УДК 004.415

**С. Д. Штовба, д. т. н., проф.; А. А. Яковенко**

# ПРОГНОЗУВАННЯ ТРУДОМІСТКОСТІ РОЗРОБКИ ПРОГРАМНИХ СИСТЕМ ЗА ДОПОМОГОЮ НЕЧІТКОЇ ГІБРИДНОЇ МОДЕЛІ

*Для прогнозування трудомісткості розробки програмних систем запропоновано використання нечіткої гібридної моделі, у якій антецеденти правил задаються нечіткими термами, а консеквенти – лінійними залежностями "входи – вихід" з нечіткими коефіцієнтами. За експериментальними даними проведено ідентифікацію залежності трудомісткості розробки програмних модулів від стажу програміста, новизни та складності завдання. Установлено, що запропонована нечітка гібридна модель забезпечує вищу точність порівняно з іншими п'ятьма конкурентними моделями.*

***Ключові слова**: нечітка гібридна модель, нечітке виведення, нечітка регресія, нечітка ідентифікація, програмна система, прогнозування трудомісткості.*

## Вступ

Під час створення програмних систем постає проблема оцінювання трудомісткості їхньої розробки. Як правило, увесь процес розробки розбивають на етапи, кожен із яких складається з конкретних завдань. Трудомісткість виконання кожного завдання оцінює лідер команди розробників. Під час оцінювання трудомісткості досить важко адекватно врахувати всі чинники впливу, що призводить до істотної похибки прогнозування та погіршує якість управління проектом.

Існує багато моделей прогнозування трудомісткості розробки програмних систем [1, 2], серед них однією з найпопулярніших є 22-факторна модель COCOMO II – Constructive Cost Model [3]. Результати практичного застосування цих моделей вказують, що вони не повною мірою враховують усі особливості процесу розробки програмних систем. Труднощі оцінювання зумовлені невизначеністю початкових даних і параметрів моделей прогнозування трудомісткості, що пов'язано із суттєвим упливом людського чинника, тому виникає зацікавленість у застосуванні нових методів прогнозування, які добре пристосовані для врахування такої невизначеності, наприклад, технологій нечіткої ідентифікації.

**Мета статті** полягає в перевірці можливості прогнозування трудомісткості розробки програмних систем за допомогою нової нечіткої гібридної моделі [4]. Ця модель складається з продукційних правил, антецеденти яких задано нечіткими множинами, а консеквенти – нечіткими регресійними рівняннями. Такий гібридний формат дозволяє описати складну залежність лише кількома правилами, які враховують невизначеність зон дії правил за допомогою нечітких антецедентів. Водночас невизначеність сили впливу факторів ураховують за допомогою нечітких коефіцієнтів у регресійних залежностях, які складають консеквенти правил. Невизначеність початкових даних ураховують шляхом представлення їх у формі нечітких множин із подальшим логічним виведенням за нечітких значень факторів упливу.

## 1. База нечітких гібридних правил

Базу нечітких гібридних правил запишемо так [4]:

$$\text{If } \left(x_1 = \widetilde{a}_{j1} \text{ and } x_2 = \widetilde{a}_{j2} \text{ and ... and } x_n = \widetilde{a}_{jn}\right), \text{ then } y = \widetilde{d}_j,\ j = \overline{1, m}, \qquad (1)$$

де $\widetilde{a}_{ji}$ – нечіткий терм, яким оцінено вхідну змінну $x_i$ у $j$-му правилі $i = \overline{1, n}$, $j = \overline{1, m}$;

$m$ – кількість правил;

$\widetilde{d}_j = \widetilde{k}_{j0} + \widetilde{k}_{j1}x_1 + \widetilde{k}_{j2}x_2 + ... + \widetilde{k}_{jn}x_n$ – консеквент *j*-го правила, який представлено лінійною функцією з нечіткими коефіцієнтами $\widetilde{k}_{j0},\ \widetilde{k}_{j1}, ..., \widetilde{k}_{jn}$.

На відміну від бази знань Сугено [5], у (1) коефіцієнти в консеквентах правил задано нечіткими числами. Тому експерт може описати ці нечіткі коефіцієнти лінгвістичними оцінками "мало впливає", "помірно впливає", "сильно впливає" тощо, які відображають його знання про ступінь впливу відповідної вхідної змінної на вихідну. За такими лінгвістичними оцінками можна визначати ядро нечіткого коефіцієнта в (1). Розмитість нечіткого коефіцієнта залежить від впевненості експерта в достовірності своїх знань, яку можна виразити термами "абсолютно достовірно", "майже достовірно", "більш-менш достовірно" тощо. Чим достовірніші знання, тим концентрованіша функція належності нечіткого коефіцієнта.

## 2. Логічне виведення за базою нечітких гібридних правил

Логічне виведення за базою правил (1) здійснимо так. Спочатку для вектора $X^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$ поточних значень вхідних змінних за правилами нечіткої арифметики розрахуємо нечіткі значення консеквентів:

$$\widetilde{d}_j = \widetilde{k}_{j0} + \widetilde{k}_{j1}x_1^* + \widetilde{k}_{j2}x_2^* + ... + \widetilde{k}_{jn}x_n^*,\ j = \overline{1, m}\,. \tag{2}$$

Це перетворить (1) у базу правил Мамдані, тому подальші кроки здійснимо за алгоритмом Мамдані [6]. Зауважимо, що для кожного вхідного вектора $X^*$ створюють базу правил Мамдані з унікальним набором нечітких консеквентів.

За алгоритмом Мамдані ступінь виконання антецедента *j*-го правила для вхідного вектора $X^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$ розрахуємо так:

$$\mu_j(X^*) = \mu_{j1}(x_1^*)\ \wedge\ \mu_{j2}(x_2^*)\ \wedge ... \wedge\ \mu_{jn}(x_n^*)\ ,\ j = \overline{1, m}\,, \tag{3}$$

де $\mu_{ji}(x_i^*)$ – ступінь належності вхідного значення $x_i^*$ нечіткому терму $\widetilde{a}_{ij}$, $i = \overline{1, n}$;

$\wedge$ – t-норма, яку в алгоритмі Мамдані зазвичай реалізують операцією мінімума або добутком.

У результаті виведення за *j*-им правилом бази знань отримаємо таку нечітку множину:

$$\widetilde{d}_j^* = imp\left(\widetilde{d}_j,\ \mu_j(X^*)\right),\quad j = \overline{1, m}\,, \tag{4}$$

де *imp* позначає імплікацію, яку реалізують операцією мінімума.

Геометричною інтерпретацією імплікації є "зрізання" графіка функції належності $\mu_{d_j}(y)$ нечіткого консеквента (2) по рівню $\mu_j(X^*)$:

$$\widetilde{d}_j^* = \int_{y \in [\underline{y},\ \overline{y}]} \min\left(\mu_j(X^*),\ \mu_{d_j}(y)\right) / y\,,$$

де $\left[\underline{y},\ \overline{y}\right]$ – діапазон зміни вихідної змінної $y$.

Результат виведення за всіма правилами знаходимо агрегуванням нечітких множин (4):

$$\widetilde{y}^* = agg\left(\widetilde{d}_1^*,\ \widetilde{d}_2^*, ..., \widetilde{d}_m^*\right), \tag{5}$$

де *agg* – агрегування, яке реалізують операцією максимума над функціями належності.

Чітке значення $y^*$ визначаємо через дефазифікацію нечіткої множини $\widetilde{y}^*$ за методом центра тяжіння.

Якщо початкові дані задано нечіткими значеннями $\widetilde{X}^* = \left(\widetilde{x}_1^*, \widetilde{x}_2^*, \ldots, \widetilde{x}_n^*\right)$, то за правилами

нечіткої арифметики розрахуємо нечіткі значення консеквентів $\widetilde{d}_j = \widetilde{k}_{j0} + \widetilde{k}_{j1}\widetilde{x}_1^* + \widetilde{k}_{j2}\widetilde{x}_2^* + ... + \widetilde{k}_{jn}\widetilde{x}_n^*$, $j = \overline{1, m}$. Подальше виведення здійснимо за формулами (3) – (4), тільки замість $\mu_{ji}(x_i^*)$ використовуватимемо $\mu_{ji}(\widetilde{x}_i^*)$ – ступінь належності нечіткого вхідного значення $\widetilde{x}_i^*$ нечіткому терму $\widetilde{a}_{ij}$, $j = \overline{1, m}$, $i = \overline{1, n}$. Ці значення розрахуємо так [6]:

$$\mu_j(\widetilde{x}_i^*) = height\left(\widetilde{x}_i^* \cap \widetilde{a}_{ij}\right), \tag{5}$$

де $height$ – висота нечіткої множини.

## 3. Параметрична ідентифікація залежностей за допомогою бази нечітких гібридних правил

Навчальну вибірку з $M$ пар "входи – вихід" запишемо так:

$$\left(X_r, y_r\right), \qquad r = \overline{1, M},$$

де $X_r$ – вхідний вектор у *r*-ому рядку вибірки та $y_r$ – відповідний вихід.

Позначимо через $y = F(P, X)$ модель на основі бази нечітких гібридних правил (1) з параметрами $P$. Параметрична ідентифікація полягає у знаходженні вектора $P$, що забезпечує:

$$RMSE = \sqrt{\frac{1}{M}\sum_{r=\overline{1, M}}\left(y_r - F\left(P, X_r\right)\right)^2} \to \min. \tag{6}$$

У (7) керовані змінні $P$ відповідають параметрам функцій належності нечітких множин з антецедентів та консеквентів правил (1). Для збереження інтерпретабельності моделі на параметри функцій нечітких множин $\widetilde{a}_{ij}$, $i = \overline{1, n}$, $j = \overline{1, m}$ накладемо обмеження згідно з [7].

## 4. Експериментальні дані для ідентифікації моделі прогнозування трудомісткості

Для синтезу моделі прогнозування трудомісткості розробки програмних систем скористаємося даними, наданими фірмою "Орнеон". Вони відповідають проекту зі створення гри типу "Quest". Вихідною змінною є $y$ – трудомісткість виконання завдання, яку вимірюють у людино-днях. Вхідними змінними є такі фактори:

$x_1$ – стаж програміста в місяцях;

$x_2$ – новизна завдання, яку оцінюють у балах від 1 до 10;

$x_3$ – складність завдання, яку оцінюють у балах від 1 до 10.

Навчальну вибірку сформовано за завданнями, які виконували протягом перших трьох місяців проекту, а тестову – за більш пізніми завданнями. У навчальну вибірку ввійшло 107 випадків (рис. 1), а в тестову – 106 (рис. 2).

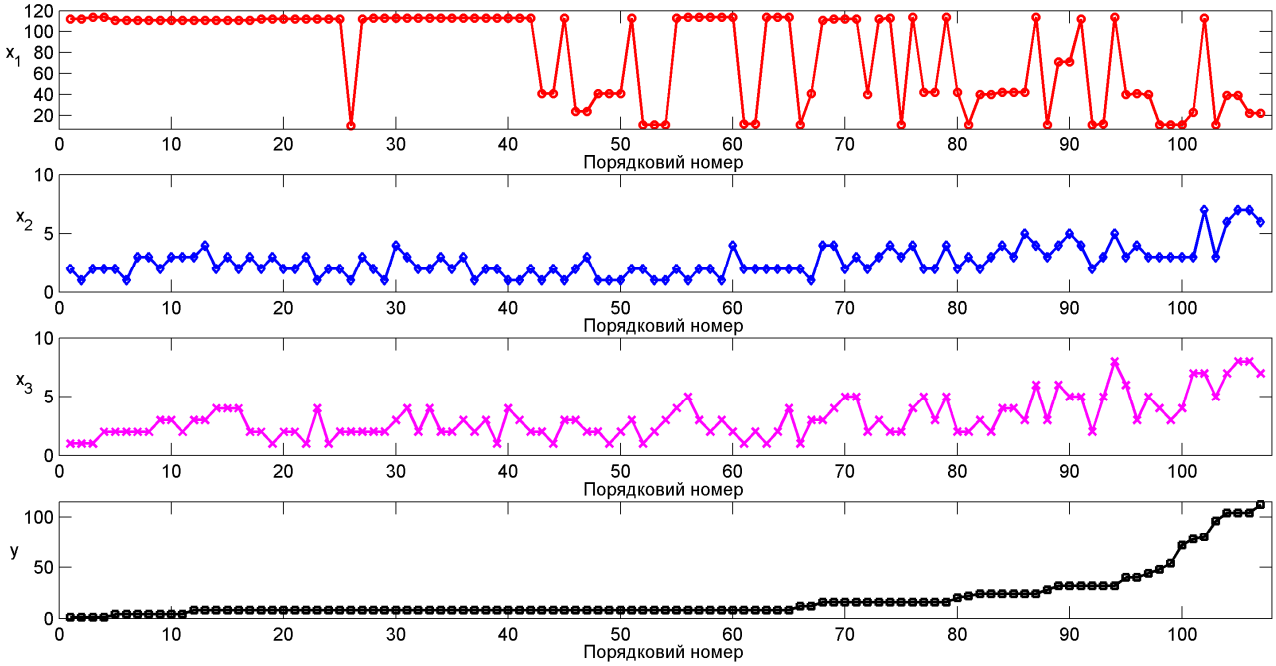


Рис. 1. Навчальна вибірка

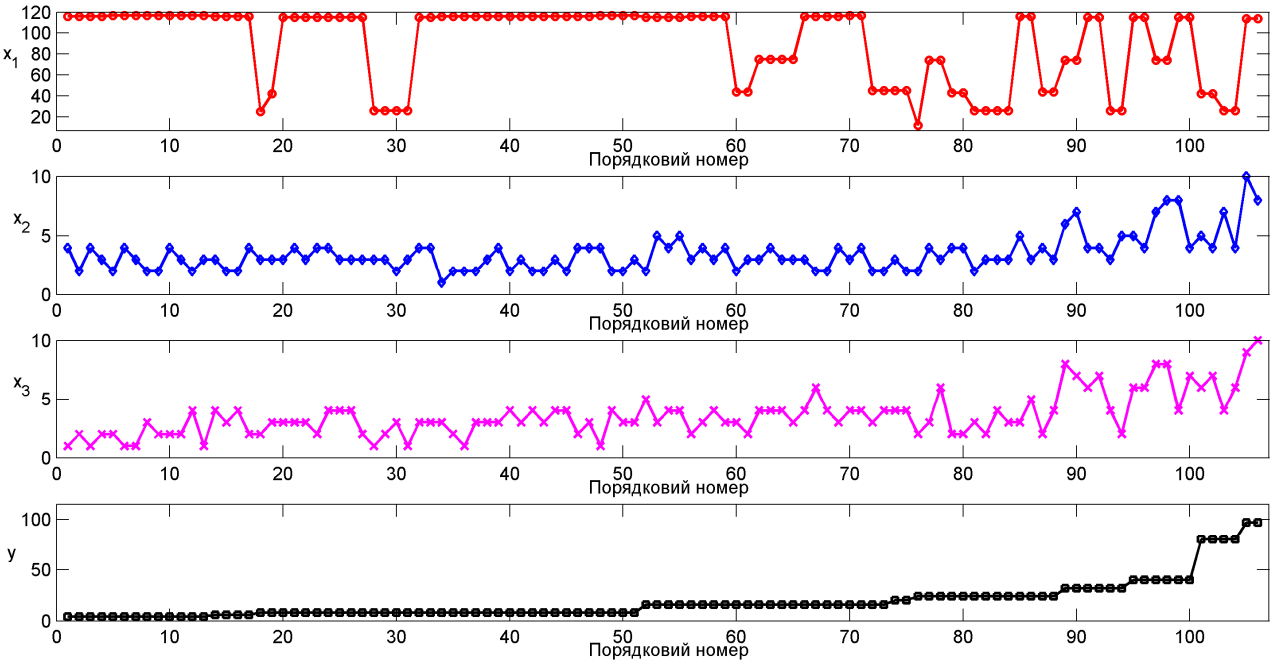


Рис. 2. Тестова вибірка

## 5. Моделі прогнозування трудомісткості розробки програмних систем

Моделювання залежності $y = f(x_1, x_2, x_3)$ здійснимо за допомогою таких трьох нечітких гібридних правил:

If $x_1 = Low$, then $y = \tilde{k}_{10} + \tilde{k}_{11}x_1 + \tilde{k}_{12}x_2 + \tilde{k}_{12}x_3$;

If $x_1 = Average$, then $y = \tilde{k}_{20} + \tilde{k}_{21}x_1 + \tilde{k}_{22}x_2 + \tilde{k}_{22}x_3$;

If $x_1 = High$, then $y = \tilde{k}_{30} + \tilde{k}_{31}x_1 + \tilde{k}_{32}x_2 + \tilde{k}_{32}x_3$.

У цих правилах терми *Low* і *Average* та нечіткі коефіцієнти консеквентів задано трикутними функціями належності. Трикутна функція належності має три параметри

$(a, b, c)$, які задають носій $(a, c)$ та ядро $(b)$ нечіткої множини. Терм *High* описано трапецієподібною функцією належності з чотирма параметрами $(a, b, c, d)$, які задають носій $(a, d)$ та ядро $(b, c)$ нечіткої множини. Параметри функцій належності термів і нечітких коефіцієнтів після навчання зведено в табл. 1.

Таблиця 1

**Параметри функцій належності термів і нечітких коефіцієнтів нечіткої гібридної моделі**

| Нечітка множина | Параметри функції належності |
|---|---|
| $\tilde{k}_{10}$ | (-3.13, 0.87, 4.97) |
| $\tilde{k}_{11}$ | (-4.84, -3.26, -2.86) |
| $\tilde{k}_{12}$ | (7.29, 9.22, 10.71) |
| $\tilde{k}_{13}$ | (11.54, 12.74, 14.04) |
| $\tilde{k}_{20}$ | (-5.95, -4.91, -3.97) |
| $\tilde{k}_{21}$ | (-1.41, -1.37, -1.37) |
| $\tilde{k}_{22}$ | (3.97, 8.32, 10.12) |
| $\tilde{k}_{23}$ | (7.99, 14.97, 16.9) |
| $\tilde{k}_{30}$ | (-2.14, 1.88, 8.09) |
| $\tilde{k}_{31}$ | (-0.46, -0.46, -0.23) |
| $\tilde{k}_{32}$ | (9.78, 10.7, 11.55) |
| $\tilde{k}_{33}$ | (2.11, 2.9, 5.29) |
| *Low* | (0, 0, 18.59) |
| *Average* | (-10.91, 43.42, 78.17) |
| *High* | (22.6, 54.64, 120, 138) |

Для порівняння якості ідентифікації ми синтезували 5 конкурентних моделей:

- лінійну

$$y = -1 - 0.22x_1 + 6.84x_2 + 6.15x_3;$$

- квадратичну

$$y = 24.51 - 0.51x_1 - 1.38x_2 + 0.46x_3 + 0.002x_1^2 + 1.165x_2^2 + 0.689x_3^2;$$

- поліном степеня 1/2

$$y = 84.1 + 0.12x_1 + 18.56x_2 + 17.4x_3 - 4.82\sqrt{x_1} - 41.39\sqrt{x_2} - 42.1\sqrt{x_3};$$

- ряд Вінера

$$\begin{aligned} y \;=\; & 0.99 + 0.87x_1 + 0.95x_2 + 0.96x_3 - 0.026x_1^2 - 0.15x_1x_2 + 0.052x_1x_3 + 0.763x_2^2 - \\ - \; & 0.792x_2x_3 + 0.823x_3^2 - 0.0002x_1^3 - 0.0001x_1^2x_2 - 0.0004x_1^2x_3 + 0.0217x_1x_2^2 + \\ + \; & 0.0112x_1x_2x_3 - 0.018x_1x_3^2 - 0.0929x_2^3 - 0.0527x_2^2x_3 - 0.0527x_2x_3^2 + 0.0674x_3^3; \end{aligned}$$

- базу нечітких правил Сугено

$$\text{If } x_1 = \textit{Low}, \text{ then } y = 22.56 - 6.92x_1 + 10.33x_2 + 12.8x_3;$$

$$\text{If } x_1 = Average, \text{ then } y = 2.05 - 0.79x_1 + 9.33x_2 + 12.11x_3;$$

$$\text{If } x_1 = High, \text{ then } y = 0.97 - 0.1x_1 + 4.81x_2 + 3.03x_3.$$

Параметри функцій належності термів з антецедентів правил Сугено зведено в табл. 2.

Таблиця 2

**Параметри функцій належності термів і нечітких коефіцієнтів нечіткої гібридної моделі**

| Нечітка множина | Параметри функції належності |
|---|---|
| *Low* | (0, 0, 29.13) |
| *Average* | (-11.49, 28.07, 67.63) |
| *High* | (21.3, 54.8, 129.68, 138) |

Порівняння результатів моделювання з експериментальними даними наведено на рис. 3 та 4. Для деяких випадків за конкурентними моделями спрогнозована трудомісткість була меншою за 1. Для цих випадків вихідне значення встановлено таким, що дорівнює 1. З рис. 3 та 4 видно, що як за середньою квадратичною нев'язкою ( *RMSE* ), так і за максимальною абсолютною нев'язкою (*MaxErr*) нечітка гібридна модель є найкращою. Модель на основі правил Сугено показала близькі результати, оскільки її можна вважати частковим випадком гібридної нечіткої моделі, у якій усі коефіцієнти в консеквентах є чіткими числами.

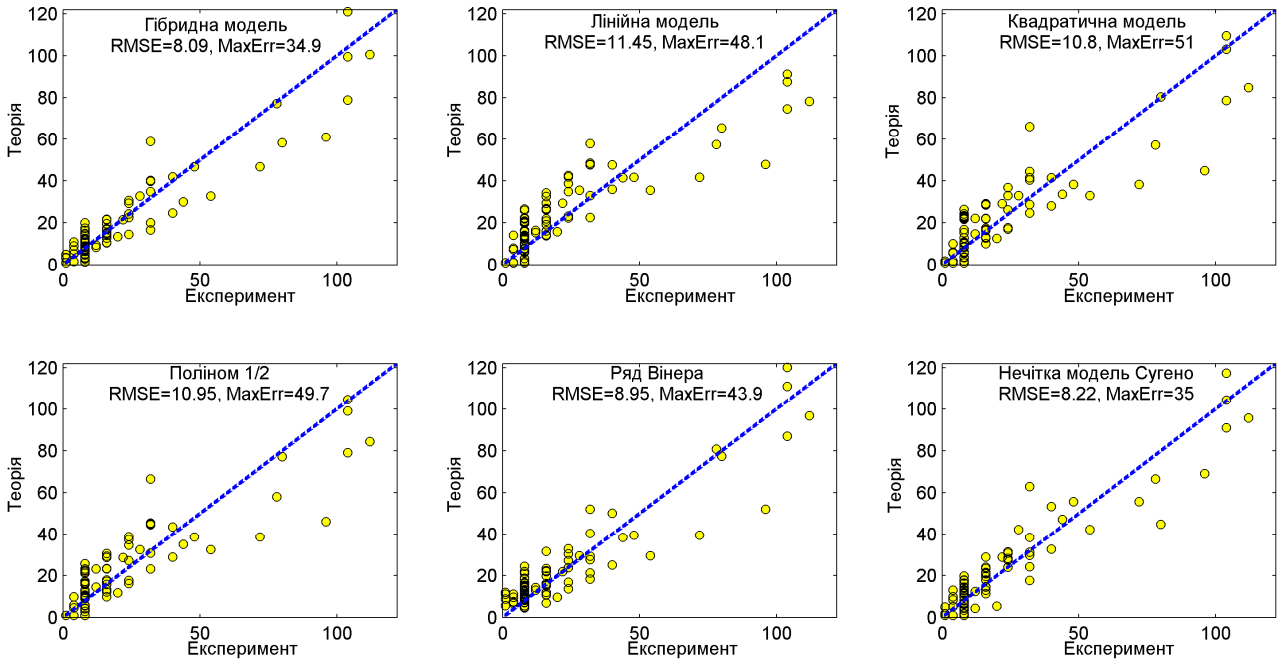


Рис. 3. Перевірка моделей на навчальній вибірці

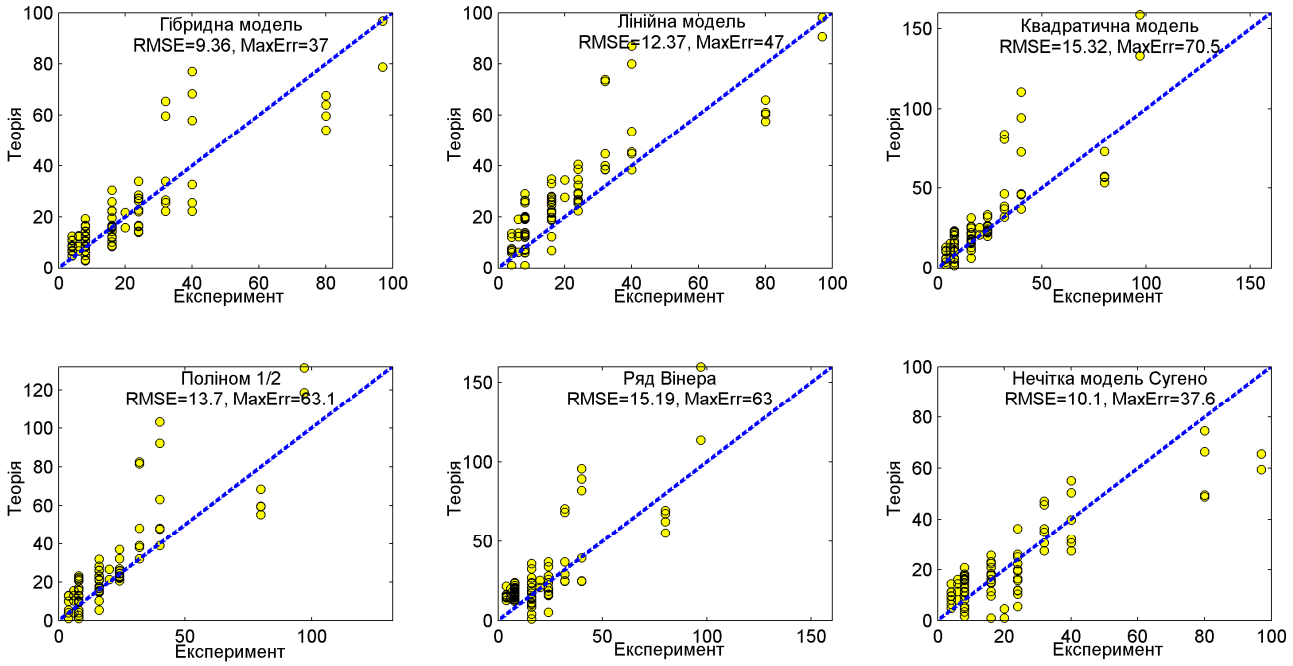


Рис. 4. Перевірка моделей на тестовій вибірці

## Висновки

Досліджено застосування нечіткої гібридної моделі для прогнозування трудомісткості розробки програмних систем. Проведено порівняння результатів з альтернативними моделями. Гібридна нечітка модель показала найкращі результати серед розглянутих моделей. Отже, її застосування для прогнозування часу трудомісткості розробки програмних систем є доцільним, оскільки це дозволить покращити проектне планування.

## СПИСОК ЛІТЕРАТУРИ

1. Липаев В. В. Программная инженерия. Методологические основы : Учеб. / Липаев В. В. – Гос. ун-т – Высшая школа экономики. – М.: ТЕИС, 2006. – 608 с.

2. Hernandez-Lopez A. Software engineering job productivity – a systematic review / A. Hernandez-Lopez , R. Colomo-Palacios, A. Garcia-Crespo // International Journal of Software Engineering and Knowledge Engineering. – 2013. – Vol. 23. – №3. – P. 387 – 406.

3. Boehm B. W. Software Cost Estimation with COCOMO II. / Boehm B. W. et al. – New Jersey: Prentice Hall, 2000. – 502 p.

4. Штовба С. Д. Моделювання залежностей за допомогою нечіткої бази знань з нечіткими регресійними рівняннями / С. Д. Штовба / Вісник Вінницького політехнічного інституту. – 2011. – №3. – С. 195 – 199.

5. Takagi T. Fuzzy Identification of Systems and Its Applications to Modeling and Control / T. Takagi, M. Sugeno // IEEE Trans. on Systems, Man, and Cybernetics. – 1985. – Vol. 15. - №1. – P. 116 - 132.

6. Штовба С. Д. Проектирование нечетких систем средствами MATLAB / Штовба С. Д. – М.: Горячая линия – Телеком, 2007. – 288 с.

7. Штовба С. Д. Обеспечение точности и прозрачности нечеткой модели Мамдани при обучении по экспериментальным данным / С. Д. Штовба // Проблемы управления и информатики. – 2007. – №4. – С. 102 – 114.

***Штовба Сергій Дмитрович*** – професор, д. т. н., професор кафедри комп'ютерних систем управління.

***Яковенко Антон Андрійович*** – студент ІнАЕКСУ.
Вінницький національний технічний університет.