

УДК 519.6:678.7

# ДОСВІД ЗАСТОСУВАННЯ КЛАСТЕРНОГО АЛГОРИТМУ ДЛЯ БАГАТО- ЕКСТРЕМАЛЬНОЇ ОПТИМІЗАЦІЇ СКЛАДУ ПОЛІМЕРНОГО КОМПОЗИТУ

Д. М. Складанний

Кандидат технічних наук, доцент

Кафедра кібернетики хіміко-технологічних процесів

Національний технічний університет України

«Київський політехнічний інститут»

пр-т Перемоги, 37, корпус 4, м. Київ, Україна, 03056

E-mail: skl\_den@ukr.net

*В роботі показано досвід застосування кластерного алгоритму для багатоекстремальної оптимізації складу полімерного композиту з використанням узагальненого показника якості. Показано, що невдале обрання алгоритму визначення кількості кластерів може призвести до спотворення результатів оптимізації. Продемонстровано, що алгоритм формального елемента найкраще підходить до вирішення подібних задач*

*Ключові слова: багатоекстремальна оптимізація, кластерний алгоритм, полімерний композит, узагальнений показник якості*

*В работе представлен опыт применения кластерного алгоритма для многоэкстремальной оптимизации состава полимерного композита с использованием обобщенного показателя качества. Показано, что неудачный выбор алгоритма определения количества кластеров может привести к искажению результатов оптимизации. Продемонстрировано, что алгоритм формального элемента наилучше подходит для решения подобных задач*

*Ключевые слова: многоэкстремальная оптимизация, кластерный алгоритм, полимерный композит, обобщенный показатель качества*

## 1. Вступ

Методи багатоекстремальної оптимізації знаходять все ширше застосування для вирішення інженерних технологічних задач. На даний час розроблено та успішно застосовується велика кількість методів багатоекстремальної оптимізації: генетичні алгоритми, алгоритми конкуруючих точок, кластерні алгоритми, тощо.

Основним недоліком усіх методів багатоекстремальної оптимізації є необхідність проведення великої кількості обчислень цільової функції. Тому для вирішення таких задач зазвичай розробляють застосовують спеціалізоване програмне забезпечення. В той же час, більшість комп'ютерних програм, що виконують розв'язок оптимізаційних задач, в кращому випадку дають можливість користувачу обрати метод розрахунку і задати його точність, після чого надають йому готовий результат, а сам процес розрахунку і проміжні результати залишається прихованими від користувача. Такий підхід, на думку автора, не лише поганий з методичної точки зору, а й може призвести до спотворення результатів.

## 2. Аналіз результатів попередніх досліджень і постановка проблеми

Дослідження в області вирішення задач оптимізації складу полімерних композитів започатковані на

кафедрі КХТП НТУУ «КПІ» під керівництвом тодішнього завідувача кафедри професора Статюхи Г.О. і їх результати представлені в ряді опублікованих робіт, наприклад [1, 2, 3]. В попередніх дослідженнях [2, 3] показано результати застосування експериментально-статистичного підходу до розв'язку задач оптимізації складу полімерного композиту за узагальненими показниками якості, що включають різні механічні властивості.

Результати експериментів на полімерному композиті подані у вищевказаних роботах, виконаних за участю автора.

У роботах [2, 3] запропоновано використовувати метод псевдокомпонентів [4] для дослідження залежності властивостей полімерного композиту за наявності обмежень на склад компонентів. Сумішева система, що досліджується в роботі, складається з трьох компонентів:  $q_1$  – доля поліуретану,  $q_2$  – доля стиролу і  $q_3$  – доля полієфіру в суміші відповідно. При цьому на вміст компонентів в суміші накладаються технологічні обмеження:  $q_1 \geq 0,1$ ;  $q_2 \geq 0,1$ ;  $q_3 \geq 0,5$ . Контроль якості одержаного полімерного композиту здійснюється за чотирма показниками: міцність на розрив, МПа (позначимо як  $Y_1$ ); подовження при розриві, % ( $Y_2$ ); ударна міцність, КДж/см<sup>2</sup> ( $Y_3$ ); твердість, OShA ( $Y_4$ ). Адекватні експериментально-статистичні моделі у вигляді поліномів Шеффе, на основі яких проведено розрахунки, побудовані за участю автора в роботі [2]. Ці моделі використовуються у цій статті, коефіцієнти цих моделей зведені в табл. 1.

Таблица 1

Коефіцієнти моделей Шеффе

Регресор	$Y_1$	$Y_2$	$Y_3$	$Y_4$
$q_1$	1,52	2,90	0,69	89,10
$q_2$	28,31	4,30	0,55	96,50
$q_3$	16,23	7,80	0,88	96,50
$q_1q_2$	18,05	2,84	0,69	13,05
$q_1q_3$	1,17	5,90	2,01	6,98
$q_3q_3$	21,31	6,37	1,43	15,75
$q_1q_2 (q_1 - q_2)$	31,05	7,24	0,42	26,10
$q_1q_3 (q_1 - q_2)$	60,25	21,91	3,13	47,47
$q_2q_3 (q_2 - q_3)$	34,54	5,65	0,85	40,50
$q_1q_2q_3$	20,91	6,68	9,10	133,5

В роботі [1, 2], використано узагальнений показник якості складу полімерного композиту – узагальнену функцію бажаності Харінгтона, який також використано у цій роботі. Оскільки функція бажаності Харінгтона носить частково емпіричний характер, досить ризиковано приймати за однозначне рішення задачі оптимізації єдиний розв'язок, знайдений як глобальний оптимум цієї функції.

В роботі [2] проведена спроба багатоекстремальної оптимізації складу полімерного композиту і подані її результати. Проте в процесі виконання цієї роботи автори зштовхнулися з труднощами, пов'язаними з неоднозначністю розв'язків задачі багатоекстремальної оптимізації.

Кластерні алгоритми дозволяють суттєво скоротити кількість процедур локальної оптимізації, є ефективні і нині мають широке практичне застосування. На даний час відомо багато методів кластерного аналізу [5, 6].

Очевидно, що методи кластеризації, які розраховані за знаходження заздалегідь відомої кількості кластерів не можуть бути використані для розв'язку задачі багатоекстремальної оптимізації, тому перед дослідником постає проблема вибору кількості кластерів [7, 8, 9].

В цій статті автор спробує виявити причини такої неоднозначності та запропонувати шляхи її подолання та продемонструвати вплив обраного методу кластеризації на результати оптимізації.

### 3. Мета і задачі дослідження

Дана стаття є продовженням досліджень, проведених за участю автора, результати яких представлені у [2, 3]. Метою статті є демонстрація досвіду вирішення інженерної задачі оптимізації складу полімерного композиту з пошуком не лише єдиного, а і усіх можливих екстремумів за узагальненим показником якості.

Може скластися враження, що використанням кластерного методу багатоекстремальної оптимізації є надмірно складним для вирішення даної задачі, проте, як буде показано нижче, розглянутий приклад достатньо характерний і дозволяє продемонструвати переваги, недоліки і точки підвищеної уваги зазначеного методу.

### 4. Обрання стартових точок та перша ітерація алгоритму

В якості методу багатоекстремальної оптимізації в роботі застосовано кластерний алгоритм А. Торна, описаний, наприклад, в [10]. Множину початкових точок в області псевдокомпонентів оберемо так, щоб вони рівномірно покривали всю область. В даній роботі початкове розташування точок обране у відповідності до плану Шеффе четвертого порядку [4], показане на рис. 1. З кожної з початкових точок виконано процедуру локальної оптимізації з початковим кроком 0,1. Для цього застосовано розроблений на кафедрі КХТП НТУУ «КПІ» програмний пакет STAT-SENS [11]. В результаті одержано ряд точок, які наближені до шуканих екстремумів. Розміщення точок на площині псевдокомпонентів після першого кроку алгоритму показано на рис. 2. Згідно алгоритму Торна, для всіх одержаних точок має бути проведена процедура кластерного аналізу, і після того, як кластер буде виявлено, в ньому залишають лише одну, зазвичай найкращу, точку.

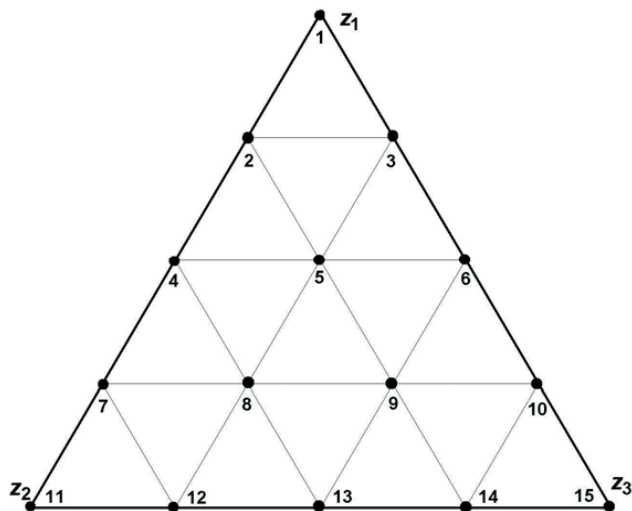


Рис. 1. Початкове розміщення точок

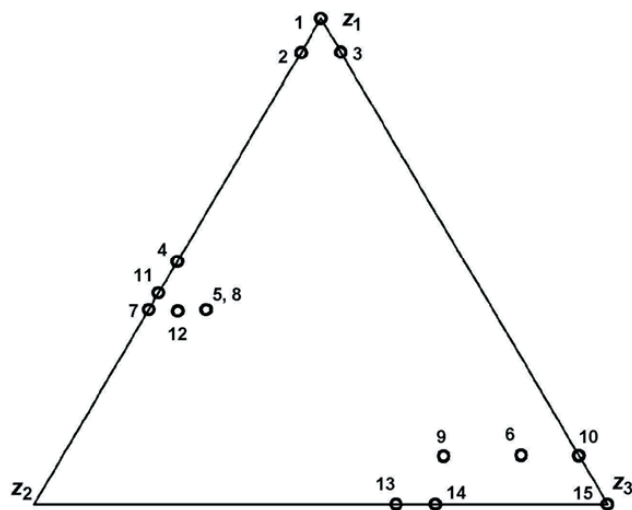


Рис. 2. Розміщення точок після першого кроку алгоритму

Слід зазначити що закінчення першої ітерації є дуже важливим етапом кластерного алгоритму. Адже кількість кластерів, що буде знайдена на цьому етапі в подальшому може бути лише зменшена, наприклад, за рахунок об'єднання кластерів на подальших кроках алгоритму, але не може бути збільшеною, адже з кожного знайденого кластеру буде залишено лише одну точку.

**5. Визначення кількості кластерів із застосуванням індексного підходу**

Одним з найбільш часто застосовуваних підходів до визначення кількості кластерів є підхід із застосуванням індексу, розробленого Р. Калінським і Дж. Хазабадом [9]. У відповідності до нього, найбільш ймовірною кількістю кластерів є та, яка максимізує запропонований ними індекс:

$$CH(k) = \frac{B(k)/(k-1)}{W(k)/(n-k)}, \tag{1}$$

де: k – кількість кластерів, n – кількість об'єктів, як потрібно кластеризувати, в даному випадку кількість точок, W(k) – сума квадратів міжкластерних відстаней за даної кількості кластерів, B(k) – відповідна сума квадратів внутрішньокластерних відстаней. Для проведення кластерного аналізу і розрахунку відстаней автором застосовано математичний пакет Statistica v.6. Результати розрахунків зведено в табл. 2.

**Таблиця 2**

Визначення індексу Калінського-Хазабада

Кількість кластерів, k	2	3	4	5
CH(k)	0,76	1,45	1,08	0,59

Таким чином, у відповідності до підходу Калінського-Хазабада найбільш ймовірною є кількість кластерів рівна трьом. Надалі з кожного кластеру залишається по одній точці в яких значення узагальненого показника якості максимально (в цьому випадку точки 1, 11, 15 на рис. 2) і процедура оптимізації доводиться до кінця. Результати подано в табл. 3.

**Таблиця 3**

Результати оптимізації

Псевдокомпоненти			Реальні компоненти			D
z <sub>1</sub>	z <sub>2</sub>	z <sub>3</sub>	q <sub>1</sub>	q <sub>2</sub>	q <sub>3</sub>	
1,000	0,000	0,000	0,400	0,100	0,500	0,701
0,487	0,513	0,000	0,246	0,254	0,500	0,708
0,000	0,000	1,000	0,100	0,100	0,800	0,648

**6. Визначення кількості кластерів із застосуванням рандомізованого підходу**

Рандомізований підхід до визначення кількості кластерів запропоновано К. Сьюгар та Г. Джеймсом

[8]. У відповідності до цього підходу множина даних розбивається послідовно на 2, 3 і т.д. кластерів з використанням будь-якого алгоритму кластеризації, наприклад k-середніх [5, 6]. Для кожного з кластерів обчислюються внутрішні дисперсії і серед них обирається мінімальна. Ця дисперсія отримує назву «спотворення» – d<sub>k</sub>.

Найбільш ймовірною кількістю кластерів вважається така, для якої «стрибок» J<sub>k</sub> = d<sub>k</sub><sup>-λ</sup> – d<sub>k-1</sub><sup>-λ</sup> буде максимальним, де λ – параметр, який автори [8] рекомендують обирати рівним половині розмірності простору даних. В даній роботі λ = 1, оскільки простір двомірний. Результати розрахунку «стрибків» зведено в табл. 4.

**Таблиця 4**

Визначення «спотворень» і «стрибків»

Кількість кластерів, k	2	3	4	5
«Спотворення» d <sub>k</sub>	1,851	0,737	0,464	0,441
Стрибок d <sub>k</sub> <sup>-λ</sup> – d <sub>k-1</sub> <sup>-λ</sup>	–	0,817	0,798	0,112

Як видно з табл. 3, найбільший «стрибок» спостерігається на переході від двох до трьох кластерів. Слід відмітити, що він незначно перевищує «стрибок» на переході від трьох до чотирьох кластерів, проте слідуючи алгоритму констатуємо, що найбільш ймовірна кількість кластерів рівна трьом. Подальша процедура оптимізації та її результати не відрізняються від наведених у попередньому пункті статті (табл. 2).

**7. Визначення кількості кластерів із застосуванням алгоритму формального елементу**

Іншим підходом до визначення кількості кластерів є алгоритми родини FOREL, тобто формального елементу [6]. Відповідно до алгоритму формального елементу слід встановити формальний центр кластеру і радіус сфери, центр якої збігається з формальним центром кластеру, і яка охоплює всі точки, що потрапили в кластер. Оскільки в роботу кластеризація застосовується як частина алгоритму оптимізації, встановимо ці величини, виходячи саме з міркувань задачі. У якості формальних центрів кластеризації приймемо точки, у яких значення узагальненого показника якості буде мінімальним. Радіус сфери оберемо відповідно до початкового кроку обраного методу оптимізації. Якщо процедура локальної оптимізації з усіх початкових точок виконана вірно, точки в околі локальних екстремумів не можуть лежати одна від одної на відстані, що перевищує зображену на рис. 3. Ця відстань, очевидно, буде подвоєною висотою рівностороннього трикутника зі стороною, рівною початковому кроку методу оптимізації, тобто r = h√3. Для включення всіх точок відповідно до алгоритму формального елементу потрібно чотири сфери з центрами в точках 1, 11, 14, 15 (рис. 2). Таким чином, до трьох кластерів знайдених вище, додається четвертий. Після залишення в кожному кластері по одній точці і до виконання процедури оптимізації, знайдено чотири екстремуми, три з яких відповідають наведеним в табл. 2, а четвертий у табл. 5:

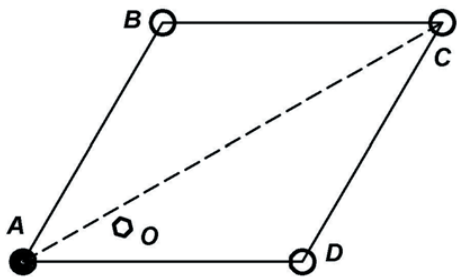


Рис. 3. Визначення центру і радіуса сфери при кластеризації: А – точка центру сфери з найближчим до оптимального значенням; В, С, D – точки кандидати на потрапляння в кластер, О – точка оптимуму, АС – максимально можлива відстань від центру сфери до точки кандидата

Таблиця 5

Додаткові результати оптимізації

Псевдокомпоненти			Реальні компоненти			D
$z_1$	$z_2$	$z_3$	$q_1$	$q_2$	$q_3$	
0,373	0,000	0,627	0,212	0,100	0,688	0,629

### 8. Порівняльний аналіз результатів

Як видно з наведеного вище, результати вирішення задачі багатоекстремальної оптимізації узагальненого показника якості загалом різні. Не залежно від обраного методу кластеризації, глобальний оптимум складу полімерного композиту збігається і відповідає складу {0,246; 0,254; 0,5}. У той же час наявні розбіжності у кількості локальних екстремумів задачі.

У зв'язку з розбіжністю результатів оптимізації за кількістю локальних екстремумів, автором вирішено перевірити результати, застосувавши відомий підхід мультистрату. Початкові точки, загальною кількістю 50, для процедури мультистрату обиралися випадковим чином і охоплювали всю область, виділену вказаними вище обмеженнями задачі. Загальна кількість

розрахунків узагальненого показника якості становила близько 15,5 тис. Задача вирішувалася модифікованим для симплексних системи безградієнтним методом Гауса-Зейделя і для вирішення задачі широко застосовувалась вищезгадана система STAT-SENS [11]. Ймовірність того, що хоча б одна випадкова точка потрапила в область притягання кожного з локальних екстремумів становить не менше 95%. У результаті застосування процедури мультистрату визначена наявність у даній задачі чотирьох екстремумів – одного глобального і трьох локальних, які збігаються з наведеними в табл. 2 і 5. Результати мультистрату збігаються з результатами, отриманими із застосуванням модифікованого алгоритму пошуку формального елементу з точністю, що не перевищує задану точність числового методу оптимізації.

### 9. Висновки

Таким чином, кластерний алгоритм багатокритеріальної багатоекстремальної оптимізації у застосуванні до оптимізації складу композиційного матеріалу має різну ефективність у залежності від обраного методу визначення кількості кластерів. Глобальний екстремум узагальненого показника якості визначений вдало не залежно від обраного алгоритму визначення кількості кластерів. Використовуючи індексний підхід Р. Калінського і Дж. Хазабада і рандомізований підхід К. Стьюгера і Г. Джеймса для визначення кількості кластерів у вирішенні даної задачі, автору не вдалося отримати всіх відомих локальних екстремумів через невірне визначення кількості кластерів після першої ітерації алгоритму. Автор вбачає причину такого визначення в тому, що в досліджуваній задачі два екстремуми узагальненого показника якості знаходились достатньо близько один до одного у порівнянні з відстанями від них до інших екстремумів. В той же час, алгоритм визначення кількості кластерів шляхом пошуку формального елементу з незначною його модифікацією, яка враховує особливості оптимізаційного пошуку, показав задовільний результат і дозволив дійсно знайти всі відомі локальні екстремуми задачі.

### Література

1. Статюха, Г.О. Використання даних псевдоактивного експерименту для вирішення інженерних задач при експериментально-статистичному дослідженні полімерних композитів [Текст] / Г.О. Статюха, Н.Є. Теліцина, А.Г. Патрань // Наукові вісті НТУУ «КПІ». – № 5. – 2004. – С. 42–47.
2. Статюха, Г.О. Багатоекстремальна оптимізація процесу синтезу полімерного композиту на основі поліуретану [Текст] / Г.О. Статюха, Д.М. Складанний, Н.Є. Теліцина, О.В. Лизюк // Східноєвропейський журнал передових технологій. – №2/6 (26). – 2007. – С. 32–34.
3. Krol, P. Application of a Method for the Mathematical Experimental Statistical Modeling Approach to Analyze Physical-Chemical Properties of Interpenetrating Polymer Network of Polyurethane and Unsaturated Polyester [Text] / P. Krol, J. Wojturska, G.A. Statyukha, D.M. Skladanny // Jornal of Polymer Science. – 2005, Vol. 97 – pp. 1855–1867.
4. Зедгинидзе, Г. П. Планирование эксперимента для исследования многокомпонентных систем [Текст] / Г.П. Зедгинидзе; Академия Наук СССР, Научный совет по комплексной проблеме Кибернетика. – Москва: Наука. – 1976. – 390 с.
5. Kaufman L. Finding Groups in Data: An Introduction to Cluster Analysis [Text] / L. Kaufman, P. Rousseeuw. – New York: Wiley. – 1990. – 368 p.
6. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний [Текст] / Н.Г. Загоруйко; Новосибирск: Издательство Института математики. – 1999. – 270 с. – ISBN 5-86134-060-9.

7. Krzanowski W. A criterion for determining the number of groups in a dataset using sum of squares clustering [Text] / W. Krzanowski, Y. Lai // Biometrics. – 1985. – № 44. – pp. 23–34.
8. Sugar C. Finding the number of clusters in a data set: An information theoretic approach [Text] / C. Sugar, G. James // J. of the American Statistical Association. – 2003. – № 98. – pp. 750–763.
9. Calinski, R. B. Dendrite method for cluster analysis [Text] / R. B. Calinski, J. A. Harabasz // Communications in Statistics. – 1974. – № 93. – pp. 1–27.
10. Семенкин, Е. С. Методы оптимизации в управлении сложными системами [Текст]: учебное пособие / Е.С. Семенкин, О.Э. Семенкина, В.А. Терсков; Россия. Министерство внутренних дел. – Красноярск: Сибирский юридический институт, 2000. – 254 с. – ISBN 5–93182–008–6.
11. Статюха Г.О. Вступ до планування оптимального експерименту [Текст]: навч. посіб. / Г.О. Статюха, Д.М. Складанний, О.С. Бондаренко. – К.: НТУУ «КПІ», 2011. – 124 с. – 300 пр. ISBN 978-966-622-408-1.

**У роботі пропонується метод прогнозування знаків приростів часових рядів, який базується на застосуванні в комплексі комбінованих моделей селективного типу, складовими яких є індикатори плинних середніх, та попередньої кластеризації часових рядів за методом К-найближчих сусідів**

**Ключові слова:** часовий ряд, прогнозування, знак приросту, кластеризація, метод найближчих сусідів, комбінована модель прогнозування, плинна середня

**В работе предлагается метод прогнозирования знаков приростов временных рядов, который базируется на применении в комплексе комбинированных моделей селективного типа, составляющими которых являются индикаторы скользящих средних, и предварительной кластеризации временных рядов по методу К-ближайших соседей**

**Ключевые слова:** временной ряд, прогнозирование, знак прироста, кластеризация, метод ближайших соседей, комбинированная модель прогнозирования, скользящая средняя

УДК 004:519.2

# МЕТОД ПРОГНОЗУВАННЯ ЗНАКІВ ПРИРОСТІВ ЧАСОВИХ РЯДІВ

О. Ю. Берзлев

Аспірант

Кафедра кібернетики і прикладної  
математикиУжгородський національний університет  
вул. Університетська 14, м. Ужгород,  
Україна, 88000

E-mail: berzlev@gmail.com

## 1. Вступ

Відомо, що більшість часових рядів, для яких виникає задача прогнозування, зокрема рядів економічної природи, як правило, характеризуються нестационарністю і нестійкістю відносно їх середнього рівня. Переважна більшість класичних статистичних моделей та відповідних методів (експоненціальні, лінійні регресійні, авторегресійні типу ARIMA [1-4]) не призначені для прогнозування нестационарних часових рядів, а ті, які для цього призначені (ARIMAX, нелінійні регресійні тощо) характеризуються складністю оцінювання численних параметрів та ідентифікації функціональних залежностей. З огляду на це, окрім задачі прогнозування майбутніх значень рядів, застосовуються інші специфічні задачі, серед яких: ідентифікація моментів локальних екстремумів [5], прогнозування знаків приростів рядів. Остання розглядається в даній роботі.

На фінансовому і валютному ринках часто виникає задача передбачення короткочасної динаміки часового ряду без розрахунку безпосередньо прогнозних значень. Тобто управління процесом прогнозування

в даному випадку передбачає вибір або побудову такої моделі, яка б розраховувала прогноз знаку приросту значення часового ряду на одну точку вперед з необхідною максимальною точністю. Моделі такого типу зазвичай застосовуються для визначення напрямку руху ціни валютних пар і можуть використовуватися для визначення екстремальних точок або точок розвороту ринку, тобто таких точок, які вказують на подальший напрямок руху ціни. На практиці для підвищення точності прогнозування знаків приростів застосовують специфічні моделі та методи.

Питання розробки моделей та методів прогнозування знаків приростів висвітлені в роботі [6], зокрема в ній пропонується модель прогнозування знаків приростів рядів з нестабільним характером коливань. Але на сьогоднішній день не розроблено універсальної методики вирішення цієї задачі, яка б повністю задовольняла цілі прогнозіста, аналітика або інвестора в частині забезпечення необхідної точності прогнозів незалежно від структури часових рядів.

Для вирішення даної задачі автором пропонується метод, який базується на використанні в комплексі