

Гнатюк І.А.

Керівник: д.т.н., професор Волошко А.В.

ІНФОРМАЦІЙНА МОДЕЛЬ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ЗА ДОПОМОГОЮ ПОПЕРЕДНЬОЇ ЇХ СЕГМЕНТАЦІЇ НА ОСНОВІ АНАЛІЗУ СКАЛОГРАМИ

Анотація. В статті досліджується можливість застосування властивостей часових рядів для підвищення ефективності прогнозування інформаційних сигналів за наявності у них трендів, сезонних коливань та шумових компонент.

Ключові слова. Інформаційна модель, вейвлет-декомпозиція, скалограма.

Abstract. The article investigates the possibility of using the properties of time series to increase the efficiency of forecasting information signals in the presence of trends, seasonal fluctuations and noise components.

Keywords. Information model, wavelet decomposition, scalogram.

Вступ

Методологічний підхід при розробці такої моделі прогнозування заключається у застосуванні властивостей частотно-впорядкованих енергій вейвлет-коефіцієнтів за рівнями вейвлет-декомпозиції для виявлення довготривалих трендів, сезонних та шумових компонент в інформаційних сигналах, сегментуванні результатів його вейвлет-декомпозиції і подальшого прогнозування кожного сегменту адаптованим до нього методом прогнозування окремо. Послідуочий синтез результатів прогнозу (зворотнє вейвлет-перетворення) і є зпрогнозованим значенням інформаційного сигналу.

Як відомо, скалограма у дискретному вейвлет-перетворенні аналогічна періодограмі із спектрального аналізу часових рядів. При цьому, періодограма проводить декомпозицію енергії сигналу відповідно до різноманітних частот Фур'є, а скалограма декомпозує енергію сигналу у відповідності до рівнів вейвлет-перетворення. Проаналізуємо особливості застосування скалограми для виявлення періодичних компонент у часовому ряді і розробимо метод попередньої обробки інформаційного сигналу для підвищення точності його прогнозування.

Дискретне вейвлет-перетворення представляється ортогональною матрицею W розмірністю $N \times N$. На практиці для цього застосовуються квадратурні дзеркальні фільтри, які характеризуються відповідними вейвлет-базисами. Тобто, сигнал $y(t)$ представляється вектором відповідного розміру

$$v = (H^n y, GH^{n-1} y, \dots, GH^2 y, GH y, G y), \quad (1)$$

де оператори G та H визначаються як $(Ha)_k = \sum_n h_{n-2k} a_n$, $(Ga)_k = \sum_n g_{n-2k} a_n$; g та h – високо- та низькочастотні фільтри відповідних вейвлет-базисів.

У (1) вектор $H^n y = \{cA_j(k)\}$ і вектор $GH^{n-1} y = \{cD_j(k)\}$ представляють собою вейвлет-коефіцієнти.

Подібно до залежності обчислимо енергію вейвлет-коефіцієнтів $E(j) = \sum_{k=0}^{2^j-1} cD_j^2(k)$ за рівнями вейвлет-декомпозиції $j = 0, \dots, J$. Скалограма буде представляти собою вектор із значень енергій на кожному рівні вейвлет-декомпозиції

$$\left[\left\{ cA_j^2(k) \right\}_{1 \leq k \leq 2^j-1}^2, E(1), E(2), \dots, E(J) \right], \quad (2)$$

де $\left\{ cA_j^2(k) \right\}_{1 \leq k \leq 2^j-1}^2$ – енергія апроксимуючих вейвлет-коефіцієнтів останнього

рівня (J) вейвлет-декомпозиції, $E(j) = \sum_{k=0}^{2^j-1} cD_j^2(k)$ – енергія деталізуючих вейвлет-коефіцієнтів рівнів вейвлет-декомпозиції від $j = 1$ до $j = J$.

З урахуванням (1) перепишемо вираз для скалограми у наступному виді

$$\left[\|H^n y\|^2, \|GH^{n-1} y\|^2, \dots, \|GH y\|^2, \|Gy\|^2 \right]. \quad (3)$$

На рис. 1 представлена формальна інформаційна модель по сегментного прогнозування часового ряду. На основі аналізу скало грами ортогонального розкладання часового ряду (ГЕН) проводиться сегментація рівнів ортогонального розкладання і до кожного сегменту застосовується свій адаптивний до його структури метод прогнозування.

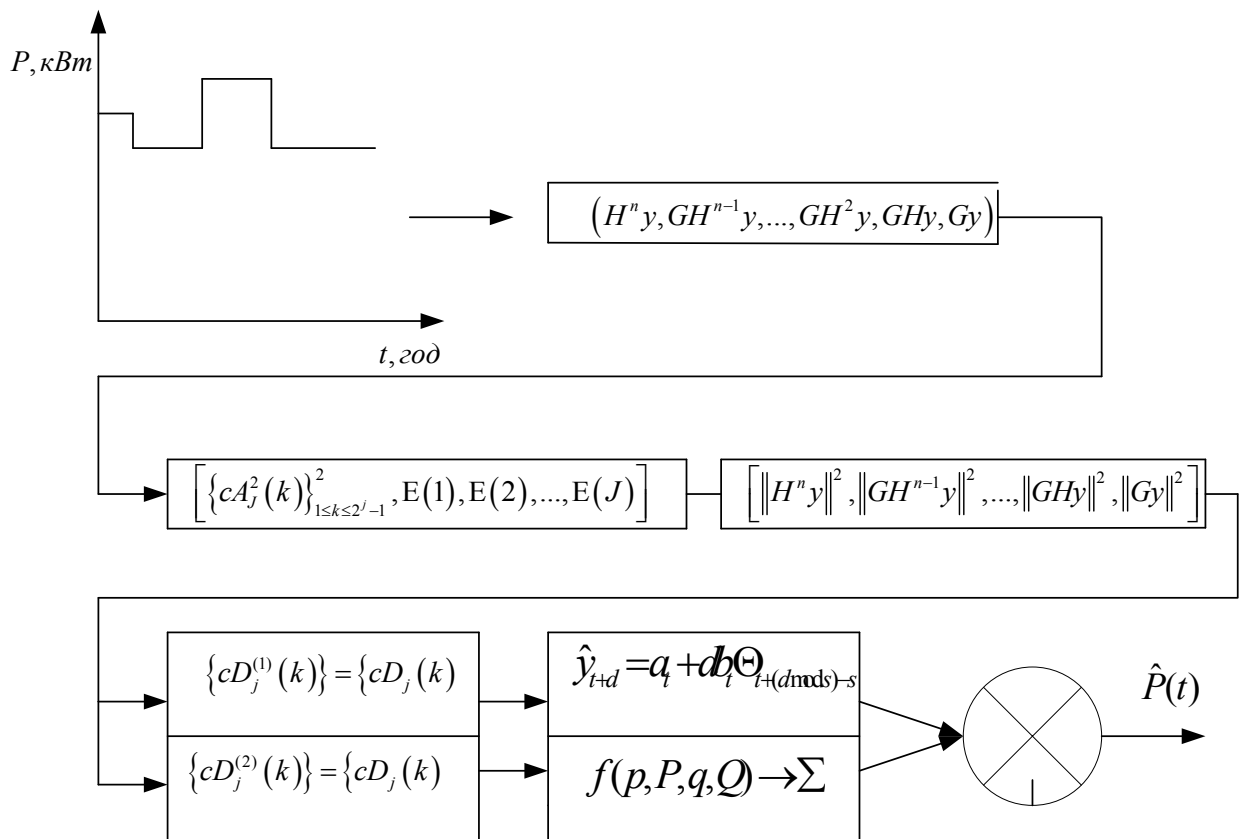


Рисунок 1 – Формальна інформаційна модель прогнозування часових рядів із попереднім їх сегментуванням

Для проведення дослідження змодельємо інформаційний сигнал (рис. 2) із наявністю тренду, циклічної компоненти та шуму.

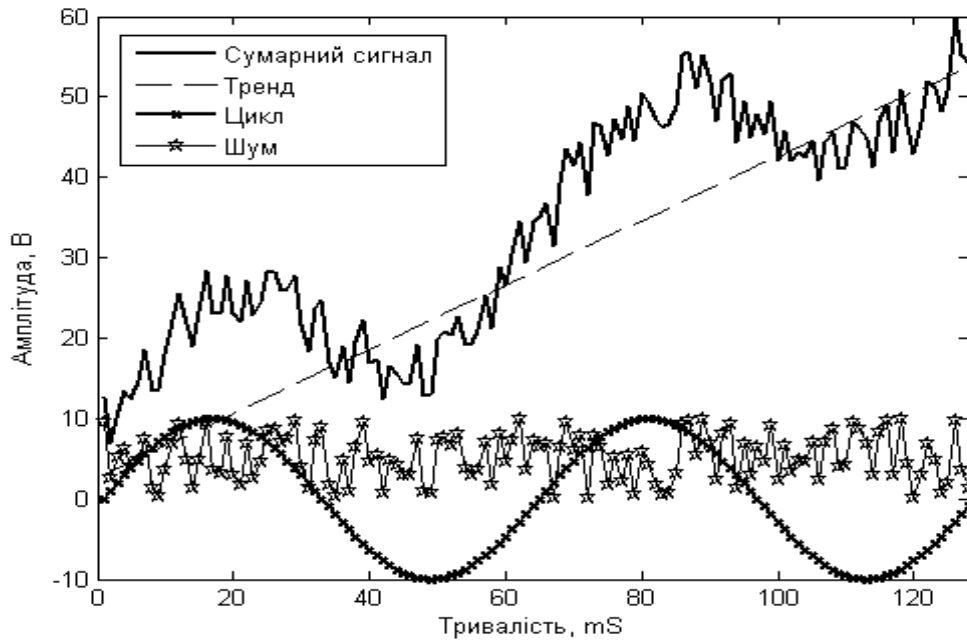


Рисунок 2 – Змодельований інформаційний сигнал

Відповідно до (1) – (3) розрахуємо та побудуємо скалограму (рис 3, а). Для видалення шумової компоненти застосуємо розроблений метод виділення шумової компоненти. На рис. 3, б представлено скалограму змодельованого інформаційного сигналу після видалення шумової компоненти.

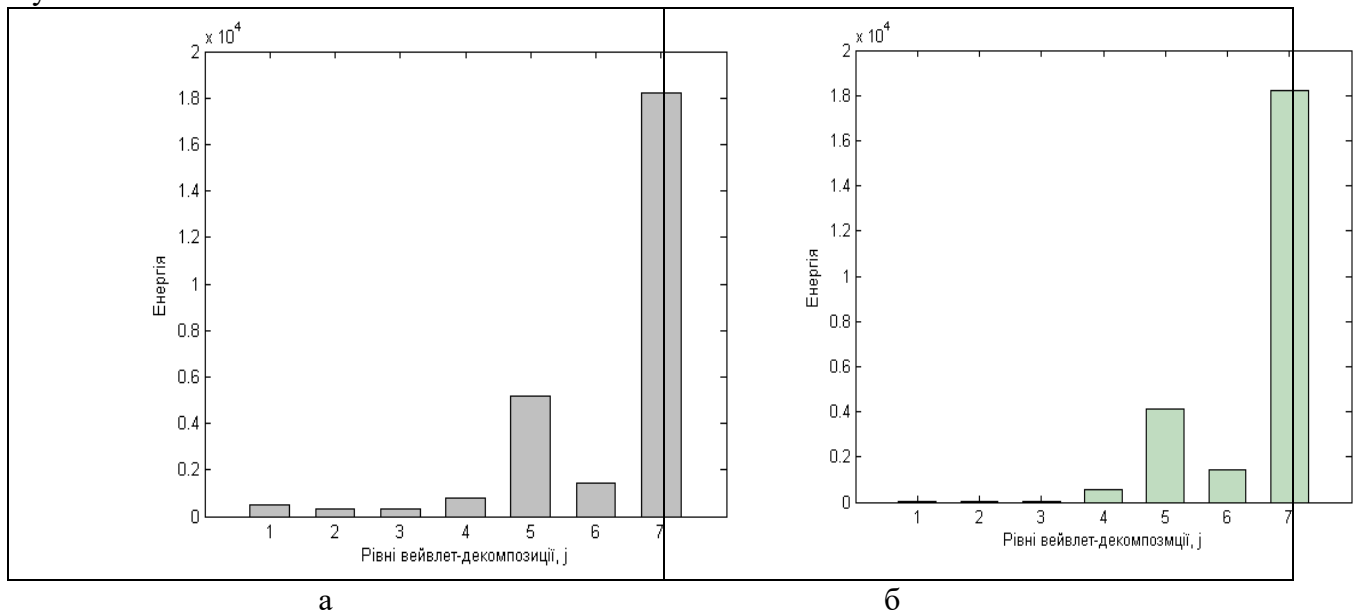


Рисунок 3 – Скалограма досліджуваного (а) та із видаленим шумом (б) інформаційного сигналу

Як слідує із скалограми рис. 3, б, видалення шумової компоненти призводить до зникнення високочастотних значень енергії вейвлет-коефіцієнтів на верхніх рівнях вейвлет-декомпозиції (рівні 1 – 3).

Проаналізуємо застосування методу прогнозу із рухливим середнім (ARMA) для прогнозування змодельованого на рис. 2 сигналу. Даний метод обраний для аналізу виходячи із результатів його застосування для ГЕН за серпень місяць (рис. 4). В якості вектору

прогнозування вибираємо вейвлет-коефіцієнти всіх рівнів вейвлет-декомпозиції (апроксимуючі вейвлет-коефіцієнти останнього рівня вейвлет-декомпозиції (J) та деталізуючі вейвлет-коефіцієнти від рівня $j = 1$ до рівня J). На рис. 4 представлений результат такого прогнозування.

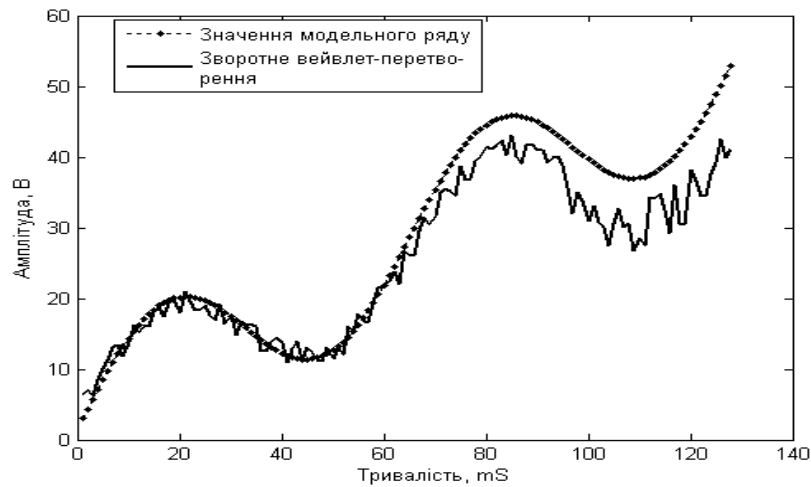


Рисунок 4 – Результат зворотного вейвлет-перетворення спрогнозованих вейвлет-коефіцієнтів

Із рис. 4 видно, що прогнозні значення цілком задовільно узгоджуються з модельним часовим рядом. Деяка розбіжність наприкінці прогнозу пояснюється відкиданням крупномасштабних вейвлет-коефіцієнтів завдяки заміні безперервного вейвлет-перетворення дискретним та недостатньою точністю прогнозування високочастотних компонент даним методом. На рис. 5 представлені значення модуля абсолютної похибки прогнозування. Відносна похибка прогнозування становить 10.46 %.

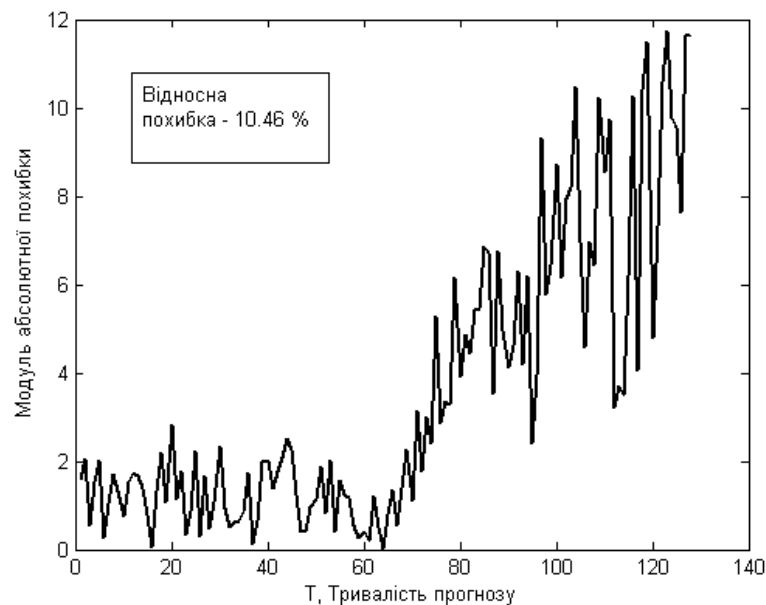


Рисунок 5 – Модуль абсолютної похибки прогнозування із застосуванням всіх вейвлет-коефіцієнтів і одного методу прогнозу

Дана похибка прогнозування є неприйнятною для практичного застосування і виникає необхідність розробки та/або вдосконалення методів підвищення точності

прогнозування.

Проаналізуємо скалограму модельного часового ряду на рис. 3, б.

Аналіз даної скалограми показує наявність двох явно виражених складових (піків). Перша складова охоплює рівні вейвлет-декомпозиції від 3 до 6, а друга – рівень 7. У зв'язку з чим пропонується ідентифікувати дані періодичні компоненти сигналу шляхом їх розбиття на дві серії (сегменти) вейвлет-коефіцієнтів

$$\{cD_j^{(1)}(k)\} = \{cD_j(k) \text{ для } j = 3, \dots, 6 \cap 0 \text{ для } j = 1 - 2, 7, \quad (4)$$

$$\{cD_j^{(2)}(k)\} = \{cD_j(k) \text{ для } j = 7 \cap 0 \text{ для } j = 1, \dots, 6. \quad (5)$$

Рівняння (4) та (5) характеризують дві вейвлет-декомпозиції, які представляють інформаційний сигнал (рис. 2) двома векторами, і які після вейвлет-реконструкції будуть представлені у спрогнозованому сигналі.

Для прогнозування низькочастотного сегменту вейвлет-декомпозиції обираємо прогнозну модель Тейла-Вейджа (модель не накладає ніяких обмежень на стохастичну структуру часового ряду і оснований на експоненціальному згладжуванні). Для прогнозування середньочастотного сегменту застосуємо модель ARMA. На рис. 6 представлені результати даного прогнозування.

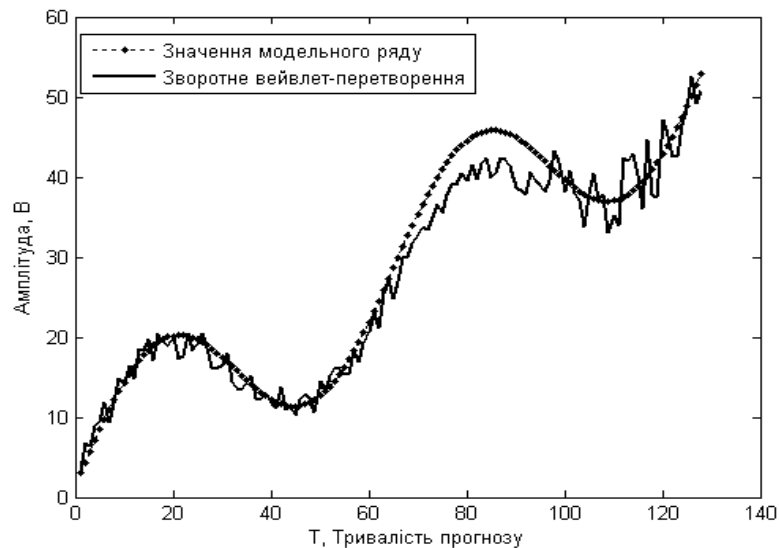


Рисунок 6 – Комбіноване прогнозування часового ряду (із сегментуванням)

Як слідує із рис. 6, сегментоване розділення вейвлет-декомпозиції і застосування до кожного сегменту «свого» методу прогнозування приводить до підвищення точності прогнозу. На рис. 7 представлені значення абсолютної похибки прогнозування. Відносна похибка прогнозування складає – 5.6 %.

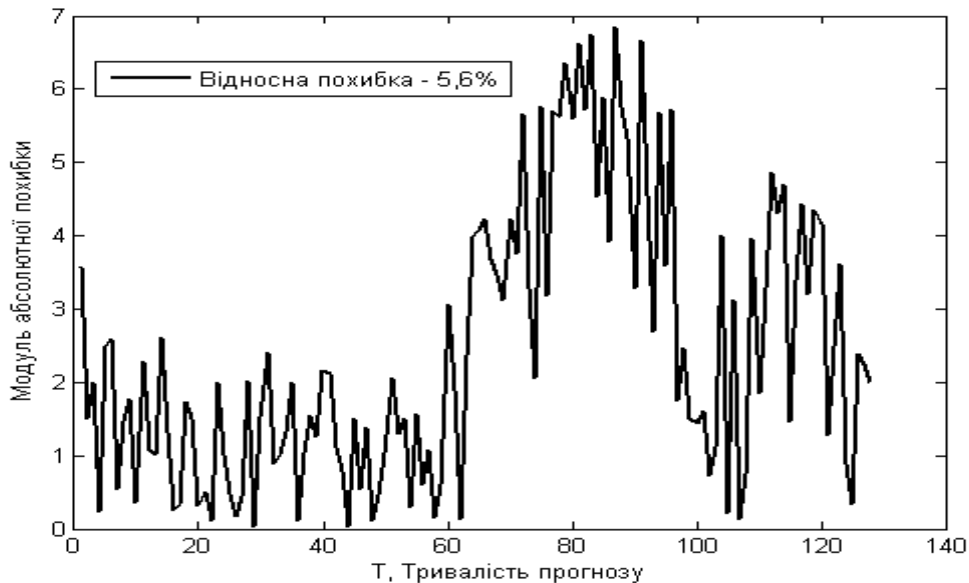


Рисунок 7 – Абсолютна похибка прогнозування (із сегментуванням)

Висновки

Преш ніж застосовувати ті чи інші методи прогнозу необхідно провести вейвлет-декомпозицію інформаційного сигналу, побудувати скалограму, і на основі результатів її аналізу сегментувати коефіцієнти вейвлет-декомпозиції за наступними правилами. У випадку наявності на скалограмі одного піку, проводиться сегментування вейвлет-коефіцієнтів на апроксимуючу та деталізуючу частину, кожна з яких підлягає окремому прогнозуванню з використанням відповідного методу. Якщо на скалограмі спостерігається декілька піків – проводиться відповідна ідентифікація періодичних компонент шляхом розробки вейвлет-коефіцієнтів на окремі секції, за кожною з яких проводиться окреме прогнозування.

Запропонована модель прогнозування доцільно застосовувати для середньострокового прогнозування інформаційних сигналів, оскільки похибка прогнозу при цьому знаходиться у межах 1%.

Література

1. Aizenberg I. Multilayer Neural Network with Multi-Valued Neurons in Time series Forecast- ing of Oil Production / I. Aizenberg, L.Sheremetov, L. Willa // Springer. — 2011. — 273 p.
2. Bennett G.J. Forecasting low voltage distribution network demand profiles using a pattern recognition based expert sysnem [Електронний ресурс] / G.J. Bennett.Режим доступу: <http://dx.doi.org/10.1016/j.energy.2014.01.032>.
3. Paulo M. Modeling and forecasting interval Time Series with Threshold models: An Applica- tion to S&P 500 Index Returns [Електронний ресурс] / M. Paulo, M. Rodrigues, N. Salish. — Режим доступу: <http://www.bportugal.pt/e>
4. Ghandehari M. A four dimensional continuous wavelet transform. Commutative and Non-commutative Harmonic Analysis and Applications, Contemporary Mathematics / M. Ghandehari, A. Syz- dykova, K. Taylor // Amer. Math. Soc., Providence. — 2013. — Vol. 603. — P. 123–136.
5. Arino M.A. Wavelet scalograms and their application in ecjnjmic time series / M.A. Arino, P.A. Morettin, B. Vidakovic // Brazilian Journal of Probability and Statistics. — 2004. — Vol. 18. — P. 37–51.
6. Santi E. Hybrid ARIMAX-NN model for forecasting Inflation / E. Santi, P.D. Dwi //

International Conference on Science, Technology and Humanity. — 2015. — P. 181–187.

7. Bennett G.J. Autoregressive with Exogenous Variables and Neural Network Short-Time Load Forecast Model for Residential Low Voltage Distribution Networks / C. Bennett, R.A.S Stewart, J. Lu // Energies. — 2014. — N 7. — P. 2938–2960.

8. Takagi H. Introduction to Fuzzy Systems, Neural Network and Genetic Algorithms [Електронний ресурс] / H. Takagi. — Режим доступу: www.teleametrika.net/reference/Software/infoToFuzzy-System