

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій  
(повна назва)

Кафедра Інформаційно-мережної інженерії  
(повна назва)

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

рівень вищої освіти другий (магістерський)

*Дослідження моделі ARIMA нестационарних випадкових процесів*  
(тема)

Виконав:

студент 2 курсу, групи ІМІм-22-2  
Старокожко Д.А.  
(прізвище, ініціали)

Спеціальність 172 Телекомунікації та  
радіотехніка  
(код і повна назва спеціальності)

Тип програми освітньо-наукова  
(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційно-мережна  
інженерія  
(повна назва освітньої програми)

Керівник: проф. Тихонов В. А.  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

Безрук В.М.

(прізвище, ініціали)

2024 р.

Не містить відомостей, заборонених до відкритого публікування

Студент \_\_\_\_\_ / Старокожко Д. А. /  
( підпис ) ( прізвище та ініціали )

Керівник \_\_\_\_\_ / Тихонов В. А. /  
( підпис ) ( прізвище та ініціали )

Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій  
Кафедра Інформаційно-мережної інженерії  
Рівень вищої освіти другий (магістерський)  
Спеціальність 172 Телекомунікації та радіотехніка  
(код і повна назва)  
Тип програми освітньо-наукова  
Освітня програма Інформаційно-мережна інженерія  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

«    » \_\_\_\_\_ 20\_\_ р.

## ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Старокожку Денису Анатолійовичу  
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження моделі ARIMA нестационарних випадкових процесів

затверджена наказом по університету « 18 » березня 2024 р. № 232 ст.

2. Термін подання студентом роботи до екзаменаційної 18.06.2024

3. Вихідні дані до роботи Імітаційні нестационарні процеси, характеристики СЦП, частоти мод та їх ширина смуг

4. Перелік питань, що потрібно опрацювати в роботі Вступ

Генерації випадкових імітаційних процесів у рамках моделей декомпозиції та ARIMA

Генерації нестационарних випадкових процесів із квадратичним трендом, сезонної та стаціонарної складової у вигляді процесу авторегресії четвертого порядку

Вирішення завдань математичної статистики за допомогою дослідження часових рядів

Передбачення нестационарних процесів

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій Слайди у форматі Power Point

---

---

---

---

---

---

---

---

---

---

---

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Ознайомлення із завданням. Уточнення ТЗ.	18.03.2024 – 31.03.2024	вик.
2	Підбір літератури за темою роботи	01.04.2024 – 05.05.2024	вик.
3	Виконання розділу 1	06.05.2024 – 12.05.2024	вик.
4	Виконання розділу 2	13.05.2024 – 19.05.2024	вик.
5	Виконання розділу 3	20.05.2024 – 26.04.2024	вик.
6	Виконання розділу 4	27.05.2024 – 02.06.2024	вик.
7	Оформлення презентаційного матеріалу, підготовка до захисту у ЕК	03.06.2024 - 11.06.2024	вик.

Дата видачі завдання 18.03.2024 р.

Студент \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_  
(підпис) \_\_\_\_\_  
(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка випускної кваліфікаційної роботи спеціаліста:  
53с., 28 рис., 23 посилання, 17 слайдів.

Об'єктом дослідження є імітаційні та реальні випадкові процеси та їх статистична обробка.

Метою роботи є розробка імітаційного моделювання нестационарних випадкових процесів із заданими характеристиками СЦП, знаходження виразів статистичних характеристик цих процесів у вигляді авторегресійних моделей лінійного передбачення, а також дослідження можливостей їх використання для підвищення якості радіоелектронних систем, статистичне моделювання імітаційних та реальних процесів.

У роботі використовувалися методи теорії множин, лінійних систем, різницевих лінійних рівнянь, методи статистичного моделювання, прикладний аналіз випадкових процесів.

Предметом дослідження є імітаційні моделі випадкового процесу у вигляді декомпозиції та моделей ARIMA.

**МОДЕЛЬ ARIMA, ІМІТАЦІЙНІ ТА РЕАЛЬНІ ВИПАДКОВІ ПРОЦЕСИ,  
ГЕНЕРАЦІЯ ВИПАДКОВИХ ІМІТАЦІЙНИХ ПРОЦЕСІВ**

## ABSTRACT

Explanatory note of the specialist's graduation thesis: 54 pages, 28 figures, 23 references, 17 slides.

The object of research is simulated and real random processes and their statistical processing.

The purpose of the work is the development of simulation modeling of non-stationary random processes with the given characteristics of the SSC, finding expressions of the statistical characteristics of these processes in the form of autoregressive models of linear prediction, as well as researching the possibilities of their use to improve the quality of radio electronic systems, statistical modeling of simulated and real processes.

The work used methods of theory of sets, linear systems, differential linear equations, methods of statistical modeling, applied analysis of random processes.

The subject of the study is simulation models of a random process in the form of decomposition and ARIMA models.

ARIMA MODEL, SIMULATED AND REAL RANDOM PROCESSES,  
GENERATION OF SIMULATED RANDOM PROCESSES

## ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	8
ВСТУП.....	9
1 ОГЛЯД ОСНОВНИХ РОБІТ В АНАЛІЗІ ЧАСОВИХ РЯДІВ.....	11
1.1 Часові ряди.....	11
1.2 Модель авторегресії – ковзного середнього.....	14
1.3 Вистовки розділу та постановка задач.....	19
2 МОДЕЛІ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ.....	22
2.1 Модель авторегресії проінтегрованого ковзного середнього нестационарного процесу.....	22
2.2 Загальні положення теорії виявлення екзопланет.....	27
3 ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ СТАЦІОНАРНИХ І НЕСТАЦІОНАРНИХ ПРОЦЕСІВ НА ОСНОВІ АВТОРЕГРЕСІЇ ТА МОДЕЛІ ARIMA.....	35
3.1 Декомпозиція нестационарного процесу.....	35
3.2 Імітаційний стаціонарний процес авторегресії.....	37
3.3 Нестационарні імітаційні процеси з авторегресійною стаціонарною складовою.....	42
3.4 Моделі імітаційних нестационарних процесів із трендом та сезонною складовою.....	48
4 МОДЕЛЬ ARIMA ДЛЯ КВАДРАТИЧНОГО ТРЕНДА.....	56
4.1 Імітація нестационарного процесу з квадратичним трендом зі стаціонарним процесом другого порядку.....	56
4.2 Імітація нестационарного процесу з квадратичним трендом зі стаціонарним процесом авторегресії четвертого порядку.....	58
ВИСНОВКИ.....	61
ПЕРЕЛІК ПОСИЛАНЬ.....	63
ДОДАТОК А.....	66
ДОДАТОК Б.....	75

## ПЕРЕЛІК СКОРОЧЕНЬ

АР – авторегресія;

КС – ковзне середнє;

БШ – білий шум;

СЩП – спектральна щільність потужності.

## ВСТУП

Аналіз випадкових процесів базується на дослідженні статистичних зв'язків, які втілені в кореляційні та моментні функції. Інформативні відліки, на відміну від білого шуму, статистично пов'язані між собою. За ними розраховуються оцінки статистичних характеристик, які описують випадковий процес.

У роботі розглянуто задачі генерації випадкових імітаційних процесів у рамках моделей декомпозиції та ARIMA. Імітаційні процеси складаються з нестационарної та стаціонарної частин. Для стаціонарних процесів задавалися параметри спектра: частоти мод та їх ширини смуг. Імітаційні нестационарні процеси генерувалися згідно з моделлю декомпозиції. Складовими процесами моделі декомпозиції були лінійний та квадратичний тренд, сезонна і стаціонарна складові. Для генерації імітаційних стаціонарних процесів використовувалися моделі авторегресії різних порядків. порушено питання генерації нестационарних випадкових процесів із квадратичним трендом, сезонної та стаціонарної складової у вигляді процесу авторегресії четвертого порядку.

Операції з відліками випадкових процесів здійснюються за допомогою складних електронних пристроїв: контролерів чи процесорів. Контролер це певний пристрій управління обчислювальної техніки. Операції, які виконує контролер, мають бути зумовлені статистичними математичними моделями. Саме вони визначають коло завдань та алгоритмів, які він має виконувати. У зв'язку з цим зростає роль алгоритмів отримання імітаційних випадкових процесів. На них пробуються методи та алгоритми обробки випадкових процесів. Випадковими процесами описуються номенклатура та ціни товарів, економічні дані, екологічні дані, вимірювані дані виробничого процесу та інші.

Єдину теорію нестационарних процесів створити дуже складно через численні проблеми. Наприклад, за наявності у нестационарного випадкового

процесу тренду дуже складно знаходити навіть прості характеристики процесу, а саме: середнє і дисперсію. Статистичні показники залежатимуть від часу. Немає гарантії, що знайдені статистичні характеристики будуть справедливі і для нових даних нестационарного процесу. Тому, під час аналізу нестационарних процесів, необхідно вказувати для якого типу нестационарних процесів справедливі отримані результати досліджень.

Випадкові часові ряди можна розділити на стаціонарні та нестационарні. Якщо зі стаціонарними процесами особливих проблем немає і вони досить добре вивчені, то нестационарні випадкові процеси дуже складно вивчати і, тим більше, обробляти. Прикладом може бути складне завдання передбачення нестационарних процесів. Котирування на біржі є прикладом нестационарних процесів, які передбачаються за історичними даними, що мають числову природу. Крім цього, котирування цінних паперів та інші інструменти біржі описуються характеристиками нечислової природи. До них відносяться думки фахівців, аналітиків, видатних економічних та політичних діячів. Дані числової та нечислової природи нестационарні, крім того, мають тренд і високий рівень шуму. Це ускладнює статистичну обробку численних даних біржі. Нестационарність ускладнює використання історичних даних. Застосування для прогнозу методів технічного та фінансового аналізу дуже спірне, і не гарантує успіху у торгівлі цінними паперами.

## 1 ОГЛЯД ОСНОВНИХ РОБІТ В АНАЛІЗІ ЧАСОВИХ РЯДІВ

### 1.1 Часові ряди

В останні роки досить різко змінилися здобутки людської діяльності. Важливе значення набуває зберігання та обробка даних. Дані можуть мати різну природу, пов'язану своїм походженням. Насамперед, це часові ряди, інтерес яких викликаний їх розвитком, еволюцією і зміною у часі. З часом змінюються як часові ряди, а також їх статистичні характеристики. Часовими рядами, тобто послідовностями спостережень, вважатимемо випадкові процеси. Ними можуть бути, наприклад, астрономічні процеси, вартість цінних паперів на біржі, показники атмосфери тощо. Випадкові часові ряди широко використовуються у цифровій обробці мовних сигналів, розпізнаванні голосу та економетриці, а також у багатьох інших додатках. Відліки часового ряду статистично пов'язані. Статистичний зв'язок описується функцією кореляції та моментними функціями порядку вище двох. До випадкових часових рядів відносяться також номенклатура та ціни товарів, економічні дані, екологічні дані, режими протікання того чи іншого виробничого процесу та ін.

Для аналізу часового ряду важлива повторюваність, хоча б на деякому інтервалі часу. Цілі дослідження часових рядів можуть бути різними. Для стаціонарних і нестаціонарних процесів можна, наприклад, проводити вимірювання характеристик часового ряду, вирішувати завдання його виявлення, робити прогноз, на підставі історичних даних минулого, управляти процесом, знаходити спектр процесу, намагатися з'ясувати механізм, що лежить в основі процесу, усунути складові нестаціонарного процесу.

Обробка послідовностей подій, вилучення інформації з тимчасових рядів викликає особливий інтерес [1]. Його пов'язують з їх обсягом та накопиченням інформації, еволюцією та зміною її в часі. Нижче розглядаються часові ряди, які моделюються автокорельованими випадковими стаціонарними та

нестационарними процесами [2, 3]. Саме з ними пов'язана передача та зберігання даних про фізичний чи інформаційний процес. Такі часові ряди використовуються у цифровій обробці мовних сигналів, розпізнаванні голосу, економетриці, а також у багатьох інших важливих комерційних або потенційно комерційних додатках [4-8].

Прикладом, який увібрав у собі різні сторони обробки випадкового процесу, є мобільний зв'язок [9-11]. Він вважається розподіленою інтелектуальною системою через свою складність і багатогранність. Для конкретизації мобільного зв'язку говоритимемо про передачу мовного сигналу [12]. Складна апаратура елементів, що забезпечують мобільний зв'язок та обробку мовлення, працює на основі великих мікросхем – контролерів, на яких реалізуються складні алгоритми обробки. Тому нині різко зросла роль статистичних математичних моделей. Контролер може не розуміти, що він робить. Але статистичною моделлю, якою він користується, повинні бути визначені математичні операції, які він виконує.

Часовим рядом називають упорядковану за часом послідовність спостережень [13, 14]. Існують два види вимірів часових рядів. Процес вимірювання може здійснюватися безперервно, а може – дискретно. У базі даних та дослідницьких програмах дані зберігаються зазвичай у дискретному вигляді. Дискретизація проводиться із частотою Найквіста. Квантування за рівнями здійснюється в залежності від завдання, передбачуваних можливостей апаратури обробки, швидкодії апаратури, обсягу електронної пам'яті. Тобто якщо час випадкового процесу вимірюється безперервно, то тимчасовий ряд називається безперервним, якщо час визначено дискретно - тимчасовий ряд дискретний. При цифровій обробці зазвичай розглядаються дискретні тимчасові ряди. Дискретні часові ряди виходять двома способами:

- як вибірки з безперервних часових рядів, утвореними через регулярні проміжки часу;
- вивченням тимчасового ряду протягом деякого періоду часу, наприклад: виявлення екзопланет, температури атмосфери Землі, процесів у

стільниковому зв'язку, обсягу виробництва якого-небудь виду продукції, кількості опадів, обсягу імпорту [1, 2, 16]. Зазвичай, такі випадкові процеси залежить від дискретного часу.

Математична статистика оцінює параметри випадкового процесу по його вибірці. Правильність методів обробки та оцінювання випадкових часових рядів визначається зазвичай спочатку, до обробки та аналізу реальних процесів. Потім ці методи перевіряються на моделі часового ряду. Тому зростає роль теорії генерування випадкових процесів із заданими характеристиками. Уявлення випадкового процесу як синусоїди з доданим білим шумом, чи у вигляді випадкових блукань згідно із законом броунівського руху, не завжди виправдано. Важливо знайти та ефективно використовувати зв'язок між параметрами випадкового процесу та параметрами генерованого процесу. Обробка в реальному часі випадкового процесу передбачає використання деякої статистичної моделі оброблюваних даних.

При нестационарних часових рядах передбачається або наявність тренду, або мінливість статистичних характеристик процесу. Хоча для нестационарного процесу, вважається, що запропонована Дженкінсом і Боксом модель AutoRegressive Integrated Moving Average (ARIMA) є узагальненням змішаної моделі авторегресії - ковзного середнього (АРКС), є різні принципи їх побудови. Для моделі випадкових вузькосмугових процесів АР та широкосмугових процесів ковзного середнього (КС) або моделі зі змішаним спектром – АРКС, вона будується як часовий ряд лінійного передбачення. Але модель ARIMA розуміється, як модель, мультиплікативні складові якої можуть бути видалені за допомогою оператора різниці. А вже отриманий залишок моделюється в загальному випадку процесом АРКС. Процес взяття різниць за допомогою відповідних операторів часто визначається як декомпозиція адитивного випадкового нестационарного процесу. Результатом декомпозиції нестационарного процесу, що описується моделлю ARIMA, є стаціонарний процес, що у загальному вигляді описується моделлю АРКС.

## 1.2 Модель авторегресії – ковзного середнього

Найбільш повною параметричною моделлю лінійного передбачення стаціонарних випадкових процесів є змішана модель АРКС. Об'єднання в однієї моделі властивостей АР і КС процесів дозволяє отримати параметричну модель випадкового процесу з нижчим порядком, ніж можна зробити з допомогою моделей АР чи КС. Ця властивість є основою теореми декомпозиції Уолда [2]. Теорема декомпозиції стверджує, що будь-який стаціонарний процес АРКС з кінцевою дисперсією можна уявити єдиними моделями АР або КС, можливо нескінченного порядку. Моделью АРКС описуються процеси зі змішаним спектром, що містять вузькосмугові та широкосмугові складові. Стаціонарна модель, що є складовою моделі нестаціонарного процесу, представляється у загальному випадку моделлю АРКС.

В операторному вигляді модель АРКС( $p, q$ ) описується виразом

$$\Phi(z)x[t] = Q(z)a[t], \quad (1.1)$$

де оператори  $\Phi(z)$  і  $Q(z)$  є операторами моделей АР і КС  $p$ -го і  $q$ -го порядків, відповідно. Різниця рівняння моделі АРКС(1.1) зручно записати у вигляді

$$\begin{aligned} x[t] = & \Phi[1]x[t-1] + \Phi[2]x[t-2] + \dots + \Phi[p]x[t-p] - Q[1]a[t-1] - \\ & - Q[2]a[t-2] - \dots - Q[q]a[t-q] + a[t]. \end{aligned} \quad (1.2)$$

Для розрахунку параметрів моделі використовується система рівнянь, яка зв'язує значення кореляційної функції процесу АРКС та невідомі параметри моделі. Систему рівнянь можна отримати, якщо помножити ліву та праву частини (1.1) на  $x[t-j]$  і взяти математичне очікування. Після нескладних перетворень отримаємо

$$R[j] = \Phi[1]R[j-1] + \Phi[2]R[j-2] + \dots + \Phi[p]R[j-p] + \dots + R_{xa}[j] - Q[1]R_{xa}[j-1] - \dots - Q[q]R_{xa}[j-q], \quad (1.3)$$

де  $R_{xa}[j]$  – взаємна кореляційна функція процесів  $x[t-j]$  і  $a[t]$ , рівна

$$R_{xa}[j] = E\{x[t-j]a[t]\}. \quad (1.4)$$

Взаємна кореляційна функція не дорівнює нулю, якщо в рівнянні для  $x[t-j]$  не міститься  $a[t]$ . Як випливає з (1.4) и (1.2), взаємна кореляційна функція задовольняє співвідношенням

$$\begin{aligned} R_{xa}[j] &= 0, \quad j > 0, \\ R_{xa}[j] &\neq 0, \quad j \leq 0. \end{aligned} \quad (1.5)$$

З урахуванням (1.3), із (1.5) випливає, що

$$R[j] = \Phi[1]R[j-1] + \Phi[2]R[j-2] + \dots + \Phi[p]R[j-p], \quad j \geq q+1,$$

або

$$r[j] = \Phi[1]r[j-1] + \Phi[2]r[j-2] + \dots + \Phi[p]r[j-p], \quad j \geq q+1.$$

Вираз для дисперсії процесу АРКС можна знайти з (1.3), поклавши  $j = 0$

$$\begin{aligned} R[0] &= \Phi[1]R[1] + \Phi[2]R[2] + \dots + \Phi[p]R[p] - Q[1]R_{xa}[-1] - \\ &- Q[2]R_{xa}[-2] - \dots - Q[q]R_{xa}[-q] + D_a. \end{aligned} \quad (1.6)$$

Тоді вираз для коефіцієнтів кореляції процесу АРКС набуває вигляду

$$r[j] = \frac{\Phi[1]R[j-1] + \dots + \Phi[p]R[j-p] + R_{xa}[j] - Q[1]R_{xa}[j-1] - \dots - Q[q]R_{xa}[j-q]}{\Phi[1]R[1] + \Phi[2]R[2] + \dots + \Phi[p]R[p] - Q[1]R_{xa}[-1] - \dots - Q[q]R_{xa}[-q] + D_a}. \quad (1.7)$$

Можна показати, що СЦП моделі АРКС визначається виразом

$$P(f) = \frac{\left| 1 - \sum_{n=1}^q Q[n] e^{-j2\pi f n T} \right|^2}{\left| 1 - \sum_{i=1}^p \Phi[i] e^{-j2\pi f i T} \right|^2} D_a. \quad (1.8)$$

Як очевидно з (1.8), процеси АРКС описуються полюсно-нульовою СЦП. Спектр такого процесу містить вузькосмугові піки, що визначаються моделлю АР, та широкосмугові максимуми, що визначаються моделлю КС.

Як приклад розглянемо модель АРКС(1.1), яка описується різницеvim рівнянням

$$x[t] = \Phi[1]x[t-1] - Q[1]a[t-1] + a[t]. \quad (1.9)$$

З (1.3) і (1.6) випливають вирази для дисперсії та кореляційної функції

$$\begin{aligned} R[0] &= \Phi[1]R[1] - Q[1]R_{xa}[-1] + D_a, \\ R[1] &= \Phi[1]R[0] - Q[1]D_a, \\ R[k] &= \Phi[1]R[k-1], \quad k \geq 2. \end{aligned} \quad (1.10)$$

Значення  $R_{xa}[-1]$  можна знайти, помножив рівняння (1.1) на  $a[t-1]$  і взявши математичне очікування

$$R_{xx}[-1] = (\Phi[1] - Q[1])D_a. \quad (1.11)$$

Підставляючи (1.11) у (1.10), після перетворень, отримаємо

$$R[0] = \frac{1 + Q^2[1] - 2\Phi[1]Q[1]}{1 - \Phi^2[1]} D_a,$$

$$R[1] = \frac{(1 - \Phi[1]Q[1])(\Phi[1] - Q[1])}{1 - \Phi^2[1]} D_a, \quad (1.12)$$

$$R[k] = \Phi[1]R[k-1], \quad k \geq 2.$$

Для коефіцієнтів кореляції система рівнянь набуває вигляду

$$r[1] = \frac{(1 - \Phi[1]Q[1])(\Phi[1] - Q[1])}{1 + Q^2[1] - 2\Phi[1]Q[1]}, \quad (1.13)$$

$$r[2] = \Phi[1]r[1].$$

Система рівнянь (1.10) нелінійна, але у цьому випадку її можна точно вирішити щодо невідомих  $\Phi[1]$  і  $Q[1]$ .

Розглянемо модель АРКС(2.1), що описується рівнянням

$$x[t] = \Phi[1]x[t-1] + \Phi[2]x[t-2] - Q[1]a[t-1] + a[t]. \quad (1.14)$$

Із (1.3) випливає, що система рівнянь для значень кореляційної функції має вигляд

$$\begin{aligned}
R[0] &= \Phi[1]R[1] + \Phi[2]R[2] - Q[1]R_{xa}[-1] + D_a, \\
R[1] &= \Phi[1]R[0] + \Phi[2]R[1] - Q[1]D_a, \\
R[2] &= \Phi[1]R[1] + \Phi[2]R[0], \\
R[k] &= \Phi[1]R[k-1] + \Phi[2]R[k-2], \quad k \geq 3.
\end{aligned} \tag{1.15}$$

Взаємну кореляційну функцію  $R_{xa}[-1]$  можна знайти, помножив (1.1) на  $a[t-1]$  і взявши математичне очікування. Після перетворень система рівнянь (1.15) представляється як

$$\begin{aligned}
R[0] &= \frac{2Q[1]\Phi[1] + (1 + Q^2[1])(\Phi[2] - 1)}{1 - \Phi^2[1] - \Phi[2](1 + \Phi^2[1])} D_a, \\
R[1] &= \Phi[1]R[0] + \Phi[2]R[1] - Q[1]R[0] \frac{1 - \Phi^2[1] - \Phi[2](1 + \Phi^2[1])}{2Q[1]\Phi[1] + (1 + Q^2[1])(\Phi[2] - 1)}, \\
R[2] &= \Phi[1]R[1] + \Phi[2]R[0], \\
R[k] &= \Phi[1]R[k-1] + \Phi[2]R[k-2], \quad k \geq 3.
\end{aligned} \tag{1.16}$$

Як видно з (1.14), навіть прості моделі АРКС описуються досить складною системою рівнянь, що включає нелінійні рівняння. Щоб визначити коефіцієнти АР і КС використовуються чисельні методи. Якщо процедура розв'язання системи нелінійних рівнянь небажана, можна використовувати апроксимацію АРКС моделлю АР високого порядку [1].

Модель АРКС може використовуватися для синтезу відбілюючого фільтра. Різницеве операторне рівняння рекурсивного відбілюючого АРКС фільтра, визначається виразом

$$a[t] = H(z)x[t] = \frac{\Phi(z)}{Q(z)}x[t]. \tag{1.17}$$

Звідси вираз для системної функції фільтра має вигляд

$$H(z) = \frac{\Phi(z)}{Q(z)}. \quad (1.18)$$

Формуючий фільтр АРКС описується різницеvim рівнянням в операторній формі

$$x[t] = H(z)a[t] = \frac{Q(z)}{\Phi(z)}a[t]. \quad (1.19)$$

Звідси випливає, що його системна функція представляється у вигляді

$$H(z) = \frac{Q(z)}{\Phi(z)}. \quad (1.20)$$

Передаточна функція формуючого фільтра містить вузькосмугові та широкосмугові складові. Фільтр стійкий, якщо коріння характеристичного рівняння лежить усередині одиничного кола комплексної площині. Умова стаціонарності процесу АРКС збігається з умовою стаціонарності процесу АР, а його зворотність – з умовою зворотності КС процесу.

### 1.3 Висновки розділу та постанови задач

У кваліфікаційній роботі магістра аналізуються стаціонарні та нестаціонарні імітаційні моделі випадкових процесів. Перевага у дослідженні надається нестаціонарним процесам, що описуються моделями декомпозиції та АRІМА. Найбільш складними процесами вважаються нестаціонарні випадкові процеси, модель яких не завжди дає можливість передбачити її майбутню

поведінку. З ряду причин корисними виявляються лише деякі нестационарні статистичні моделі.

У роботі аналізуються моделі нестационарних процесів, які включають в себе складові стаціонарних АР процесів. Нестационарні дані, що розглядаються, відрізняються за величиною і типом тренду, сезонною складовою, моделями стаціонарних складових. Таким чином, у загальному випадку, розглянута модель нестационарного випадкового процесу складається з адитивної суми тренду, сезонної складової та моделі стаціонарної статистичної складової. Модель декомпозиції, справедлива деяких класів нестационарних випадкових процесів, будується з допомогою операторів взяття різниці. Це так звана модель ARIMA.

Операція взяття різниці це дискретне диференціювання. Тому у випадковому процесі, тільки з деякими видами тренда, шляхом взяття різниці, можна позбутися тренда. Замість дослідження самого процесу доводиться досліджувати його похідну, бо взяття різниці у дискретному випадку, еквівалентне взяття дискретної похідної. У такий спосіб аналізуються процеси з трендом, у яких властивості похідної не сильно відрізняються від властивостей самого процесу.

Від сезонної складової позбавляються тим краще, чим ближче амплітуди та періоди коливань до детермінованого процесу. Тоді взяття різниці через інтервал часу, що дорівнює періоду сезонної складової, дає облік сезонної складової.

У кваліфікаційній роботі магістра розглянуто питання щодо побудови нестационарної моделі ARIMA. Ця модель будується як декомпозиція складових із використанням операторів взяття різниць. Як модель стаціонарної складової використовується АРКС. Таким чином, модель ARIMA можна вважати узагальненням моделі АРКС на нестационарний випадок. У роботі показані переваги та недоліки цієї моделі. Компенсація тренду та сезонної складових можливі в разі їх детермінованості та суворої періодичності. У разі нездійсненності цих вимог до моделі ARIMA, помилки методу взяття різниць

враховуються у моделі стаціонарної складової АРКС. Оцінка параметрів моделі може здійснюватись методом максимальної правдоподібності. Показано, що модель ARIMA може успішно використовуватись для пошуку транзитивних планет. Вони визначаються змінним світінням деяких зірок. Транзитивні планети знаходяться поза сонячною системою.

Проведений огляд наукової літератури на тему кваліфікаційної роботи дозволяє вирішувати такі завдання:

1. Проаналізувати використання стаціонарних та нестаціонарних процесів у статистичній радіотехніці для нестаціонарних моделей змінної складової.
2. Для часових рядів, що описують зміни реальних процесів, будується модель ARIMA.
3. Знайти імітаційні моделі нестаціонарних процесів, що описуються процесами ARIMA з трендом, сезонною та стаціонарною складовими.
4. Дослідити завдання отримання імітаційних моделей стаціонарних складових для гаусових та негаусових породжуючих процесів.

## 2 МОДЕЛІ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

### 2.1 Модель авторегресії проінтегрованого ковзного середнього нестационарного процесу

Для моделі нестационарного процесу, декомпозицію на її складові, можна розглядати як подання нестационарного процесу у вигляді адитивної суми

$$\omega[t] = m[t] + c[t] + y[t], \quad (2.1)$$

де  $\omega[t]$  – нестационарний процес,  $m[t]$ ,  $c[t]$ ,  $y[t]$  – тренд, сезонна складова, корельована випадкова стаціонарна складова [17]. Зауважимо, що стаціонарна складова змінюється після видалення тренду та сезонної складової. Вона починає враховувати відмінність тренду та сезонної складової від невіддаленого до кінця детермінованого тренду та некомпенсованої сезонної складової операторами взяття різниці. У стаціонарний залишок  $y[t]$  включається стаціонарний корельований процес, що спочатку присутній у моделі декомпозиції (2.1). При видаленні тренду, сезонної складової, представлених у моделі декомпозиції (2.1), стаціонарний залишок  $y[t]$  представляється у випадку моделлю АРКС. Вузкосмугова частина процесу залишку описується моделлю АР, а широкосмугова частина цього процесу моделюється ковзним середнім (КС).

Тренд – це нестационарний процес, що описується різними видами функцій: лінійними, квадратичними, логарифмічними, степінними та іншими. У нестационарного випадкового процесу тренд часто описується довільною функцією. Крім функції, яка описує тренд, потрібно задати деякі параметри, що визначають швидкість зміни цих функцій. Тому опис тренду за допомогою відповідних функцій не завжди є прийнятним.

Емпіричне знаходження тренду нестационарного випадкового процесу часто не означає, що він повторюватиметься надалі. Це одна з причин суттєвих помилок нестационарних моделей. Пророцтво в цьому випадку можливе тільки на однорідних ділянках тренду. Там, де тренд не змінює своїх характеристик.

З деякими припущеннями, можна вважати, що тренд визначає зміну математичного очікування випадкового нестационарного процесу. Наприклад, лінійний тренд нестационарного процесу може з часом залишатися лінійним, але змінювати кут нахилу. Тому успішне передбачення на деякій ділянці, навіть при правильному визначенні виду тренду, не означає, що на інших ділянках часового ряду воно також буде вірним.

У моделі ARIMA існує сильне обмеження. Трендова складова в цій моделі видаляється за допомогою відповідного оператора віднімання  $\nabla^d = (1 - z)^d$ . Це можливо, якщо тренд представляється лінійною, квадратичною, кубічною функцією. Процес без тренда можна подати у вигляді  $\omega_1[t]$ . Зазвичай, обмежуються найпростішими випадками лінійного та квадратичного трендів. Для цих випадків  $d = 1$ ,  $d = 2$  відповідно. Тоді

$$\omega_1[t] = \nabla \omega[t] = \omega[t] - \omega[t-1], \quad (2.2a)$$

$$\omega_1[t] = \nabla^2 \omega[t] = \nabla(\omega[t] - \omega[t-1]) = (\omega[t] - 2\omega[t-1] + \omega[t-2]). \quad (2.2b)$$

Дію операторів взяття різниці можна уявити в рівнянні

$$x[t] = \nabla^d \nabla_s \omega[t] = (1 - z)^d (1 - z^s) \omega[t], \quad (2.3)$$

де  $x[t]$  – стаціонарний процес без тренду та сезонної складової,  $\Delta_s = (1 - z^s)$  – різницевий оператор, що видаляє сезонну складову,  $\nabla^d = (1 - z)^d$  – різницевий оператор, що видаляє тренд,  $d$  – постійна, яка залежить від виду тренду,  $z^s$  – оператор зсуву назад,  $s$  – період сезонної складової. Рівняння дозволяє

розраховувати параметри моделі мультиплікативно. Залишковий стаціонарний процес без тренда та сезонної складової описуватимемо стаціонарною моделлю АР. Як очевидно з (2.3), модель АRІМА [2] нестаціонарного процесу  $x[t]$  представляється як мультиплікативний процес.

Щоб видалити сезонну складову в рівнянні (2.1), застосовується оператор  $\nabla_s = 1 - z^s$ , де  $z^s$  – оператор зсуву, дія якого визначається виразом  $z^s x[t] = x[t - s]$ . Видалення сезонної складової за допомогою оператора  $\nabla_s = 1 - z^s$  описується виразом

$$x[t] = \nabla_s \omega_1[t] = (1 - z^s) \omega_1[t]. \quad (2.4)$$

де випадкова змінна  $\omega_1[t]$  містить тренд та не містить сезонну складову.

У моделі АRІМА сезонна складова видаляється оператором  $\nabla_s = 1 - z^s$ , а різницевий оператор, що видаляє тренд, має вигляд  $\Delta^d = (1 - z)^d$ . Декомпозицію (2.1) моделі нестаціонарного процесу АRІМА, при дії операторів взяття різниці, можна записати у вигляді

$$\Delta^d \nabla_s \omega[t] = m[t] - m[t - s] + c[t] - c[t - s] + y[t] - y[t - s]. \quad (2.5)$$

Якщо сезонна складова суворо детермінована, то  $c[t] = c[t - s]$  та декомпозиція (2.5) спрощується [17]

$$\omega_1[t] = m[t] - m[t - s] + y[t] - y[t - s]. \quad (2.6)$$

Однак сезонна складова не детермінована і тому вираз (2.6) не зовсім точно видаляє сезонну складову цих даних.

Детермінований тренд враховується в  $\omega_1[t]$  оператором взяття різниці  $\Delta^d = (1 - z)^d$ . Видалення тренду можна подати у вигляді

$$x[t] = (1-z)^d \omega_1[t]. \quad (2.7)$$

Нижче вважатимемо, що тренд описується найпростішими випадками: лінійним і квадратичним законом.

Прикладом нестационарного процесу може бути трафік передачі даних у стільниковому зв'язку, представлений на рис. 2.1. Як очевидно з графіка, трафік згодом зростає, так як має тренд. У нестационарних процесах, залежно від ділянки спостереження, спостерігається лінійний тренд [18, 19], так і нелінійний. Лінійна швидкість його зростання різна.

На тлі тренду спостерігаються сезонні коливання трафіку. Період коливань становить у середньому тиждень. Тобто така циклічність процесу. Протягом тижня трафік то зростає, то зменшується. Проте складно сказати, що така періодичність сезонного процесу є стійкою і завжди справедливою. Це позначається на точності прогнозу процесу.

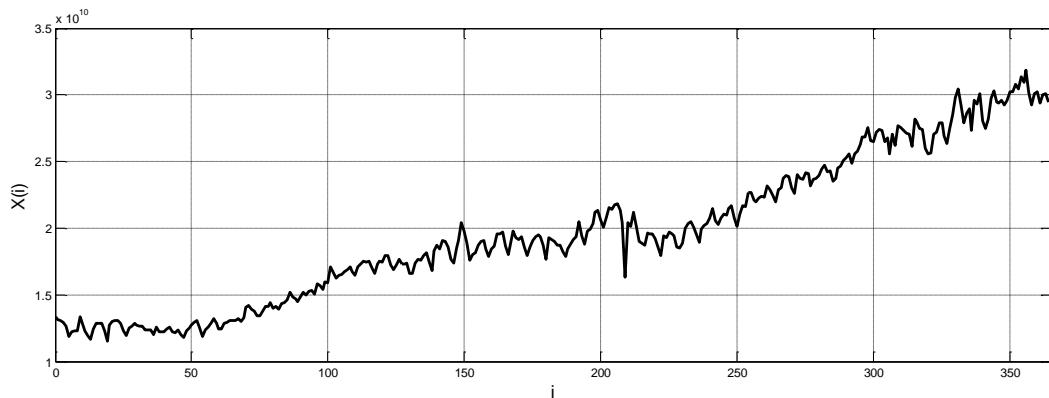


Рисунок 2.1 – Дані трафіку передачі в стільниковому зв'язку

Прикладом нестационарного процесу також може бути зміна температур, що фіксуються метеообсерваторією. Циклічні зміни температур рік у рік, мають слабкі тренди, які найцікавіші деяким дослідникам. Але слід зазначити, що вони залежать від різних факторів, пов'язаних як з розташованою до місця вимірювання температури промисловістю, так і впливом теплих течій Ель-Ніньо, що виникають в океані, зміною складу атмосфери тощо. Існує безліч

факторів, що впливають на температуру, що призводить не тільки до вікового зростання тренду температур, а й навіть до його падіння.

На рис. 2.2 показаний нестационарний процес, пов'язаний із періодичною зміною температур, зафіксований метеообсерваторією. Інтерес представляє також довгостроковий тренд. Його можна отримати згладжуванням процесу. Як видно з графіка, знайдені тренди мають довільну форму, характеризуються різними статистичними залежностями (лінійними, параболічними та ін.). У цілому тренди неоднорідні, так як у них є ділянки з різною швидкістю зміни. Аналіз тренду на рис. 2.2 показує, що з 1752 до 1803 року температура падала з  $4,05^{\circ}\text{C}$  до  $3,45^{\circ}\text{C}$ . Спад температури склав  $0,6^{\circ}\text{C}$ . З 1804 року спостерігалось зростання температури з  $3,45^{\circ}\text{C}$  до  $5,96^{\circ}\text{C}$  у 2010 році. Приріст температури за цей термін становив  $2,51^{\circ}\text{C}$ .

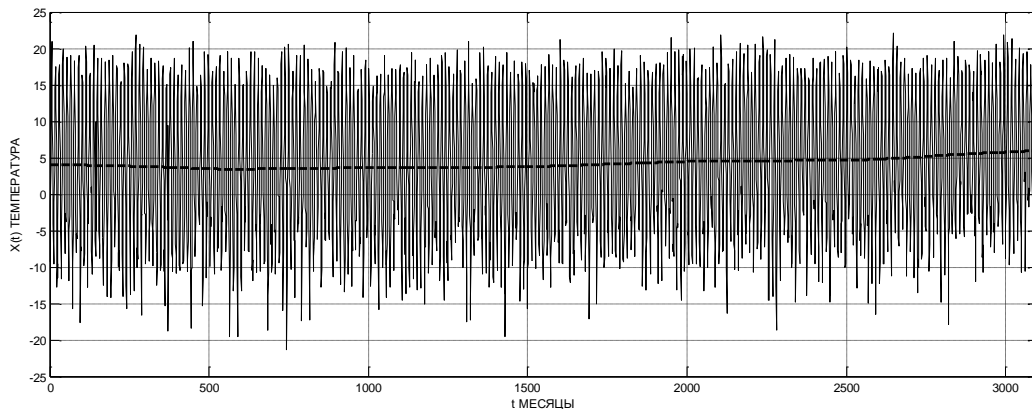


Рисунок 2.2 – Дані середньомісячних температур, зафіксованих у метеообсерваторії міста Пулково

Оператори взяття різниці, що входять до (2.3), враховують сезонну складову та тренд шляхом їх віднімання. Якщо тренд або сезонна складова видаляється оператором взяття різниці, цей оператор добре враховує ці складові. Зауважимо, що за такого підходу до побудови моделі ARIMA нестационарного процесу  $\omega[t]$  випадковий процес  $x[t]$  можна вважати стаціонарним. Навіть якщо оператори видалення сезонної складової та тренду не повністю їх враховують.

Конструктивна модель (2.3), що дозволяє безпосередньо визначити її параметри, є мультиплікативною. Найприроднішою є адитивна модель нестационарного процесу. Такою моделлю є декомпозиція нестационарного процесу (2.1). Вона представляє випадковий нестационарний процес у вигляді його складових [17]. Однак ця модель не дає способів розрахунку складових нестационарного процесу. Деякі автори для розрахунку параметрів цієї моделі використовують методи, які впливають з моделі декомпозиції.

## 2.2 Загальні положення теорії виявлення екзопланет

В астрофізиці часто буває корисною модель AutoRegressive Fractional Integrated Moving Average (ARFIMA) [15] із дробовим  $d$ . У цьому випадку  $d$  є дробом, а не цілим числом. Цей процес називається «фракційно інтегрованим» АРКС, скорочено ARFIMA або FARIMA. Коли  $0 < d < 0,5$ , результатом є стаціонарна автокореляція з довгою пам'яттю зі степінною поведінкою як в автокореляційній функції. Варіант, який представляє інтерес для астрофізики – це коли  $d$  є дробом, а не цілим числом, а автокореляційні функції залежать від зрушень  $K^{2d-1}$ . Процес ARFIMA, можна вважати нестационарним при  $d > 0,5$ . Параметр ARFIMA  $d$  арифметично пов'язаний з  $\alpha$  у компоненті «червоного шуму» фізиків  $1/f^\alpha$ , де  $\alpha = 2d$ , і пов'язаний з параметром Херста в економетриці,  $H = d + 0,5$ .

Математично глибоким підґрунтям ефективності моделей типу ARMA є теорема розкладання Волда, яка стверджує, що будь-який часовий ряд може бути розкладений на детерміновану частину плюс нескінченну суму (загалом), стаціонарних процесів АРКС. Модель ARIMA ( $p, d, q$ ) є економним наближенням до цього розкладання.

Моделювання перешкод за методом ARIMA може зменшити автокореляцію та шум на кривій блиску зірок, які не вдалося усунути лише за допомогою оператора різниці. ARIMA та ARFIMA дуже ефективні у зменшенні

непланетної зоряної змінності для великої частини зірок, знайдених за допомогою телескопа Кеплера.

Це відповідає широким можливостям моделей ARIMA для моделювання та прогнозування часових рядів в інших областях. Моделі ARIMA широко застосовуються для моделювання дорожньо-транспортних пригод, цін на акції, коливань вітру та погоди, сонячної активності та інших випадкових процесів у людських та фізичних системах. Моделі ARFIMA використовуються рідше, але застосовуються для моделювання інфляції та інших макроекономічних показників біржі, нестійкого серцебиття у кардіології та різних ситуацій прогнозування.

Існує кілька методів виявлення екзопланет. Наприклад, транзитний метод виявлення планет поза нашою Сонячною системою. Цей метод заснований на спостереженні зменшення світності зірки при проходженні на її тлі транзитивної планети. В останні роки, особливо коли ширше стали використовуватися високогірні телескопи і супутники, почали відкривати планети поза Сонячною системою. Виявляються лише ті планети, орбіта яких проходить на тлі зірки. Цей метод дозволяє визначити розміри планети, а у поєднанні з методом Доплера – густину планети. Він також дає інформацію про наявність атмосфери та її склад. Зрозуміло, що цим методом можна виявити ті планети, орбіта яких лежить в одній площині з точкою спостереження.

Планети виявляються шляхом аналізу світності, що змінюється по спектру зірки. Виявлення періодичних сигналів від транзитних екзопланет, часто відбувається при аперіодичній змінності яскравості зірки, яка властива зірці, або виникає в процесі вимірювань. Зміни світності зірки є автокорельованими. При цьому пізніші значення миттєвих значень світності зірки корелюють із попередніми.

У кваліфікаційній роботі магістра представлена методологія проекту авторегресійного пошуку планет, названа ARPS, яка базується на авторегресійній моделі даних і використовує відому модель нестационарного процесу ARIMA і пов'язані з нею статистичні моделі [15]. Метод, що

використовується, розглядає широкий ряд випадкових процесів, використовує їх нестаціонарність для поліпшення виявлення нових планетарних транзитів.

Для пошуку транзитів на часових рядах, що представляють криві блиску зірок, виявився дуже успішним підходом до виявлення екзопланет. Внаслідок цього підходу на сьогоднішній день з'явилися кілька тисяч кандидатів у екзопланети.

Однією з основних проблем, із якими стикається телескоп Кеплера та інші дослідження виявлення планет із космосу, є шум і мінливість зірок. Вони відбуваються разом із коливаннями світності зірок із великими амплітудами і мають іншу природу. Їх можна порівняти зі зміною блиску зірок через переміщення планет, але часто перевищує зміна глибини блиску зірки за рахунок планетарних транзитів. Істотний вплив на аналіз вимірів відіграє аперіодичний «червоний шум», який особливо поширений при виявленні планет.

Наземні дослідження стикаються з аналогічними проблемами. Але причина мінливості яскравості зірки інша. Вона виникає переважно через зміни інструментальних та атмосферних умов вимірювання світності. Кількість знайдених транзитивних планет значно збільшиться за допомогою нових статистичних методів обробки даних, здатних відновлювати слабші планетарні сигнали.

Зазначимо, що зміни зоряної яскравості можуть виникати внаслідок затемнення, пульсації та інших явищ, що відбуваються з зірками. Однак найчастіше вона обумовлена магнітною активністю, включаючи фотосферні зоряні плями, хромосферні плями та спалахів при перез'єднанні.

Сонячна та зоряна активності часто обумовлені авторегресійною закономірністю, за якої майбутні фотометричні значення залежать від поточних та минулих значень. Так, сонячні та зоряні спалахи моделюються як «лавинні» процеси, які пов'язані з наявністю шуму типу  $1/f$ . Вони відповідальні за степінний статистичний розподіл властивостей сонячної активності. Як буде

показано нижче, зоряна світність є процесом із «довгою пам'яттю»  $1/f$ -типу. Його можна розглядати за допомогою певних авторегресійних моделей.

Кореляційні характеристики зоряної фотометричної «корельованої перешкоди» призводять до ідеї, що можуть бути застосовані стохастичні авторегресійні статистичні моделі-фільтри для зменшення їх впливу. Це також допоможе виявити слабкі періодичні планетарні транзити. І тому використовується стаціонарна модель АРКС. Вона широко використовується для визначення характеристик випадкового процесу, при обробці сигналів, розпізнаванні голосу та економетриці, а також у багатьох інших програмах.

Модель АРКС використовується для боротьби із корельованими перешкодами. Модель нестационарних процесів АRІMA може зменшити автокореляцію та шум на кривій блиску Кеплера, які не вдалося усунути за допомогою оператора різниці. Виявлено, що моделі АRІMA та АRFIМА дуже ефективні у зменшенні зоряної змінності для великої частини зірок, виявленої телескопом Кеплера. Це відповідає успіхам моделей АRІMA для моделювання та прогнозування широкого спектру часових рядів в інших областях. Моделі АRІMA застосовуються для моделювання дорожньо-транспортних пригод, цін на акції, коливань вітру та погоди, сонячного випромінювання та незліченних інших випадкових процесів у людських та фізичних системах. Моделі АRFIМА використовуються рідше. Вони застосовуються для моделювання інфляції та інших макроекономічних показників, нестійкого серцебиття у кардіології та різних ситуацій прогнозування.

В астрофізиці нестационарними можуть бути процеси, пов'язані з життям зірок, а також періодичність світності зірок, що виникає через транзитивний рух планет. Часовий ряд є стаціонарним, якщо його властивості не змінюються з часом. Тому їх статистичні характеристики, такі як середнє значення, дисперсія та автокореляційні функції однаково незалежно від того, в який час він спостерігається. У нестационарних процесів в астрофізиці ці показники залежатимуть від того, де за часом вони спостерігаються. Формально сильно стаціонарні процеси мають постійний спільний розподіл ймовірностей. Моделі

ARMA використовуються для слабких стаціонарних процесів. Перші два моменти – середнє значення та автоковаріація, що використовуються в моделі, залишаються приблизно постійними у часі, а кореляційна функція залежить лише від часу зрушень спостережень.

Процеси нестационарні, якщо є тренд даних. Тренди можуть бути різними. У процесі може бути тренд середнього значення. При цьому зберігається деяка однорідність процесу. Така нестационарність процесу описується одиничним корінням характеристичного рівняння. У разі одиничного кореня нестационарність можна усунути, якщо застосувати різницевий оператор до процесу.

Нестационарність типу аперіодичності може виникати у разі обертання зірки або спостереження проходження планетарної орбіти на тлі зірки. Нестационарні процеси, що описуються моделлю ARMA, часто створюють квазіперіодичність, яка змінюється або зникає за тривалих часових масштабів. Процеси з довгою пам'яттю, що описуються функцією  $1/f^\alpha$  можуть бути стаціонарними чи нестационарними залежно від величини  $\alpha$ .

Оператор взяття різниці призначений для усунення нестационарності. Для моделей лінійного передбачення характерно безліч інтерпретацій. Так, цей оператор, що виконує операцію диференціювання, також можна розглядати як фільтр верхніх частот. Він видаляє довгомасштабні зміни, залишаючи при цьому короткострокові коливання. Після взяття різниці, нову серію відліків даних, можна моделювати за допомогою методів авторегресії, докладно описаних у розділі 1.

Диференціювання рядів було запропоновано деякими авторами лише для усунення деяких видів нестационарності з трендом. Як показує практика, застосування різницевого оператора до нестационарного часового ряду буває достатнім, щоб зробити його стаціонарним. Зауважимо, що диференціювання або іншими словами взяття різниці, може підвищити рівень шуму, якщо вихідний часовий ряд є гаусівським білим шумом.

Модель нестационарного процесу ARIMA ефективна для отримання характеристик процесу зоряної світності, і навіть видалення, небажаних у разі, астрофізичних змінних зірки. Стаціонарна модель AP може використовуватись для пошуку слабких сигналів, пов'язаних із транзитом планет на тлі зірки. Їх пошук здійснюється після аналізу стаціонарних залишків моделі ARIMA. Транзити планет, на відміну інших зіркових явищ, суворо періодичні і виявляються на основі спектрального аналізу залишків нестационарної моделі. Залишками є стаціонарні процеси, які залишаються після обліку дії операторів віднімання, які усувають тренд та сезонну складову. Вони кількісно визначають зміни інтенсивності випромінювання зірок.

Використовувані методи виявлення планет поза сонячною системою аналізують періодограми та алгоритми, отримані за допомогою перетворення Фур'є. До них відносяться періодограма Lomb-Scargle (LS) або алгоритм узгоджених фільтрів, таких як алгоритм *boxfitting least-squares* (BLS) [15].

Пошук планет за допомогою транзитного гребінчастого фільтра (TCF), як і у випадку з іншими періодограмами, ускладнена. Він здійснюється за оцінкою частоти помилкових спрацьовувань піків на періодограмі. Виняток становлять нереалістичні ідеальні ситуації. Аналітичні розрахунки рівнів значущості періодограми ненадійні для LS-періодограм нерівномірно розташованих даних і навіть для періодограм Фур'є з регулярно розташованими даними.

При практичному використанні періодограми методом TCF, було виявлено, що розподіл потужності без періодичного сигналу є негаусовим зі змінним середнім значенням. На періодограмах даних, отриманих телескопом Кеплера, спостерігається нелінійний тренд. Усунути цей тренд можна за допомогою плавної локальної регресії, що відповідає періодограмі TCF. Для цього використовується алгоритм LOESS, що добре зарекомендував себе. Подібна поведінка спостерігається в періодограмах кривих блиску Кеплера, аналізованих МНК.

Пік періодограми можна оцінити, знайшовши граничну ймовірність параметра глибини транзиту  $A$  [15]. Так само оцінюється ставлення

сигнал/шум. Величину  $A$  можна знайти з фільтра і легко визначити як функцію інших параметрів. Для заданих параметрів  $p$ ,  $\varphi$  і  $d$ , величина  $A$  лінійно залежить від періоду, фази та тривалості  $d$ . Вона повністю описує розташування кожного фільтра.

Для цього використовують нестационарну модель ARIMA. Вона включає просту прямокутну модель транзиту планет, що відповідає оптимальному періоду TCF. Завдання, що розглядається, нагадує ситуацію з аналогічною економетричною задачею моделювання роздрібних продажів зі стохастичними характеристиками авторегресії, або з детермінованим тижневим циклом. У ній період і фаза відомі, а амплітуда циклічної складової невідома. Таку статистичну модель часто називають ARIMAX для ARIMA з «екзогенними» змінними або «динамічною регресією» [15].

Ідея, що використовується, формалізується шляхом визначення індикаторної функції  $I(t)$ , яка дорівнює 1, якщо визначені  $p$ ,  $\varphi$  і  $d$ . В іншому випадку функція дорівнює 0. Значення 1 та 0 оцінюються за оптимальним періодом періодограми TCF. Для моделі авторегресії даних, розрахованої для цієї індикаторної функції, відповідний коефіцієнт лінійної регресії, дорівнює «глибині ящика». Тобто

$$x[t] = \beta_0 + \beta_1 I[t] + a[t], \quad (2.8)$$

де  $\beta_0$  – відповідає глобальному середньому значенню,  $\beta_1$  – коефіцієнт індикаторної функції, а  $a[t]$  – нормальна помилка передбачення. Значення  $\beta_1$  являє собою середнє усунення під час транзиту планети.

Наведена детермінована регресія відповідає стохастичній моделі АРКС при відмові від обмеження, що помилка передбачення являє собою білий гаусівський шум. Натомість є автокорельований процес ARIMA. Модель ієрархічної регресії з помилками АРКС (тимчасово ігноруючи різницевий оператор для спрощення запису) тоді виглядає так:

$$x[t] = \beta_0 + \beta_1 I[t] + \eta[t], \quad (2.9)$$

$$\eta[t] = \sum_{i=1}^p \Phi[i]x[t-i] - \sum_{q=1}^q Q[j]a[t-j] + a[t], \quad (2.10)$$

де  $a[t]$  – нормальна помилка передбачення. Таким чином для оцінки ставлення сигналу до шуму, ми обчислюємо значення максимальної правдоподібності в моделі, оцінюємо значення параметра  $\beta_1$  з моделі лінійної регресії та обчислюємо відношення сигнал-шум.

Використання моделі ARIMAX може бути складним, у порівнянні з більш простою моделлю ARIMA. Однак показано, що вона може виявитися ціннішою для оцінки значущості піку методом періодограми TCF. У додатку цього методу оцінки ставлення сигнал-шум, до даних телескопа Кеплера, оцінка A при використанні моделі ARIMAX виявилася найбільш важливою «функцією» в класифікаторі машинного навчання. Особливо для ідентифікації планет кандидатів у транзити та відокремлення їх від помилкових результатів, таких як змішані подвійні зірки.

## 3 ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ СТАЦІОНАРНИХ І НЕСТАЦІОНАРНИХ ПРОЦЕСІВ НА ОСНОВІ АВТОРЕГРЕСІЇ ТА МОДЕЛІ ARIMA

### 3.1 Декомпозиція нестационарного процесу

Тут досліджено питання створення випадкових процесів. Для стаціонарних процесів вихідними даними були параметри спектрів. Імітаційні нестационарні процеси генерувалися згідно з моделлю декомпозиції. Складовими процесами моделі декомпозиції були лінійний тренд, сезонна і стаціонарна складові. Для генерації імітаційних стаціонарних процесів використовувалися моделі авторегресії першого – четвертого порядків. порушено також питання генерації нестационарних випадкових процесів із квадратичним трендом.

Відліки часового ряду пов'язані статистично. Цілями досліджень часових рядів є вирішення завдань математичної статистики. Для випадкових процесів можна проводити контрольні вимірювання характеристик тимчасового ряду, вирішувати завдання виявлення фізичного явища, робити прогноз, виходячи з минулих історичних даних, управляти процесом, досліджувати спектри та інші характеристики процесу. Можна спробувати з'ясувати механізм, що лежить в основі процесу, усунути деякі складові нестационарного процесу, вирішувати завдання декомпозиції випадкового процесу.

У цій роботі аналізується генерація нестационарних моделей випадкового процесу. Найбільшу популярність мають моделі нестационарних процесів, що допускають лінійне розкладання на адитивні складові. Ця модель більше відома як декомпозиція нестационарного процесу на складові. Нестационарна модель сприймається як уявлення процесу адитивної суми, тобто декомпозиції (2.1).

Наведені приклади реального нестационарного процесу показують, що ці процеси мають тренд, сезонна складова, і навіть стаціонарний процес. Залежно від виду нестационарного процесу роль стаціонарної складової може сильно змінюватися. Мається на увазі її важливість при дослідженні та вплив на

формування процесу. Якщо для стаціонарного процесу, його статистична модель відіграє вирішальну роль, то у разі нестаціонарного процесу, важливе значення набуває сезонна складова та тренд.

Як очевидно з графіка, представлених на рис. 2.1 та рис. 2.2 тренд має довільну форму. Він допускає лінійне представлення. Тренд нестаціонарного процесу описується різними функціями: лінійними, квадратичними, логарифмічними, степінними та іншими. У нестаціонарного процесу вони часто мають вигляд довільної функції. Емпірично знайдений тренд випадкового процесу часто не означає, що він повторюватиметься надалі. Це є однією з причин суттєвих помилок моделей нестаціонарних процесів.

З деякими припущеннями, можна вважати, що тренд визначає зміну математичного очікування випадкового нестаціонарного процесу. Часто тренд має вигляд деякої функції, але з різною швидкістю зміни. Наприклад, лінійний тренд нестаціонарного процесу може з часом залишатися лінійним, але змінювати кут нахилу. Це ще одна причина, чому правильне пророцтво на деякій ділянці не означає, що на інших ділянках часового ряду воно також буде успішним.

Приклад лінійного тренду, швидкість зміни якого стала, представлений на рис. 3.1. Цей лінійний тренд буде використаний нижче для отримання нестаціонарних процесів із трендом.

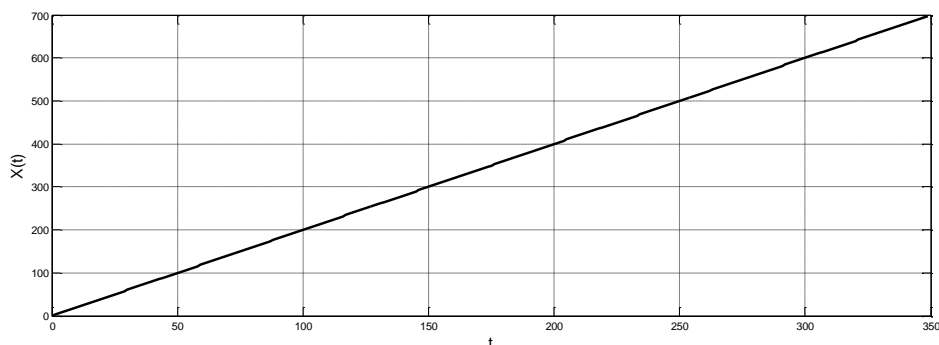


Рисунок 3.1 – Лінійний тренд імітаційного процесу

Сезонна складова характеризує циклічні процеси. Чим точніше у них повторюється період і чим ближче у них миттєві значення через період, тим ближчі вони до детермінованих процесів. У цьому випадку вони добре описуються методами нестационарних процесів, наприклад оператором взяття різниці.

У деяких моделях нестационарних процесів відсутні або тренд або сезонна складова. У багатьох нестационарних процесів вони присутні. У деяких випадках вони відіграють важливу роль.

На рис. 3.2 представлена сезонна складова у вигляді дискретної синусоїди. Дія дискретної сезонної складової на стаціонарний процес у вигляді моделі АР змінює як період сезонної складової, так і вносить зміни до її миттєвих значень, амплітуди сезонних коливань. Наведена на рис. 3.2 сезонна складова, що використовується нижче для отримання імітаційного нестационарного процесу.

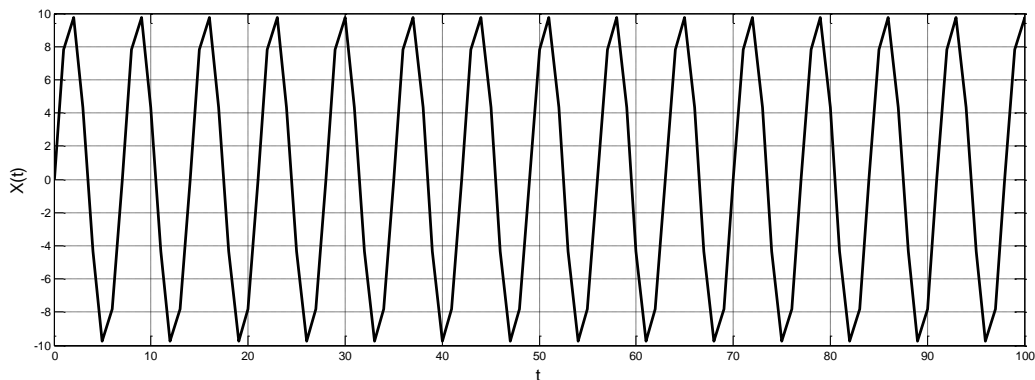


Рисунок 3.2 – Ділянка сезонної складової процесу з періодом 7 відліків

### 3.2 Імітаційний стаціонарний процес авторегресії

Для отримання стаціонарної складової імітаційного нестационарного процесу найбільш успішною є модель АР. Саме ця модель дозволяє найбільш гнучко отримувати стаціонарні випадкові процеси із заданими характеристиками.

Вузькосмуговий стаціонарний процес АР описується рівнянням [2]

$$x[t] = \sum_{j=1}^p \Phi[j]x[t-j] + a[t], \quad (3.1)$$

де  $\Phi[j]$  – коефіцієнти АР,  $a[t]$  – некорельовані випадкові відліки,  $p$  – порядок моделі АР. В операторній формі модель АР( $p$ ), що описується рівнянням (3.1), може бути представлена у вигляді

$$\Phi(z)x[t] = a[t], \quad (3.2)$$

де в (3.2) оператор АР  $\Phi(z)$  введено наступним чином

$$\Phi(z) = 1 - \Phi[1]z^{-1} - \dots - \Phi[p]z^{-p}. \quad (3.3)$$

Дія оператора зсуву  $z^{-i}$  на поточний відлік  $x[t]$  описується виразом

$$z^{-i}x[t] = x[t-i]. \quad (3.4)$$

Характеристичне рівняння АР можна записати в такий спосіб

$$c^p - \Phi[1]c^{p-1} - \dots - \Phi[p] = \prod_{i=1}^p (c - c[i]) = 0. \quad (3.5)$$

Порівнюючи в (3.5) коефіцієнти при однакових степенях знайдемо зв'язок між коефіцієнтами АР та коріннями  $c[i]$ . Впливає, що коефіцієнти АР пов'язані з корнями характеристичного рівняння [2]. Наведемо деякі формули для  $p = 1 \div 4$ :

$$\begin{aligned}
\Phi[1,1] &= \pm c[1], \\
\Phi[2,1] &= c[2,1] + c[2,2], \\
\Phi[2,2] &= -c[2,1]c[2,2], \\
\Phi[3,1] &= c[3,1] + c[3,2] + c[3,3], \\
\Phi[3,2] &= -(c[3,1]c[3,2] + c[3,1]c[3,3] + c[3,2]c[3,3]), \\
\Phi[3,3] &= c[3,1]c[3,2]c[3,3], \\
\Phi[4,1] &= c[4,1] + c[4,2] + c[4,3] + c[4,4], \\
\Phi[4,2] &= -(c[4,3]c[4,4] + c[4,2]c[4,3] + c[4,1]c[4,3] + c[4,4]c[4,1] + c[4,2]c[4,3] + c[4,2]c[4,4]), \\
\Phi[4,3] &= c[4,1]c[4,3]c[4,4] + c[4,2]c[4,3]c[4,4] + c[4,1]c[4,2]c[4,3] + c[4,1]c[4,2]c[4,4], \\
\Phi[4,4] &= -c[4,1]c[4,2]c[4,3]c[4,4].
\end{aligned} \tag{3.6}$$

У цих виразах перший індекс у квадратних дужках означає порядок моделі, для яких розраховується коефіцієнт АР.

Корені характеристичного рівняння задаються центральною частотою (частотою моди) та її шириною смуги [18]. Для цього використовується зв'язок між корнями характеристичного рівняння та параметрами спектру: частотами мод  $f_i$  та їх ширинами смуг  $\Delta f_i$  СЦП. Тоді дійсні корні характеристичного рівняння  $p$ -го порядку та параметри АР, пов'язані співвідношеннями (3.6), виражаються через ширину смуги моди  $\Delta f_i$

$$c[i] = e^{-\pi \Delta f_i T}. \tag{3.7a}$$

Дійсні корні описують моди на нульовій частоті або на максимальній частоті дискретного спектру, що дорівнює половині частоти дискретизації.

Так як модель АР дійсна, то комплексно-сполучене корні характеристичного рівняння описується виразами [18]

$$c[i] = e^{-\pi \Delta f_i T - j 2\pi f_i T}, \quad c[i+1] = e^{-\pi \Delta f_i T + j 2\pi f_i T}. \tag{3.7b}$$

Кожній  $i$ -й моді відповідають два корені характеристичного рівняння. У найпростіших випадках вузькосмугових СЦП на один дійсний корінь припадає одна мода на нульовій частоті чи на максимальній частоті. На два комплексно-

сполучених кореня характеристичного рівняння припадає два порядки моделі АР або одна спектральна мода на ненульовій частоті або не на максимальній частоті.

Параметричні спектри процесів АР описуються коефіцієнтами АР їх моделей. Оцінювання коефіцієнтів АР отримують із оптимального рівняння Юла-Уокера. Його одержують із рівняння (3.1), із урахуванням того, що помилки передбачення мають характеристики білого шуму (БШ). Параметрична СЦП АР процесу, у випадку багатомодова, описується виразом

$$P(f) = \frac{D_a}{\left| 1 - \sum_{i=1}^p \Phi[p, i] e^{-j2\pi f i T} \right|^2}, \quad (3.8)$$

де  $T=1/F$ , де  $F$  – частота дискретизації. Як очевидно з формули (3.6), лише кілька параметрів моделі АР несуть повну інформацію про СЦП випадкового процесу. У параметричній СЦП безліч переваг. Головні з них – це відсутність бічних пелюсток, висока роздільна здатність при коротких реалізаціях, залежність СЦП від одного параметра – коефіцієнтів АР, простота у обчисленні СЦП за формулою (3.8) та інше.

Проте є недоліки параметричної оцінки СЦП. Вони пов'язані з необхідністю використання хорошої оцінки моделі АР аналізованого процесу. Модель АР, найкраще підходить для вузькосмугових випадкових процесів. Необхідно також правильно визначити порядок моделі АР [1].

Побудова АР моделі випадкового процесу зводиться до знаходження коефіцієнтів АР та визначення порядку. Знайдемо рівняння до розрахунку оптимальних оцінок коефіцієнтів АР. Помноживши праву та ліву частини (3.1) на  $x[t-i]$  і усереднивши, можна отримати систему  $p+1$  рівнянь Юла-Уокера

$$R[i] - \sum_{j=1}^p \Phi[j] R[j-i] = 0, \quad i = 1, \dots, p, \quad (3.9a)$$

$$R[0] - \sum_{j=1}^p \Phi[j]R[j] = D_a, \quad i=0, \quad (3.9б)$$

де  $R[i] = E\{x[t]x[t-i]\}$  – значення функції кореляції випадкового процесу,  $D_a$  – дисперсія помилок передбачення моделі АР,  $R[0]$  – дисперсія випадкового процесу  $x[t]$ .

Представимо систему рівнянь (3.9а,б) у матричному вигляді

$$\vec{R} = R_p \vec{\Phi}, \quad (3.10)$$

де

$$\vec{\Phi} = \begin{vmatrix} \Phi[1] \\ \Phi[2] \\ \vdots \\ \Phi[p] \end{vmatrix}, R_p = \begin{vmatrix} R[0] & R[1] & \dots & R[p-1] \\ R[1] & R[0] & \dots & R[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ R[p-1] & R[p-2] & \dots & R[0] \end{vmatrix}, \vec{R} = \begin{vmatrix} R[1] \\ R[2] \\ \vdots \\ R[p] \end{vmatrix}. \quad (3.11)$$

Кореляційна матриця  $R_p$  має теплицеву структуру, у якій елементи, розташовані на діагоналях, ідентичні.

Рішення матричного рівняння, класичним методом Юла-Уокера, можна записати так

$$\vec{\Phi} = R_p^{-1} \vec{R}. \quad (3.12)$$

Класичний метод обчислення коефіцієнтів АР при довгих реалізаціях дає, як правило, стійкі рішення. Для генерації імітаційного процесу АР використовуються коефіцієнти АР, що задовольняють умову стаціонарності.

### 3.3 Нестационарні імітаційні процеси з авторегресійною стаціонарною складовою

Наведемо приклади використання виразів (3.6) для синтезу стаціонарних моделей АР. Потім будемо використовувати стаціонарні моделі для імітації нестационарних моделей з трендом та сезонною складовою. Для моделі низькочастотних коливань АР(1), будемо вважати, що корінь характеристичного рівняння задається характеристиками СЦП:  $f=0$  и  $\Delta f_i=10$ , при частоті дискретизації  $1/T=1000$ . Для цих параметрів корінь характеристичного рівняння дорівнює  $c[1,1]=0.9691$ . Йому відповідає коефіцієнт АР(1)=0,9691. Графік низькочастотного процесу АР(1) представлений на рис. 3.3.

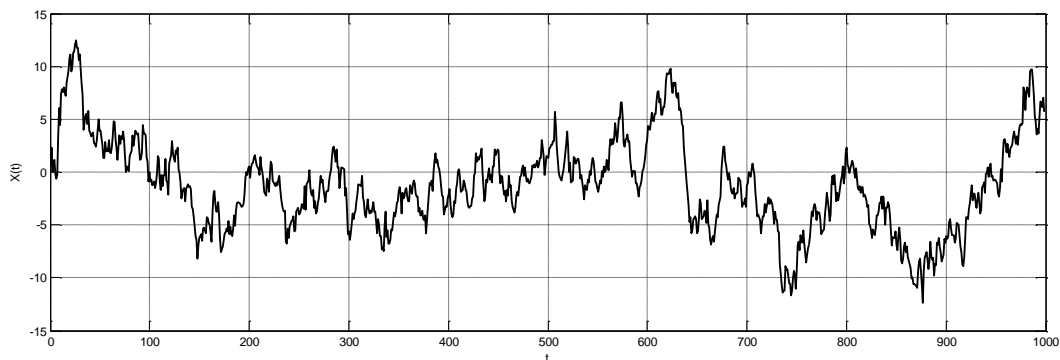


Рисунок 3.3 - Низькочастотна вибірка випадкового процесу АР(1)

Імітаційні процеси АР(1) були отримані методом породжувального процесу. В якості породжувального процесу використовувався БШ з нормальним розподілом. Нормальність БШ є необов'язковою умовою для породжувального процесу. Обробка випадкового процесу, що зводиться до визначення його параметрів, ведеться лише на рівні кореляційних функцій. Як гаусові процеси, так і негаусові процеси описуються головною статистичною залежністю у вигляді кореляційної функції.

На рис. 3.4а представлена теоретична СЩП, знайдена за формулою (3.6), після підстановки до неї заданого коефіцієнта  $AR(1)$ . Оціночна СЩП процесу, знайдена за вибіркою (рис. 3.3), з використанням оцінок кореляційної функції, рівнянь Юла-Уокера (3.9) показана на рис. 3.4б. Довжина вибірки, за якою оцінювалися значення кореляційної функції, становила 1000 відліків.

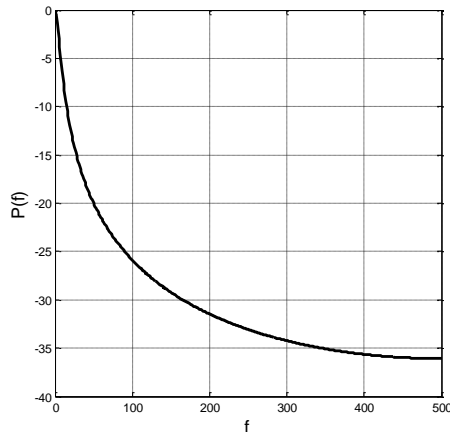


Рисунок 3.4а – Теоретична СЩП для моделі  $AR(1)$  із параметрами:  $f_1 = 0$ ,  $df_1 = 10$

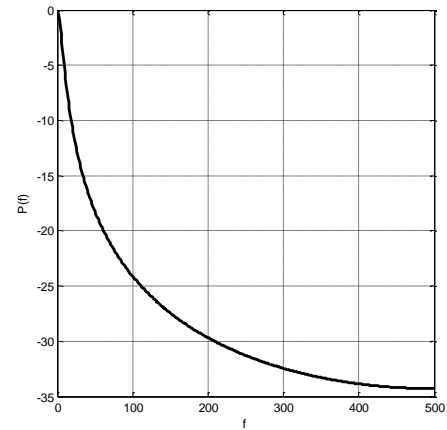


Рисунок 3.4б – Вибіркова оцінка СЩП для моделі  $AR(1)$  із параметрами:  $f_1 = 0$ ,  $df_1 = 10$

Розглянемо задачу генерації імітаційного нестационарного процесу. Модель нестационарного процесу із трендом, представленим на рис. 3.1 та сезонної складової (рис. 3.2), низькочастотної стаціонарної складової  $AR(1)$  (рис. 3.3), показана на рис. 3.5. Графік сезонної складової спотворений стаціонарною складовою  $AR(1)$ .

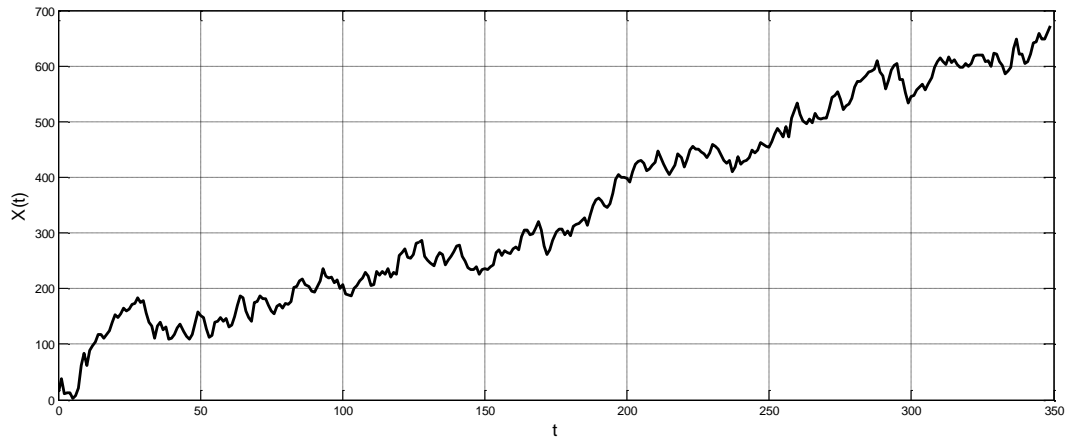


Рисунок 3.5 – Нестационарний імітаційний процес з низькочастотною стаціонарною складовою, що описується моделлю  $AR(1)$

Розглянемо високочастотну модель  $AR(1)$  для  $F=1000$  і  $\Delta f_i=10$ . Для імітаційної моделі  $AR(1)$  знайдемо коефіцієнт  $AR(1)$ . Для вибраних параметрів СЩП він дорівнює  $AR(1)=-0,9691$ . Тоді високочастотна вибірка має вигляд, що представлений на рис. 3.6. З огляду на особливості представленої моделі, коливання відліків максимальні. Вони змінюють свій знак із кожним новим відліком.

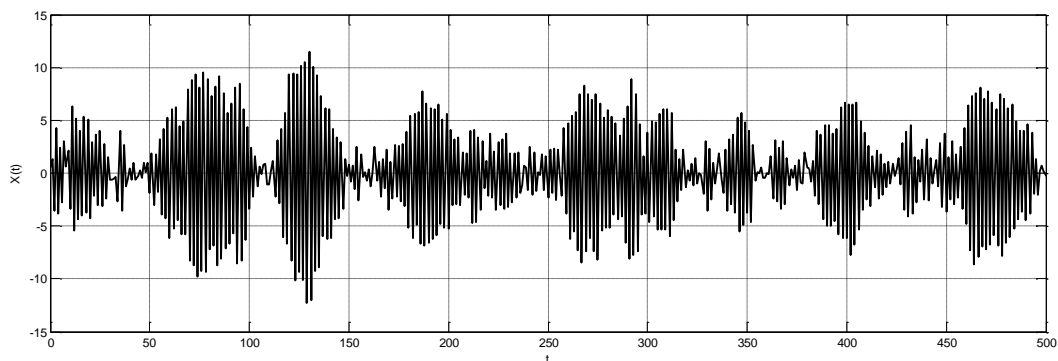


Рисунок 3.6 – Високочастотна вибірка випадкового процесу  $AR(1)$

На рис. 3.7а представлена теоретична СЩП, знайдена за формулою (3.6), після підстановки в неї коефіцієнта  $AR$ , що задається:  $AR(1)=-0,9691$ . Оцінювальна СЩП, знайдена за вибіркою, представленою на рис. 3.6, показана

на рис. 3.7б. Довжина вибірки, за якою оцінювали значення кореляційної функції, становила 1000 відліків.

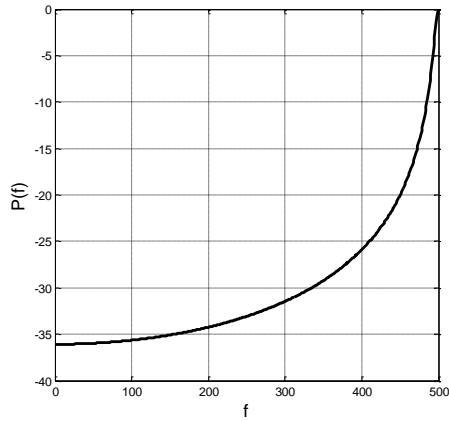


Рисунок 3.7а – Теоретична СЦП для високочастотної моделі  $AR(1)$  з параметрами:  $f_1 = 500$ ,  $df_1 = 10$

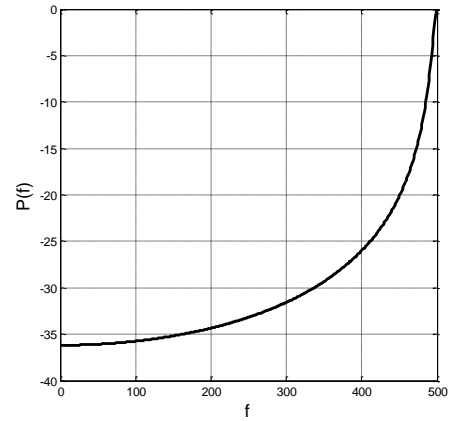


Рисунок 3.7б – Вибіркова оцінка СЦП для моделі  $AR(1)$  із параметрами:  $f_1 = 500$ ,  $df_1 = 10$

Вибірка імітаційного нестационарного процесу з трендом та сезонною складовою показана на рис. 3.8. В якості високочастотної стаціонарної складової, використовувався високочастотний процес, показаний на рис. 3.6.

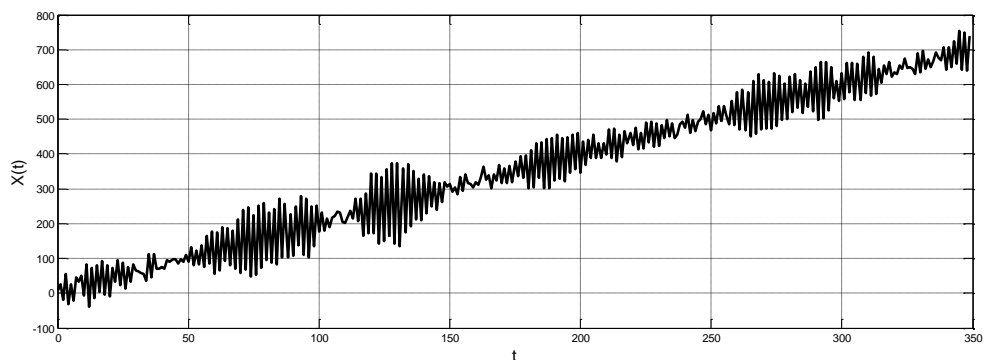


Рисунок 3.8 – Нестационарний імітаційний процес із високочастотною стаціонарною складовою, що описується моделлю  $AR(1)$

Розглянемо важливий випадок імітаційної нестационарної моделі, що включає стаціонарний процес  $AP(2)$ , при довільних параметрах СЦП:  $f_1 = 100$  і  $df_1 = 10$ . Імітаційний процес  $AP(2)$  був отриманий методом породжуючого процесу. В якості породжувального процесу використовувався БШ із нормальним розподілом. Якщо породжуючий процес негаусів, то при генерації відбувається часткова нормалізація. Але повної нормалізації за допомогою наявності підсумовування у рівнянні  $AP(2)$ , домогтися складно. Умови центральної граничної теореми в повному обсязі справедливі для  $AP(2)$ . Чим вище порядок моделі  $AP$ , тим більша нормалізація процесу. Другого порядку явно замало для повної нормалізації.

Негаусові процеси також характеризуються кореляційною функцією. Генерація корельованого випадкового процесу, як і раніше, здійснювалася за допомогою однієї і тієї ж вибірки БШ (рис. 3.9). Коріння характеристичного рівняння, отримане за допомогою виразів (3.7б), дорівнює

$$\begin{aligned} c[2,1] &= 0.7840 + 0.5696i, \\ c[2,2] &= 0.7840 - 0.5696i. \end{aligned} \quad (3.13a)$$

Їм відповідають коефіцієнти  $AP(2)$ , які дорівнюють 1.5680 та -0.9391.

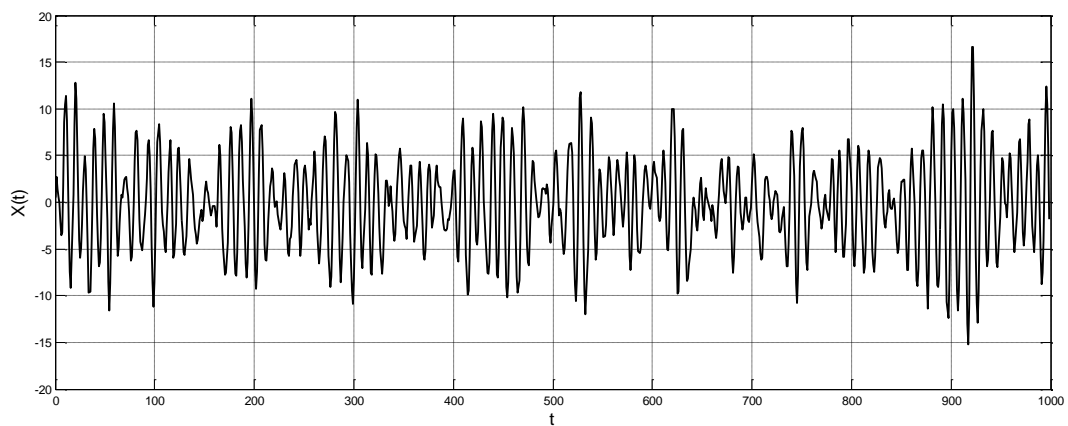


Рисунок 3.9 – Вибірка стаціонарного випадкового процесу  $AP(2)$

На рис. 3.10а представлена теоретична СЩП імітаційного процесу, знайдена за формулою (3.8), після підстановки в неї коефіцієнтів моделі  $AR(2)$ . Оціночна СЩП, що знайдена за вибіркою, наведеною на рис. 3.9, показана на рис. 3.10б. Довжина вибірки, за якою оцінювалися значення кореляційної функції, становила 1000 відліків.

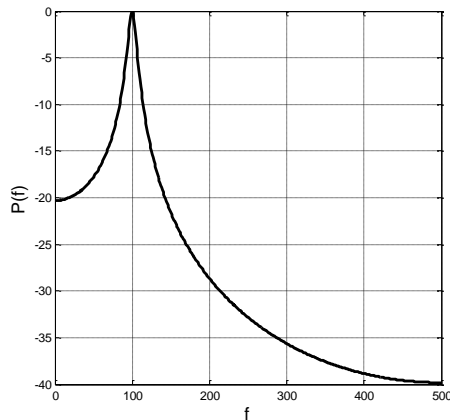


Рисунок 3.10а – Теоретична СЩП для моделі  $AR(2)$  із параметрами:  $f_1 = 100$ ,  $df_1 = 10$

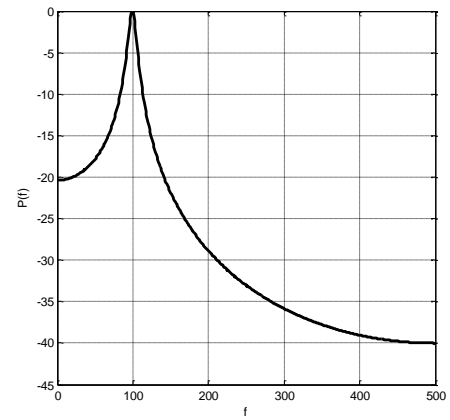


Рисунок 3.10б – Вибіркова оцінка СЩП для моделі  $AR(2)$  із параметрами  $f_1 = 100$ ,  $df_1 = 10$

На рис. 3.11 показано модель імітаційного нестационарного процесу, отриманого за допомогою моделі декомпозиції (2.1). Стационарна складова описується моделлю  $AR(2)$ . Нестационарний процес має, описані вище, тренд (рис 3.2) та сезонну складову (рис 3.3).

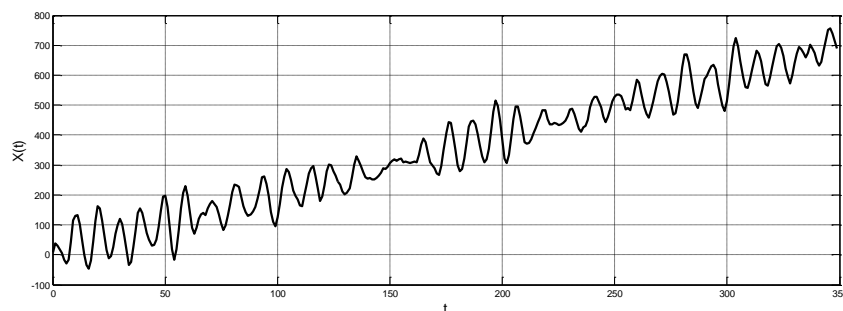


Рисунок 3.11 – Нестационарний імітаційний процес зі стаціонарною складовою, що описується моделлю  $AR(2)$

### 3.4 Моделі імітаційних нестационарних процесів із трендом та сезонною складовою

У наступній серії імітаційних експериментів використовувався інший лінійний тренд та сезонна складова. Моделі АР стаціонарних складових також змінювались. Їхній порядок був третім і четвертим. Генерована модель при застосуванні декомпозиції, є уявленням нестационарного процесу як адитивної суми складових [17].

Якщо детерміновані тренди і сезонні процеси можна порівняно легко генерувати, певні труднощі викликає формування стаціонарного процесу. У роботі розглянуто процедуру формування імітаційного АР випадкового процесу. Цей процес можна згенерувати за заданими параметрами СЦП. При генерації дуже важливо вибрати вихідні параметри процесу. Це мають бути досить очевидні параметри СЦП.

Як і раніше, емпіричне знаходження тренду випадкового процесу часто не означає, що він повторюватиметься надалі. Це одна з причин суттєвих помилок нестационарних моделей. Пророцтво в цьому випадку можливе тільки на однорідних ділянках тренду.

На рис. 3.12 показано графік лінійного тренду. На цьому малюнку показаний приклад тренда, швидкість зміни якого стала. Цей зростаючий тренд буде використаний нижче для отримання нестационарних процесів.

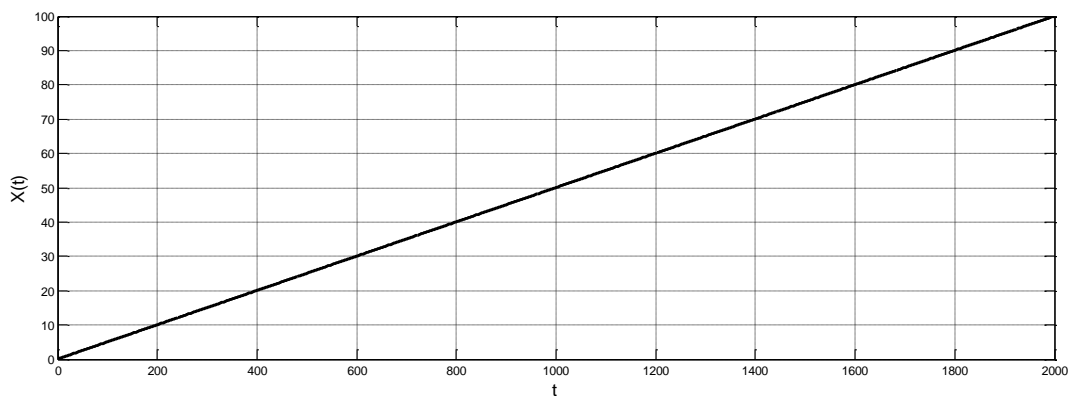


Рисунок 3.12 – Лінійний тренд імітаційного процесу

Іншою важливою складовою нестационарного процесу вважається сезонний процес. Він є моделлю циклічного процесу нестационарного часового ряду. Циклічність може виявлятися у різних нелінійних функціях. Необов'язково вона проявляється у вигляді найвідомішого виду – синусоїди. Її апроксимація у вигляді синусоїди найчастіше використовується через поширеність, простоту завдання та використання тощо. Крім цього синусоїда часто є рішенням диференціальних і різницевих рівнянь.

Сезонну складову можна описувати інакше. Однак часто інший облік сезонної складової має недоліки. Наприклад, сезонна складова створює яскраво виражений спектр, що є сильною стороною і його недоліком. Неможливо досить гнучко одночасно використовувати частотний та часовий опис. Маскування СЩП спектрів стаціонарних складових сильним спектром сезонного процесу, створює перешкоди при виявленні слабких сигналів.

Модель сезонної складової у вигляді дискретної синусоїди з періодом 12 відліків представлена на рис. 3.13. Ця сезонна складова використовується надалі для отримання нестационарного процесу. У нестационарній моделі (2.1) сезонна складова дещо змінює період і свою амплітуду під дією стаціонарної частини процесу у вигляді процесу АР.

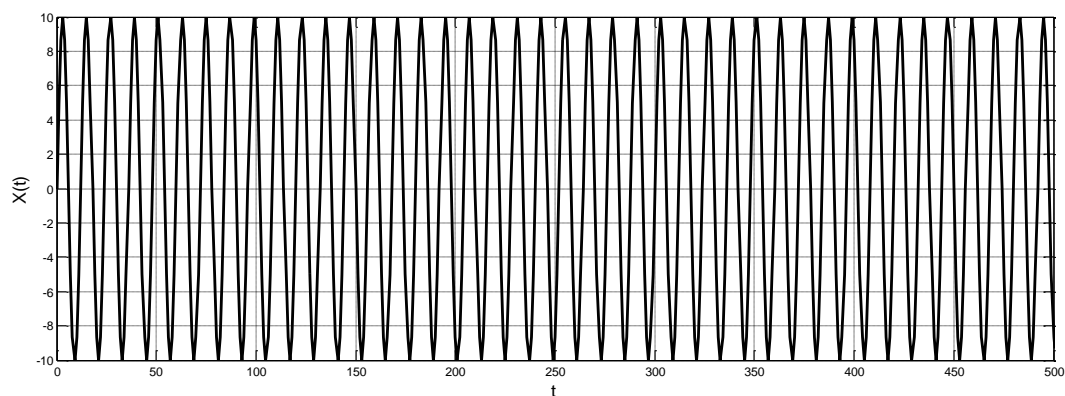


Рисунок 3.13 – Сезонна складова процесу з періодом 12 відліків

Зупинимося на моделі імітаційної стаціонарної складової, що є процесом АР(3) третього порядку. Параметри СЩП виберем рівні  $f_{1,2} = 0, 100$ ,  $df_1 = 10, 15$ .

Імітаційний процес  $AR(3)$  був отриманий методом породжуючого процесу. Породжувальним процесом був БШ із нормальним розподілом, із математичним очікуванням рівним нулю та одиничною дисперсією.

БШ може мати і негаусовий розподіл. Негаусові процеси існують у величезній кількості і їх також треба обробляти. Чекати, доки з'являться критерії і для негаусових процесів не варто. Вони можуть не з'явитися, чи не існувати зовсім. Негаусові процеси теж треба імітувати. Однак тут проявляються нові проблеми. Наприклад, складно імітувати негаусові процеси з такими простими характеристиками як ексцес і асиметрія. Доводиться підбирати ці характеристики у породжуючого негаусового БШ, щоб імітаційний процес мав бажаний ексцес або асиметрію.

Корні характеристичного рівняння, отримані за допомогою виразів (3.7a) і (3.7б), дорівнюють

$$\begin{aligned} c[3,1] &= 0,9691, \\ c[3,2] &= 0.7718 + 0.5607i, \\ c[3,3] &= 0.7718 - 0.5607i. \end{aligned} \tag{3.14}$$

Цим корням відповідають коефіцієнти моделі  $AR(3)$ , які дорівнюють: 2.5126, -2.4059, і 0.8819. На рис. 3.14 показано частину отриманої вибірки імітаційного стаціонарного процесу, що описується моделлю  $AR(3)$ . Повна довжина породжуючого процесу, становила 2000 відліків.

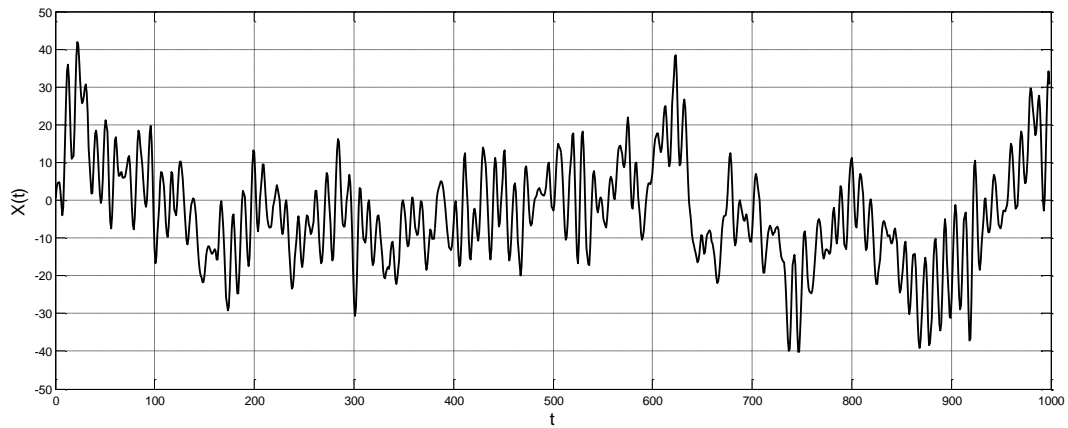


Рисунок 3.14 – Вибірка випадкового процесу  $AR(3)$

Зазвичай, чим вище порядок моделі  $AR$ , тим більше довжина генерованого випадкового процесу. При зменшенні довжини породжуючого процесу, погіршується роздільна здатність, спотворюється вибірковий спектр. Так при генерації імітаційних процесів  $AR(1)$  та  $AR(2)$  можна використовувати вибірку завдовжки 1000 відліків. Але для імітації стійких процесів  $AR(3)$  бажано використовувати породжувальний процес, довжиною 2000 відліків БШ.

Теоретична СЦП, знайдена для моделі  $AR(3)$ , показана на рис. 3.15а. Вона отримана за формулою (3.8), яка застосовується для знаходження параметричного спектру. Коефіцієнти  $AR(3)$ , розраховані по значенням коренів характеристичного рівняння, задаються, виходячи з форми бажаного спектру. Значення параметрів спектра вибиралися довільними. Корінь характеристичного рівняння знайдено з використанням виразів (3.7б).

На рис. 3.15б представлено вибіркову оцінку СЦП імітаційної моделі  $AR(3)$  стаціонарного процесу. Вибіркова СЦП була отримана за формулою (3.8). При цьому використовувалися вибіркові значення коефіцієнтів  $AR(3)$ , отримані за допомогою системи рівнянь Юла-Уокера (3.9). Значення коефіцієнтів кореляції оцінювалися за імітаційною вибіркою стаціонарного процесу.

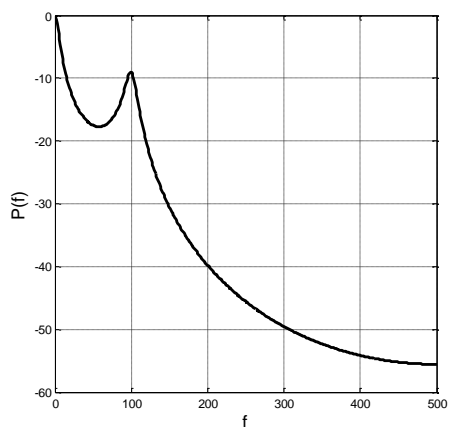


Рисунок 3.15а – Теоретична СЦП для моделі  $AR(3)$  із параметрами:

$$f_{1,2} = 0, 100, \quad df_{1,2} = 10, 15$$

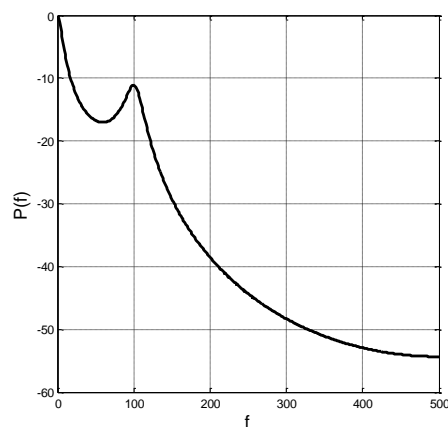


Рисунок 3.15б – Вибіркова оцінка СЦП для моделі  $AR(3)$  із параметрами:

$$f_{1,2} = 0, 100, \quad df_{1,2} = 10, 15$$

Для генерації імітаційного нестационарного процесу з лінійним трендом, сезонною та стаціонарною складовими, будемо використовувати тренд та сезонну складові, представлені на рис. 3.12 та рис. 3.13. Стаціонарну складову у вигляді процесу  $AR(3)$  показано на рис. 3.14. Імітаційний нестационарний процес представлений на рис. 3.16. Як видно з графіка, тренд та сезонна складові сильно змінені внеском стаціонарного  $AR(3)$  процесу.

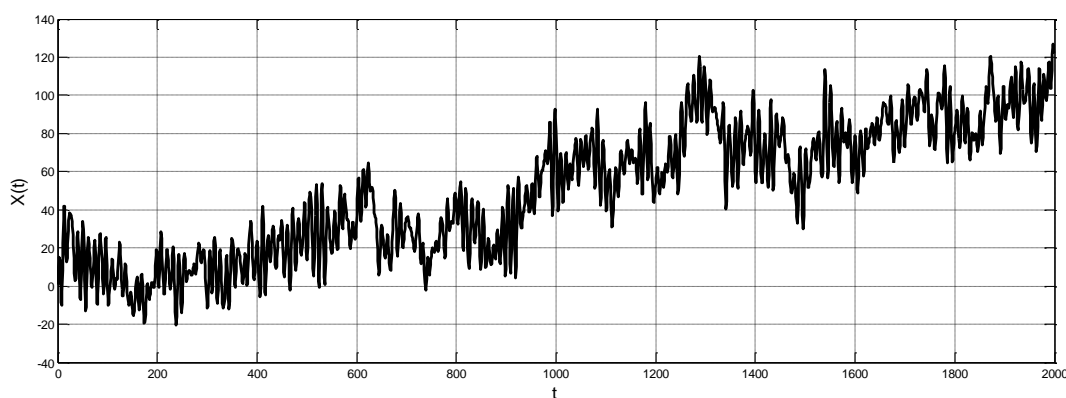


Рисунок 3.16 – Модель нестационарного процесу з лінійним трендом та сезонною складовими

Розглянемо випадок імітаційної моделі, коли стаціонарна складова описується моделлю  $AR(4)$ . Корні характеристичного рівняння, отримані за допомогою виразу (3.76), дорівнюють

$$\begin{aligned} c[4,1] &= 0,9216 + 0,2995i, \\ c[4,2] &= 0,9216 - 0,2995i, \\ c[4,3] &= 0,7718 + 0,5607i, \\ c[4,4] &= 0,7718 - 0,5607i. \end{aligned} \tag{3.15}$$

Цим корням відповідають коефіцієнти моделі  $AR(4)$ : 3.3868, -4.6944, 3.1270, -0.8546. Довжина породжувального процесу, становила 2000 відліків.

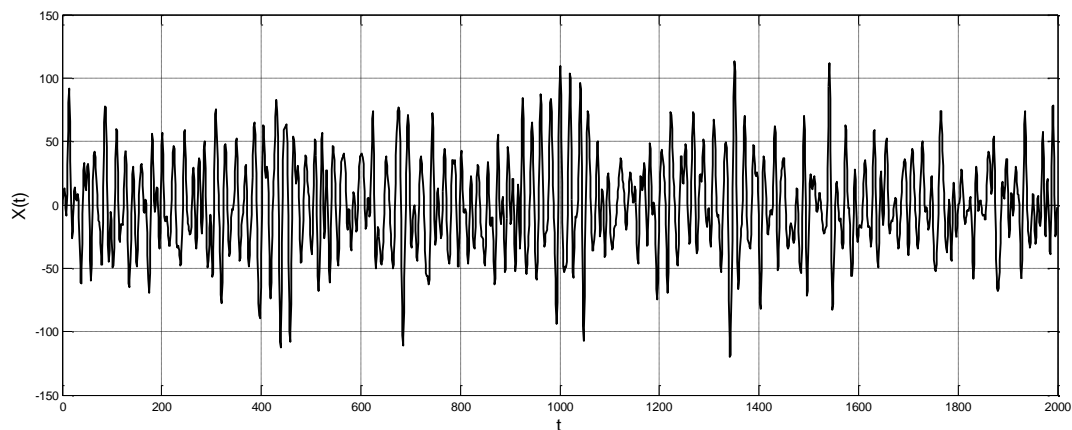


Рисунок 3.17 – Вибірка випадкового процесу  $AR(4)$

При малій довжині породжувального процесу, вибірковий спектр не відповідає заданому. Однак ці спотворення, швидше за все, виникають не через похибку методу генерації стаціонарної вибірки, а через вибірковий метод оцінки. У тому числі використання рівнянь Юла-Уокера може вносити свої похибки. Для генерації стаціонарного імітаційного процесу  $AR(4)$  використовувалася вибірка завдовжки 2000 відліків. На рис. 3.18а показано теоретичну СЩП для моделі  $AR(4)$ , отриману за формулою (3.96). Коефіцієнти

AR(4) розраховані за значеннями коренів характеристичного рівняння. Значення параметрів спектра вибиралися довільними.

На рис. 3.18б представлена вибіркова оцінка СЩП імітаційної моделі AR(4) стаціонарного процесу. Для отримання теоретичної та вибіркової СЩП використовувався вираз (3.8). Вибіркові значення коефіцієнтів AR(4) отримані за допомогою системи рівнянь Юла-Уокера (3.9). Попередньо необхідні значення коефіцієнтів кореляції оцінювалися за імітаційною вибіркою стаціонарного процесу. Із порівняння графіків видно, що представлені на рис. 3.18а та рис. 3.18б теоретична та вибіркова СЩП, практично збігаються. Відмінність вибіркової СЩП від теоретичної можна віднести на похибки, що виникають при використанні оцінок.

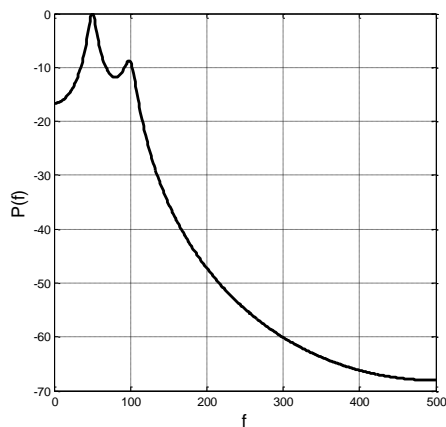


Рисунок 3.18а – Теоретична СЩП для моделі AR(4) із параметрами:

$$f_{1,2} = 50, 100, \quad df_{1,2} = 10, 15$$

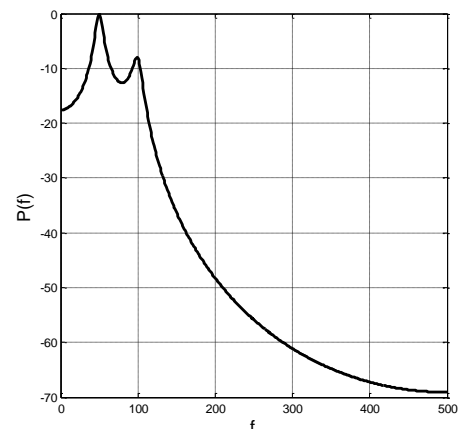


Рисунок 3.18б – Вибіркова оцінка СЩП для моделі AR(4) із

$$\text{параметрами: } f_{1,2} = 50, 100, \quad df_{1,2} = 10, 15$$

Нестационарна імітаційна модель показана на рис. 3.19. Вона складається з лінійного тренду (рис. 3.12), сезонної складової (рис. 3.13) та моделі стаціонарної AR(4) складової. Тренд та детермінована сезонна складова спотворені імітаційним процесом AR(4), показаним на (рис. 3.17).

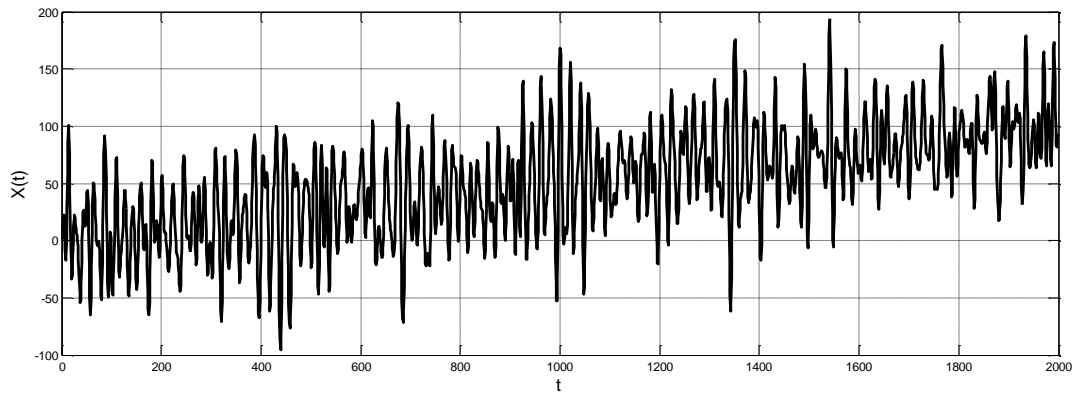


Рисунок 3.19 – Нестационарна модель процесу

## 4 МОДЕЛЬ ARIMA ДЛЯ КВАДРАТИЧНОГО ТРЕНДА

4.1 Імітація нестационарного процесу з квадратичним трендом зі стаціонарним процесом другого порядку

Розглянемо спосіб отримання нестационарного процесу з параболічним трендом. Зв'яжемо його з моделлю авторегресії проінтегрованого ковзного середнього (ARIMA). В операторній формі, що формує стаціонарний імітаційний процес AP, він являє собою рекурсивний фільтр. Білий шум використовується як породжувальний процес. Формуючий фільтр стаціонарного процесу описується рівнянням

$$x[t] = H(z)a[t] = \frac{a[t]}{\Phi(z)}. \quad (4.1)$$

Нехай стаціонарна складова описується моделлю AP(2) другого порядку. AP модель нестационарного процесу часто називається проінтегрованою AP. Нестационарна модель з лінійним трендом і без сезонної складової представляється у вигляді

$$(1 - \Phi[1]z^{-1} - \Phi[2]z^{-2})(1 - z^{-1})\omega_1[t] = a[t]. \quad (4.2)$$

Модульований імітаційний нестационарний процес може мати детермінований квадратичний тренд. Для моделі нестационарного процесу проінтегрованої авторегресії другого порядку з квадратичним трендом маємо

$$(1 - \Phi[1]z^{-1} - \Phi[2]z^{-2})(1 - z^{-1})^2\omega_1[t] = a[t]. \quad (4.3)$$

На рис. 4.1 наведено приклад такого процесу. Як сезонний і стаціонарний процес які використовувалися в імітаційній моделі, показані на рис. 3.2 та рис. 3.9.

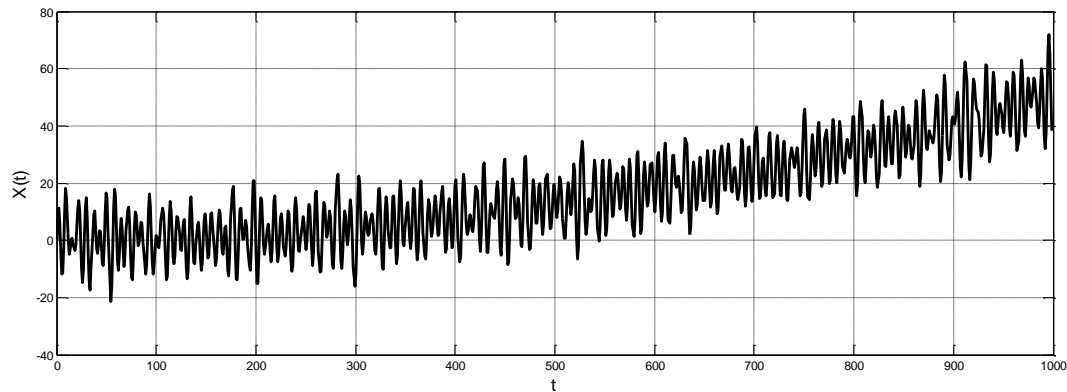


Рисунок 4.1 – Нестационарний імітаційний процес із квадратичним трендом, сезонною складовою та зі стаціонарною складовою, що описується моделлю  
AR(2)

Завданням побудови моделі ARIMA є знаходження способу видалення з вихідних даних тренду та сезонної складової за допомогою відповідних операторів взяття різниці. Зрозуміло, що це можна зробити, якщо модель, яка отримується, досить точна. Після цього виконується будівництво моделі лінійного передбачення стаціонарного залишку. Для реальних процесів модель стаціонарного процесу може бути складнішою. Загалом використовується модель ARKS (див. розділ 2).

Модель ARIMA [2] нестационарного процесу представляється у вигляді мультиплікативного процесу з трендом та сезонною складовою

$$x[t] = \nabla^d \nabla_s \omega[t] = (1-z)^d (1-z^s) \omega[t]. \quad (4.4)$$

Оператори взяття різниці враховують сезонну складову та тренд шляхом їх віднімання. Для видалення сезонної складової застосовувався оператор

$\nabla_s = 1 - z^s$ . Процес видалення лише сезонної складової за допомогою оператора зсуву за виразом, складає

$$x[t] = \nabla_s \omega_1[t] = (1 - z^s) \omega_1[t]. \quad (4.5)$$

4.2 Імітація нестационарного процесу з квадратичним трендом зі стаціонарним процесом авторегресії четвертого порядку

Другий отриманий нестационарний процес із квадратичним трендом мав стаціонарну складову у вигляді моделі AR(4). Для його формування використовувалася модель ARIMA. Формуючий імітаційний AR процес, був, по суті, рекурсивним фільтром. Помилка передбачення типу БШ, використовувалася як породжувальний процес. Формуючий фільтр стаціонарного процесу можна записати операторним рівнянням

$$x[t] = H(z)a[t] = \frac{a[t]}{\Phi(z)}, \quad (4.6)$$

де  $a[t]$  – вхідний процес,  $x[t]$  – вихід формуючого фільтра. Нехай залишок складеного нестационарного процесу описується стаціонарною моделлю AR(4)

$$(1 - \Phi[1]z^{-1} - \Phi[2]z^{-2} - \Phi[3]z^{-3} - \Phi[4]z^{-4})x[t] = a[t]. \quad (4.7)$$

Для моделі нестационарного процесу ARIMA, який був імітований, стаціонарний процес був AR процесом четвертого порядку з лінійним трендом

$$(1 - \Phi[1]z^{-1} - \Phi[2]z^{-2} - \Phi[3]z^{-3} - \Phi[4]z^{-4})(1 - z^{-1})\omega_1[t] = a[t], \quad (4.8)$$

де  $\omega_1[t]$  – нестационарний процес із трендом. Нестационарний процес можна одержати, використовуючи модель декомпозиції (2.1). Для моделі

нестационарного процесу – проінтегрованої авторегресії четвертого порядку з квадратичним трендом, маємо

$$(1 - \Phi[1]z^{-1} - \Phi[2]z^{-2} - \Phi[3]z^{-3} - \Phi[4]z^{-4})(1 - z^{-1})^2 \omega_1[t] = a[t]. \quad (4.9)$$

Приклад такого процесу представлено на рис. 4.2. Сезонна складова, що використовувалась як детермінований процес, показана на рис. 3.12. Залишок стаціонарної складової моделювався процесом AP(4) (рис. 3.17).

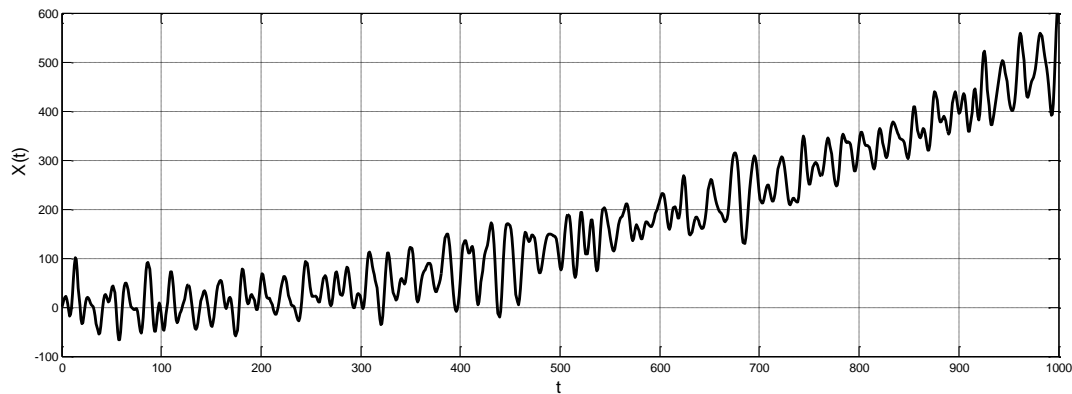


Рисунок 4.2 – Нестационарний імітаційний процес із квадратичним трендом, сезонною складовою та зі стаціонарною складовою, що описується моделлю AP(4)

Завдання побудови моделі нестационарного процесу, зводиться до знаходження операторів взяття різниць тренду і сезонної складовій нестационарного процесу. Для процесів, що зустрічаються в реальності, модель стаціонарного процесу описується моделлю авторегресії – ковзного середнього.

Використовуючи оператори взяття різниць, модель ARIMA [2] нестационарного процесу подається у вигляді процесу з трендом та сезонною складовою. Тільки у разі правильно обраних операторів взяття різниць, враховують сезонну складову та тренд шляхом їх віднімання. Для видалення сезонної складової застосовується оператор  $\nabla_s = 1 - z^s$ , де  $z^s$  – оператор зсуву через період, дія якого визначається виразом

$z^s x[t] = x[t - s]$ . Для видалення сезонної складової використовується оператор взяття різниці

$$x[t] = \nabla_s \omega_1[t] = (1 - z^s) \omega_1[t]. \quad (4.10)$$

Чим точніше в циклах повторюється період і чим ближче їх значення, що повторюються через період, тим краще вони описуються різницеvim оператором. Отже, чим ближче сезонні складові до детермінованих процесів, тим краще вони компенсуються різницеvими операторами. У деяких моделях сезонна складова або не враховується, тому що дуже слабка або її немає. Але в ряді випадків вона відіграє важливу роль і її слід моделювати. Періодичні коливання сезонної складової часто змінюються коливаннями, що у стаціонарної складової. Все залежить від амплітуди сезонної та стаціонарної складових.

## ВИСНОВКИ

У роботі показано деякі проблеми, що виникають при генерації імітаційних нестационарних моделей. Ці процеси можуть бути представлені у вигляді декомпозиції. При генерації нестационарних процесів дані містили трендову, сезонну і стационарну складові. Хоча в загальному, стационарні складові можуть моделюватися процесами АРКС, основна увага приділялася процесам АР. Саме ці процеси можна гнучкіше імітувати. В якості стационарних складових використовувалися моделі АР першого, другого, третього і четвертого порядків.

Розглянуто проблеми, пов'язані із генерацією стационарних процесів АР. Насамперед, вони стосуються питання, що виникають при генерації. По перше, це стосується параметрів, які необхідно задавати. У роботі такі параметри є характеристики СЦП. Дуже зручно та конструктивно задавати ці параметри спектру випадкового процесу. У роботі використовується зв'язок параметрів СЦП з корнями характеристичного рівняння. Корні цього рівняння мають однозначний зв'язок із коефіцієнтами АР. А вже коефіцієнти АР використовуються у конструктивній моделі для генерації випадкових процесів.

Важливим чинником є стационарність імітаційних моделей. У роботі показано, що знайдені для імітації коефіцієнти АР завжди задовольняють умові стационарності цієї моделі.

Для створення процесу зазвичай використовувався лінійний тренд. У роботі наведено приклади отримання нестационарного процесу з лінійним і квадратичним трендом. Залежно від умов завдання, тренди можуть мати більшу кількість параметрів.

В імітаційній моделі сезонна складова відіграє важливу роль. Її роль була обумовлена частотою зустрічі в реальних процесах. Показано, що властивості стационарної складової впливають на період та можливість компенсації сезонної складової при використанні моделі ARIMA.

З декомпозицією нестационарного процесу тісно пов'язана нестационарна модель ARIMA. Використання в моделі ARIMA операторів взяття різниці, що враховують у такий спосіб складові нестационарного процесу, можливе лише у разі досить точної моделі. В цьому випадку можна звести модель нестационарного процесу до стаціонарної моделі залишків складових.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Марпл. –мл. С. Л. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов: Пер. с англ. – М.: Мир, 1974. – Вып.1. – 406с.
3. Адаптивные фильтры: Пер. с англ. / Под ред. К.Ф.Н. Коуэна, П.М. Гранта. – М.: Мир, 1988. – 392 с.
4. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М.В. Назарова, Ю.Н. Прохорова. - М.: Радио и связь, 1981. - 495 с.
5. Маркел Д.Д., Грэй А.Х. Линейное предсказание речи. - М.: Связь, 1980. - 308 с.
6. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. – М.: Радио и связь, 2000. – 454 с.
7. Natalija V. Chmelarova (Kudriavtseva), Vyacheslav A. Tykhonov, Valerij M. Bezruk, Pavel Chmelar, Lubos Rejfeek, Speaker Verification Using Autoregressive Spectrum of Speech Signal in Composite Vector Stochastic Processes Model Representation". Journal of Mechanic of Continua and Mathematical Sciences (JMCMS) SP. ISSUE 4 NOVEMBER 2019.  
<https://doi.org/10.26782/jmcms.spl.4/2019.11.00018>
8. Доугерти, К. Введение в эконометрику / К. Доугерти. – М.: ИНФРА-М, 1997. – 402 с.
9. Громаков Ю.А. Стандарты и системы подвижной связи. - М.: Мобильные ТелеСистемы - Эко-Трендз, 1997. - 239с.
10. Ратынский М.В. Основы сотовой связи. - М.: Радио и связь, 2000. - 248 с.
11. Коротаев Г.А. Анализ и синтез речевого сигнала методом линейного предсказания // Зарубежная радиоэлектроника. - 1990. - №3. - С. 31-51.

12. Сукачев З.А. Сотовые сети связи с подвижными объектами: Учебн. пособ. - Одесса: УГАС им. Попова, 2000.
13. Тихонов В.И. Статистическая радиотехника. – М.: Радио и связь, 1982. – 624 с.
14. Рытов С.М. Введение в статистическую радиофизику. Случайные процессы. – М.: Наука, 1976. – 496 с.
15. Autoregressive Planet Search: Methodology. Gabriel A. Caceres and othes. The Astronomical Journal, 158:57 (21pp), 2019 August <https://doi.org/10.3847/1538-3881/ab26b8>.
16. Коротаев Г.А. Эффективный алгоритм кодирования речевого сигнала на скорости 4,8 кбит/с и ниже // Зарубежная радиоэлектроника. – 1996. – №3. С. 57–68.
17. Brockwell P.J., Davis R.A. Introduction to Time Series and Forecasting. – Springer, 2002. – P. 434.
18. Тихонов В.А., Безрук В.М., Корсун И.В., Кудрявцева Н.В. Исследование возможностей прогнозирования трафика сети мобильной связи // Восточно-Европейский журнал передовых технологий. – 2010 – №4(46). – С. 58–62. (фахове видання)
19. Тихонов В.А., Чеботарёва Д.В. Прогнозирование потока данных в сетях мобильной связи // Матеріали ІІІ Міжнародної науково-практичної конференції «Наукоємі технології в інфокомунікаціях» (23-25 травня 2019 р., Харків, Кам'янець-Подільський, Україна). – Харків: Друкарня «Мадрид», 2019, с. 126-127.
20. Тихонов В.А., Русановский Д.Е., Тихонов Д.В. Генерирование узкополосных имитационных случайных процессов // Радиоэлектроника и информатика. – 1999. – № 4. – С. 83-85.
21. Быков В.В. Цифровое моделирование в статистической радиотехнике. – М.: Сов. Радио, 1971. – 326 с.

22. Омельченко В.О., Безрук В.М., Драган Я.П., Колесников О.О., Омельченко А.В. Імовірнісні моделі випадкових сигналів та полів прикладах та задачах: Навч. посібник / За ред. В.О. Омельченка.–К.:ІСДО, 1996.– 272 с.

23. Старокожко Д.А., Перспективи розвитку інфокомунікацій та інформаційно-вимірювальних пристроїв// Матеріали 28-го Міжнародного молодіжного форум «Радіоелектроніка і молодь у ХХІ столітті» Зб. Матеріалів форуму. Т.4. – Харків: ХНУРЕ. 2024. – 159 с.