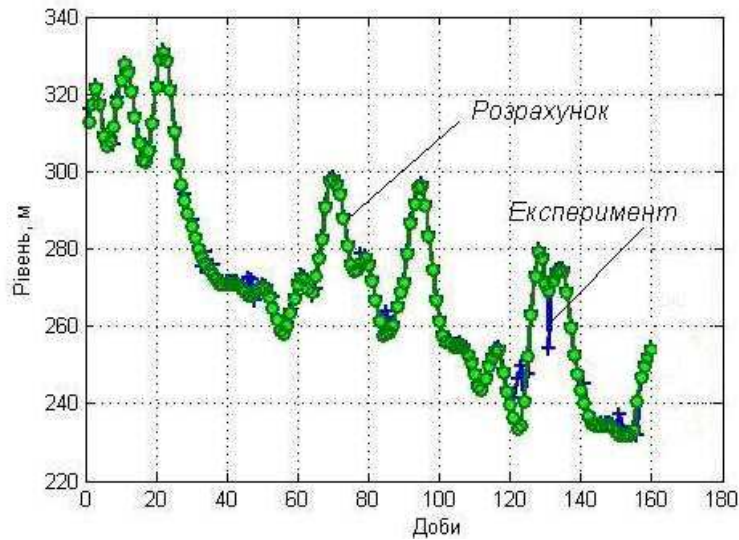


Рисунок 3 – Залежність рівня води у р. Дністер від погодних умов

сом до моменту N включно при припущенні про незмінність структури моделі (19) та її коефіцієнтів;

y_{N+k} – дійсне значення процесу в момент дискретного часу t_{N+k} .

У випадку точного прогнозу $\delta_{N+k} = 0$. Насправді, прогнозовані і фактичні значення будуть відрізнятися між собою. Щоб охарактеризувати відмінність, побудуємо лінію регресії між фактичними значеннями y_{N+k} і відповідними розрахунковими значеннями \tilde{y}_{N+k}

$$y_{N+k}^{(pr)} = a_p + b_p y_{N+k} \quad (21)$$

Довірчий інтервал прогнозованого значення \tilde{y}_{N+k} для оцінки точності моделі можна обчислити за такою формулою [15]:

$$y_{N+k}^{(pr)} - t_{1-\alpha/2} s_k \leq \tilde{y}_{N+k} \leq y_{N+k}^{(pr)} + t_{1-\alpha/2} s_k, \quad (22)$$

де $t_{1-\alpha/2}$ – безрозмірна статистика Стюдента, яка підпорядкована t -розподілу з $N_{pr} - 2$ степенями вільності і для якої рівень значущості α ;

$$s_k^2 = \tilde{s}_{N_{pr}}^2 \left(\frac{1}{N_{pr}} + \frac{(\tilde{y}_{N+k} - y_{N+k}^{(pr)})^2}{\sum_{i=1}^{N_{pr}} (\tilde{y}_{N+i} - y_{N+i}^{(pr)})^2} \right) -$$

оцінка дисперсії відносно лінії регресії;

$$\tilde{s}_{N_{pr}}^2 = \frac{1}{N_{pr} - 1} \sum_{i=1}^{N_{pr}} (\tilde{y}_{N+i} - y_{N+i}^{(pr)})^2.$$

Розроблений метод синтезу моделей оптимальної складності та оцінка точності прогнозу за $\Delta^2(S) \leq 0,5$ - критерієм та за довірчими інтервалами був застосований для прогнозування рівня води у р. Дністер. Математична модель, що описує зміну рівня води за часом, була вибрана у вигляді співвідношення (1) і включає у

себе лінійний тренд $h(t) = a_0 + a_1 t$, гармонічну складову $G(t)$ (7) та залишкову складову $H(\bar{x})$. Залишкова складова апроксимується поліномом (19) і залежить від погодних умов, які представлені такими параметрами - T_t - середньодобова температура повітря; $f_t, f_{t-1}, f_{t-2}, f_{t-3}$ - середньодобова кількість опадів в поточний момент та напередодні; v_t - середньодобова швидкість вітру; p_t - середньодобовий барометричний тиск. Вибір структури та визначення коефіцієнтів математичної моделі (1) здійснювалося на основі навчальної та перевірної вибірок, що містить $N_R + N_Q = 160$ спостережень за рівнем води та погодними умовами. Результат синтезу моделі оптимальної складності відображає рис. 3.

Оцінка адекватності математичної моделі (1) здійснювалась за довірчими інтервалами [16], для яких рівень значущості $\alpha = 0,01$. Аналіз отриманих результатів показав, що переважна більшість точок (96,2 %) потрапляють у довірчий інтервал. Це дає змогу з високим ступенем імовірності стверджувати про адекватність побудованої математичної моделі. Вихід невеликого числа точок за межі довірчого інтервалу можна пояснити можливими промахами при вимірюваннях рівня, які ведуться візуально, або друкарськими помилками при підготовці звіту і звідки були взяті експериментальні дані.

Для оцінки якості прогнозу за допомогою моделі (1) з вхідних даних крім навчальної N_R і перевірної N_Q вибірок виділялась частина даних на екзаменаційну вибірку $N_S = 24$. На рис. 4 показані отримані на екзаменаційній множині експериментальні значення та значення рівня води обчислені за побудованою моделлю (1).

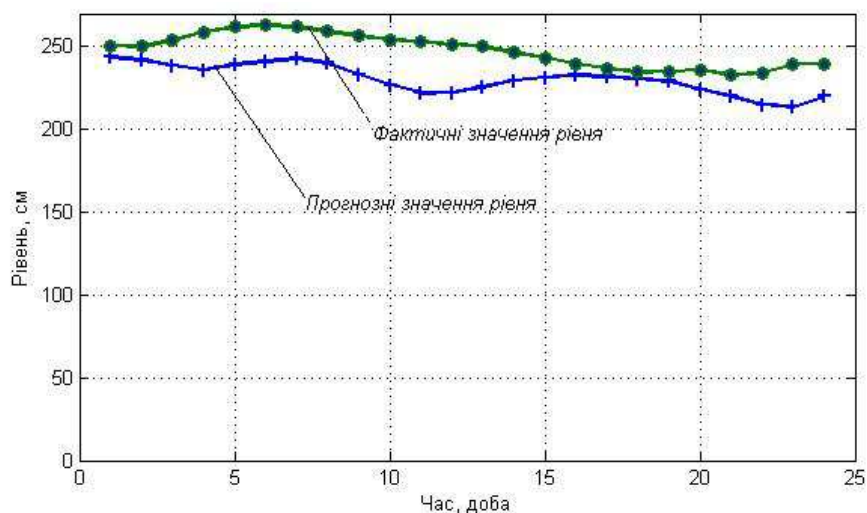


Рисунок 4 – Результати прогнозування рівня води на экзаменаційній множині

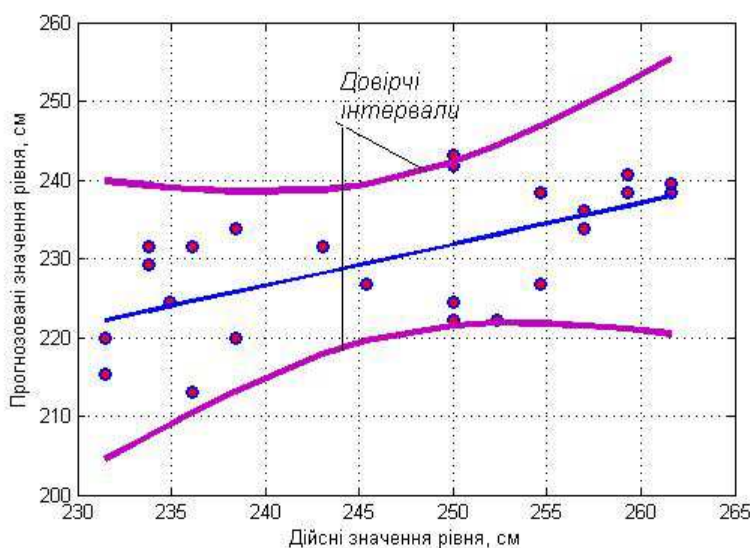


Рисунок 5 – Оцінка точності прогнозу за довірчими інтервалами

На множині точок N_s визначали значення $\Delta^2(S)$ -критерію за формулою (20). Отримане значення $\Delta^2(S)=0,0058$ свідчить про достатню точність прогнозування рівня ріки Дністер за даним критерієм.

Оцінка точності прогнозування за довірчими інтервалами здійснювалась у відповідності до співвідношення (22). Результати обчислень довірчого інтервалу відображає рис. 5.

Аналіз отриманих результатів свідчить, що з ймовірністю 0,99 можна стверджувати про достовірність зробленого прогнозу на вибраному інтервалі екстраполяції.

Висновки

Розроблено метод побудови математичних моделей складних процесів на засадах генетичних алгоритмів, які можна подати як композицію трьох складових – тренду, коливної складової з некротними частотами та залишкової складової, що описується рівнянням регресії

вибраного порядку. В порівнянні з багаторядним алгоритмом групового урахування аргументів даний метод дає змогу синтезувати математичні моделі будь-якої складності без попереднього вибору числа рядів селекції та вибрати оптимальну за структурою адекватну модель і значно зменшити об'єм обчислень.

Розроблена методика синтезу математичних моделей оптимальної складності на засадах генетичних алгоритмів апробована на прикладі прогнозування рівня води в ріці Дністер. Точність результатів прогнозування, отриманих за побудованою математичною моделлю зміни рівня води ріки Дністер, перевірено на множині даних, що не використовувались при побудові моделі. Обчислене значення критерію $\Delta^2(S)=0,0058$, а також побудовані довірчі інтервали свідчать проте, що синтезована математична модель з вірогідністю до 0,99 здатна прогнозувати зміну рівня води в ріці Дністер відповідно до погодних умов з випередженням до 24 діб.

Література

- 1 Чуев Ю. В. Прогнозирование количественных характеристик процессов / Ю. В. Чуев, Ю. Б. Михайлов, В. И. Кузьмин. – М.: Сов. радио, 1975. – 400 с.
- 2 Ермаков С. М. Математическая теория оптимального эксперимента: учебное пособие / С. М. Ермаков, А. А. Жиглявский. – М.: Наука, 1987. – 320 с.
- 3 Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем / А. Г. Ивахненко. – К.: Наукова думка, 1981. – 286 с.
- 4 Ивахненко А. Г. Справочник по типовым программам моделирования / А. Г. Ивахненко, Ю. В. Коппа, В. С. Степашко и др. – К.: Техніка, 1980. – 180 с.
- 5 Ивахненко А. Г. Помехоустойчивость моделирования: монография / А. Г. Ивахненко, В. С. Степанко. – Киев: Наук. думка, 1985. – 216 с.
- 6 Гершилов А. А. Математические методы построения прогнозов / А. А. Гершилов, В. А. Стакун, А. А. Стакун. – М.: Радио и связь, 1997. – 112 с.
- 7 Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс; пер. с англ. А. Л. Левшина под ред. В. Ф. Писаренко. – М.: Мир, 1974. – 402 с.
- 8 Hyndman R. J. Forecasting: principles and practice / R. J. Hyndman, G. Athanasopoulos. – OTexts, 2013. – Режим доступу: <https://www.otexts.org/book/fpp>.
- 9 Вержбицкий В. М. Основы численных методов: учебник для вузов / В. М. Вержбицкий. – М.: Высшая школа, 2002. – 840 с.
- 10 Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польск. И. Д. Рудинского. – М.: Горячая линия-Телеком, 2004. – 452 с.
- 11 Горбійчук М. І. Метод побудови математичних моделей складних процесів на засадах генетичних алгоритмів / М. І. Горбійчук, М. А. Шуфранович // Штучний інтелект. – 2010. – № 4. – С. 50 – 57.
- 12 Горбійчук М. І. Індуктивний метод побудови математичних моделей газоперекачувальних агрегатів природного газу / М. І. Горбійчук, М. І. Когутяк, Я. І. Заячук // Нафтова і газова промисловість. – 2008. – № 5. – С. 32 – 35.
- 13 Горбійчук М. І. Метод синтезу емпіричних моделей на засадах генетичних алгоритмів / М. І. Горбійчук, М. І. Когутяк, О. Б. Василенко, І. В. Щупак // Розвідка та розробка нафтових і газових родовищ. – 2009. – № 4(33). – С. 72-79.
- 14 Ивахненко А. Г. Самоорганизация прогнозирующих моделей / А. Г. Ивахненко, Й. А. Мюллер. – К.: Техніка, 1984. – 223 с.
- 15 Химмельблау Д. Анализ процессов статистическими методами / Д. Химмельблау; пер. с англ. В. Д. Скаржинского; под ред. В. Г. Горского. – М.: Мир, 1975. – 957 с.
- 16 Горбійчук М. І. Метод синтезу математичних моделей рівня води у р. Дністер залежно від погодних умов / М. І. Горбійчук, О. В. Пендерецький // Науковий вісник Івано-Франківського національного технічного університету нафти і газу. – 2010. – № 1 (23). – С. 105 – 112.

Стаття надійшла до редакційної колегії

11.09.14

Рекомендована до друку

*професором **Заміховським Л.М.***

(ІФНТУНГ, м. Івано-Франківськ)

*професором **Мещеряковим Л.І.***

(Національний гірничий університет України,

м. Дніпропетровськ)