5th International Scientific Conference

Science progress in European countries: new concepts and modern solutions

Hosted by the ORT Publishing and

The Center for Scientific Research "Solution"

Conference papers

February 28, 2019

Stuttgart, Germany

5th International Scientific Conference

"Science progress in European countries: new concepts and modern solutions": Papers of the 5th International Scientific Conference.
February 28, 2019, Stuttgart, Germany. 994 p.

Edited by **Ludwig Siebenberg**

Technical **Editor: Peter Meyer**

ISBN **978-3-944375-22-9**

Published and printed in Germany by ORT Publishing (Germany) in association with the Center For Scientific Research "Solution" (Ukraine)
February 28, 2019.

ORT Publishing

Schwieberdinger Str. 59

70435 Stuttgart, Germany

ISBN **978-3-944375-22-9**

ВИЯВЛЕННЯ ІНФОРМАЦІЇ У СИЛЬНОЗАШУМЛЕНИХ ЧАСОВИХ РЯДАХ ЗА ДОПОМОГОЮ ВИДАЛЕННЯ ВИПАДКОВИХ СПЛЕСКІВ

КОЗУБ П.А.,
кандидат технічних наук, доцент,
доцент кафедри природничих наук
Харківський національний університет радіоелектроніки
Харків, Україна

ВАМБОЛЬ В.В.
доктор технічних наук, професор,
професор кафедри організації та технічного забезпечення аварійно-рятувальних робіт
Національний університет цивільного захисту України
Харків, Україна

КОВАЛЕНКО С.
викладач кафедри прикладної механіки та технологій захисту навколишнього середовища
Національний університет цивільного захисту України
Харків, Україна

КОЗУБ С.М.
кандидат технічних наук, доцент,
доцент кафедри медичної та біоорганічної хімії
Харківський національний медичний університет
Харків, Україна

Часові ряди є одним із найбільш важливих елементів математики та математичної статистики, який є за своєю суттю математичним відображенням реальних процесів, які традиційно представляються як поєднання двох складових – інформації та шуму. При цьому завдання розподілу цих складових і є однією з основних математичних завдань при роботі з часовими рядами [1-7].

Розвиток статистичних методів дозволив виділити різні види шуму, які дуже відрізняються за своїми статистичними параметрами, а тому для їх видалення з числового ряду повинні використовуватись різні математичні методи.

Білий шум з математичної точки зору є послідовність чисел з постійним математичним очікуванням, дисперсією та нульову автоковариаційну функцію. Це означає, що білий шум є стаціонарним процесом, оскільки його показники майже не змінюються з часом, а нульова автокореляційна функція означає, що значення не залежні один від одного.

Ще одним з видів шуму є випадкове блукання яке відрізняється від білого шуму тим, що з часом може змінюватись один або декілька параметрів – математичне очікування, дисперсія, або автокореляціна функція.

При зміні в часі середнього значення часовий ряд називають моделлю із ковзним середнім, в цьому випадку часовий ряд може бути розподілений на часовий ряд із зміною середнього значення та ряду з білим шумом.

При наявності залежностей між попередніми значеннями часовий ряд може бути представлений у вигляді авто регресійної моделі – тобто розділений на функціональну залежність від попередніх значень, часовий ряд з білим шумом.

Таким чином, майже всі методи основані на розділенні часового ряду на якусь функціональну залежність та білий шум, який принципово не може бути приведений до будь-якої функціональної залежності.

Чисто візуально особливістю будь якого шуму є наявність сплесків які помітні у вигляді так званих піків – максимумів. Якщо піки знаходяться досить близько один від одного то між ними утворюються мінімуми подібні за формою до максимумів, а коли ці максимуми та мінімуми повторюються через однакову відстань один від одного то вони стають схожі на тригонометричні функції.

Слід зазначити що ці спостереження є не тільки відображення математичних залежностей, але й відображення законів математичної статистики.

Так можна з показати (навіть теоретично довести), що форма сплеску повинна описуватись для великої кількості даних законом нормального розподілу (що випливає з центральної граничної теореми математичної статистки).

Оскільки координати сплесків також розподілені рівномірно, то вони будуть накладатись один на одного в результаті чого будуть утворюватись сплески з більшою шириною та висотою. Але внаслідок тієї ж теореми, розподіл розмірів та відстані між сплесками також буде відповідати нормальному закону розподілу.

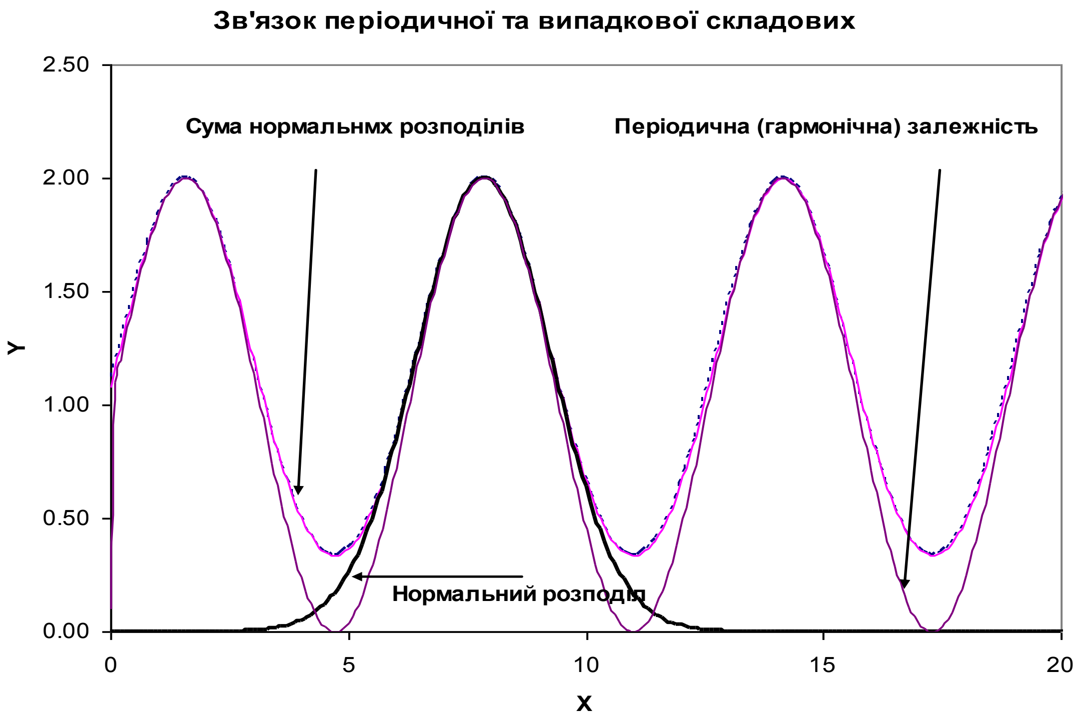

Рис.1. Зв’язок між тригонометричними функціями (sin(x), cos(x)) та функцією нормального розподілу.

Додатково, будь який сплеск є незалежним від інших, тому його вилучення не буде змінювати вид «чистої» залежності, тому послідовність та місце видалення сплесків не має значення для отримання кінцевого результату.

Таким чином будь який шум може бути видалений за допомогою віднімання функції нормального розподілу з загальної залежності.

$$f = y - h \cdot \exp\left[-\left(\frac{x - x_0}{d}\right)^2\right]$$

де

h – висота сплеску;

x_0 – значення максимуму;

d – ширина сплеску.

Таким чином, як видно з рисунку періодична функція є по суті сумою функцій нормального розподілу, які мають сталу ширину та повторюються через рівну відстань по осі абсцис.

Рис.2. Часовий ряд отриманий сумою сплесків нормального розподілу з d = 1/□

У свою чергу, будь який шум є сумою функцій нормального розподілу, які повторюються через нерівну відстань один від одного, але мають середньовирогідну ширину, та відстань між собою.

Слід зазначити, що таке визначення шуму дозволяє запропонувати і критерій його оцінки, оскільки попередні критерії (наприклад дисперсія) основані на вимірюванні відстані від деякої «ідеальної» залежності, яку ми якраз і шукаємо.

Часовий ряд з більшою кількістю шуму буде мати більший шлях від початкової до кінцевої точки, таким чином мінімальна сума різниць між найближчими ординатами ряду буде у ряду з найменшою кількістю шуму.

Це дозволяє запропонувати метод, який заснований двох основних шагах

1 - видалення сплеску з ряду;

2 – якщо цей сплеск призводить до зменшенню відстані між граничними точками, то такий сплеск є найбільш вірогідно випадковим і належить до шуму.

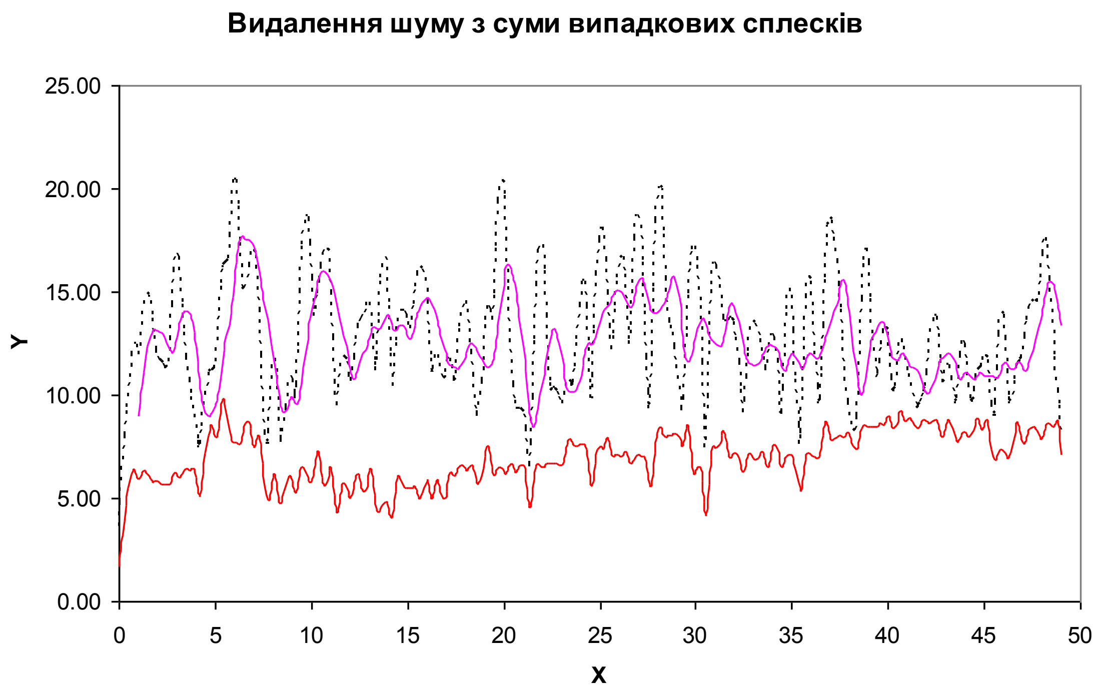

Рис.3. Очищений від шуму часовий ряд видаленням випадкових сплесків (червона лінія) та за допомогою методу ковзкого середнього (рожева лінія)

Рис.3. Видалені випадкові сплески з часового ряду з різними значеннями на півширини випадкової функції

Перевірка цього методу на тестовій функції дозволила видалити до 90% всіх сплесків, у той час як метод ковзкого середнього майже не змінив вид часового ряду. Більше того, видалені сплески можливо розділити за розмірами для подальшого аналізу на наявність додаткової інформації в них.

З метою визначення можливостей методу було проведено його перевірку на реальних часових рядах, взятих з різних галузей науки та техніки

Так одним із найбільш складних об’єктів є дані спектроскопії, які дуже часто дають спектрограми з великою кількістю шумів. Їх обробка є дуже складним завданням, яке навіть зараз не повністю вирішене.

Але використання цього методу дозволило майже повністю видалити шум без видалення інформаційних піків, що видно з наступного рисунку.

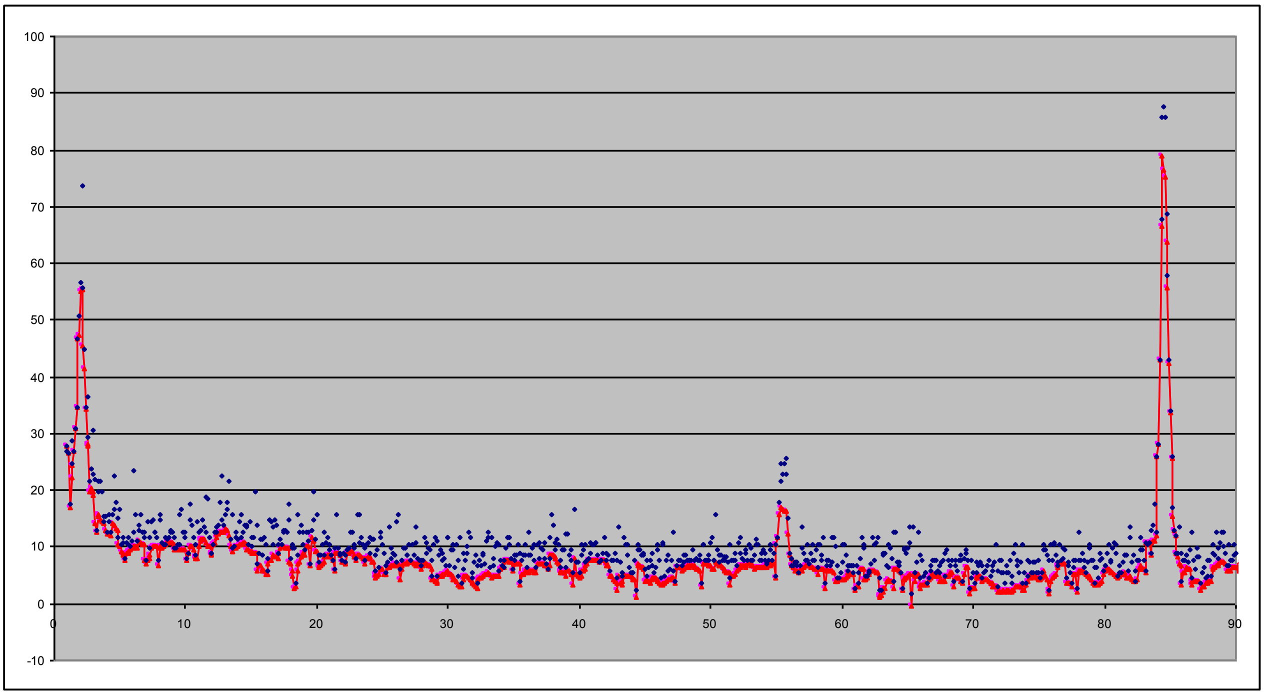

Рис. 4. Приклад видалення шумів із спектрограми

Не менш ефективним виявилось використання цього методу для видалення добових коливань температури повітря, яка залежить від багатьох факторів.

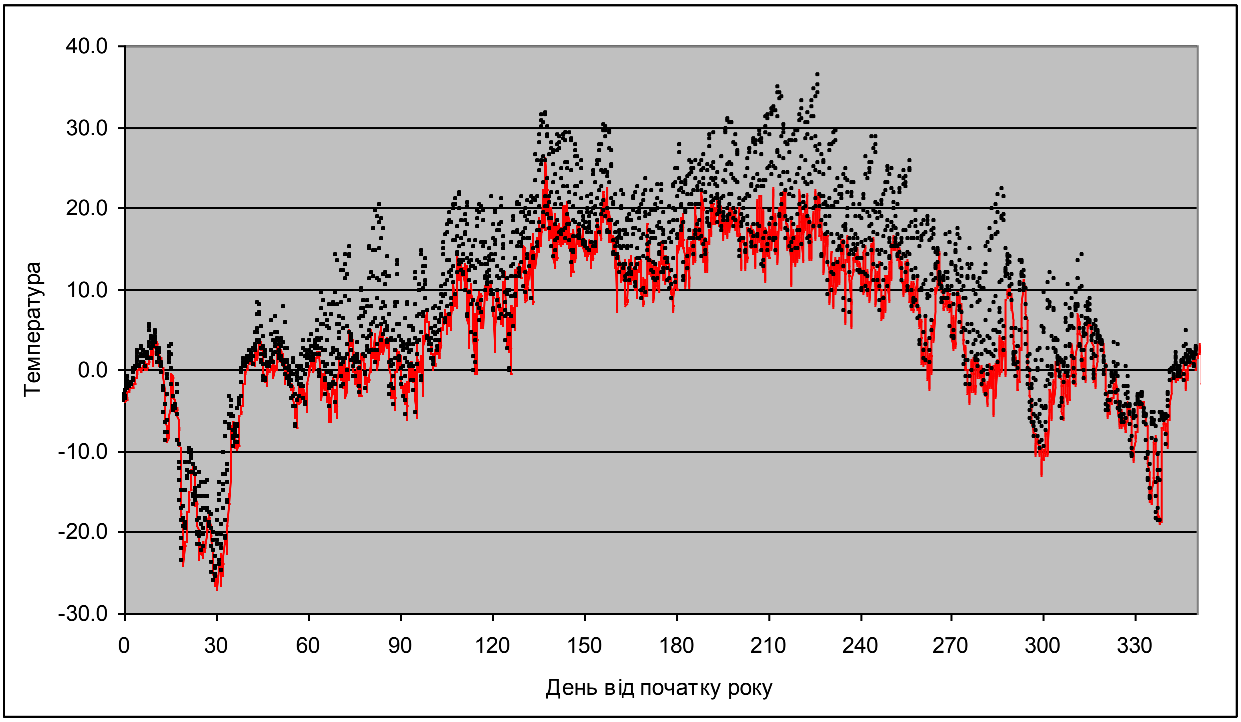

Рис. 5. Приклад видалення добових коливань з річних температурних даних

Ще одним прикладом використання методу є видалення шуму, та виявлення чистої залежності для економічних даних, наприклад коливання ціни на золото.

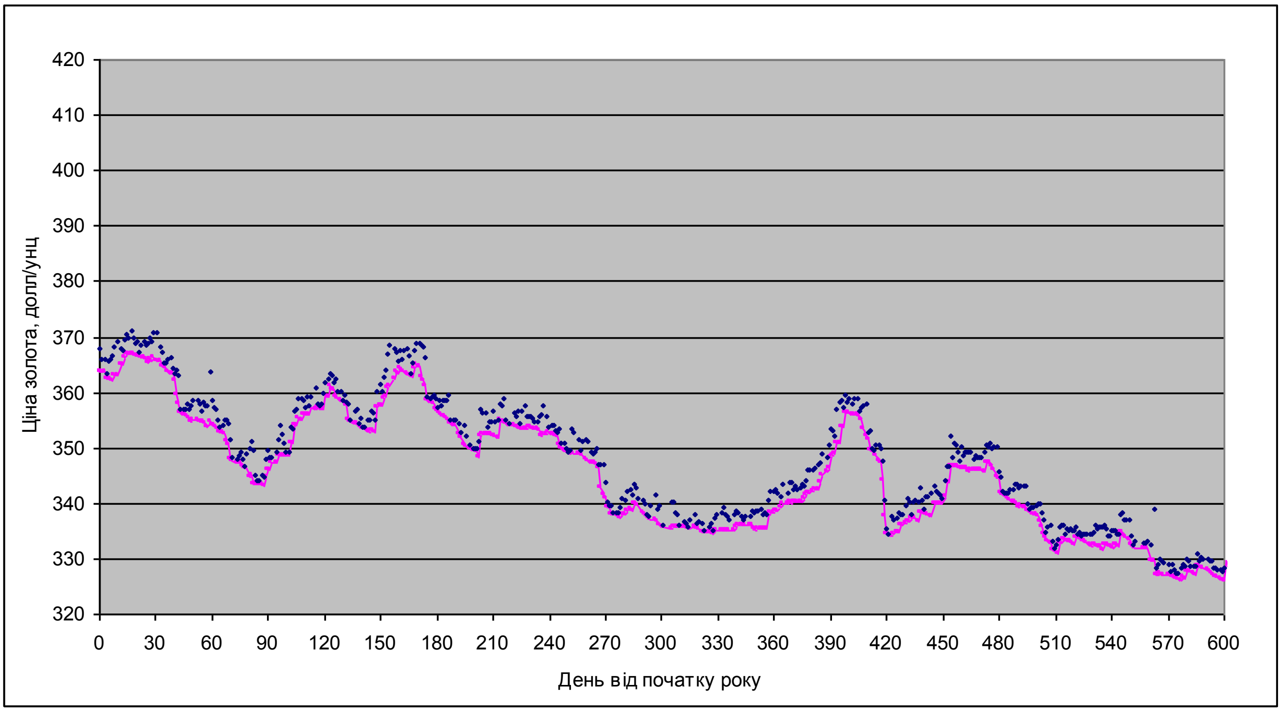

Рис. 6 Приклад видалення випадкових коливань ціни золота з 1991 р.

Таким чином навіть при видаленні тільки однієї складової шуму можливо досягти майже вдесятеро меншої амплітуди випадкових сигналів. Але при цьому вкрай важливим є вибір показників випадкової функції для видалення шуму, який має бути об'єктивним процесом, основаним на теорії математичної статистики.

Тому наступним кроком вдосконалення методу має бути об'єктивізація оцінки параметрів випадкової функції, що дозволить знизити ризик видалення інформаційної складової. Крім того, важливим є збереження даних по видаленим сплескам, оскільки їх послідовність та амплітуда сама по собі може стати наступним рівнем аналізу часового ряду.

Попередні розрахунки показують, що впровадження тільки цих вдосконалень може підвищити ефективність цього методу майже в 10 разів.

Використана література

1. Бокс Д., Дженкинс Г. Анализ временных рядов: прогноз и управление. М: МИР, 1974. С. 406.

2. Вайнштейн Л. А., Вакман Д. Е. Разделение частот в теории колебаний и волн. М: Наука, 1983. С. 287.

3. Добеши И. Десять лекций по вейвлетам. Ижевск: РХД, 2001. С. 464.

4. Ефимов В. М., Галактионов Ю. К., Шушпанова Н. Ф. Анализ и прогноз временных рядов методом главных компонент. Новосибирск: Наука, Сибирское отделение, 1988. С. 70.

5. Кендалл М., Стюарт А. Многомерный статистический анализ и временне ряды. М: НАУКА, 1976. С. 375.

6. Макс Ж. Методы и техника обработки сигналов при физических измерениях. М: Мир, 1983. Т. 2. С. 568.

7. Марпле С. Л. Цифровой спектральный анализ и его приложения: Пер. с англ. Мир, 1990. С. 265. ISBN: 9785030011912.