

# Інтелектуальний аналіз в задачах прогнозування тенденцій розвитку технологічного комплексу цукрового заводу

*В.Д. Кишенько, кандидат технічних наук, професор кафедри автоматизації та інтелектуальних систем керування, Національний університет харчових технологій*

*М.А. Сич, аспірант кафедри автоматизації та інтелектуальних систем керування, Національний університет харчових технологій*

*В статті розглянуті проблеми прогнозування поведінки технологічного комплексу цукрового заводу як складного об'єкта керування, що функціонує в умовах невизначеності. Запропонований метод, заснований на інтелектуальному аналізі часових рядів технологічних змінних, який дозволяє здійснювати оперативне прогнозування тенденцій розвитку складного об'єкта керування, забезпечуючи прийняття ефективних керуючих дій превентивного характеру.*

*Ключові слова: інтелектуальний аналіз, цукровий завод, часовий ряд, об'єкт керування.*

*В статье рассмотрены проблемы прогнозирования поведения комплекса сахарного завода как сложного объекта управления, функционирующего в условиях неопределенности. Предложенный метод, основанный на интеллектуальном анализе временных рядов технологических переменных, который позволяет осуществлять оперативное прогнозирование тенденций развития сложного объекта управления, обеспечивая принятие эффективных управляющих воздействий превентивного характера.*

*Ключевые слова: интеллектуальный анализ, сахарный завод, временной ряд, объект.*

*The article deals with the problem of predicting the behavior of complex sugar factory as a complex control facility operating in an uncertain environment. The proposed method is based on the intelligent analysis of time series of process variables, which allows for the operational forecasting of trends in the development of complex control object, ensuring that effective control preventive actions.*

*Keywords: intelligent analysis, sugar refinery, time series, object.*

Технологічний комплекс цукрового заводу є складною організаційно-технічною системою, характерними ознаками якого є наявність і тісна взаємодія типових технологічних процесів різної фізико-хімічної природи, багатофакторність, не лінійність, що викликає складну поведінку об'єкта з утворенням складних дисипативних просторово-часових структур, включаючи фрактальні структури хаотичного характеру, що ускладнює прийняття рішень по керуванню [1].

Суттєве значення при керуванні технологічними процесами цукрового заводу набуває врахування невизначеностей в оцінці технологічних параметрів, особливо показників якості буряку, дифузійного соку, сиропу, багато із яких оцінюється органолептичним шляхом або інструментальними методами, що характеризуються низькою точністю та оперативністю вимірювань [2].

При синтезі стратегій керування технологічними комплексами важливою проблемою є оцінка стану об'єкта та прогнозування його поведінки в умовах невизначеності [3]. Особливо ця проблема загострюється в задачах прогнозування пер-

спектив розвитку із значною глибиною прогнозу [4] на основі інтелектуальних інформаційних технологій.

Для інтелектуального аналізу розвитку складних організаційно-технічних систем, до яких належить технологічний комплекс цукрового заводу, необхідно в максимальній мірі використовувати знання про поведінку об'єкта у вигляді часових рядів технологічних змінних у, отриманих шляхом спостереження та вимірювань, а також неформальну інформацію, як правило у вербальній формі, про характерні регулярні особливості станів (патернів) на основі знань та досвіду експертів [5]. В такому випадку експертам необхідно оцінювати: часові області різної тривалості та наявності характерних проявів поведінки, ранжування рівнів параметрів, формування ознак патернів поведінки, визначення атрибутів, відношень і зв'язків в часових рядах, що забезпечить достатність знань для розв'язання задач прогнозування і прийняття рішень по керуванню [6]. Таку впорядковану в часі послідовність оцінок та характеристик називають нечітким часовим рядом [5]. При

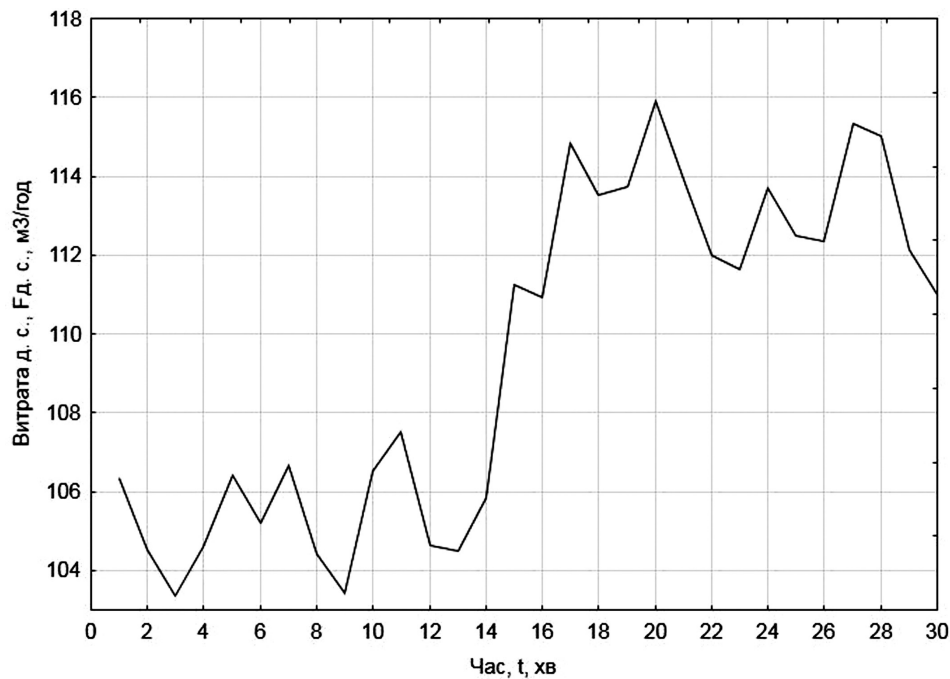


Рис. 1. Часовий ряд «Витрата дифузійного соку»

цьому основна увага зосереджується на виявленні небажаних та небезпечних тенденцій змінювання показників. Всі дані про системні поведінки об'єкта, в тому числі і про тривалості збереження тенденції виражаються в лінгвістичній формі, тобто на відміну від традиційного часового ряду значеннями нечіткого часового ряду є нечіткі множини, а не дійсні числа спостережень.

Сучасні інтелектуальні методи аналізу часових рядів спираються на визначення інформаційної гранули, що об'єднує тенденцію і її тривалість в єдине поняття [6]. В рамках технології DATA MINING розв'язуються задачі сегментації, класифікації, індексації, виявлення аномалій, частотно-часовий аналіз, включаючи прогнозування, формування асоціативних правил [6].

На основі нової інтелектуальної методології вирішуються традиційні завдання аналізу часових рядів:

сегментація - розбивка часових рядів на значущі сегменти;

кластеризація - пошук угруповань часових рядів або їх патернів;

класифікація - призначення часових рядів або їх патернів до одного із заздалегідь визначених класів;

прогнозування - прогноз чергового значення на базі історії часових рядів;

визначення асоціативних правил - пошук правил, що відносяться до патернів часових рядів.

Даний науковий напрям розвивається в працях таких вчених: Х.Танаки, К. Сонга [7], К. Хіרותи, Я. Капржика [8], В. Педріча [9], І. Батиршіна [10], школи Ярушкіної Н.Г. [4-6].

**Мета дослідження** полягає в розробці алгоритмів прогнозування тенденцій розвитку технологічного комплексу цукрового заводу на основі

інтелектуальних технологій.

Центральним поняттям у використанні теорії нечітких множин в інтелектуальному аналізі часових рядів є поняття нечіткого часового ряду. Нечітким часовим рядом називається впорядкована послідовність спостережень, якщо значення, які приймає деяка величина в момент часу, виражена за допомогою нечіткої мітки.

Для опису розвитку об'єкта керування в лінгвістичних термінах введено поняття часового ряду нечітких тенденцій [4].

Якщо  $\tilde{y}_\Delta = \{\tilde{y}_1, \dots, \tilde{y}_m\}$  – нечіткий часовий ряд лінгвістичної змінної  $(Y, T_y, U_y, G_y, M_y)$ , де  $\tilde{Y}_\Delta = \{\tilde{y}_\Delta\}$  – множина нечітких часових рядів однакової довжини;  $T_y$  - термножина нечітких значень у вербальній формі;  $U_y$  - універсальна множина, на якій визначається  $y$ ;  $G_y$  - процедура визначення вербальних значень нечітких змінних із  $T_y$ ;  $M_y$  - експертне оцінювання функцій належності окремих значень часового ряду  $y$ .

Якщо говорити про тенденцію як лінгвістичну змінну, термножиною якої є множина різних тенденцій, спостережуваних на часовому ряді, а універсумом – множина усіляких функцій, нечітка тенденція визначається як  $(\tau, \{\tau^i\}, \tilde{Y}_\Delta, G_\tau, M_\tau)$ ,  $i=1 \dots p$ , де  $p$  – кількість видів нечіткої тенденції, визначених на часовому ряді;  $\tilde{Y}_\Delta$  – множина нечітких часових рядів змінної довжини. Визначаючи нечітку тенденцію на всіх інтервалах  $[t-m+1, t]$  часових рядів і позиціонуючи початок або закінчення інтервалу до часової шкали, одержимо часовий ряд нечіткої тенденції.

Нехай  $\{\tilde{Y}_\Delta\}$  – множина нечітких часових рядів довжиною  $m$ , де  $\tilde{Y}_{\Delta t} = \{\tilde{y}_{t-m+1}, \dots, \tilde{y}_t\}$ ,  $\tilde{y}_t \in \tilde{Y}_t$ . Тоді тенденція  $\tau$ , визначена на  $\tilde{Y}_\Delta$ , є сукупністю упорядкованих пар  $\tau = \{\tilde{Y}_\Delta, \mu_\tau(\tilde{Y}_\Delta)\}$ , де  $\mu_\tau(\tilde{Y}_\Delta)$  являє собою

ступінь належності  $\tilde{y}_\Delta$  до нечітких тенденцій [4].

Для лінгвістичної змінної «тенденція» часовий ряд визначається сукупністю значень всіх видів нечітких тенденцій:  $\tau_t = \{\tau_t^1, \dots, \tau_t^p\}$ , де  $p$  – кількість видів нечітких тенденцій, визначених на часовому ряді.

У припущенні про розвиток системи як результату попередніх станів, допускаємо наявність залежності нечітких тенденцій від значень тенденцій у попередні моменти часу у вигляді так званих моделей нечітких тенденцій.

Модель нечітких тенденцій це сукупність компонент та рівнянь [4, 5]:

$$\tilde{y}'_t = Fuzzy[y_t], y'_t = deFuzzy[\tilde{y}'_t],$$

$$\tau_t^j = Tend[\tilde{y}'_{t-m_j+1}, \dots, \tilde{y}'_t],$$

$$\tilde{y}''_t = deTend[\tau_t^1, \dots, \tau_{t+m-1}^p],$$

$$\tau_t = f(\tau_{t-1}, \dots, \tau_{t-l}), i=1..n, j=1..p, m = \max(m_j),$$

де  $\tilde{y}_t = \{\tilde{y}_t^j\}$  – нечіткий часовий ряд;  $n$  – кількість термів нечіткого часового ряду;  $Fuzzy$  – функціонал фазифікації (приведення до нечіткості);  $\tau_t = \{\tau_t^j\}$  – часовий ряд нечіткої тенденції;  $p$  – кількість термів нечіткої тенденції;  $m_j$  – інтервал визначення нечіткої тенденції;  $Tend$  – функціонал розпізнавання нечіткої тенденції;  $f$  – функціональна залежність;  $\tilde{y}'_t = \{\tilde{y}'_t^i\}$  – нечіткий часовий ряд, отриманий із часового ряду нечітких тенденцій;  $deTend$  – функціонал отримання нечіткого часового ряду із оцінки нечіткої тенденції;  $\tilde{y}_t$  – дефазифікований (приведений до чіткості) часовий ряд;  $deFuzzy$  – функціонал отримання чіткого часового ряду із нечіткого часового ряду.

Побудова моделей часових рядів - це ітеративний процес ідентифікації, оцінки та діагностики моделі. Ідентифікація моделі нечітких тенденцій полягає в описі нечітких змінних і побудові відповідних функціоналів. На етапі оцінки для моде-

лі знаходяться функціональні залежності і оцінюються параметри обраної функції. При діагностиці перевіряється здатність моделі до прогнозування. При цьому використовується ряд правил рівнів R1-R3.

Сукупність всіх правил визначення видів нечітких тенденцій складає перший рівень загальної системи логічного виведення моделі нечіткої тенденції (МНТ), яка реалізує функціонал  $Tend$  для одержання часового ряду нечітких тенденцій [5]:

$$R_1: \tau_t = \{\tau_t^i = \bigcup_{j=1}^{m_j} (T(\tilde{y}_{t-k+1}^{a(i,j,k)}))\} \quad (1)$$

Множина правил рівнянь моделей нечітких тенденцій складає другий рівень виведення:

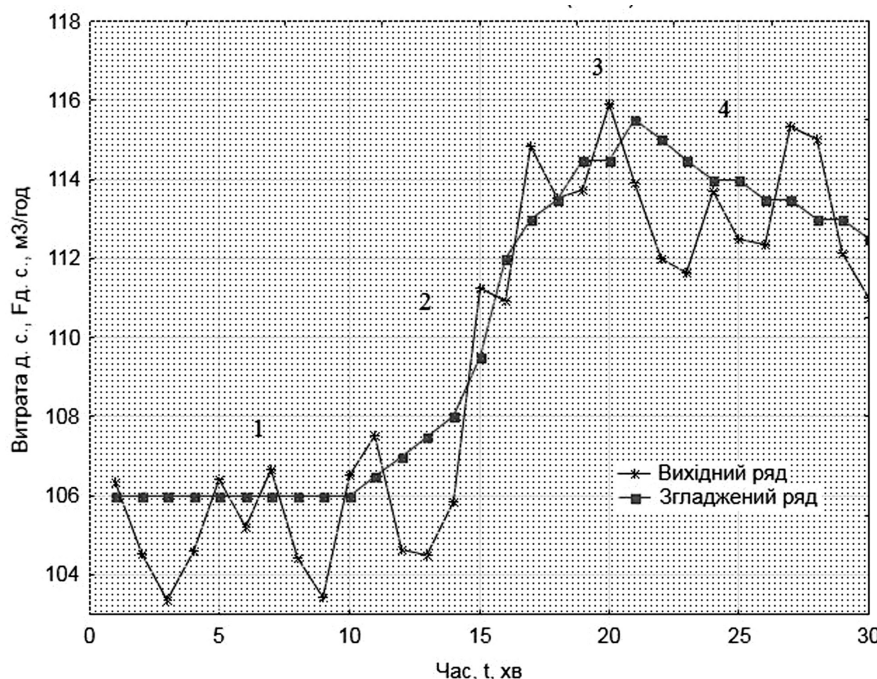
$$R_2: \tau_t = \{\tau_t^i = \bigcup_{j=1}^l (T(\tau_{t-k}^{c(i,j,k)}))\} \quad (2)$$

Для аналізу і побудови чіткого часового ряду за нечіткою моделлю (функціонал  $deTend$ ) кожному виду нечіткої тенденції співставляється нечіткий часовий ряд, що має найбільшу ступінь належності. Множина таких правил, що реалізує функціонал перетворення часового ряду нечітких тенденцій в нечіткі часові ряди  $deTend$ , формує третій рівень:

$$R_3: \tilde{y}_t = \{\tilde{y}_t^k = \bigcup_{i=1}^p (\bigcup_{\substack{j=1, \\ b(i,j)=k}}^{m_i} \tau_{t+j-1}^i)\} \quad (3)$$

Таким чином, модель нечіткої тенденції повністю реалізується багаторівневою системою правил:  $R_1 \Rightarrow R_2 \Rightarrow R_3$ , де виходи у вигляді нечітких змінних одного набору правил подаються на входи наступного набору правил.

При використанні моделей часових рядів на основі нечітких елементарних тенденцій для прогнозування часових рядів використаний підхід, при якому не здійснюється попередня обробка ви-



- 1 – незначне зростання
- 2 – стрімке зростання
- 3 – зміна тенденцій (збільшення/зменшення)
- 4 – помірне зниження

Рис. 2. Вихідний та згладжений нечіткий часовий ряд «Витрата дифузійного соку»



хідних часових рядів. На практиці необхідно виявляти основну тенденцію та прогноз саме цієї тенденції. Для побудови моделей основної тенденції використовувались методи згладжування часового ряду [5].

Побудова моделей прогнозування системних тенденцій здійснювалась в декілька етапів: побудова лінгвістичної шкали, де виходячи із точності вимірювань та розмаху часового ряду визначали параметри і кількість градацій нечітких шкал; нечіткого перетворення часового ряду в нечіткий ряд значень та нечітких елементарних тенденцій; сегментація часових рядів та ідентифікація нечітких локальних тенденцій; ідентифікація основної нечіткої тенденції часового ряду.

Для прикладу, наведені результати прогнозування тенденцій поведінки об'єкта керування (технологічного комплексу цукрового заводу) за часовими рядами витрати дифузійного соку (рис. 1).

На **рис. 2** наведено графіки вихідного часового ряду (витрата дифузійного соку) та результат його згладжування на основі методу нечітких перетворень у вигляді нечітких тенденцій.

Основна ідея лінгвістичної апроксимації полягає в формалізації причинно-наслідкових зв'язків між змінними «Входи (фактори) - Вихід» за допомогою опису цих зв'язків природною мовою з застосуванням теорії нечітких множин та лінгвістичних змінних.

Формально, лінгвістична змінна описується такими параметрами:  $X$  - ім'я змінної;  $T$  - термножина, кожен елемент якої задається нечіткою множиною на універсальній множині  $U$ ;  $G$  - синтаксичні правила, що утворюють назви термів;  $M$  - семантичні правила, що задають функції належності нечітких термів [7].

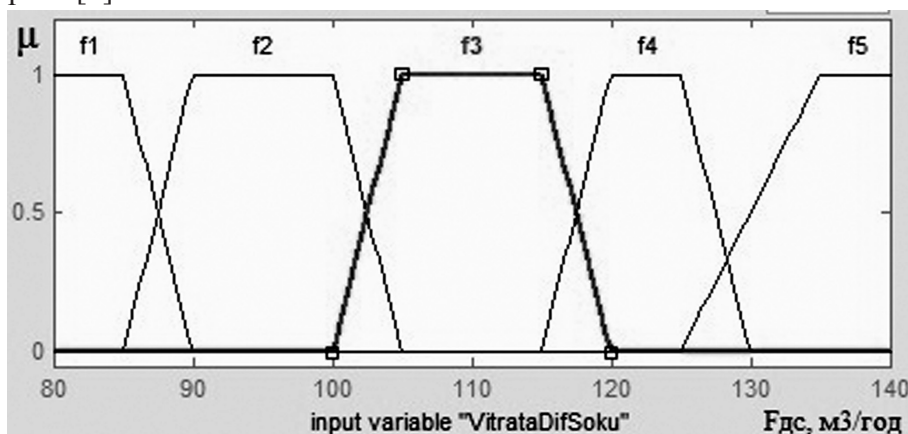
Кожен експерт заповнив анкету, в якій вказав свою думку про наявність у елементів  $u_i (i = \overline{1, n})$  властивостей нечіткої множини  $\bar{l}j (j = \overline{1, m})$ . Нехай  $K$  - кількість експертів,  $b_{i,j}^k$  - думка  $k$ -го експерта про наявність у елемента  $u_i$  властивостей нечіткої множини  $\bar{l}j, k = \overline{1, K}, i = \overline{1, n}, j = \overline{1, m}$ . Вважалося, що експертні оцінки бінарні, тобто  $b_{i,j}^k \in \{0;1\}$  де 1 - наявність властивостей нечіткої множини  $\bar{l}j$  у елемента  $u_i$ , а 0 - відсутність властивостей (ознак). За результатами анкетування ступінь належності множині був розрахований за формулою:

$$\mu_{l_j}(u_i) = \frac{1}{K} \sum_{k=1, K} b_{j,i}^k, i = \overline{1, n}. \quad (8)$$

Функції належності зручно задавати в параметричній формі. Найбільшу популярність отримали трикутна, трапецевидна, гаусова, сигмоїдальна та Пі-подібна функції належності [8].

Так як, Пі-подібна функція дозволяє здійснити плавний перехід від песимістичної до оптимістичної оцінки нечіткого числа, тому для побудови функції належності змінної «Витрата дифузійного соку» було використано саме її. На **рис. 3** показано побудову даних функцій, а в табл.1 наведені визначення терм - множин змінної.

В основі нечітких моделей лежить сукупність нечітких правил «ЯКЩО-ТОДІ» логічного виведення, які описують залежності між нечіткими змінними предметної області, композиційне правило виведення і спосіб обчислення значень нечітких змінних (спосіб нечіткого виведення). До системи нечіткого логічного виведення входять



**Рис. 3.** Функції належності лінгвістичної змінної «Витрата дифузійного соку».

Для лінгвістичної апроксимації змінних використовувався аналітичний підхід на основі інформації, отриманої в результаті експертного опитування.

Для побудови функцій належності було використано метод, що базується на статистичній обробці думок групи експертів [7].

наступні об'єкти:

- 1) сукупність нечітких продукційних правил (база правил);
- 2) набір функцій належності бази нечітких змінних (база змінних);
- 3) блок фазифікації;

Характеристика термів змінної «Витрата дифузійного соку»

Позначення факторів	Витрата дифузійного соку	Координати функції належності Фд.с., м <sup>3</sup> /год
f1	Незначна	[80 85 90]
f2	Близька до нормальної	[85 90 100 105]
f3	Нормальна	[100 105 115 120]
f4	Висока	[115 120 125 130]
f5	Занадто висока	[125 135 140]

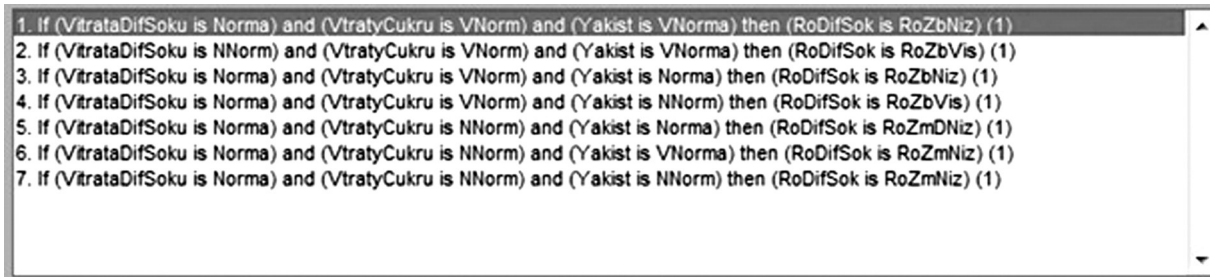


Рис. 4. Фрагмент бази знань

- 4) блок дефазифікації;
- 5) блок виведення.

База правил містить множину логічних правил виведення, а також їх порядок у вигляді ієрархічної структури застосування. База нечітких змінних містить назви лінгвістичних термів та параметри їх функцій належності. Бази правил і нечітких змінних утворюють базу знань системи нечіткого виведення (рис.4).

**Висновки**

Характерною особливістю цукрового заводу є великий ступінь невизначеності; це стосується насамперед якісних показників продукту. Часові ряди технологічних змінних є основною інформацією про стан технологічного комплексу цукрового заводу.

Інтелектуальний аналіз даних дозволяє визначити стан технологічного об'єкта, зробити висновок про якість його функціонування, дати рекомендації з пошуку і усуненню проблемних ситуацій. Застосування інтелектуальних технологій дозволяє усунути різноманітні невизначеності, що позитивно впливає на ефективність прийняття рішень по управлінню. ■

**Список використаних джерел**

1. Інноваційні технології в управлінні складними біотехнологічними об'єктами агропромислового комплексу/А. П. Ладанюк, В. М. Решетюк, В. Д. Кишенько, Я. В. Смітюх. – К. : Центр учбової літератури, 2014.– 280 с.
2. Бойко Р.О. Система управління технологіч-

ним комплексом цукрового заводу з використанням комбінованого критерію ефективності / Р.О. Бойко, Л.Г. Загоровська // Восточно-Европейский журнал передовых технологий. – 2011. – №6/9 (54). – С. 58–60.

3. Минаев Ю. Н. Методы и алгоритмы идентификации и прогнозирования в условиях неопределенности в нейросетевом логическом базисе / Ю. Н. Минаев, О. Ю. Филимонова, Бенамеур Л. – М. : Горячая линия - Телеком. – 2003. – 205 с.

4. Афанасьева Т. В. Моделирование нечетких тенденций временных рядов / Т. В. Афанасьева. – У. : УлГТУ, 2013. – 215 с.

5. Ярушкина Н. Г. Интеллектуальный анализ временных рядов: Учебн. пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. – У. : УлГТУ, 2010. – 320 с.

6. Ярушкина Н. Г. Основы теории нечетких и гибридных систем / Н.Г. Ярушкина. – М. : Финансы и статистика, 2004. – 243 с.

7. Song Q. Forecasting enrollments with fuzzy time series Part I. / Song Q., Chissom B. // Fuzzy Sets and Systems. 1993. №54. P. 1-9.

8. Kacprzyk J. Linguistic summarization of time series by using the Choquet integral / J. Kacprzyk, A. Wilbik, S. Zadrozny // Theoretical Advances and Applications of Fuzzy Logic. 2007. IFSA'07

9. Pedrycz W. Granular correlation analysis in data mining / W. Pedrycz, M. Smith // Proc. IEEE Int. Fuzzy Systems Conf. 1999. III-1235 IH-1240.

10. Батыршин И. З. Основные операции нечеткой логики и их обобщения / И.З. Батыршин.– К. : Отечество, 2001. – 100 с.