

Міністерство освіти та 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 декомпозиції нестационарного випадкового процесу»  
(тема)

Виконав: студент 2 курсу, групи ІМІм21-1

Спеціальність 172 "Телекомунікації та радіотехніка"

(код і повна назва спеціальності)

Освітньо-професійна програма \_\_\_\_\_

інформаційно-мережна інженерія

(повна назва освітньої програми)

Федючок Б.А.

(прізвище, ініціали)

Керівник проф. Тихонов В.А.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри \_\_\_\_\_  
(підпис)

Безрук В.М  
(прізвище, ініціали)

2022 р.

Не містить відомостей, заборонених  
до відкритого публікування

Керівник  / *В.А. Тихонов*

Студент  / *Б.А. Федючок*

Харківський національний університет радіоелектроніки

(повне найменування вищого навчального закладу)

Факультет інфокомунікацій

Кафедра інформаційно-мережної інженерії

Рівень вищої освіти другий (магістерський)

Спеціальність 172 "Телекомунікації та радіотехніка"

(код і назва)

Освітньо-професійна програма інформаційно-мережна інженерія

(назва)

**ЗАТВЕРДЖУЮ**

Зав. кафедри «ІМІ»

\_\_\_\_\_ проф. Безрук В.М.

«\_\_\_» \_\_\_\_\_ 2022 р

## ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ

Федючку Борису Анатолійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження моделі декомпозиції нестационарного випадкового процесу

керівник роботи Тихонов Вячеслав Анатолійович, доц.

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджена наказом ВНЗ від «21» жовтня 2022 р. № 1376 Ст

2. Строк подання студентом роботи 27 грудня 2022 р.

3. Вихідні дані до роботи: нестіціонарні часові ряди

модель декомпозиції часових рядів

модель авторегресії проінтегрованного ковного середнього

складові моделі декомпозиції

оператори взяття різниці для усунення тренда

оператори взяття різниці для усунення сезонної складової

модель стаціонарної складової  $AR(1)$

виділення довгострокових трендів температур

виділення довгострокового тренда трафіка передчі

прогнозування часових рядів з трендом, сезоною складовою, з стаціонарним

процесом

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)

Вступ

1. Аналітичний огляд методів представлення випадкових процесів лінійними

різничними стохастичними рівняннями

2. Моделі авторегресі-ковного середнього і декомпозиції нестационарних даних

3. Оцінка тренду нестационарного процесу

4. Використання моделі декомпозиції для прогнозування часових ряд

5. Перелік графічного матеріалу (із зазначенням обов'язкових креслень, плакатів)

Слайди у форматі Power Point (Дані середньомісячних температур, зафіксованих у метеообсерваторіях міст та оцінка їх довготривалого тренду, трафік передачі даних у стільниковому зв'язку та оцінка його тренду, зміни середньорічних температур, часовий ряд середньомісячних температур після видалення сезонної складової. зміни середньорічних температур у різних містах, прогнозування вибірки щоденних даних трафіку)

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	25.10.2022	Виконано
2	Огляд літературних джерел	02.11-13.11.2022	Виконано
3	Розробка теоретичної частини	14.11-31.11.2022	Виконано
4	Алгоритми моделювання АРПКС	01.12-05.12.2022	Виконано
5	Алгоритми моделювання декомпозиції	05.12-09.12.2022	Виконано
7	Оформлення пояснювальної записки	10.12-19.12.2022	Виконано
8	Перевірка роботи	20.12-22.12.2022	Виконано
9	Підготовка до захисту	15.12-20.12.2022	Виконано

Дата видачі завдання 24 жовтня 2022 р.

Студент



(підпис)

Федючок Б. А.

(Прізвище та ініціали)

Студент



(підпис)

Тихонов В. А.

(Прізвище та ініціали)

## РЕФЕРАТ

Пояснювальна записка випускної кваліфікаційної роботи спеціаліста:

73 сторінок, 31 рисуноків.

В кваліфікаційній роботі запропонована методика виделення довгострокового тренду з середньомісячних температур, виявлені недоліки моделі авторегресії проінтегрованного ковзного середнього та декомпозиції.

Мета даної кваліфікаційної роботи полягає в розробці алгоритму виявлення трендів, сезонних складових, їх взаємного впливу, а також вирішення задачі прогнозування деяких видів нестационарних процесів.

Метод проектування базується на використанні численних алгоритмів розрахунку оптимальних параметрів нестационарних процесів та їх моделей.

Галузь застосування – аналіз нестационарних процесів та їх прогноз.

Ключові слова: НЕСТАЦІОНАРНИЙ ПРОЦЕС, МОДЕЛЬ АВТОРЕГРЕСІЇ ПРОІНТЕГРОВАННОГО КОВЗНОГО СЕРЕДНЬОГО, ДЕКОМПОЗИЦІЯ ПРОЦЕСУ, ТРЕНД, СЕЗОНА СКЛАДОВА, СТАЦІОНАРНИЙ ПРОЦЕС, ТРЕНДИ ТЕМПЕРАТУР, ТРЕНДИ ТРАФІКА, ПРОГНОЗУВАННЯ ПРОЦЕСУ.

## THE ABSTRACT

Explanatory note of graduation qualification work of a specialist:

73 sides, 31 drawings.

In the qualification work, the technique of seeing a long-term trend from average monthly temperatures, revealing a small part of the autoregression model of the integrated average and decomposition was revealed.

The metadata of this qualification work is used in the analysis of the algorithm for revealing trends, seasonal warehouse, their mutual inflow, as well as the solution of the task of forecasting certain types of non-stationary processes.

The design method is based on various numerical algorithms for the analysis of optimal parameters for non-stationary processes of those models.

Scope of application- analysis of non-stationary processes and their forecast.

**Key words: NON-STATIONARY PROCESS, AUTO-REGRESSIVE INTEGRATED AVERAGE MODEL, PROCESS DECOMPOSITION, TREND, WAREHOUSE SEASON, STATIONARY PROCESS, TEMPERATURE TREND, TRAFFIC TREND, PROCESS FORECASTING.**

## ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	11
ВСТУП.....	12
1. Огляд основних понять в аналізі часових рядів.....	14
1.1 Складові моделі декомпозиції.....	14
1.2 Стаціонарність, автоковаріації та автокореляції .....	18
1.3 Тренди.....	22
1.4 Сезонні коливання.....	23
1.5 Модель авторегресії – ковзного середнього стаціонарного складника декомпозиції нестационарного процесу.....	25
1.6 Висновки до розділу та постановка задачі .....	27
2 Моделі авторегресії-проінтегрованого скільчувачого середнього і декомпозиції нестационарних даних.....	29
2.1 Модель авторегресії проінтегрованого ковзного середнього нестационарного процесу .....	29
2.2 Модель декомпозиції нестационарних даних.....	32
2.3 Неточності, що виникають під час використання операторів взяття різниці ..	32
3 Оцінка тренду нестационарного процесу .....	37
3.1 Нестационарні процеси з трендом .....	37
3.2 Виділення тренду температур .....	41
3.3 Чисельна оцінка тренду .....	42
4. Використання моделі декомпозиції для прогнозування часових рядів.....	47
4.1 Завдання прогнозування у стільниковому зв'язку .....	47
4.2 Модель авторегресії .....	51
4.3 Спектр процесу авторегресії .....	57
4.4 Характеристичне рівняння моделі авторегресії .....	59
4.5 Вплив щільності забудови на розповсюдження радіохвиль.....	59
4.6 Прогнозування на основі моделі декомпозиції .....	61

ВИСНОВОК.....	70
ПЕРЕЛІК ПОСИЛАНЬ .....	71
Додаток А - Слайди презентації.....	73

## ПЕРЕЛІК СКОРОЧЕНЬ

АР – авторегресія

КС – ковзне середнє

МНК – метод найменших квадратів

СГМ – спектральна густина потужності

СЩП – спектральна щільність потужності

АЧХ – амплітудно-частотна характеристика

## ВСТУП

До нестационарних випадкових процесів відносяться всі випадкові процеси, що не мають властивості стаціонарності хоча б у широкому сенсі. Характеристики нестационарного процесу є деякі функції часу, визначити які можна лише усередненням по ансамблю реалізацій, що утворюють процес. У практичних завданнях часто є принципово неможливими отримати досить велику кількість реалізацій для відшукування показників процесу з необхідною достовірністю. Ця обставина перешкоджає розвитку практичних методів оцінювання та аналізу нестационарних випадкових процесів.

У багатьох випадках у класі нестационарних процесів, відповідних реальним фізичним явищам, можна назвати особливі типи нестационарності, коли завдання оцінювання та аналізу спрощується. Наприклад, деякі випадкові явища що описуються нестационарним випадковим процесом, кожна реалізація якого містить тренд, сезонну складову та стаціонарний випадковий процес.

Процеси такого типу мають загальний детермінований тренд та детерміновану сезонну складову. Якщо нестационарний процес відповідає конкретній моделі такого типу, то для його опису немає необхідності проводити усереднення ансамблю: будь-які необхідні характеристики можна оцінити по одній реалізації, як для ергодичних процесів.

Якість роботи будь-якої нестационарної моделі, зрештою, визначається величиною помилки, що дорівнює різниці між необхідним і дійсним значеннями модельованого процесу. Насправді найімовірнішими є випадки, коли впливи змінюються в часі. Тому розрахункова і реальна оцінки помилки не рівні. Наявність різноманітних впливів, що виникають, обумовлена взаємопов'язаністю і взаємовпливом виникаючих через нестационарність процесів у складних багатовимірних системах, що функціонують у різних експлуатаційних режимах, а також природною зміною в часі, наприклад, характеристик об'єкта управління і параметрів системи управління. Випадковий характер процесів, що протікають у різних системах, призводить до того, що виконані за допомогою вимірювальних засобів спостереження спотворені не тільки перешкодами, а також містять зміни параметрів процесів, що змінюються в часі. Тому з метою покращення показників якості нестационарних процесів потрібна розробка методів, алгоритмів та засобів, що забезпечують підвищення достовірності оцінки моделей випадкового процесу на тлі перешкод.

Найчастіше використовується модель лінійного передбачення випадкових процесів Дженкінса і Бокса, далеко ще не універсальна. Її ефективність залежить від кількох факторів. Зокрема, від умов застосування цієї моделі. Використання для

адекватного застосування операторів взяття різниць, що застосовується для детермінованих моделей тренду та сезонної складової. У кваліфікаційній роботі показано взаємний вплив цих операторів як на характеристики трендів, так і на сезонну складову нестационарних процесів. Модель декомпозиції випадкових нестационарних процесів, як правило, використовує підхід Дженкінса-Бокса, не вирішує описані вище проблеми нестационарності. Тому необхідні подальші дослідження цього методу обліку нестационарності для уточнення параметрів моделі. Вирішенню цих проблем присвячено кваліфікаційну роботу.

Мета даної кваліфікаційної роботи полягає у розробці засобу підвищення ефективності моделей лінійного передбачення для уточнення оцінки трендів, сезонної, а також стаціонарної складових.

Об'єктом дослідження – є нестационарні випадкові процеси та сигнали та їх статистична обробка.

Предмет дослідження – властивості декомпозиції нестационарної моделі авторегресії лінійного передбачення процесів.

## 1 ОГЛЯД ОСНОВНИХ ПОНЯТЬ В АНАЛІЗІ ЧАСОВИХ РЯДІВ

### 1.1 Складові моделі декомпозиції

У кожній сфері людської діяльності зустрічаються явища, які цікаво і важливо вивчати в їх розвитку, так як вони еволюціонують і флюктуують у часі. З часом змінюються статистичні характеристики випадкових процесів, змінюються ціни, економічні умови, режим перебігу того чи іншого виробничого процесу. Сукупність змін подібних показників протягом деякого періоду часу і представляє часовий ряд. Цілі аналізу часових рядів можуть бути різними. Для нестационарних процесів можна, наприклад, прагнути передбачити майбутнє на підставі знань минулого, керувати процесом, що породжує ряд, намагатися з'ясувати механізм, що лежить в основі процесу, очистити часовий ряд від компонентів, які затемнюють його динаміку, або стисло описати характерні особливості ряду.

Часовим рядом називають послідовність спостережень, зазвичай впорядковану в часі [1, 2], хоча можливе впорядкування і за іншим параметром. Основною рисою, що виділяє аналіз часових рядів серед інших видів статистичного аналізу, є суттєвість порядку, у якому проводяться спостереження. Розрізняють два види часових рядів. Вимірювання деяких величин (температури, напруги і т.д.) проводиться безперервно принаймні теоретично. При цьому спостереження можна фіксувати у вигляді графіка. Але навіть у тому випадку, коли величини, що вивчаються, реєструються (або можуть реєструватися) безперервно, практично при їх обробці використовуються тільки ті значення, які відповідають дискретній безлічі моментів часу. Отже, якщо час вимірюється безперервно, часовий ряд називається безперервним, якщо час фіксується дискретно (тобто через фіксований інтервал часу), то часовий ряд дискретний. Дискретні часові ряди отримуються двома способами:

- Вибіркою з безперервних часових рядів через регулярні проміжки часу (наприклад, чисельність населення, величина власного капіталу фірми, обсяг грошової маси, курс акції). Такі часові ряди називаються моментними.

- Накопиченням змінної протягом деякого періоду часу, наприклад: температура атмосфери Землі [3, 4], процеси в стільниковому зв'язку [5], обсяг виробництва якогось виду продукції, кількість опадів, обсяг імпорту. У цьому випадку часові ряди називаються інтервальними.

У математичній статистиці прийнято моделювати часовий ряд як випадковий процес, який називається також стохастичним процесом, під яким розуміється статистичне явище, що розвивається у часі згідно із законами теорії ймовірностей. Випадковий процес – це випадкова послідовність. Зазвичай припускають, що ця послідовність йде від мінус до плюс нескінченності:  $\{x[t]\}_{t=-\infty, \dots, +\infty}$ . Часовий ряд – це лише одна окрема реалізація такого теоретичного стохастичного процесу:  $x[t]$ ,  $t=1, \dots, T = (x_1, \dots, x_T)$ , де  $T$  - це довжина часового ряду.

Часовий ряд  $x = (x[1], \dots, x[T])$  також часто неформально називають вибіркою. Зазвичай стоїть завдання по даному ряду зробити якісь висновки про властивості випадкового процесу, що лежить в його основі, оцінити параметри, зробити прогнози і т.п. У літературі з часових рядів існує деяка неоднозначність. Іноді часовим рядом називають сам випадковий процес, чи його відрізок  $t = 1, \dots, T$ , котрий іноді складає статистичну модель, яка породжує даний випадковий процес.

Можливі значення часового ряду в даний час  $t$  описуються за допомогою випадкової величини  $x[t]$  і пов'язаного з нею розподілу ймовірностей  $p(x[t])$ . Тоді значення  $x[t]$  часового ряду, що спостерігається, в момент  $t$  розглядається як одне з безлічі значень, які могла б прийняти  $x[t]$  випадкова величина в цей момент часу. Хоча за формальним визначенням вибірка має складатися із незалежних однаково розподілених випадкових величин.

Слід зазначити, проте, що, зазвичай, спостереження часового ряду взаємопов'язані, й у коректного його опису слід розглядати спільну ймовірність  $p(x[1], \dots, x[T])$ . Для зручності можна провести класифікацію випадкових процесів та відповідних ним часових рядів на детерміновані та випадкові процеси (часові ряди). Детермінованим називають процес, який набуває заданого значення з ймовірністю одиниця. Наприклад, його значення можуть точно визначатися якоюсь математичною функцією від часу  $t$ , як у наступному прикладі:  $x[t] = A \cos(2\pi f t - \theta)$ . Коли ж ми говоритимемо про випадковий процес і випадковий часовий ряд, то, як правило,

матимемо на увазі, що він суттєво випадковий, тобто не є детермінованим. Стохастичні процеси поділяються на стаціонарні та нестаціонарні. Стохастичний процес є стаціонарним, якщо він у певному сенсі у статистичній рівновазі, тобто його властивості з ймовірної точки зору не залежать від часу. Процес нестаціонарний, якщо умови порушуються.

Як правило, термін «часовий ряд» і сам по собі має на увазі, що цей ряд є одновимірним (скалярним). Часто буває важливо розглянути спільну динаміку набору часових рядів  $x[t] = (x[1], \dots, x[k])$ . Такий набір називають багатовимірним часовим рядом або векторним часовим рядом. Відповідно, говорять також про багатовимірні (векторні) випадкові процеси.

При аналізі часових рядів у математичній статистиці, в радіотехніці, економіки традиційно розрізняють різні види еволюції (динаміки). Ці види динаміки можуть, власне кажучи, комбінуватися. Тим самим задається розкладання часового ряду на складові (компоненти), які з економічної точки зору несуть різне змістовне навантаження. Перерахуємо найважливіші [6]:

Тренди (тенденції) відповідають повільній зміні, що проходить у певному напрямку, яке зберігається протягом значного проміжку часу. Тенденцію називають також трендом чи довготривалим рухом.

Циклічні коливання - це швидша, ніж тенденція, квазіперіодична динаміка, в якій є фаза зростання і фаза спадання. Найчастіше цикл пов'язаний з флуктуаціями економічної активності.

Сезонні коливання відповідають змінам, які відбуваються регулярно протягом року, тижня чи доби. Вони пов'язані з сезонами та ритмами людської активності. Календарні ефекти – це відхилення, пов'язані з певними передбачуваними календарними подіями, такими як святкові дні, кількість робочих днів за місяць, високосність року тощо.

Випадкові флуктуації – безладні рухи щодо великої частоти. Вони породжуються впливом різнорідних подій на величину, що вивчається (несистематичний або випадковий ефект).

Викиди — це аномальні рухи часового ряду, пов'язані з подіями, що рідко відбуваються, які різко, але лише дуже короткочасно відхиляють ряд від загального закону, за яким він рухається.

Структурні зрушення — це аномальні рухи часового ряду, пов'язані з подіями, що рідко відбуваються, мають стрибкоподібний характер і змінюють тенденцію.

У кваліфікаційній роботі в модель декомпозиції [7] будуть включені лише найважливіші складові нестационарних випадкових процесів: тренди, сезонні коливання, стаціонарні флуктуації випадкових процесів. Деякі часові процеси можна вважати такими, що представляють ті чи інші види таких рухів майже в чистому вигляді. Але більшість їх має дуже складний вигляд. Вони можуть проявлятися, наприклад, як загальна тенденція зростання, і деякі сезонні зміни, куди можуть накладатися випадкові флуктуації. Часто для аналізу часових рядів виявляється корисним ізольований розгляд окремих компонентів. Для того, щоб можна було розкласти конкретний ряд на ці складові, потрібно зробити якісь припущення про те, які властивості повинні мати ці складові.

Однак побудова формальної моделі є складним завданням. Зокрема, зі змістовного опису який завжди ясно, як моделювати ті чи інші компоненти. Наприклад, тренд може бути детермінованим чи стохастичним. Аналогічно, сезонні коливання можна комбінувати за допомогою детермінованих змінних або стохастичного процесу певного виду. Компоненти часового ряду можуть входити до нього адитивно чи мультиплікативно. Більше того, далеко не всі часові зміни мають досить просту структуру, щоб можна було розкласти їх на ці складові.

Існують дві основні категорії методів розкладання часових рядів на компоненти. Перша категорія використовує множинні регресії з факторами, що є функціями часу, друга заснована на застосуванні лінійних фільтрів.

## 1.2 Стаціонарність, автоковаріації та автокореляції

Статистичний процес називається строго стаціонарним, якщо взаємний розподіл ймовірностей  $m$  спостережень інваріантний стосовно загального зсуву часового аргументу, тобто. спільна щільність розподілу випадкових величин  $x[1]$ ,  $x[2]$ , ...,  $x[t]$

така сама, як для величин  $x[1+k]$ ,  $x[2+k]$ ,... , $x[t+k]$  за будь-яких цілих значеннях зсуву  $k$ . Коли при  $m=1$  витікає з припущення стаціонарності слід, вважати що безумовний розподіл величини  $x[t]$ ,  $p(x[t])$ , однаковий для всіх  $t$  і може бути записаний як  $p(x)$ . Вимога стаціонарності, визначена цими умовами, є досить жорсткою [8].

Насправді, щодо випадкових процесів, обмежуються моментами першого і другого порядку і тоді говорять про слабку стаціонарність чи стаціонарність другого порядку. У цьому випадку процес має постійне середнє значення  $\mu = E(x[t])$  для всіх  $t$ , що визначає рівень, щодо якого він флюктує, постійну дисперсію  $\sigma^2 = E(x[t]-\mu)^2$  для всіх  $t$  та постійну автоковаріацію  $R[k] = E(x[t]-\mu)(x[t+k]-\mu)$  всім  $t$ , тобто. коваріація між  $x[t]$  та  $x[t+k]$  залежить тільки від величини зсуву  $k$  і не залежить від  $t$ . Слід мати на увазі, що два процеси, що мають однакові моменти першого та другого порядку, можуть мати різний характер розподілу.

Зрозуміло, що згідно з даним визначенням стаціонарним може бути тільки нескінченно триваючий процес, але не кінцевий часовий ряд  $x[1], \dots, x[T]$ . Однак ми можемо назвати часовий ряд  $x[1], \dots, x[T]$  стаціонарним, якщо він може бути частиною нескінченного стаціонарного процесу [9]  $\{x[t]\}$   $t=-\infty, \dots, +\infty$ . Розглянемо властивості слабко стаціонарних процесів. Автоковаріацією із затримкою  $k$  (автоковаріацією  $k$ -го порядку) стаціонарного процесу називається величина  $R[k] = E(x[t]-\mu)(x[t+k]-\mu)$ . Автокореляцією із затримкою  $k$  (автокореляцією  $k$ -го порядку) стаціонарного процесу з ненульовою дисперсією називається величина

$$r[k] = \frac{E(x[t]-\mu)(x[t+k]-\mu)}{\sqrt{E(x[t]-\mu)^2 E(x[t+k]-\mu)^2}}$$

У нашій літературі сувору стаціонарність також називають стаціонарністю у вузькому сенсі, а слабку стаціонарність — стаціонарністю у широкому сенсі.

Оскільки у стаціонарного процесу дисперсія у часі незмінна, то  $r[k] = \frac{R[k]}{R[0]} = \frac{R[k]}{\sigma^2}$ . Автоковаріаційною функцією стаціонарного процесу називають послідовність автоковаріацій  $\{R[k]\}$   $k=-\infty, \dots, +\infty$ . Оскільки автоковаріаційна функція симетрична щодо нуля  $R[k] = R[-k]$ , достатньо розглядати  $k = 0, \dots, +\infty$ .

Автокореляційною функцією (АКФ) називають послідовність автокореляцій  $\{r[k]\}$ ,  $k=-\infty, \dots, +\infty$ . Оскільки автокореляційна функція симетрична, та  $r[0]=1$ , Досить розглядати  $k = 1, \dots, +\infty$ . Автоковаріаційна матриця  $R_T = \sigma^2 r_T$  для стаціонарного ряду  $x[1], \dots, x[T]$  має вигляд [10]:

$$R_T = \sigma^2 r_T = \sigma^2 \begin{pmatrix} r(0) & r(1) & r(2) & \dots & r(k-2) & r(k-1) \\ r(1) & r(0) & r(1) & \dots & r(k-3) & r(k-2) \\ r(2) & r(1) & r(0) & \dots & r(k-4) & r(k-3) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r(k-2) & r(k-3) & r(k-4) & \dots & r(0) & r(1) \\ r(k-1) & r(k-2) & r(k-3) & \dots & r(1) & r(0) \end{pmatrix}. \quad (1.1)$$

Особливість автоковаріаційної матриці (1.1) і відповідної автокореляційної матриці (1.1) у разі стаціонарності полягає в тому, що вони мають ті самі елементи на будь-якій діагоналі. Такі матриці прийнято називати теплицівими матрицями.

Як відомо, будь-яка коваріаційна матриця є симетричною і позитивно напіввизначеною. Крім того, якщо компоненти розглядаємого випадкового вектора  $x[t]$  лінійно незалежні в тому сенсі, що не існує ненульовий вектор коефіцієнтів  $\lambda$ , такий що  $\lambda^T x[t]$ - детермінована величина, то коваріаційна матриця є позитивно визначеною. Нагадаємо, що за визначенням симетрична  $T \times T$  матриця  $A$  називається позитивно напіввизначеною, якщо для кожного вектора виконується нерівність  $\lambda^T A \lambda \geq 0$ ; матриця  $A$  називається позитивно визначеною, якщо для кожного ненульового вектора виконується нерівність  $\lambda^T A \lambda > 0$ . Автоковаріаційна та автокореляційна матриця є коваріаційними матрицями, тому вони мають зазначені властивості. З іншого боку, якщо матриця має зазначені властивості, то вона може бути автоковаріаційною матрицею деякого часового ряду. З цих міркувань випливає, що умова слабкої стаціонарності процесу, компоненти якого лінійно незалежні у зазначеному вище сенсі, накладає низку обмежень на вигляд автокореляційної та автоковаріаційної функції. Вони впливають із того, що головні мінори позитивно певної матриці, у тому числі її визначник, мають бути позитивними. Зокрема, для  $T = 2$  маємо

$$\begin{vmatrix} 1 & r[1] \\ r[1] & 1 \end{vmatrix} = 1 - r^2[1] > 0,$$

$$-1 < r[1] < 1.$$

Аналогічно, для  $T = 3$  повинні виконуватися такі співвідношення:

$$-1 < r[1] < 1,$$

$$2r^2[1] - 1 < r[2] < 1.$$

Серед стаціонарних процесів теорії часових рядів особливу роль грають процеси типу «білий шум» [11]. Це неавтокорельовані слабо стаціонарні процеси з нульовим математичним очікуванням. Таким чином, процес  $\{a[t]\}$  є «білим шумом», якщо

$$\mu_a = E\{a[t]\} = 0,$$

$$R_a = E\{a^2[t]\} = \sigma^2,$$

$$R_a[k] = E\{a[t]a[t+k]\} = 0, \quad \forall k \neq 0.$$

Отже, для білого шуму  $R_a[k] = \sigma^2 I$ , де  $I$  одинична матриця. Назва «білий шум» пов'язана з тим, що спектральна щільність такого процесу постійна, тобто він містить в однаковій кількості всі частоти, подібно до того, як білий колір містить у собі всі інші кольори. Докладніше про це буде розказано у розділі, присвяченому спектральному аналізу.

Для виявлення наявності автокореляції випадкових величин  $a[t]$  може бути використаний метод Боксу та Пірсу [12], відповідно до якого статистика

$$Q(R_a) = n(n+2) \sum_{k=1}^m \frac{R_k^2}{n-k},$$

де

$$R_a = \frac{\text{cov}(a[t]a[t-k])}{\text{var}(a[t])},$$

має асимптотичний розподіл  $\chi_m^2$ .

Оскільки асимптотичний розподіл реальних коефіцієнтів  $R_a = (R[1], \dots, R[m])'$  кореляції - багатовимірне нормальне з нульовим середнім та дисперсією

$$\text{var } R_a[k] = \frac{n-k}{n(n+2)},$$

де, використовуючи апроксимацію дисперсії  $\text{var } R_a[k] = \frac{1}{n} \text{var}$ , Бокс та Пірс запропонували статистику

$$Q(R_a) = n \sum_{k=1}^m R_k^2,$$

має асимптотичний  $\chi_m^2$  розподіл. При подальшому вивченні було показано, що

вибіркові значення статистики Бокса-Пірса  $Q(R_a) = n \sum_{k=1}^m R_k^2$ , можуть сильно відхилитися

від розподілу  $\chi_m^2$ , тому краще уявлення дає статистика Льюнга-Бокса:

$$Q(R_a) = n(n+2) \sum_{k=1}^m \frac{\hat{R}_k^2}{n-k},$$

котра має асимптотичний  $\chi_m^2$  розподіл. Якщо розрахункове значення статистики [12] Льюнга-Бокса більше 95%-го квантилю  $\chi_m^2$  розподілу, то визнається наявність автокореляції (до m-го порядку) у послідовності  $a[t]$ . Тест не втрачає своєї

спроможності навіть за відхилення гіпотези про нормальність  $a[t]$ . Результати Андерсена та Уокера показали, що для асимптотичної збіжності  $R_a$  до багатовимірного нормального розподілу нормальності спостережень  $a[t]$  не потрібно, потрібна лише кінцівка дисперсії. Аналогічні визначення стаціонарності можна дати для векторного стохастичного процесу  $x[t]$ . Слабко стаціонарний векторний процес характеризуватиметься вже не скалярними автоковараціями та автокореляціями, а аналогічними за змістом матрицями. Поза діагоналі таких матриць стоять так звані крос-коваріації та крос-кореляції.

### 1.3 Тренди

Вивчаючи тренди, слід пам'ятати, що є, взагалі кажучи, кілька їх різновидів [13]. Першим і самим очевидним типом тренда є тренд середнього, коли часовий ряд виглядає як коливання біля повільно зростаючої або спадної величини. Другий тип трендів – це тренд дисперсії. І тут у часі змінюється амплітуда коливань змінної. Іншими словами, процес гетероскедастичний. Часто економічні процеси із зростаючим середнім мають зростаючу дисперсію. Третій, і більш тонкий тип тренду, візуально який не завжди спостерігається, — зміна значимості одного з компонент часового ряду. Наприклад, зменшення величини сезонних коливань, чи, скажімо, зміна величини кореляції між поточним і попереднім значеннями низки, тобто тренд автоковаріації та автокореляції. Проводячи розкладання ряду на компоненти, ми, як правило, маємо на увазі під трендом зміну середнього рівня змінної. Прийнято виділяти чотири основні способи апроксимації часових рядів і відповідно чотири види трендів.

- Поліноміальний тренд:

$$m[t] = m_0 + m_1 t + \dots + m_p t^p .$$

- Експоненціальний тренд:

$$m[t] = e^{m_0 + m_1 t + \dots + m_p t^p}$$

- Гармонійний тренд:

$$m[t] = A \cos(2\pi f t + \phi).$$

- Тренд, що виражається логістичною функцією:

$$m[t] = \frac{k}{1 + b e^{-m t}}$$

Оцінювання параметрів поліноміального та експоненціального трендів не становить складності. Після заміни змінних у першому випадку та логарифмування функції у другому випадку використовується звичайний метод найменших квадратів.

#### 1.4 Сезонні коливання

Якщо у складі часового ряду чітко простежуються періодичні коливання, то описи цієї періодичної складової можна використовувати функцію  $A \cos(2\pi f t + \phi)$ . Тут  $A$  - амплітуда коливань,  $f$  - лінійна частота,  $\phi$  - зсув по фазі. Якщо розглядати  $A$ ,  $f$  та  $\phi$  як невідомі параметри, то функція виявляється нелінійною і не може використовуватись у лінійній регресії. Однак, якщо частота  $f$  відома, то не складно уявити цю функцію як лінійну комбінацію синуса та косинуса:

$$A \cos(2\pi f t + \phi) = \alpha \cos(2\pi f t) + \beta \sin(2\pi f t),$$

і, розрахувавши вектори  $\cos(2\pi f t)$  і  $\sin(2\pi f t)$ , скористатися МНК для оцінювання параметрів  $\alpha$  і  $\beta$

Для моделювання сезонної складової  $s[t]$  можна використовувати формулу:

$$s[t] = \lambda_1 \delta[t] + \dots + \lambda_h \delta[ht],$$

Де  $\delta[ht]$  - сезонні фіктивні змінні, що відповідають сезонам. Ці змінні будуються в такий спосіб:  $\delta[jt]=1$ , коли спостереження відноситься до сезону  $j$   $\delta[jt]=0$  в іншому випадку. Використання лінійної регресії повного набору таких змінних пов'язані з однією особливістю. У сумі вони дають одиницю:  $\delta[t]+\dots+\delta[jt]=1$ . Тому, якщо в регресії є константа, то матиме місце лінійна залежність, і  $\delta[t], \dots, \delta[jt]$  не можна буде оцінити однозначно. Таким чином, потрібно накласти на коефіцієнти  $\lambda_1, \dots, \lambda_j$  будь-яке нормуюче обмеження. Зокрема, можна покласти один із коефіцієнтів рівним нулю, що еквівалентно невикористанню відповідної змінної при побудові регресії. Однак більш вдале нормування полягає в тому, щоб покласти  $\lambda_1 + \dots + \lambda_j = 0$ . При цьому сезонна компонента центрується, тобто в середньому вплив ефекту сезонності на рівень ряду виявляється рівним нулю.

Підставимо це обмеження у сезонну компоненту, виключивши коефіцієнт  $\lambda_1$ :

$$-(\lambda_2 + \dots + \lambda_j)\delta_1 + \lambda_2\delta_2 + \dots + \lambda_j\delta_j = \lambda_2(\delta_2 - \delta_1) + \dots + \lambda_j(\delta_j - \delta_1).$$

Нові змінні  $(\delta_2 - \delta_1), \dots, (\delta_j - \delta_1)$  будуть вже лінійно незалежними і їх можна використовувати в лінійній регресії як фактори. При необхідності оцінку  $\lambda_1$  можна отримати з оцінок інших коефіцієнтів за формулою  $\lambda_1 = -(\lambda_2 + \dots + \lambda_j)$ . Це дасть оцінку структурі сезонності  $\lambda_1, \dots, \lambda_j$ . Трактувати її слід так: у  $j$ -му сезоні сезонність призводить до відхилення від основної динаміки ряду на величину  $\lambda_j$ . Якщо для опису тренда взяти поліноміальну функцію, то, використовуючи адитивну схему, можна уявити часовий ряд у вигляді наступної лінійної регресії:  $x[t] = \Phi_0 + \Phi_1 + \dots + \Phi_p + \lambda_1\delta_1 + \dots + \lambda_j\delta_j + a[t]$ , де  $\lambda_2 + \dots + \lambda_j = 0$ . У цій регресії  $\Phi_i$  і  $\lambda_j$  є невідомими коефіцієнтами. Цю модель можна оцінити за допомогою звичайного МНК, оскільки виконані всі звичайні умови його застосування. МНК дає оцінки  $p+h+1$  невідомих коефіцієнтів і, таким чином, призводить до виділення складових декомпозиції [7].

### 1.5 Модель авторегресії – ковзного середнього стаціонарного складника декомпозиції нестаціонарного процесу

До складу моделі декомпозиції входить стаціонарна корельована складова. Зазвичай її представляють параметричною моделлю лінійного передбачення. Найбільш повною параметричною моделлю лінійного прогнозу стаціонарних випадкових процесів є змішана модель АРКС [14]. Об'єднання 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 [15]. 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.

В операторному вигляді модель АРКС(p,q) описується виразом [14]

$$\Phi(z)x[t] = Q(z)a[t], \quad (1.2)$$

де оператори  $\Phi(z)$  і  $Q(z)$  є операторами моделей АР і КС р-го та q-го порядків, відповідно. Різниця рівняння моделі АРКС (1.2) зручно записати у вигляді

$$\begin{aligned} x[t] = & \Phi[1]x[t-1] + \Phi[2]x[t-2] + \dots + \Phi[p]x[t-p] - Q[1]a[t-1] - \\ & - Q[2]a[t-2] - \dots - Q[q]a[t-q] + a[t]. \end{aligned} \quad (1.3)$$

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 (1.3) на  $x[t-j]$  та взяти математичне очікування. Після нескладних перетворень отримаємо

$$R[j] = \Phi[1]R[j-1] + \Phi[2]R[j-2] + \dots + \Phi[p]R[j-p] + \dots + R_{xa}[j] - Q[1]R_{xa}[j-1] - \dots - Q[q]R_{xa}[j-q], \quad (1.4)$$

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

$$R_{xa}[j] = E\{x[t-j]a[t]\}. \quad (1.5)$$

Взаємна кореляційна функція не дорівнює нулю, якщо в рівнянні для  $x[t-j]$  не міститься  $a[t]$ . Як випливає з (1.5) та (1.3), взаємна кореляційна функція задовольняє співвідношенням

$$R_{xa}[j] = 0, j > 0, \quad (1.6a)$$

$$R_{xa}[j] \neq 0, j \leq 0. \quad (1.6b)$$

З урахуванням (1.4), з (1.6) випливає, що

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

або

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

Вираз для дисперсії процесу АРКС можна знайти з (1.4), поклавши  $j=0$

$$\begin{aligned}
 R[0] = & \Phi[1]R[1] + \Phi[2]R[2] + \dots + \Phi[p]R[p] - Q[1]R_{xa}[-1] - \\
 & - Q[2]R_{xa}[-2] - \dots - Q[q]R_{xa}[-q] + D_a.
 \end{aligned} \tag{1.7}$$

Тоді вираз (1.7) для коефіцієнтів кореляції процесу АРКС набуває вигляду

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

Можна показати, що СГП моделі АРКС визначається виразом [14]

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

Як очевидно з (1.8), процеси АРКС описуються полюсно-нульової СЦП. Спектр такого процесу містить вузькосмугові піки, що визначаються моделлю АР, та широкосмугові максимуми, що визначаються моделлю КС. Для прогнозу найчастіше використовується модель АР. Подобиці цієї моделі будуть використані під час подальшого викладу матеріалу кваліфікаційної роботи магістра.

## 1.6 Висновки до розділу та постановка задачі

У кваліфікаційній роботі аналізуються деякі класи випадкових нестационарних процесів, що допускають можливість прогнозування. Нестационарні дані, що розглядаються, відрізняються за величиною тренду, періодом і потужністю сезонних коливань, моделями стаціонарної складової. Під декомпозицією моделі нестационарного випадкового процесу розумітимемо її подання сумою тренду,

сезонної складової та стаціонарної статистичної моделі. Модель декомпозиції випадкового процесу з одного боку дозволяє вирішувати завдання визначення довільного тренду, сезонної складової та стаціонарної складової. З іншого боку, ця модель дозволяє вирішувати завдання прогнозування, тобто на один або кілька кроків наперед передбачати значення нестационарного процесу.

У кваліфікаційній роботі розглянуто також питання побудови моделі авторегресії-проінтегрованого ковзного середнього та її зв'язок із декомпозицією нестационарного процесу. Показано вплив операторів взяття різниці, що враховують складові нестационарного процесу, на точність прогнозу та декомпозицію. Точна компенсація тренду та сезонної складових можлива лише у разі їх детермінованості та суворої періодичності. У разі нездійсненності цих вимог модель АРПКС та декомпозиція неточності та їх похибки компенсуються моделлю стаціонарної складової, що представляється у загальному випадку моделлю АРПКС.

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

1. Сформулювати завдання аналізу нестационарних процесів у статистичній радіотехніці для подальшого прогнозу.
2. Для часових рядів зміни температур атмосфери та трафіку в мобільному радіозв'язку побудувати моделі декомпозиції та АРПКС складових у вигляді тренду, сезонного та стаціонарного процесу.
3. Визначити взаємний вплив операторів зсуву, що використовуються для обліку складових декомпозиції.
4. Знайти способи оцінювання тренду та сезонної складової нестационарного процесу.
5. Дослідити завдання прогнозування на прикладі трафіку нестационарного процесу мобільного радіозв'язку.

## 2 МОДЕЛІ АВТОРЕГРЕСІЇ-ПРОІНТЕГРОВАНОГО КОВЗНОГО СЕРЕДНЬОГО І ДЕКОМПОЗИЦІЇ НЕСТАЦІОНАРНИХ ДАНИХ

### 2.1 Модель авторегресії проінтегрованого ковзного середнього нестационарного процесу

У кваліфікаційній роботі магістра аналізуються деякі класи випадкових нестационарних процесів, що допускають можливість прогнозування [14]. Нестационарні дані, що розглядаються, відрізняються за величиною тренду, періодом і потужністю сезонних коливань, моделями стаціонарної складової. Під декомпозицією моделі нестационарного випадкового процесу будемо розуміти її подання сумою тренду, сезонної складової та стаціонарної статистичної моделі [7]. Модель декомпозиції випадкового процесу з одного боку дозволяє вирішувати завдання визначення довільного тренду, сезонної складової та стаціонарної складової. З іншого боку, ця модель дозволяє вирішувати завдання прогнозування, тобто на один або кілька кроків наперед передбачати значення нестационарного процесу.

У кваліфікаційній роботі магістра розглянуто також питання побудови моделі авторегресії-проінтегрованого ковзного середнього (АРПКС) та її зв'язок із декомпозицією нестационарного процесу. Показано вплив операторів взяття різниці, що враховують складові нестационарного процесу, на точність прогнозу та декомпозицію. Точна компенсація тренду та сезонної складових можлива лише у разі їх детермінованості та суворої періодичності. У разі нездійсненності цих вимог модель АРПКС та декомпозиція неточні та їх похибки компенсуються моделлю стаціонарної складової, що представляється у загальному випадку моделлю АРПКС.

Для аналізу нестационарних процесів використовують два підходи. Перший, запропонований у роботах [14], заснований на побудові моделі АРПКС. Він природний, проте, складно реалізований підхід, з урахуванням декомпозиції нестационарного випадкового процесу. Складність полягає в тому, що подання

нестационарного процесу у вигляді суми тренду, сезонної складової та стаціонарної складової при оцінці стикається з деякими проблемами, які не дозволяють виключити вплив оцінки однієї складової на іншу.

Розглянемо особливості побудови моделей АРПКС нестационарних процесів. Згідно [7], нестационарний процес, що моделюється, може мати детермінований тренд. В рамках моделі нестационарних процесів сезонна складова повинна бути періодичною функцією з певним періодом. Завданням побудови моделі АРПКС є видалення з вихідних даних тренду та сезонної складової за допомогою відповідних операторів взяття різниці та побудова моделі лінійного передбачення АРКС стаціонарного залишку. Після цього можна вирішувати завдання прогнозу шляхом відновлення тренду та сезонної складової нестационарного процесу, а також побудови моделі АРКС стаціонарної складової.

Модель АРПКС [7] нестационарного процесу  $x[t]$  представляється як мультиплікативний процес

$$\omega[t] = \nabla^d \nabla_s x[t] = (1-z)^d (1-z^s)x[t] = \nabla^d \omega_1[t]. \quad (2.1)$$

Оператори взяття різниці, що входять до (2.1), враховують сезонну складову та тренд шляхом їх віднімання. Якщо тренд або сезонна складова видаляються оператором взяття різниці, цей оператор добре враховує ці складові. Зауважимо, що за такого підходу до побудови моделі АРПКС нестационарного процесу  $x[t]$  випадковий процес  $\omega[t]$  вже можна вважати стаціонарним. Навіть якщо оператори видалення сезонної складової та тренду не повністю їх враховують. Стаціонарний випадковий процес описується у випадку моделлю АРКС.

Для видалення сезонної складової (2.1) застосовується оператор  $\nabla_s = 1 - z^s$ , де  $z^s$  - оператор зсуву, дія якого визначається виразом  $z^s x[t] = x[t-s]$ . Процес видалення сезонної складової за допомогою оператора зсуву є виразом

$$\omega_1[t] = \nabla_s x[t] = (1 - z^s)x[t]. \quad (2.2)$$

Детермінований тренд у  $\omega_1[t]$ , визначається оператором  $\nabla^d = (1-z)^d$ . Тоді процес без тренда можна уявити у вигляді

$$\omega[t] = (1-z)^d \omega_1[t]. \quad (2.3)$$

Як правило, обмежуються найпростішими випадками лінійного та квадратичного трендів. Для цих випадків вважають  $d = 1, d = 2$  відповідно. Тоді

$$\omega[t] = \nabla \omega_1[t] = \omega_1[t] - \omega_1[t-1],$$

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

Таким чином, тренд (2.2) та сезонна складова (2.3) враховуються операторами зсуву шляхом їх видалення. За прогнозом ці детерміновані складові відновлюються. Нагадаємо, що повністю їх врахувати можна, якщо описані тут оператори взяття різниці, діють на сезонну складову з чітко вираженими коливаннями з детермінованими частотою та амплітудою, а тренд змінюється за поліноміальним законом.

## 2.2 Модель декомпозиції нестационарних даних

При ширшому аналізі нестационарних процесів інтерес представляє також процедура знаходження довільного тренду та сезонної складової, що виходить за межі побудови АРПКС моделі. При цьому форма тренду та властивості сезонної складової можуть змінюватися у ширших межах, ніж у моделі АРПКС. Це завдання називають декомпозицією моделі [7] нестационарного часового ряду, яка полягає у його розкладанні на довільні складові.

Декомпозиція в моделі нестационарного процесу сприймається як уявлення нестационарного процесу у вигляді

$$x[t] = m[t] + c[t] + \omega[t], \quad (2.4)$$

де  $x[t]$ - нестационарний процес  $m[t]$ ,  $c[t]$ ,  $\omega[t]$ - тренд, сезонна складова, корельована випадкова стаціонарна складова відповідно. У силу того, що при видаленні однієї складової, оператор узяття різниці діє і на інші складові нестационарного процесу, виникають похибки декомпозиції моделі нестационарного процесу.

Зауважимо, що оператор  $\nabla_s = 1 - z^s$ , що усуває сезонні коливання в (2.1), діє не тільки на цю складову процесу, а й на трендову складову. При цьому стаціонарна складова нестационарного процесу також змінюється. Після дії оператором  $\nabla^d = (1 - z)^d$  на тренд, інші складові моделі декомпозиції нестационарного процесу можуть також змінитися. Найбільше впливає на нестационарний процес оператор видалення сезонної складової. Після видалення сезонної складової оператором  $\nabla_s = 1 - z^s$ , декомпозиція (2.4) моделі нестационарного процесу набуває вигляду

$$\nabla_s x[t] = m[t] - m[t - s] + c[t] - c[t - s] + y[t] - y[t - s], \quad (2.5)$$

де  $\nabla_s x[t]$  - нестационарний процес без сезонної складової. Якщо сезонна складова суворо детермінована, то  $c[t] = c[t - s]$  та декомпозиція (2.5) спрощується [7]

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

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

### 2.3 Неточності, що виникають під час використання операторів взяття різниці

Для аналізу впливу операторів взяття різниці при усуненні сезонної складової чи трендів, використовувалися дані середньомісячних температур атмосфери (3108 відліків) (рис. 2.1), отримані Пулковською обсерваторією за 259 років з 1752 по 2010 роки [4].

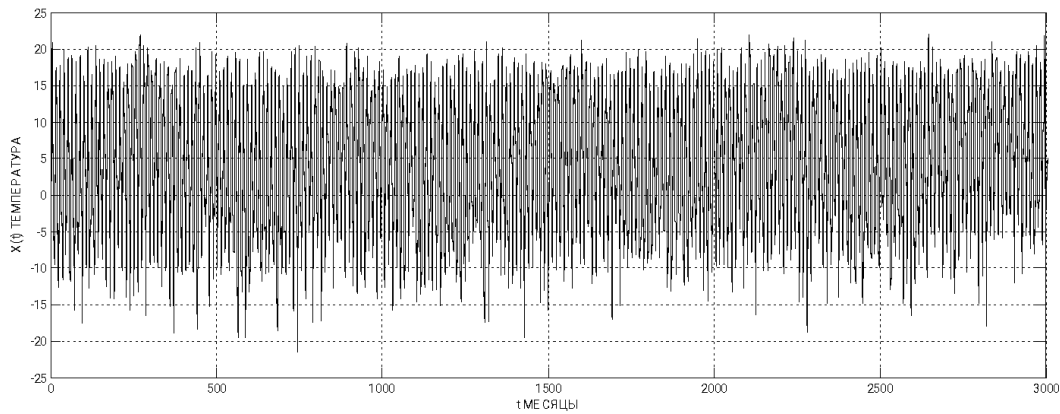


Рисунок 2.1 - Дані середньомісячних температур, зафіксованих у метеообсерваторії міста Пулково (Санкт-Петербург)

Також досліджувався процес (рис. 2.2), який є трафіком передачі даних у стільниковому зв'язку [5]. Довжина цього ряду складає 365 відліків, що відповідає 365 дням. Він має значний зростаючий тренд. Дані містять сезонну складову, що дорівнює, очевидно, 7 днів і стаціонарний процес, який описується, в загальному випадку, моделлю АРКС.

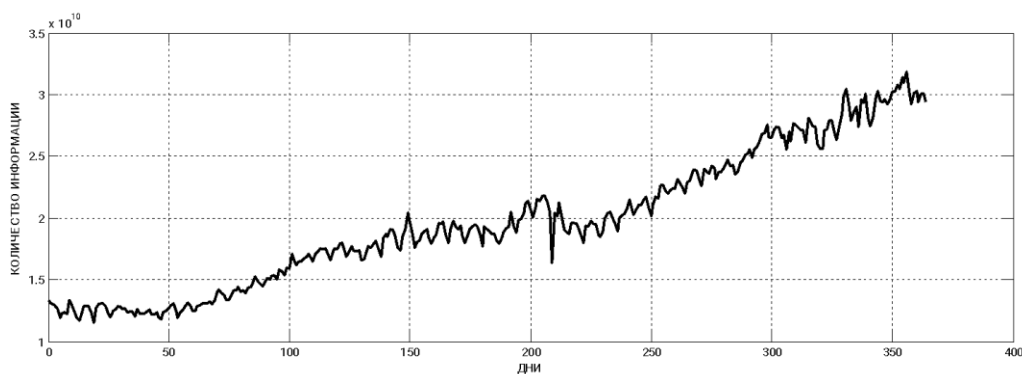


Рисунок 2.2 - Трафік передачі даних у стільниковому зв'язку

На рис. 2.3 представлений графік сезонного процесу, отриманого після усунення тренду оператором  $\nabla^d = (1-z)^d$ . Наявність сезонної складової не можна пояснити адитивним детермінованим процесом. Складові складніша, і має всі ознаки деякої випадковості. Як видно з графіка, в процесі присутні досить різні періоди коливань, а також помітні амплітуди, що відрізняються в сусідніх періодах.

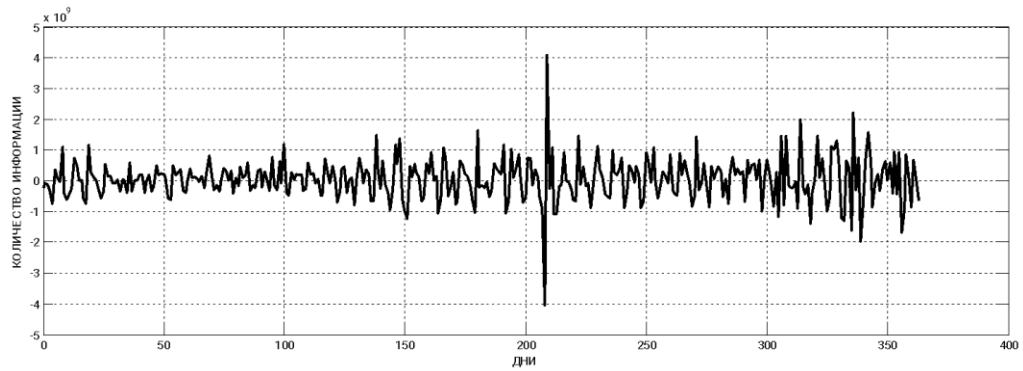


Рисунок 2.3 - Трафік передачі даних у стільниковому зв'язку із сезонною складовою

Безтрендовий процес має кореляційну функцію з вираженою незатухаючою циклічністю (рис. 2.4). Від періоду до періоду у кореляційній функції спостерігаються не строго періодичні коливання. Аналіз реальних даних та кореляційної функції показав, що період сезонних коливань, швидше за все, дорівнює 7 днів.

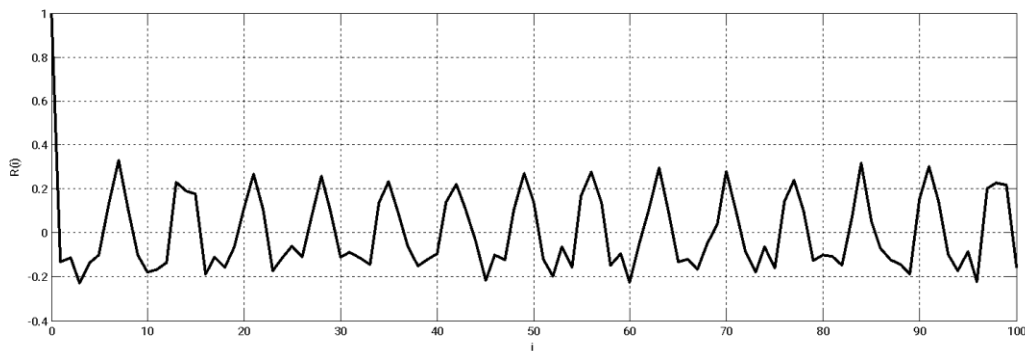


Рисунок 2.4 - Кореляційна функція трафіку передачі даних у стільниковому зв'язку із сезонною складовою

Таким чином, рівність  $c[t] = c[t - s]$  суворо не виконується і формула (2.6) не є точною. Декомпозиція випадкового нестационарного процесу (2.5) часто припускає відхилення, що є предметом досліджень. При цьому знайдене рішення може бути неоптимальним, але більш простим, що задовольняє умовам дослідження.

На рис. 2.5 та рис. 2.6 показані дані трафіку та середньомісячних температур після видалення сезонних складових. На даних, що представлені на рис. 2.5, не видно тренд, хоча він має бути присутнім. Адже видаляється лише сезонна складова моделі декомпозиції, а трендова, дуже важлива, залишається.

Під час аналізу графіка на рис. 2.6 також складно визначити довготривалий тренд температур. Таким чином, видалення сезонної складової оператором  $\nabla_s = 1 - z^s$  з вихідних даних також суттєво впливає і на характеристики тренду.

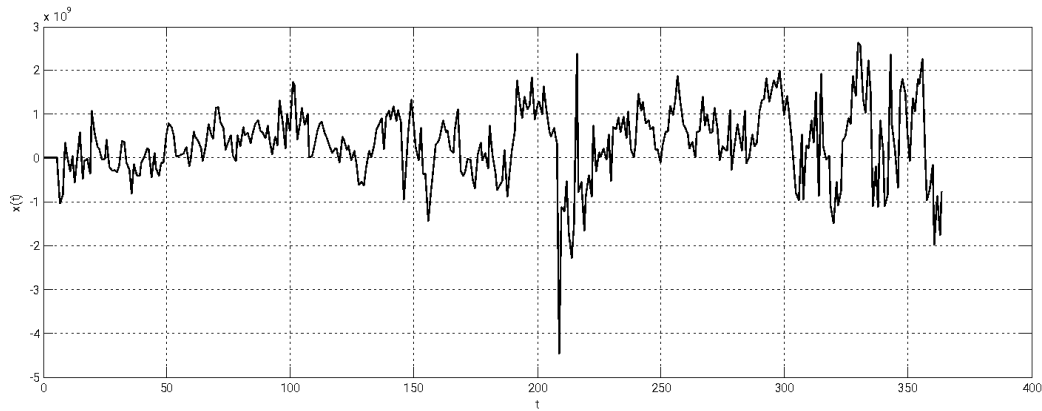


Рисунок 2.5 - Трафік передачі після видалення сезонної складової

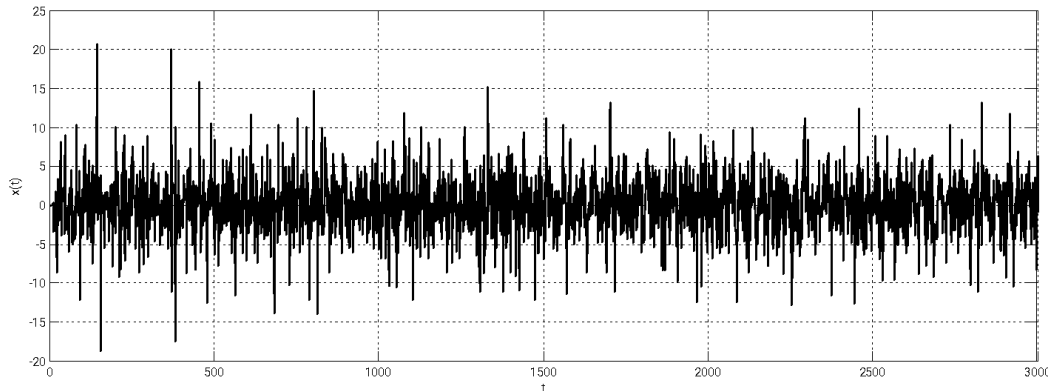


Рисунок 2.6 - Вибір даних середньомісячних температур після видалення сезонної складової

Зауважимо, що оператори, що усувають сезонні коливання, діють не лише на сезонність процесу, а й на інші складові нестационарного процесу [4]. Як показали експерименти, видалення сезонної складової сильно впливає тренд і слабо впливає на стаціонарну складову процесу АРПКС. Операція видалення тренду слабо впливає на властивості сезонної складової і на стаціонарну складову процесу.

Наявність сезонної складової даних, можна визначити за графіками параметричних спектрів [3]. Вони показані на рис. 2.7 та рис. 2.8. Аналіз графіка спектра, представленого на рис. 2.7 показує, що отримана сезонна складова трафіку даних менш виражена. Форма спектру дає підставу вважати, що в отриманих графіках,

після диференціювання даних трафіку, є постійна складова, сезонна складова з періодом 6,8 днів, а також слабкі коливання з періодами 3,4 і 2,3 дня.

Як видно із рис. 2.8 сезонна складова більш явно виражена в даних температур. З аналізу спектра, показаного на рис. 2.8 слід зазна, що період коливань сезонної складової дорівнює 11,9 днів. Зазначимо, що для деяких форм сезонної складової, диференціювання, що використовується, не змінює період коливань.

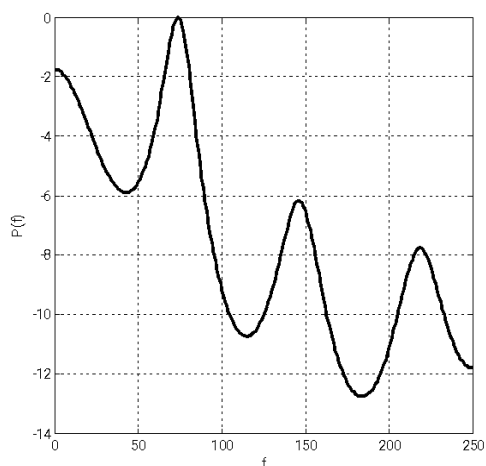


Рисунок 2.7 - СЦП після видалення тренду з даних трафіку

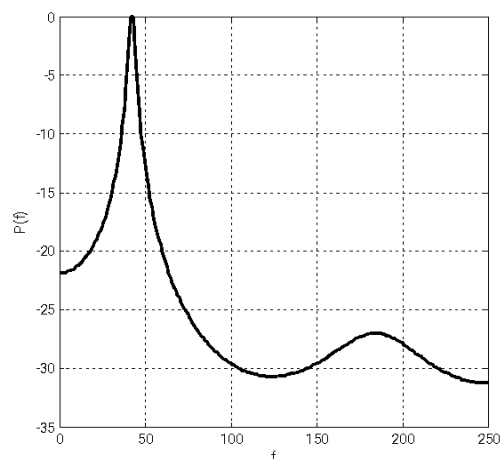


Рисунок 2.8 - СЦП після видалення тренду із даних температур

### 3. ОЦІНКА ТРЕНДУ НЕСТАЦІОНАРНОГО ПРОЦЕСУ

#### 3.1 Нестационарні процеси з трендом

У кваліфікаційній роботі магістра аналізувалась можливість знаходження параметрів слабких довгострокових змін випадкового процесу за наявності сезонного періодичного процесу, тренду та стаціонарного випадкового процесу. Модель декомпозиції була представлена саме цими складовими. Проблема, запропонована в моделі АРПКС [14], полягає в тому, що використання для опису тренду і сезонної складових, операторів зсуву, тягне за собою зміни цих складових. Так, використання оператора зсуву, що враховує тренд, змінює також сезонну складову. А використання оператора зсуву, що описує сезонну складову, змінює тренд. У розділі 2 було це показано. Нижче, з прикладу трьох реальних процесів, показано виділення трендів. Тренди відображають довготривалу зміну процесів, яка останнім часом, дедалі більше цікавить дослідників. Наприклад, завдання оцінки довгострокової зміни клімату Землі, обліку антропогенного чинника на цей процес.

Для аналізу довгострокової зміни температур використовувалися дані середньомісячних температур атмосфери (3108 відліків) (рис. 3.1), отримані Пулковською обсерваторією за 259 років з 1752 по 2010 роки.

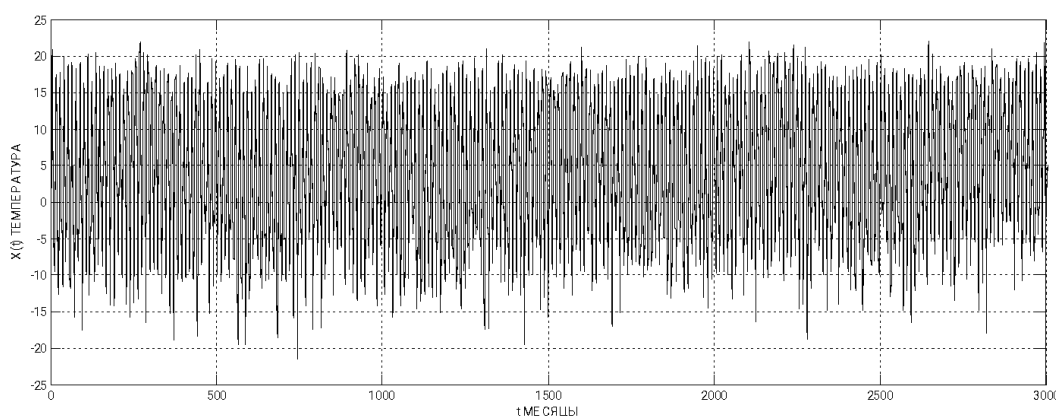


Рисунок 3.1 - Дані середньомісячних температур, зафіксованих у метеообсерваторії міста Пулково (Санкт-Петербург)

На рис. 3.2 подано дані середньомісячних температур, зафіксованих метеообсерваторією міста Сент-Луїс, шт. Міссурі [3] з 1845 по 1978 р.р. Подані на рис.

3.1 та рис. 3.2 дані містять слабкий тренд, пов'язаний із довготривалою зміною температури Землі в цих точках, сезонні коливання з періодом 12 місяців та стаціонарний процес змін середньомісячних температур.

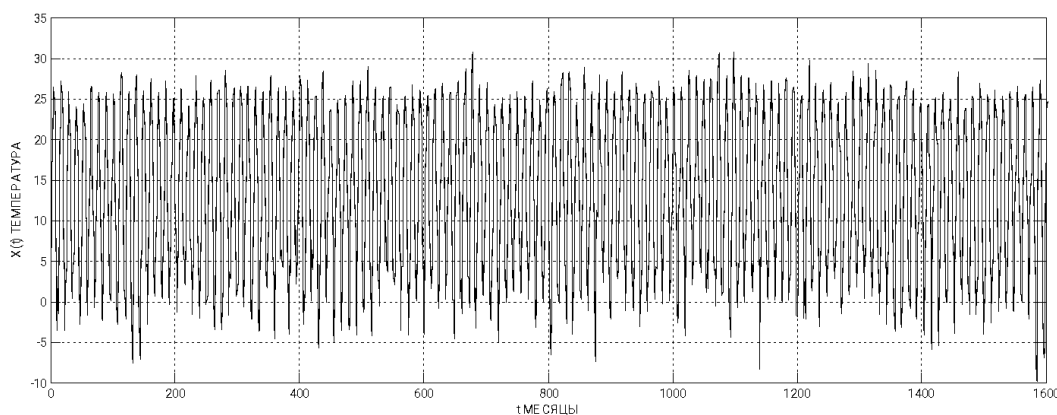


Рисунок 3.2 - Дані середньомісячних температур, зафіксованих у метеообсерваторії міста Сент-Луїс

Процес (рис. 3.3) є трафік передачі даних у стільниковому зв'язку. Довжина ряду становить 365 відліків, що відповідає 365 днів. Він має значний зростаючий тренд. Дані містять сезонну складову, що дорівнює 7 днів і стаціонарний процес, який описується у загальному випадку моделлю АРКС.

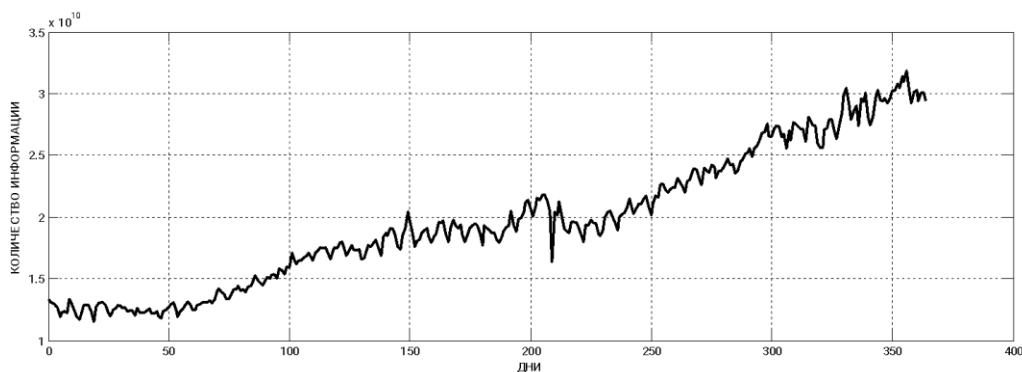


Рисунок 3.3 - Трафік передачі даних у стільниковому зв'язку

Як видно із графіків на рис. 3.1 рис. 3.2 та рис. 3.3 тренди нестационарних процесів мають довільну форму, а не детерміновану (лінійну, параболічну та ін), як це визначено в моделі АРКС. У цілому тренди неоднорідні, т.я. у них є ділянки з різною швидкістю зміни.

Операція обліку тренду шляхом його видалення оператором  $\nabla^d = (1-z)^d$ , при  $d = 1$  менше спотворює дані. Однак слід мати на увазі, що дія оператора  $\nabla = (1-z)$  суттєво зводиться до диференціювання випадкового процесу, тобто нам доводиться досліджувати властивості не безтрендового процесу, а його дискретної похідної.

На рис. 3.4 рис. 3.5 та рис. 3.6 показані дані середньомісячних температур та трафіку даних у стільниковому зв'язку після видалення сезонної складової. Аналіз даних на рис. 3.4 рис. 3.5 та рис. 3.6 показує, що при використанні зазначеного вище оператора тренд дійсно видаляється.

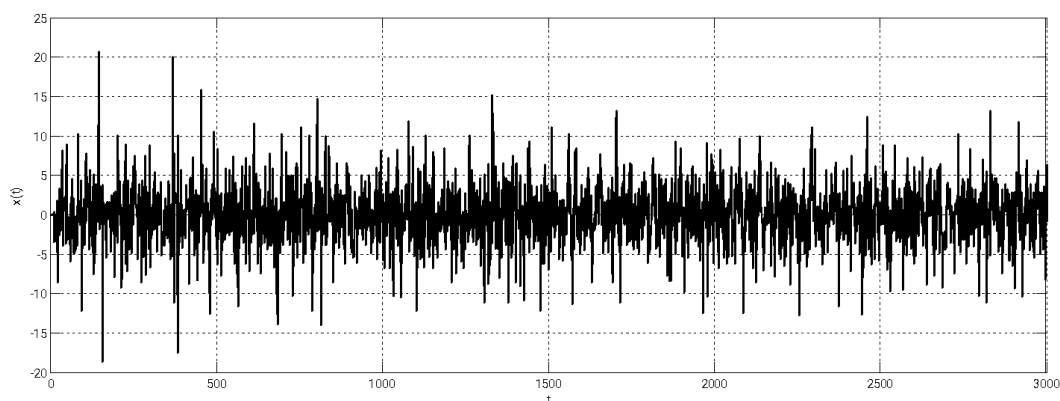


Рисунок 3.4 - Часовий ряд даних середньомісячних температур, зафіксованих у метеообсерваторії міста Пулково після видалення сезонної складової

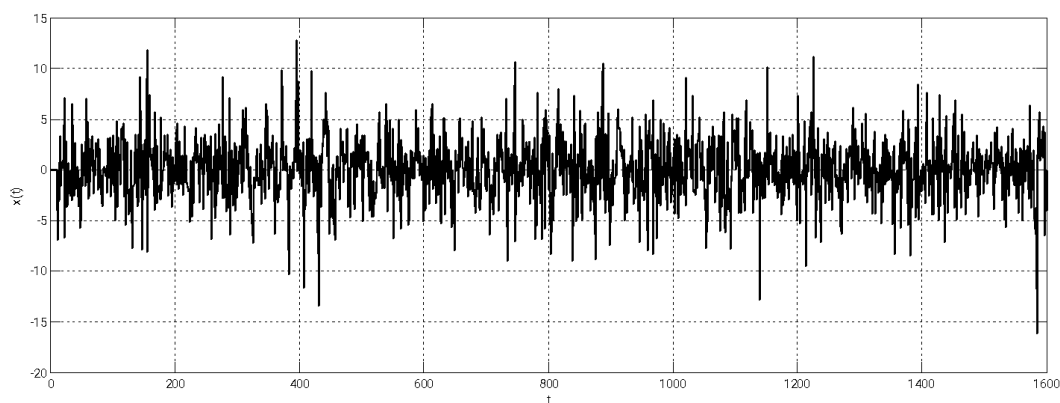


Рисунок 3.5. - Часовий ряд даних середньомісячних температур, зафіксованих у метеообсерваторії міста Сент-Луїс після видалення сезонної складової

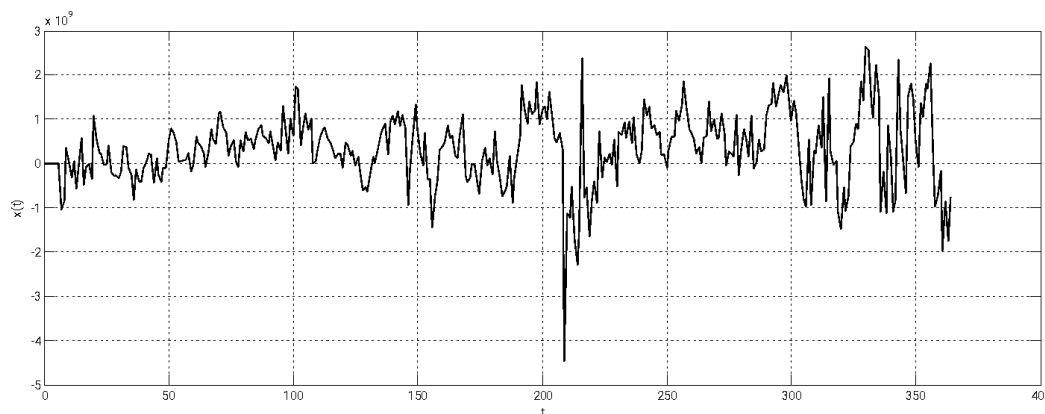


Рисунок 3.6 - Трафік передачі після видалення сезонної складової

Проблема довготривалої зміни клімату Землі під впливом природних і антропогенних чинників цікавить не тільки кліматологів, а й політиків, економістів та інших експертів і фахівців. З'ясувалося, що на температуру атмосфери впливає багато чинників: склад атмосфери; промислові викиди у повітря; активність Сонця; періодичні похолодання на Землі, що призводять до льодовикових періодів; зіткнення Землі з астероїдами та метеоритами; викиди у повітря великих вулканів; зміни у нерівномірному нагріванні земної поверхні та інших. Особливо спорним виявився вплив людини через викиди промисловими підприємствами, які впливають склад атмосфери Землі, отже, і температуру Землі. На думку експертів, подолання кліматичних аномалій коштуватиме трильйони доларів на рік протягом 30 років. Тому актуальним є облік впливу людини на температуру Землі та запобігання льодовикових періодів. Для цього необхідно, зокрема, удосконалення аналізу змін температури, що фіксуються метеостанціями у різних точках Землі.

Однією з цілей кваліфікаційної роботи магістра були дослідження, чи був аналіз методу оцінювання статистичних характеристик, довгостроковою складовою середньомісячних температур. Отримані результати можуть використовуватися для аналізу середньострокових і довгострокових змін атмосферних явищ, уточнення результатів, отриманих традиційними методами математичної статистики.

Аналізувалася можливість знаходження параметрів слабких довгострокових змін випадкового процесу за наявності потужного короткострокового сезонного періодичного процесу. Для аналізу довгострокової зміни температур використовувалися по-перше дані середньомісячних температур атмосфери (3108

відліків) (рис. 3.1), отримані Пулковською обсерваторією за 259 років з 1752 по 2010 роки. По-друге дані середньомісячних температур атмосфери, отримані у місті Сент-Луїс (1608 відліків) (рис. 3.2).

### 3.2 Виділення тренду температур

За даними миттєвих значень середньомісячних температур показаних на рис. 3.1 та рис. 3.2, важко визначити, чи є тренд температур. Видалення сезонної складової з використанням моделі АРПКС не дозволяє виділити довготривалий тренд температур.

Знайдемо тренди процесів, поданих на рис. 3.1 та рис. 3.2 шляхом усереднення. Для цього щоденний трафік передачі даних (рис. 3.1) представимо у вигляді ряду середньотижневих даних

$$x^n[t] = \overline{x_0[t]}, \overline{x_1[t]} \dots \overline{x_{m-1}[t]}, \quad (3.1)$$

де

$$\overline{x_k[t]} = \frac{1}{n} \sum_{i=1}^n x_i[t], \quad n=7. \quad (3.2)$$

У (3.1) та (3.2) час визначено як номер даних середнього значення трафіку за тиждень.

Для оцінки довготривалого тренду температур, середньомісячні температури, представлені на рис. 3.2, також перетворюються за допомогою виразу (3.1), де середньорічні значення температур обчислюються як

$$\overline{x_k[t]} = \frac{1}{n} \sum_{i=1}^n x_i[t], \quad n=12. \quad (3.3)$$

Для цих даних (3.1) і (3.3) час визначено як послідовність років.

Для оцінки тренду, знайдемо спочатку часовий ряд ковзним усередненням по кожному з відрізків, що складається з 12 місяців

$$\omega_1[t'] = \frac{1}{n} \sum_{i=1}^n x[i+t'-1], \quad n=12, \quad (3.4)$$

Де  $t' = 1, \dots, m$ , а  $m$  – кількість років спостережень середньомісячних температур. На рис. 3.7 представлені дані середньорічних змін температури, показаних на рис. 3.1.

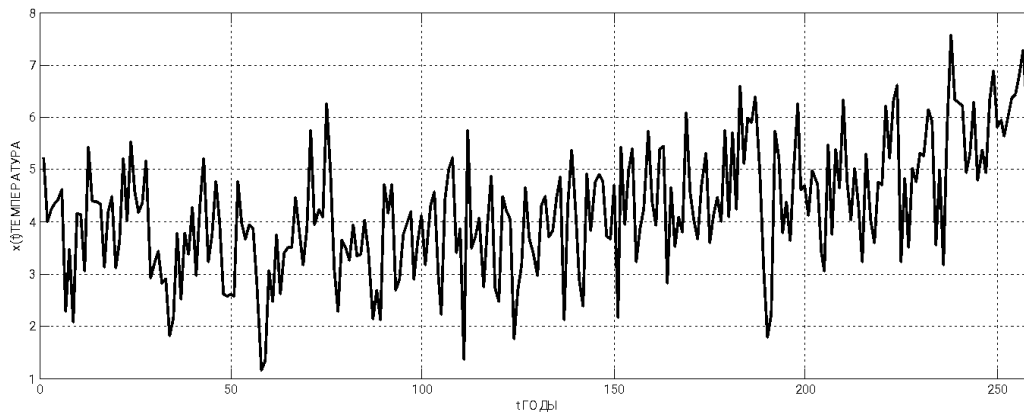


Рисунок. 3.7 - Зміни середньорічних температур у м. Пулкове

На рис. 3.8 представлені дані середньорічних змін температури, показаних на рис. 3.2, знайдені за допомогою (3.4). Як показує аналіз даних, представлений на графіках (рис. 3.7 та рис. 3.8), спостерігається тренд середньорічних температур.

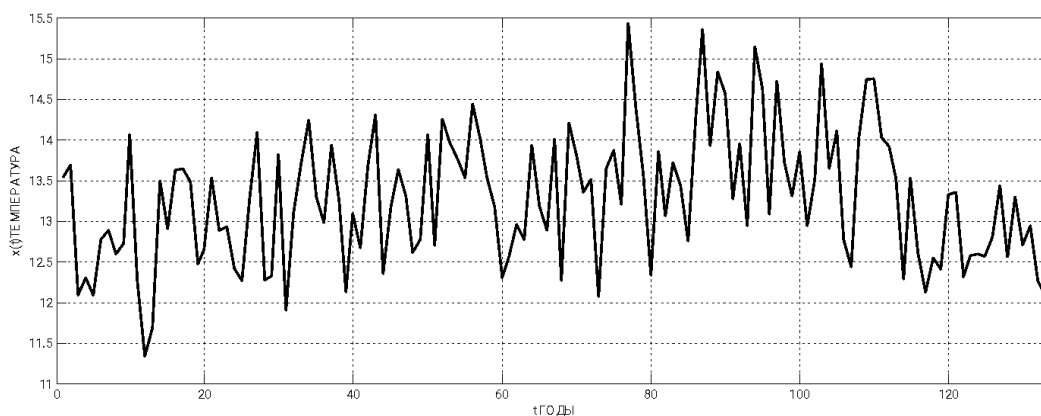


Рисунок. 3.8 - Зміни середньорічних температур у місті Сент-Луїс

### 3.3 Чисельна оцінка тренду

Щоб отримати тренд довготривалого зміни температур протягом періоду спостереження, згладимо середньорічні значення, показані рис. 3.5 та рис. 3.6, за допомогою низькочастотного фільтра. Для зручності представлення даних

продовжимо кожне значення тренда на 12 відліків. Тоді отримаємо часові ряди первісної довжини.

Зауважимо, що процедура ковзного усереднення не дає можливість отримати на початку часового ряду у його кінці таке ж точне усереднення, як і всередині часового ряду. Це є очевидними особливостями процедури використовуваного усереднення, коли на початку часового ряду відсутні необхідні у методі усередненні попередні значення часового ряду, а кінці ряду – наступні значення.

Аналіз тренду рис. 3.9 показує, що з 1752 до 1803 року температура падала з 4,045 °C до 3,446 °C. Спад температури становив 0,599 °C. З 1804 року спостерігалось зростання температури з 3,446 °C до 5,961 °C у 2010 році. Приріст температури за цей термін становив 2,515 °C.

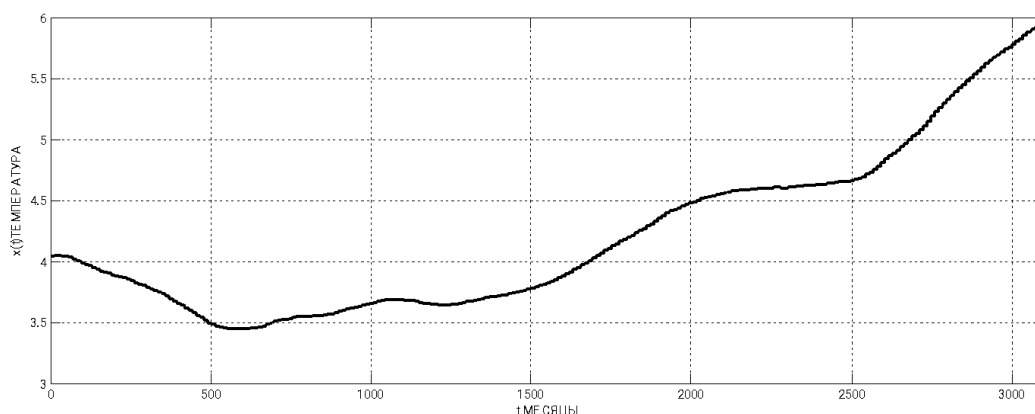


Рисунок 3.9 - Тренд температур у Пулковому, отриманий ковзним середнім від зміни середньорічних температур

Аналіз тренду на рис. 3.10 показує, що з 1845 до 1935 року температура зросла з 12,82 °C до 13,76 °C. Зростання температури склало 0,947 °C. Але з 1936 спостерігалось падіння температури з 13,76 °C до 12,79 °C в 1978 році. Падіння температури за цей термін становило 0,974 °C. тобто зростання температури у цій місцевості спостерігалось лише з 1845 року до 1935 року, а потім, до кінця спостережень у 1978 року, температура падала. Це вказує на те, що останнім часом температура повітря не для всіх місць Землі має тенденцію до зростання.

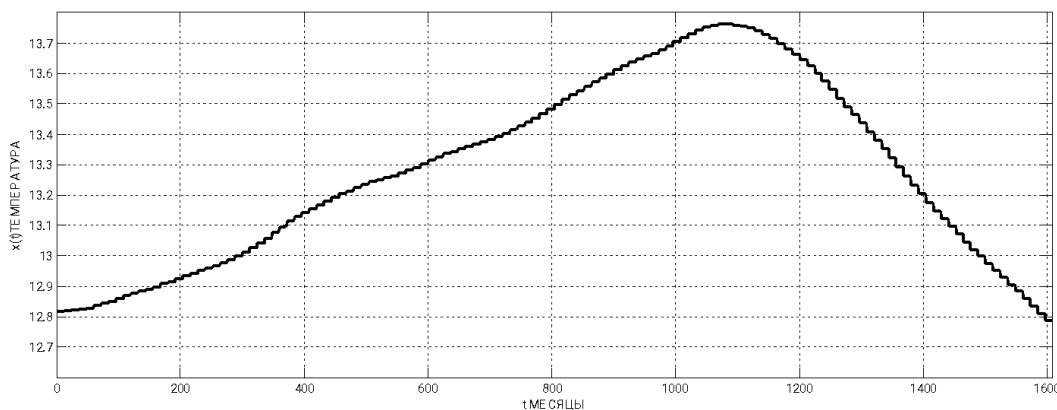


Рисунок 3.10 - Тренд температур у місті Сент-Луїс, отриманий ковзним середнім від зміни середньорічних температур

Приклади нестационарних процесів, що досить сильно відрізняються, з трендом представлені на рис. 3.11 та рис. 3.12. Перший процес (рис. 3.11) являє собою дані про середньомісячні температури, зафіксовані в метеообсерваторії міста Пулково. Повна довжина ряду становить 3108 значень температури, отримані обсерваторією за 259 років з 1752 по 2010 роки. Дані містять слабкий тренд, яскраво виражені сезонні коливання температур із періодом 12 місяців та стаціонарний процес.

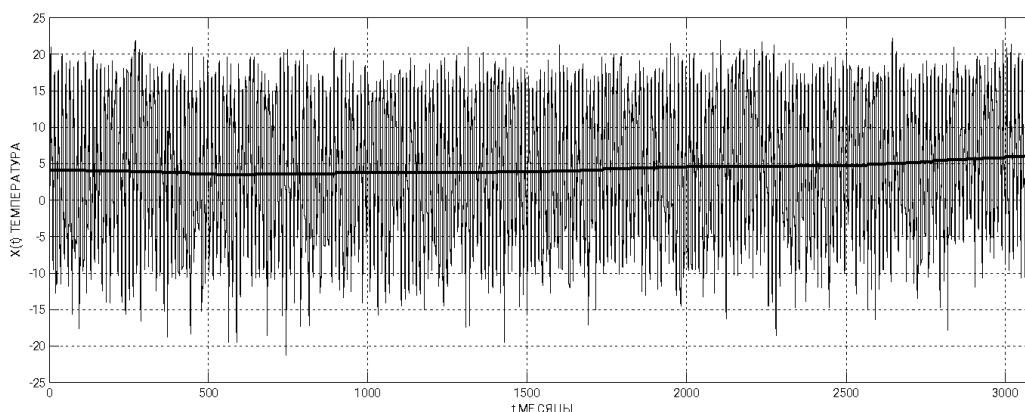


Рисунок 3.11 - Дані середньомісячних температур, зафіксованих у метеообсерваторії міста Пулково та оцінка їх довготривалого тренду

Другий процес (рис. 3.12) являє собою дані про середньомісячні температури, зафіксовані в метеообсерваторії міста Сент-Луїс. Повна довжина ряду становить 1608 значень температури, отримані обсерваторією за 133 роки з 1845 по 1978 роки. Дані містять слабкий тренд, яскраво виражені сезонні коливання температур із періодом 12 місяців та стаціонарний процес.

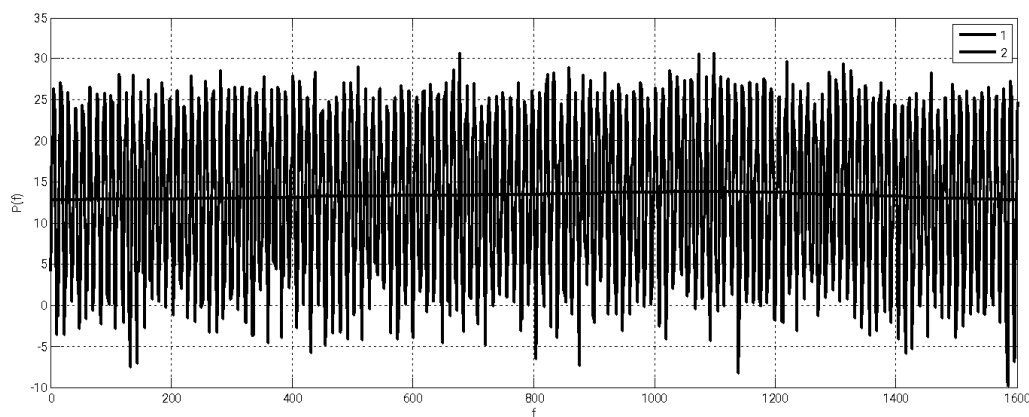


Рисунок 3.12 - Дані середньомісячних температур, зафіксованих у метеообсерваторії міста Сент-Луїс та оцінка їх довготривалого тренду

Для наочності проблеми виділення тренда, на графіки процесів зміни температур накладено отримані тренди. Як видно із графіків на рис. 3.11 та рис. 3.12, знайдені тренди мають довільну форму, а не детерміновану (лінійну, параболічну та ін), як це визначено в моделі АРПКС. У цілому нині тренди неоднорідні, так як у них є ділянки з різною швидкістю зміни.

За даними фахівців, протягом останніх 140 років середньорічна температура Землі підвищилась приблизно на  $1^{\circ}\text{C}$ . У заяві вчених також зазначається, що глобальні середні температури у 2019 році були на  $1,1^{\circ}\text{C}$  вищими порівняно з доіндустріальним періодом (1850–1900 рр.).

Третій процес (рис. 3.13) є трафік передачі даних у стільниковому зв'язку [16, 17]. Довжина ряду складає 365 відліків, що відповідає 365 дням. Він має значний зростаючий тренд та слабкі сезонні коливання. Сезонна складова має період, що дорівнює приблизно 7 днів. Стаціонарний процес, описується у випадку моделлю АРКС. Для отримання тренду згладимо середньотижневі дані (3.13) низькочастотним ковзаючим фільтром. Виділений у такий спосіб тренд також представлений на рис. 3.13.

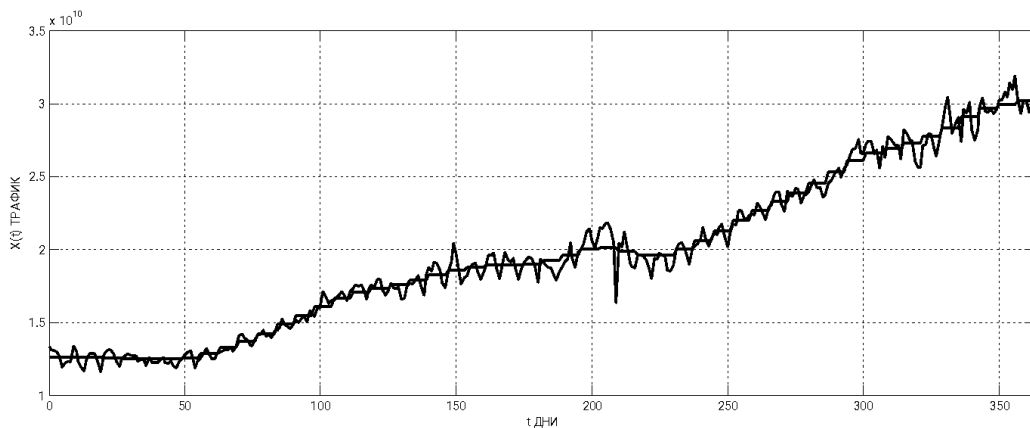


Рисунок 3.13 - Трафік передачі даних у стільниковому зв'язку та оцінка його тренду

Для аналізованих даних температур, неможливо за моделлю АРПКС визначити форму і величину тренда температур. Для оцінки тренду температур попередньо було отримано їх середньорічні значення. З отриманих середньорічних значень температур можна виділити ковзним усередненням тренд. Як показали результати досліджень, у деякі інтервали часу тренд температур не лише зростає. За даними метеообсерваторії у Пулковому, з 1804 року до 2010 року спостерігалось загальне зростання. Приріст температури за цей термін становив  $2,515\text{ }^{\circ}\text{C}$ . За даними вимірювання температури в місті Сент-Луїс, з 1936 по 1978 рік, спостерігалось падіння температури. Падіння температури за цей термін становило  $0,974\text{ }^{\circ}\text{C}$ .

Аналогічним способом було виділено тренд трафіку даних. Для його отримання використовувалося ковзне усереднення по 7 днях. Тренд даних більш виражений, порівняно із сезонними коливаннями.

## 4 ВИКОРИСТАННЯ МОДЕЛІ ДЕКОМПОЗИЦІЇ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

### 4.1 Завдання прогнозування у стільниковому зв'язку

Основним призначенням нестационарних моделей є вирішення завдань, пов'язаних із прогнозуванням процесів. При плануванні розподілу мережевих ресурсів, а також для ефективного керування та оптимізації часто необхідно прогнозувати можливі сценарії розвитку поведінки мережі за зміни параметрів експлуатації. До таких завдань можуть належати: введення в експлуатацію нового обладнання каналів зв'язку, зміна маршрутизації трафіку зональності (кластеризації) контролерів базових станцій, введення в обслуговування нових мультимедійних послуг та додаткових сервісів. У деяких випадках, що погано прогножуються, зміна складу телекомунікаційного обладнання мережі призводить до часткової або повної зупинки будь-якого сервісу, а введення нових послуг викликає лавиноподібне зростання трафіку або абонентської бази оператора.

Сам по собі аналіз отриманих вимірювань є завданням окремого великого і важливого для оператора стільникового зв'язку процесу OSS (Operation Support System), визначеного міжнародним стандартом ITU-T. Функціонування системи підтримки операційної та бізнес діяльності операторів зв'язку здійснюється за допомогою повної або часткової автоматизації цієї діяльності. Завдання системи OSS у разі полягає у цьому, що з максимальною точністю передбачити динаміку зміни параметрів мережі, і навіть можливість виникнення нештатних ситуацій. Однак системи управління мережею дозволяють лише сповістити інженерний склад про проблему, що виникла, і не мають інструменту, який би хоч у малому наближенні зміг оцінити наслідки для мережі в разі зміни вхідних параметрів (склад обладнання, його конфігурація,

У цій ситуації цікавим є застосування математичних методів, що дозволяють на основі статистики, зібраної або системою O&M, або автоматизованими вимірювальними комплексами, виконати статистичне моделювання процесів, що відбуваються в мережі. Побудована статистична модель повинна, на основі обчислень

за відомими даними параметрів моделі, давати можливість ефективно прогнозувати динаміку зміни характеристик мережі.

Для прогнозування стаціонарних та нестаціонарних випадкових процесів успішно застосовується модель лінійного передбачення [14, 18, 19]. Нестационарність таких процесів проявляється як тренд певної форми. Цей клас нестаціонарних процесів можна визначити як стаціонарні однорідні процеси, які мають постійного середнього значення. У [14] розглядаються нестаціонарні процеси, тренд яких можна видалити за допомогою взяття  $d$ -кратної різниці відліків процесу. Операція знаходження  $d$ -кратної різниці еквівалентна взяття  $d$ -ї похідної від значень випадкового процесу. Таким чином, ця операція замінює випадковий процес відповідної похідної, усуваючи при цьому тренд деякого порядку. Після видалення тренду випадкового процесу його похідна  $d$ -я стає стаціонарною і її можна аналізувати методами, що застосовуються для стаціонарних моделей лінійного передбачення. Для таких нестаціонарних процесів, що описуються моделями АРПКС [14],

У [14] розвинена теорія лінійного прогнозування цього класу нестаціонарних процесів. У кваліфікаційній роботі розроблено та подано методи прогнозування нестаціонарних процесів за наявності сезонних випадкових коливань з постійним періодом. Вибір типу моделі лінійного передбачення, її порядку, параметра  $d$  визначають властивості прогнозованого процесу, характер тренда, період сезонної складової. Вид стаціонарної моделі лінійного передбачення визначається характером зміни кореляційної функції, раціональності використуваних обчислень.

У кваліфікаційній роботі представлені теоретичні основи побудови моделей лінійного передбачення часових рядів, що характеризують функціонування системи стільникового зв'язку стандарту GSM [5]. Адекватність даним синтезованих моделей перевірялась щодо ефективності прогнозування аналізованих часових рядів. Аналізувалися такі дані: запити на встановлення високошвидкісної сесії передачі даних, успішно встановлені високошвидкісні сесії передачі даних, спроби встановлення виклику, відповіді виклики, підключені виклики, трафік передачі даних абонентів вгору і вниз, статистика передачі даних по радіо підсистемі, на одній несучій частоті сектора однієї базової станції. Нижче представлені результати дослідження та прогнозування лише деяких із аналізованих даних.

Попередній аналіз даних про функціонування системи за 222 дні, виміряних через кожні 30 хвилин, показує, що вони мають нестационарність у вигляді тренду та дві сезонні складові. У даних яскраво виражена сезонна складова, пов'язана із добовою періодичністю інтенсивності переговорів (рис. 4.1). Спостерігається також стійка сезонна складова із тижневим періодом. Зростаючий тренд інтенсивності переговорів добре помітний на графіку підсумованих за кожну добу інтенсивності переговорів (рис. 4.2). Проте загалом тренд неоднорідний, так як у ньому є ділянки зростання з різною швидкістю і навіть спаду.

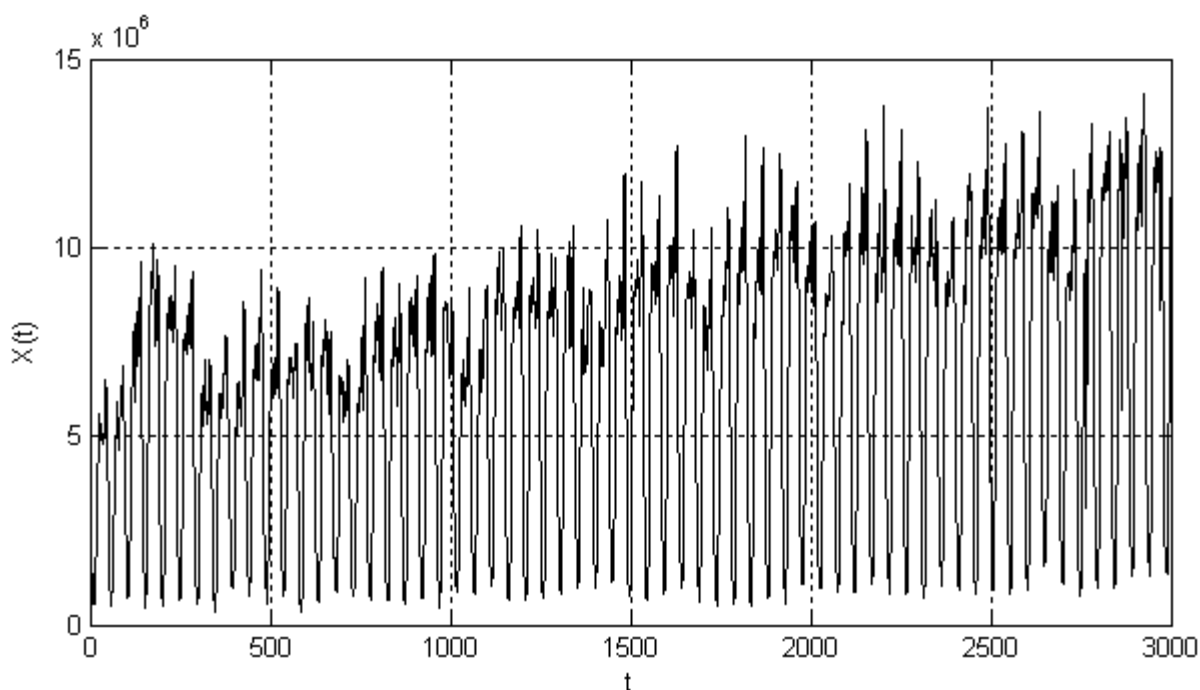


Рисунок 4.1 - Трафік передачі даних абонентів вниз за кожні 30 секунд

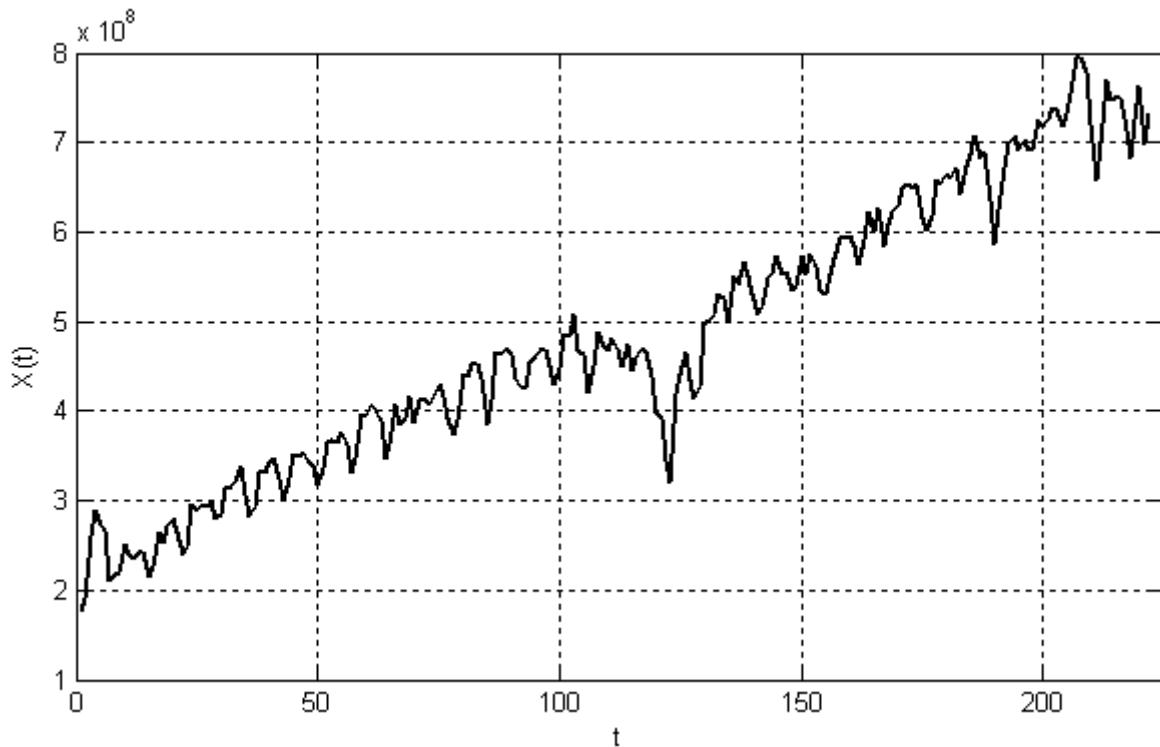


Рисунок 4.2 - Трафік передачі даних абонентів донизу підсумованих за кожен добу

Для синтезу моделі лінійного передбачення інтерес становлять статистичні характеристики корельованих стаціонарних даних, що характеризують функціонування системи. Аналіз цих даних утруднений наявністю сезонних коливань процесу і трендом. Внутрішньодобові коливання інтенсивності призводять до того, що кореляційна функція процесу характеризує в основному сезонні коливання та тренд. Для отримання моделі стаціонарного процесу необхідно спочатку усунути тренд та сезонні коливання за допомогою операторів взяття різниці. Повна модель нестаціонарного процесу, яка використовується для прогнозування, мультиплікативно включатиме моделі стаціонарної, трендової та сезонної складової.

Нестаціонарні дані, що характеризують функціонування системи стільникового зв'язку, раціонально моделювати процесами АРПКС [14]. Аналіз впливу виключення сезонної складової та застосовується оператор взяття різниці був проведений вище.

Виняток із нестаціонарного процесу сезонної складової та тренду дозволяє синтезувати моделі лінійного передбачення стаціонарного процесу. На графіку кореляційної функції процесу, представленого на рис. 4.3а видно, що у процесі є значні кореляції малих зрушень часу. Крім цього, поблизу зсуву 48 відліків

спостерігаються негативні значення кореляції. Аналіз кореляційної функції при великих зрушеннях вказує на циклічність кореляційної функції, яка спостерігається для зсуву, що дорівнює тижневому інтервалу  $T_i = 48 \times 7$  днів = 366 відліків. При довгостроковому прогнозі для усунення цієї сезонної складової можна використовувати також оператор, що спрощує взяття різниці  $\nabla_s$ , де  $s = 336$ . Однак, для довгострокового прогнозу будемо використовувати підсумовані за добу дані (рис. 4.2), для яких  $s = 7$ .

## 4.2 Модель авторегресії

Для аналізу та прогнозу стаціонарної складової моделі декомпозиції будемо використовувати модель АР. Модель АР раціонально використовувати для вузькосмугових випадкових процесів. В основу моделі АР покладено кореляцію відліку випадкового процесу в даний час з деяким кінцевим або нескінченним числом відліків в попередні моменти часу. Кореляційні зв'язки дозволяють здійснювати регресію поточного відліку процесу попередні відліки цього процесу. Такий вид регресії називається авторегресією. Вважатимемо, що корельований випадковий процес  $x[t]$  дійсний, стаціонарний із нульовим середнім. У рівнянні АР поточний відлік є виваженою сумою попередніх відліків з деякими коефіцієнтами ваги [20, 21]

$$x[t] = \sum_{j=1}^p \Phi[j]x[t-j] + a[t], \quad (4.1)$$

де  $\Phi[j]$  - коефіцієнти АР,  $a[t]$  - некорельовані випадкові відліки,  $p$  - порядок моделі АР. Модель (4.1) оптимальний, якщо

$$E\{a[t]a[t-i]\} = 0, i > 0, \quad (4.2)$$

де  $E\{\}$  - позначає статистичне усереднення. Умова оптимальності дозволяє отримати рівняння та визначити критерії для отримання параметрів моделі – коефіцієнтів АР та порядку моделі. Величина

$$\tilde{x}[t] = \sum_{j=1}^p \Phi[j]x[t-j], \quad (4.3)$$

називається передбаченням випадкової величини. Різниця між поточним значенням відліку та його передбаченням називається помилкою передбачення

$$a[t] = x[t] - \tilde{x}[t]. \quad (4.4)$$

Величина  $a[t]$  характеризує максимальну точність передбачення поточного відліку, та її статистичні властивості визначають вибір порядку моделі АР.

Модель АР, що описується виразом (4.1), має кілька інтерпретацій. (4.1) це лінійне стохастичне різницеве рівняння, що описує спосіб зміни випадкової функції  $x[t]$ , при випадковому впливі  $a[t]$ . Різничне рівняння є дискретним наближенням уявленням безперервного лінійного диференціального рівняння. Природно, що коефіцієнти різницевого та відповідного диференціального рівняння не збігаються, але їх можна зв'язати, прирівнявши коефіцієнти при відповідних відліках моделі. Випадковий процес, який описується різницевим рівнянням (4.1) зазвичай заданий. Тому вирішується зворотне завдання, що зводиться до знаходження параметрів різницевого рівняння за відомої випадкової функції  $x[t]$ .

Рівняння (4.1) можна представити як вихід  $x[t]$  лінійної системи зі зворотним зв'язком, що складається з  $P$  ланок, на вхід якої подається білий шум  $a[t]$ . Параметри системи збігаються із параметрами моделі АР. Вони досить просто розраховуються за кореляційною функцією та дисперсією виходу. Різничне рівняння (4.1) описує формуючий фільтр, на вхід якого подається гаусів або негаусів білий шум. Як параметри фільтра використовуються коефіцієнти моделі АР, розраховані для

реального процесу. Параметри фільтра можна розрахувати за заданими характеристиками СЦП.

Якщо входом системи вважати корельований процес, а виходом є білий шум, вираз (4.1) може описувати систему у вигляді фільтра, що обіляє. Використовуючи параметри моделі, можна визначити передавальну функцію системи, її імпульсну характеристику та перехідну функцію. На основі амплітудно-частотної характеристики системи одержують вираз для параметричної спектральної оцінки.

Рівняння (4.1) також інтерпретується як лінійне передбачення поточного відліку випадкового процесу  $x[t]$  за відомими  $p$  попередніми відліками  $x[t - j]$ . Коефіцієнти ваги знаходяться з урахуванням кореляційних статистичних зв'язків між відліками процесу. Передбачення є оптимальним, тому коефіцієнти ваги визначаються з умови мінімуму дисперсії помилки передбачення  $a[t]$ . У рамках статистик другого порядку мінімум дисперсії помилок передбачення еквівалентний їх некорельованості [22]. Знайдені таким чином коефіцієнти враховують всю інформацію про статистичні зв'язки процесу, так як помилка передбачення некорельована і не містить неврахованої інформації. Така інтерпретація важлива у завданні лінійного прогнозування випадкових процесів.

До моделей лінійного передбачення можна дійти, використовуючи метод найменших квадратів (МНК). Метод мінімізує суму квадратів відхилень випадкового процесу, що спостерігається, від відповідних за часом, розрахованих значень регресійного лінійного рівняння. Мінімізація здійснюється шляхом диференціювання за коефіцієнтами регресії та прирівнювання похідних до нуля. Для моделі АР рівняння регресії записується у вигляді авторегресії (4.3), а відхилення є помилками передбачення (4.4). Рівняння для оптимальних за МНК оцінок отримують з виразу

$$\frac{dE\{a^2[t]\}}{d\Phi[j]} = \frac{dE\{(x[t] - \tilde{x}[t])^2\}}{d\Phi[j]} = \frac{dE\{(x[t] - \sum_{j=1}^p \Phi[j]x[t-j])^2\}}{d\Phi[j]} = 0. \quad (4.5)$$

У літературі з лінійного передбачення найчастіше використовується інший спосіб отримання рівнянь для розрахунку оптимальних коефіцієнтів АР, який використовує

умову статистичної незалежності помилок передбачення. Цю умову оптимальності можна використовувати при побудові моделей лінійного передбачення на основі статистики вищих порядків. МНК має обмеження лише статистики другого порядку.

З (4.1) видно, що побудова АР моделі випадкового процесу зводиться до знаходження коефіцієнтів АР та визначення порядку  $P$ . Знайдемо рівняння до розрахунку оптимальних оцінок коефіцієнтів АР. Помноживши праву та ліву частини (4.1) на  $x[t-i]$  і усереднивши, можна отримати систему  $P+1$  рівнянь

$$R[i] - \sum_{j=1}^P \Phi[j]R[j-i] = 0, \quad i = 1, \dots, P, \quad (4.6a)$$

$$R[0] - \sum_{j=1}^P \Phi[j]R[j] = D_a, \quad i = 0, \quad (4.6b)$$

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  $x[t]$ . набір рівнянь (4.6a) та (4.6b) називається повною системою рівнянь Юла-Уокера. Рішенням цієї системи є коефіцієнти АР та дисперсія помилок передбачення. При виведенні рівнянь (4.6a) було враховано умову оптимальності (4.2), з якої випливає, що

$$E\{a[t]x[t-i]\} = 0, \quad i > 0.$$

Рівняння (4.6a,б) можна також отримати за допомогою МНК. Використовуючи (4.5), маємо

$$E\{(x[t] - \sum_{j=1}^P \Phi[j]x[t-j])x[t-i]\} = 0.$$

Після перемноження та усереднення приходимо до рівнянь Юла-Уокера (4.6а,б). Вираз (4.6б) визначає потужність процесу через параметри моделі та потужність помилки передбачення.

Представимо систему рівнянь (4.6а,б) у матричному вигляді

$$\vec{R} = R_p \vec{\Phi}, \quad (4.7a)$$

де

$$\vec{\Phi} = \begin{bmatrix} \Phi[1] \\ \Phi[2] \\ \vdots \\ \Phi[p] \end{bmatrix}, \quad R_p = \begin{bmatrix} R[0] & R[1] & \cdots & R[p-1] \\ R[1] & R[0] & \cdots & R[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ R[p-1] & R[p-2] & \cdots & R[0] \end{bmatrix}, \quad \vec{R} = \begin{bmatrix} R[1] \\ R[2] \\ \vdots \\ R[p] \end{bmatrix}. \quad (4.7b)$$

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

Розв'язання матричного рівняння методом Юла-Уокера (4.6а) можна записати так

$$\vec{\Phi} = R_p^{-1} \vec{R}. \quad (4.8)$$

Класичний метод обчислення коефіцієнтів АР при довгих реалізаціях дає, як правило, стійкі рішення. Однак через різні причини, наприклад, наявність похибок у вимірах, при коротких реалізаціях, рішення рівняння (4.8) можуть бути нестійкими. У цьому випадку малі зміни значень функції кореляції призводять до великих змін коефіцієнтів АР. При коротких реалізаціях метод Юла-Уокера дає неточну оцінку параметрів, що призводить до зниження роздільної здатності параметричних спектральних оцінок. Крім цього, зі зростанням порядку моделі суттєво зростає обсяг обчислень, пов'язаних із обігом кореляційної матриці. У зв'язку з цим часто застосовуються точніші рішення

рівняння (4.7а), що дозволяють скоротити кількість операцій використанням рекурентних алгоритмів [3].

Порядок процесу АР визначається з використанням різних критеріїв, як правило, заснованих на мінімізації певної теоретико-інформаційної функції. Для визначення порядку моделі користуються методами Бартлетта, Акайке, Парзена та ін. Порядок моделі можна знаходити з умови не зменшення дисперсії помилки передбачення при подальшому підвищенні порядку. Досить ефективним методом визначення порядку моделі АР є метод, що ґрунтується на перевірці близькості кореляційної функції помилки передбачення до кореляційної функції  $\delta$ -функції білого шуму. Спосіб оцінки порядку моделі лінійного передбачення тут ґрунтується на близькості до нуля останнього коефіцієнта АР.

Як видно з (4.6а), рівняння не зміниться, якщо замість  $R[i]$  використовувати нормовані значення функції кореляції  $r[i] = R[i] / R[0]$ , які називатимемо коефіцієнтами кореляції. Очевидно, що при цьому параметри моделі залишаються колишніми. Ця важлива властивість інваріантності параметрів моделі до потужності процесу є корисною при вирішенні низки завдань статистичної радіотехніки.

Відповідно до (4.7а,б), для першого порядку моделі АР

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

Для моделі АР другого порядку коефіцієнти АР дорівнюють

$$\Phi[1] = r[1](1 - r[2]) / (1 - r^2[1]),$$

$$\Phi[2] = (r[2] - r^2[1]) / (1 - r^2[1]).$$

Процеси АР можна характеризувати кінцевим числом значень функції, що визначається за кореляційною функцією процесу. Така функція має назву приватної автокореляційної функції. Її можна висловити через  $P$  коефіцієнтів АР, рівних останнім коефіцієнтам АР для порядків від одиниці до  $P$ .

### 4.3 Спектр процесу авторегресії

Модель АР випадкового процесу найчастіше серед інших моделей лінійного передбачення використовується для параметричного спектрального оцінювання. Це пов'язано, по-перше, з тим, що вузькосмугові процеси поширені у природі і техніці. По-друге, параметри моделі АР простіше обчислювати. По-третє, більшість розроблених ефективних методів оцінювання стосуються саме обчислення коефіцієнтів АР.

Для виведення формули АР параметричного спектрального оцінювання розглядатимемо процес АР як реакцію  $x[t]$  формуючого фільтра, на вхід якого подаються некорельовані відліки  $a[t]$ .  $z$ -перетворення передавальної функції АР фільтра має вигляд

$$H(z) = 1 / \Phi(z),$$

де оператор АР

$$\Phi(z) = 1 - \sum_{i=1}^p \Phi[i] z^{-i} \quad (4.10)$$

$z$ -перетворення кореляційної функції вихідного та вхідного процесів у лінійній системі пов'язані співвідношенням [3]

$$P(z) = |H(z)|^2 P_a(z) = P_a[z] / |\Phi(z)|^2 \quad (4.11)$$

Щоб знайти СЩП вихідного АР процесу необхідно у (4.11) зробити заміну  $z = e^{j2\pi fT}$  і покласти, що для білого шуму  $P_a[z] = D_a$  є стала величина. Тоді з (4.11), з урахуванням (4.10), випливає

$$P(f) = D_a \left| 1 - \sum_{i=1}^p \Phi[i] e^{-j2\pi f i T} \right|^2. \quad (4.12)$$

До цього виразу можна прийти враховуючи, що СГМ входу та виходу лінійної системи пов'язані співвідношенням

$$P(f) = |H(f)|^2 P_a(f) = D_a / |\Phi(f)|^2,$$

де  $|H(f)|^2 = |\Phi(f)|^{-2}$  – квадрат амплітудно-частотної характеристики (АЧХ) системи, що описується моделлю АР. З (4.12) випливає, що справжні процеси АР повинні мати вузькосмугову форму СЦП. Багато реальних вузькосмугових процесів досить добре оцінюються параметричними СЦП. Точність параметричних оцінок СЦП істотно залежить від того, наскільки добре описується моделлю лінійного передбачення.

Вираз (4.12) широко використовують у параметричному методі спектрального оцінювання випадкових процесів. Як параметри, що повністю характеризують спектральну оцінку випадкового процесу, виступають коефіцієнти АР і порядок моделі. Параметричне спектральне оцінювання має низку переваг у порівнянні з традиційними методами спектрального оцінювання. До них відносяться: більш висока спектральна роздільна здатність при використанні коротких вибірок, відсутність бічних пелюсток, наявність простої формули для обчислення. Недоліком параметричного спектрального аналізу є вимога адекватності моделі та процесу, чутливість до вибору порядку моделі. Якщо порядок моделі менший за істинний, то форма спектра буде грубим наближенням істинного спектра. При цьому також знижується роздільна здатність оцінки, а слабкі компоненти спектра може бути не виявлено. При підвищеному порядку моделі виникають зайві моди діапазону, роздвоюються спектральні піки.

За допомогою моделі АР можна отримати спектральні оцінки процесів зі складною формою СЦП. Для цього можливо доведеться використовувати моделі АР нескінченного порядку. На основі моделі АР легко синтезуються оптимальні фільтри

придушення, узгоджені не тільки за частотою та смугою спектра, але і формою спектра випадкового процесу. Перевагою формули (1.12) є можливість подання СГМ у вигляді аналітичного виразу. За допомогою операцій над цим виразом, наприклад, можна знайти формули для визначення частоти максимумів і мінімумів СГМ.

#### 4.4 Характеристичне рівняння моделі авторегресії

Модель АР, що описується рівнянням (4.1), може бути представлена в операторній формі

$$\Phi(z)x[t] = a[t], \quad (4.14)$$

де оператор АР  $\Phi(z)$  (4.10) зручно подати у вигляді

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

Дія оператора зсуву  $z^{-1}$  на поточний відлік описується так

$$z^{-i}x[t] = x[t - i].$$

З умови стійкості формуючого АР фільтра з раціональною функцією передачі (4.8), впливає умова стаціонарності АР процесу. Для визначення стаціонарності випадкового АР процесу використовується характеристичне рівняння

$$\Phi(c) = c^p - \Phi[1]c^{p-1} - \dots - \Phi[p] = 0. \quad (4.15)$$

Якщо корені характеристичного рівняння (4.15)  $c[i]$  лежать всередині одиничного кола на комплексній площині, процес АР задовольняє умови стаціонарності. Характеристичне рівняння (4.15), іноді представляється також як

$$1 - \Phi[1]c^{-1} - \dots - \Phi[p]c^{-p} = 0. \quad (4.16)$$

Тоді умова стаціонарності полягає в тому, що коріння характеристичного рівняння (4.16) має лежати поза одиничним колом на комплексній площині.

#### 4.5 Модель АРПКС для прогнозу

Для сезонних процесів із трендом існує проблема вибору відповідної моделі з безлічі можливих. Як показали дослідження, для аналізованих даних при короткостроковому прогнозі у якості нестаціонарної моделі, що відповідає кореляційній функції (рис. 4.3 а, б), найбільш підходяща модель має вигляд

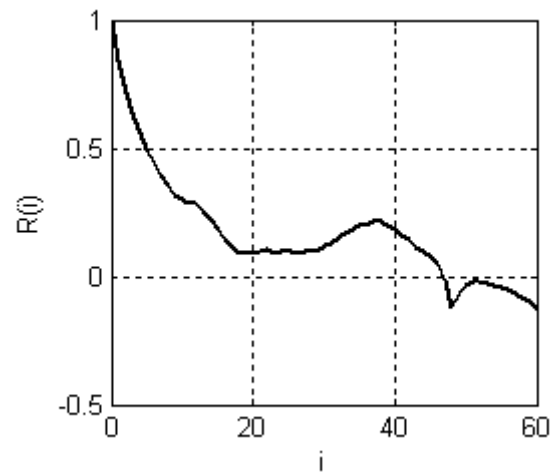
$$(1 - \Phi z^{-1})\omega[t] = (1 - \Phi z^{-1})\nabla_{48}(1 - z^{-1})x[t] = a[t],$$

де  $\Phi$  - коефіцієнт моделі АР першого порядку стаціонарного процесу  $\omega[t]$ . Тоді сезонна модель описується різницеvim рівнянням

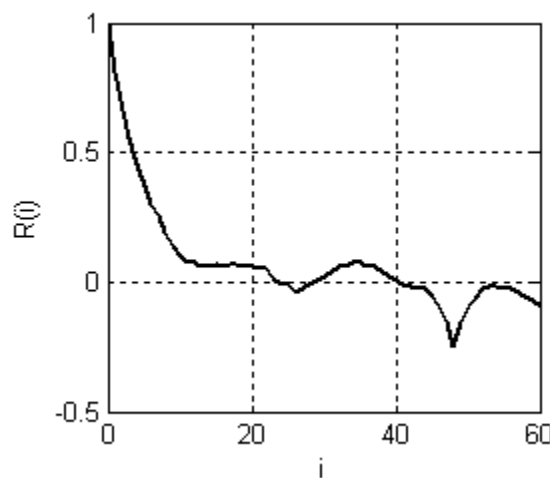
$$x[t] = (\Phi + 1)x[t-1] - \Phi x[t-2] + x[t-48] - (\Phi + 1)x[t-49] + \Phi x[t-50] + a[t]. \quad (4.17)$$

Вираз для розрахунку  $\Phi$  отримаємо, помноживши (4.17) на  $x[t-1]$  і усереднивши,

$$\Phi = \frac{R[1] - R[0] - R[47] + R[48]}{R[0] - R[1] - R[48] + R[49]}.$$



а



б

Рисунок 4.3 - Кореляційна функція процесу  $\omega[t]$  трафіку передачі даних абонентів донизу:

а - для даних без попереднього логарифмування,

б - для логарифмів даних

Для прогнозу даних часто використовується попереднє нелінійне перетворення. Як правило, використовується таке перетворення, яке зменшує амплітуду коливань великих значень даних. Однак така процедура часто не призводить до великих змін їх статистичних характеристик. Це показують дані графіків, представлених на рис. 4.3.

#### 4.6 Прогнозування на основі моделі декомпозиції

У ряді випадків для аналізу часових рядів, а також при вирішенні завдання прогнозування, застосовується попереднє нелінійне перетворення даних. У цій кваліфікаційній роботі прогнозування проводилося як за вихідними даними, так і за їх нелінійно перетвореними значеннями. Зауважимо, що нелінійне перетворення, що застосовується, зменшує розкид даних для великих інтенсивностей, але збільшує розкид даних при низьких інтенсивностях. На графіку кореляційної функції прологарифмованого процесу  $x[t]$ , представлена на рис. 4.3б видно, що після попереднього нелінійного перетворення шляхом логарифмування зменшилися коливання кореляційної функції щодо нульової лінії, а при зсуві 48 її значення стали більш значущі. Проте отримані результати дозволяють ігнорувати подібні зміни кореляційної функції при оцінці моделі АР.

Прогнози способом різницевого рівняння одержують з мінімальною середньоквадратичною помилкою на момент  $t$  з попередженням  $l$ . Показано [14], що такий прогноз є умовним математичним очікуванням  $x[t+l]$  у момент  $t$ , за умови, що всі  $x[t]$  до моменту  $t$  відомі. Умовні математичні очікування випадкових величин, що входять до (4.17), рівні

$$\{x[t+i]\} = \begin{cases} x[t+i], & i \leq 0, \\ \hat{x}[t+i], & i > 0, \end{cases} \quad \{a[t+i]\} = \begin{cases} a[t+i], & i \leq 0, \\ 0, & i > 0, \end{cases} \quad (4.18)$$

Де  $\hat{x}[t+i]$ - прогноз  $x[t]$  на момент  $t$  на  $i$  кроків.

Різницеве рівняння, яке застосовується для прогнозу на момент  $t$  на  $l$  кроків, отримаємо (4.17)

$$x[t+l] = (\Phi + 1)x[t+l-1] - \Phi x[t+l-2] + x[t+l-48] - (\Phi + 1)x[t+l-49] + \Phi x[t+l-50] + a[t+l] \quad (4.19)$$

Прогноз на попередження  $l$  отримуємо як умовне математичне очікування від (4.19). Висловлювання для прогнозу для різних попереджень  $l$  з урахуванням (4.18) набуває вигляду [5]:

$$\begin{aligned}
 \hat{x}[t+1] &= (\Phi+1)x[t] - \Phi x[t-1] + x[t-47] - (\Phi+1)x[t-48] + \Phi x[t-49], \quad l=1; \\
 \hat{x}[t+2] &= (\Phi+1)\hat{x}[t+1] - \Phi x[t+1] + x[t-46] - (\Phi+1)x[t-47] + \Phi x[t-48], \quad l=2; \\
 \hat{x}[t+l] &= (\Phi+1)\hat{x}[t+l-1] - \Phi \hat{x}[t+l-2] + x[t+l-48] - (\Phi+1)x[t+l-49] + \Phi x[t+l-50], \quad 3 \leq l \leq 48 \\
 \hat{x}[t+49] &= (\Phi+1)\hat{x}[t+48] - \Phi \hat{x}[t+47] + \hat{x}[t+1] - (\Phi+1)x[t] + \Phi x[t-1], \quad l=49; \\
 \hat{x}[t+50] &= (\Phi+1)\hat{x}[t+49] - \Phi \hat{x}[t+48] + \hat{x}[t+2] - (\Phi+1)\hat{x}[t+1] + \Phi x[t], \quad l=50; \\
 \hat{x}[t+l] &= (\Phi+1)\hat{x}[t+l-1] - \Phi \hat{x}[t+l-2] + \hat{x}[t+l-48] - (\Phi+1)\hat{x}[t+l-49] + \Phi \hat{x}[t+l-50], \quad 51 \leq l.
 \end{aligned} \tag{4.20}$$

Формули для корекції та дисперсії помилки прогнозу виражаються через імпульсну характеристику лінійної системи, що описується різницеvim рівнянням (4.17). Процес на виході лінійної системи з імпульсною характеристикою  $h[i]$  може бути представлений у вигляді згортки

$$x[t] = h(z)a[t-i] = \sum_{i=0}^{\infty} h[i]p[t-i]. \tag{4.21}$$

Підставляючи (4.21) (4.17) і прирівнюючи коефіцієнти при однакових  $z^{-1}$ , отримуємо вирази для обчислення імпульсної характеристики лінійної системи

$$\begin{aligned}
 h[1] &= \Phi, \\
 h[2] &= \Phi^2, \\
 &\vdots \\
 h[47] &= \Phi^{47}, \\
 h[48] &= 1 + \Phi^{48}, \\
 h[49] &= (1 + \Phi^{48})\Phi, \\
 &\vdots \\
 h[i] &= (1 + \Phi^{48})\Phi^{i-48}.
 \end{aligned} \tag{4.22}$$

Помилка прогнозу попередження  $l$  дорівнює [14]

$$e[t+l] = a[t+l] + h[1]a[t+l-1] + \dots + h[l-1]a[t+1], \quad (4.23)$$

де значення імпульсної характеристики визначаються (4.22). Прогноз буде незміщений, оскільки умовне математичне очікування помилки прогнозу задовольняє рівності  $\hat{e}[t+l] = 0$ . Дисперсія помилки прогнозу дорівнює

$$D_e(l) = (1 + h^2[1] + h^2[2] + \dots + h^2[l-1])D_a,$$

де  $D_a$  - дисперсія помилки передбачення.

За довірчої ймовірності  $1 - \varepsilon$ , інтервальна оцінка прогнозу  $\hat{x}[t+l]$  визначається співвідношенням [6]

$$\hat{x}[t+l](\pm) = \hat{x}[t+l] \pm u_{1-\varepsilon/2} \left( 1 + \sum_{i=1}^{l-1} h_i^2 \right)^{1/2} \sigma_a,$$

Де  $\sigma_a$  - середньоквадратична помилка передбачення,  $u_{1-\varepsilon/2}$  - квантиль нормального розподілу рівня значимості  $\varepsilon$ .

Нижче наведено результати короткострокових прогнозів даних функціонування стільникового зв'язку з урахуванням сезонної складової з періодом 48 відліків. На рис. 4.4 представлені графіки прогнозів логарифмів даних трафіку передачі даних абонентів униз, але в рис. 4.5 графік прогнозів даних без попереднього логарифмування. Для прогнозу використовувалися моделі АР першого порядку з коефіцієнтами  $\Phi = 0.7728$  і  $\Phi = 0.7633$  відповідно. Порівняння графіків показує, що прогнози з використанням попереднього нелінійного передбачення більш точні у найцікавішій галузі – за високої інтенсивності переговорів.

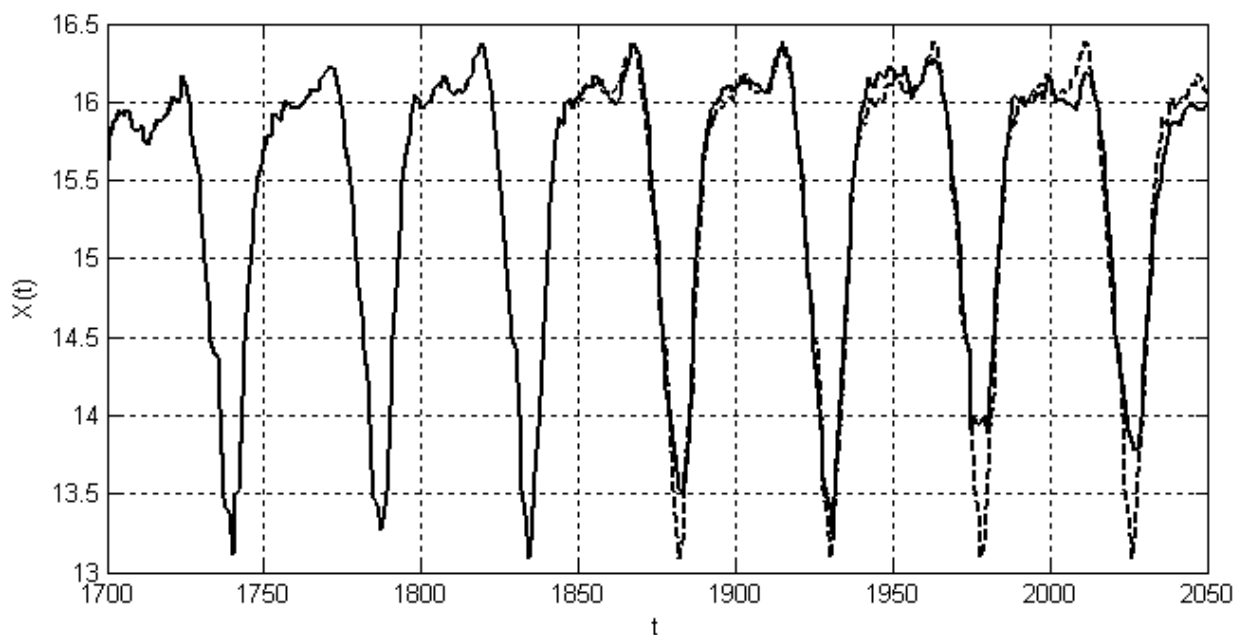


Рисунок 4.4 - Прогнозування прологарифмованої вибірки даних трафіку передачі абонентів вниз

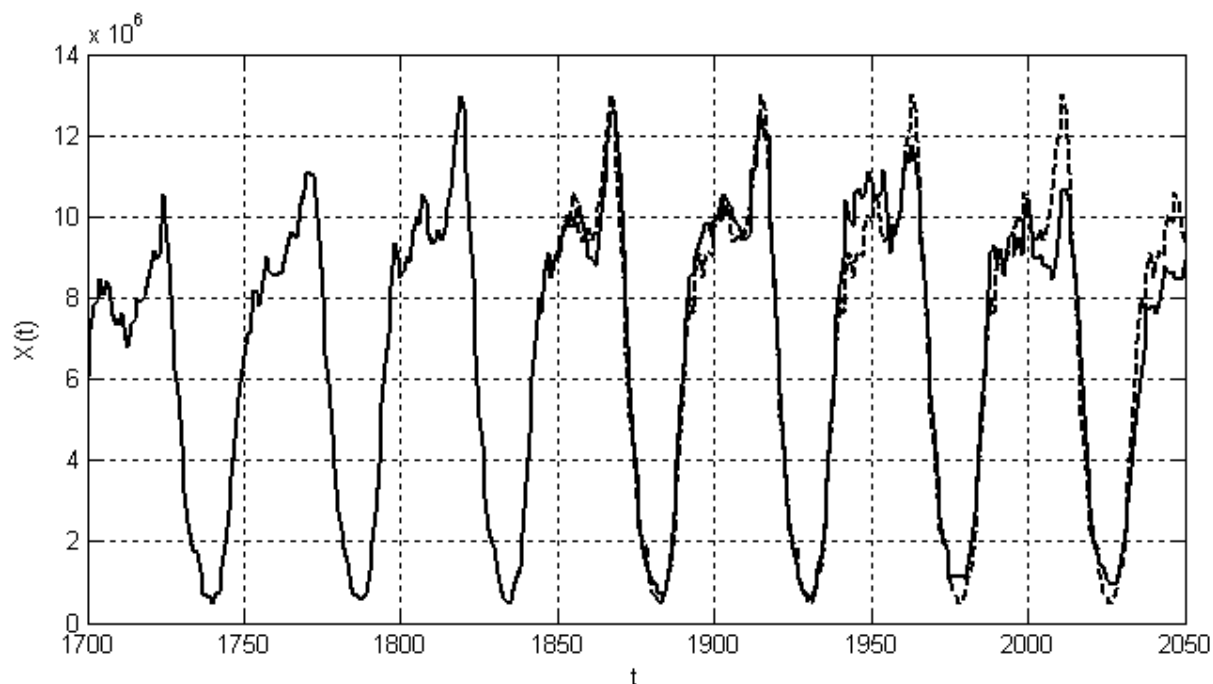


Рисунок 4.5 - Прогнозування вибірки даних трафіку передачі даних абонентів униз

На рис. 4.6 представлений графік прогнозів логарифмів даних трафіку передачі даних абонентів нагору, а на рис. 4.7 графік прогнозів прологарифмованої вибірки даних про промкнені виклики. Для прогнозу використовувалися моделі АР першого порядку з коефіцієнтами  $\Phi = 0.7848$  і  $\Phi = 0.7227$  відповідно. Аналіз графіків показує, що

прогнози з використанням попереднього нелінійного передбачення більш точні у найцікавішій галузі – за високої інтенсивності переговорів.

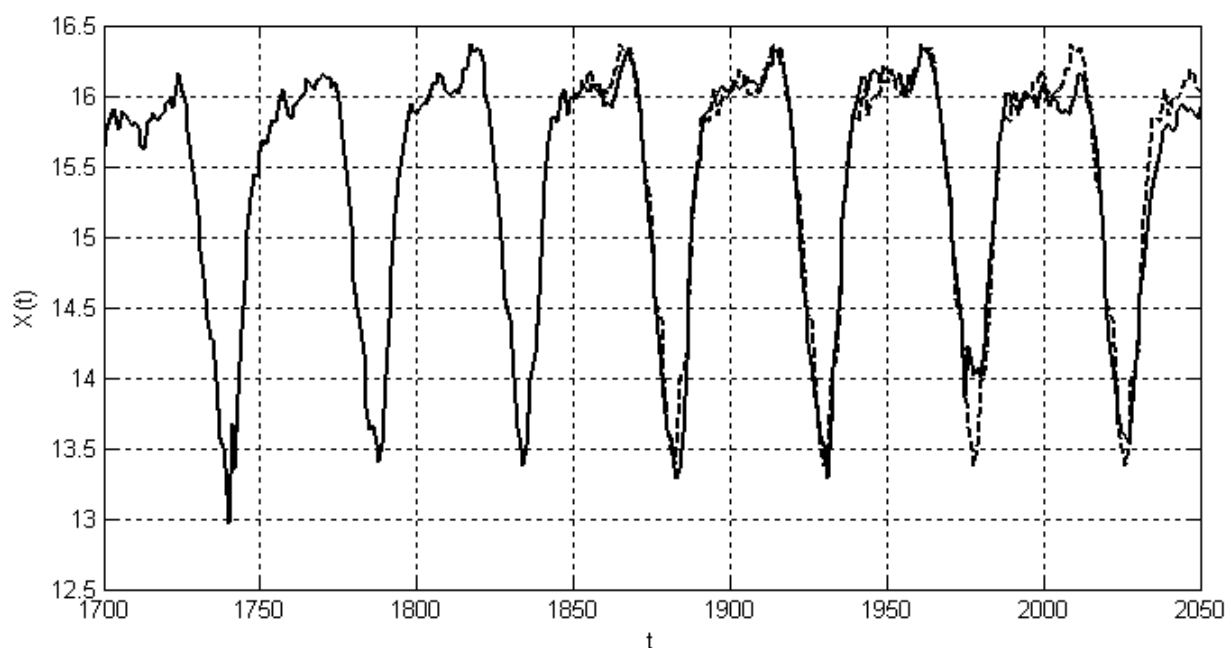


Рисунок 4.6 - Прогнозування прологарифмованої вибірки трафіку передачі абонентів вгору

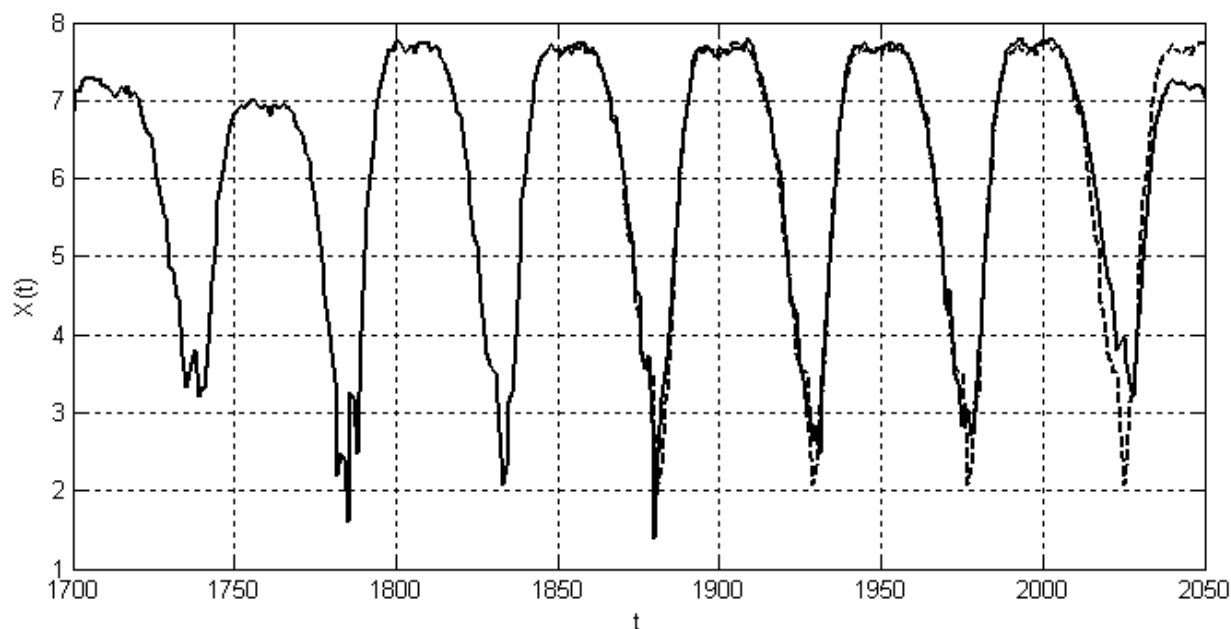


Рисунок 4.7 - Прогнозування прологарифмованої вибірки даних про здійснені виклики

Крім короткострокового прогнозу для управління системою стільникового зв'язку необхідно проводити довгостроковий прогноз даних, що характеризують функціонування системи зв'язку. Аналіз довгострокових часових рядів, що є

підсумовані за добу дані, показує наявність у тренді сезонних складових з квазідетермінованим періодом. Циклом є щотижневі коливання трафіку.

Для вирішення завдання прогнозування будувалися моделі стаціонарних процесів лінійного передбачення даних трафіку після усунення лінійного тренду та сезонної складової. Для усунення лінійного тренду застосовувалася операція взяття різниці. Сезонна складова усувалася операцією

$$\omega_1[t] = \nabla_s x[t] = x[t] - x[t-7].$$

Довгострокове прогнозування викликає труднощі, пов'язані, насамперед, з неоднорідністю тренду, з малою довжиною реалізації та вибором відповідної стаціонарної моделі. Як показали дослідження трендів аналізованих даних, вони характеризуються неоднорідними швидкостями зростання та спаду тренду. Тому довгостроковий прогноз можливий лише для невеликих випереджень на однорідній ділянці тренду.

На рис.4.8 представлені дані прогнозування вибірки щоденних даних трафіку передачі абонентів вниз. Для стаціонарного випадкового процесу, отриманого після видалення тренду, застосовувалася модель AP(1) з коефіцієнтом авторегресії рівним 0,5663. Початок прогнозу вибирався у точці відносної однорідності тренду. Випередження прогнозу становило три цикли, що відповідають трьом тижням. Для порівняння проводилося також прогнозування прологарифмованої вибірки щоденних даних трафіку передачі абонентів вниз (рис. 4.9). Стаціонарна модель AP(1) цих даних описувалася коефіцієнтом авторегресії рівним 0,6241.

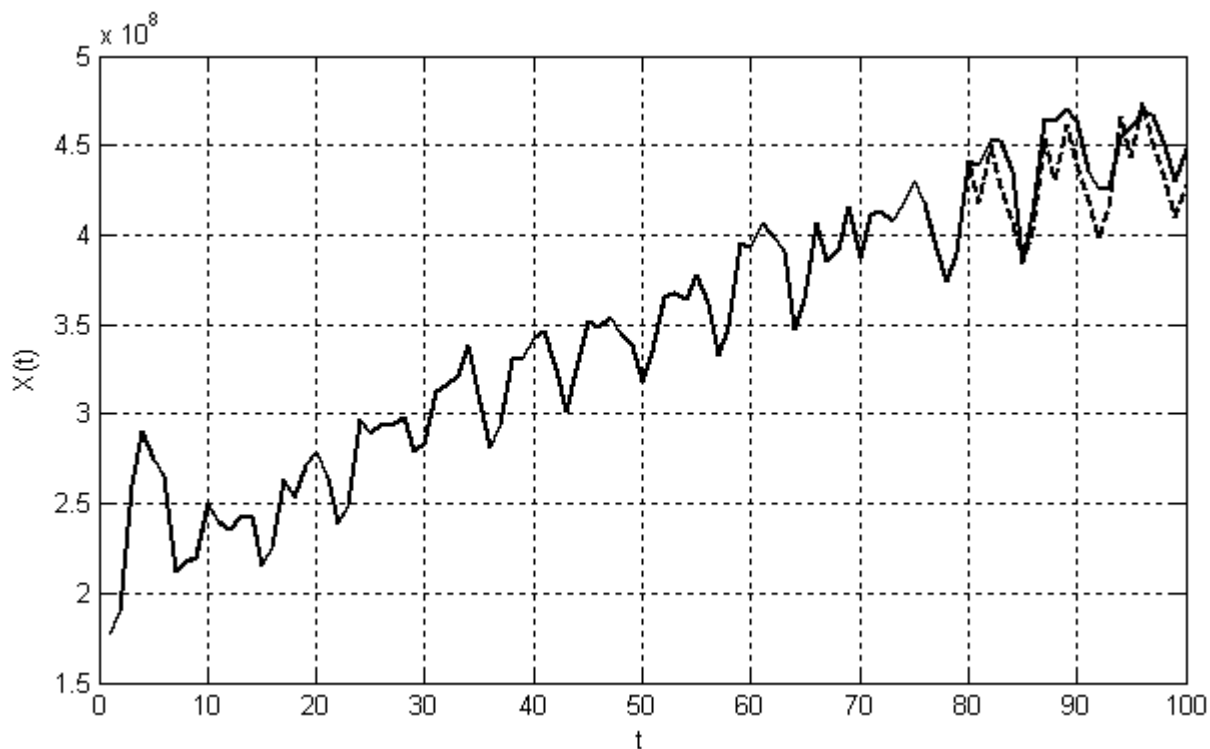


Рисунок 4.8 - Прогнозування вибірки щоденних даних трафіку передачі абонентів вниз

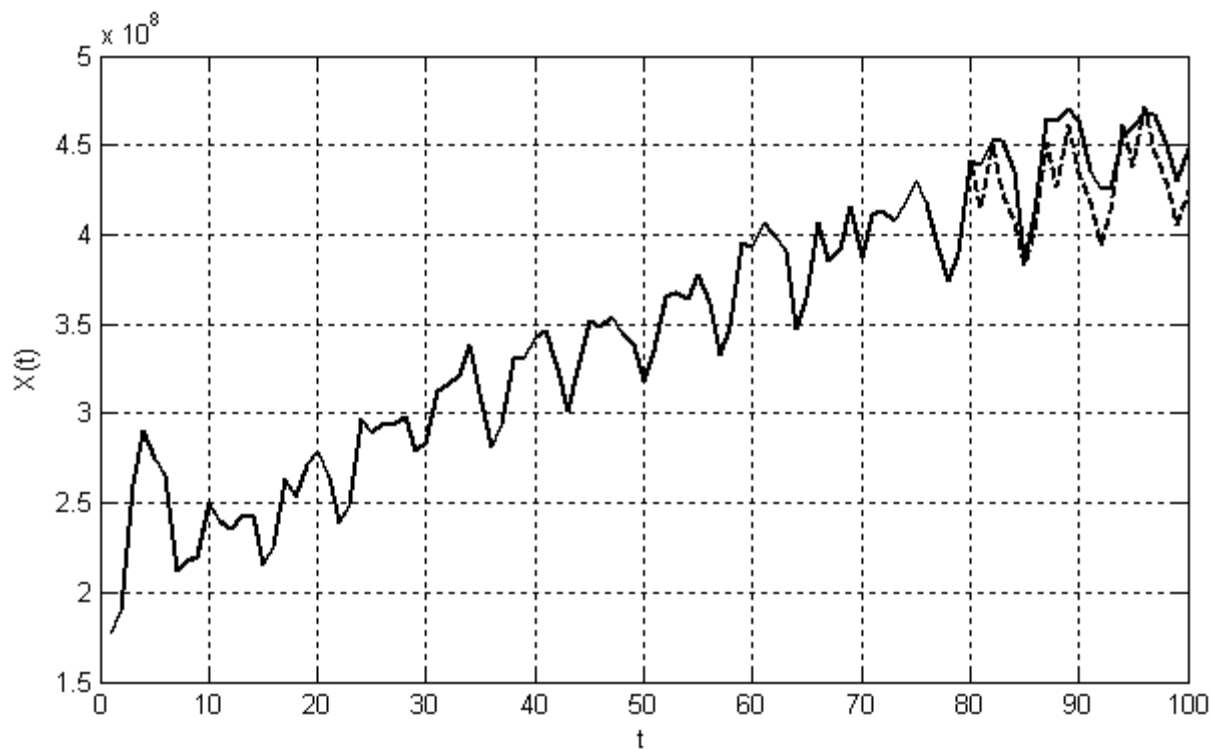


Рисунок 4.9 - Прогнозування прологарифмованої вибірки щоденних даних трафіку передачі абонентів вниз

Як очевидно з порівняння графіків прогнозів, поданих на рис. 4.8 та рис. 4.9 застосування попереднього нелінійного перетворення даних у цьому випадку помітного поліпшення не дає. Вочевидь, це пояснюється тим, що відсотковий розкид

даних (відмінність мінімального значення від максимального значення) невеликий, на відміну даних, прогноз яких наведено на рис. 4.4 – 4.7.

Як зазначалося вище, точність прогнозу істотно залежить від початкової точки прогнозу у разі неоднорідного тренду. На рис. 4.10 представлений графік прогнозу прологарифмованої вибірки щоденних даних трафіку передачі даних абонентів вниз з точкою початку прогнозу, що дорівнює 60 відліків, а не 80 відліків, як у прогнозах, поданих на рис. 4.8 – 4.9. Стационарний процес у цьому випадку, отриманий після усунення циклу та тренду описувався моделлю  $AR(1)$  з коефіцієнтом авторегресії рівним 0,6241.

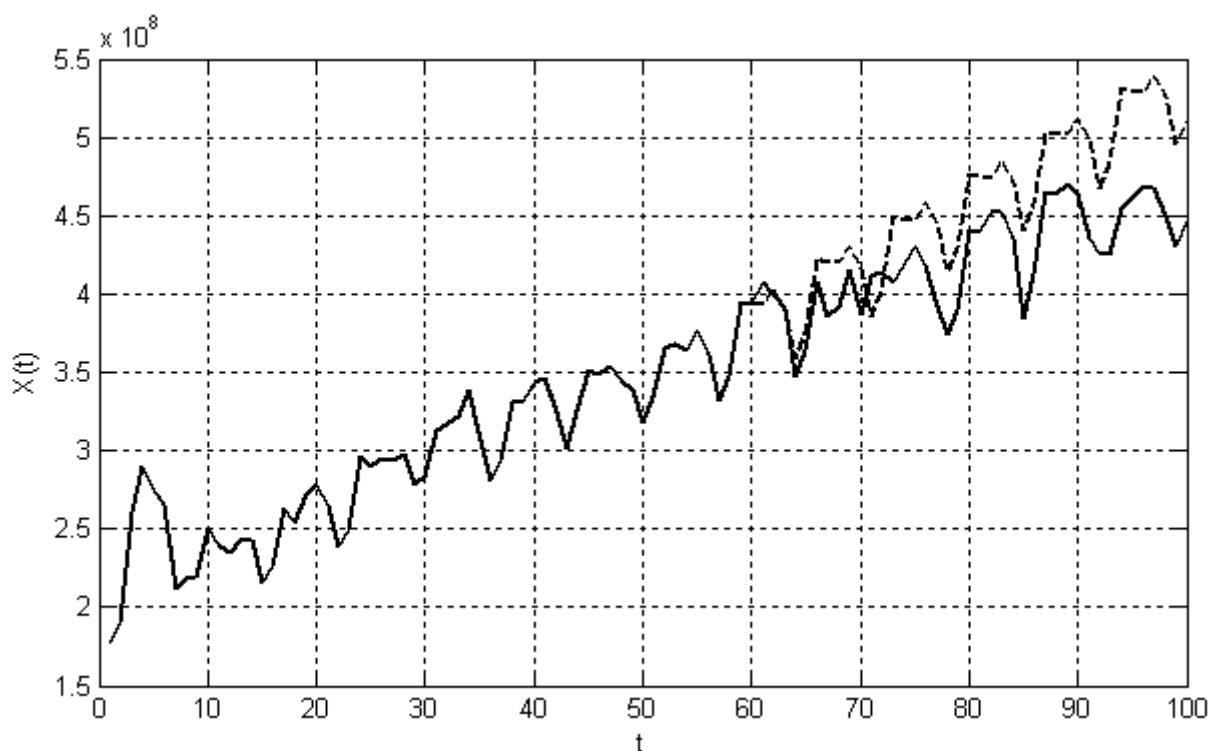


Рисунок 4.10 - Прогнозування прологарифмованої вибірки щоденних даних трафіку передачі абонентів вниз

Як зазначалося, вища точність прогнозу залежить від характеристик тренду. Так, із графіка, представленого на рис. 4.11 видно, що тренд не може бути описаний ні лінійною, ні параболічною, ні гіперболічною функцією. Тому прогнозування даних трафіку з такими трендами за методом, описаним у [14], дуже складний і потребує більш ретельного підбору моделей, а також використання складніших класів моделей лінійного передбачення.

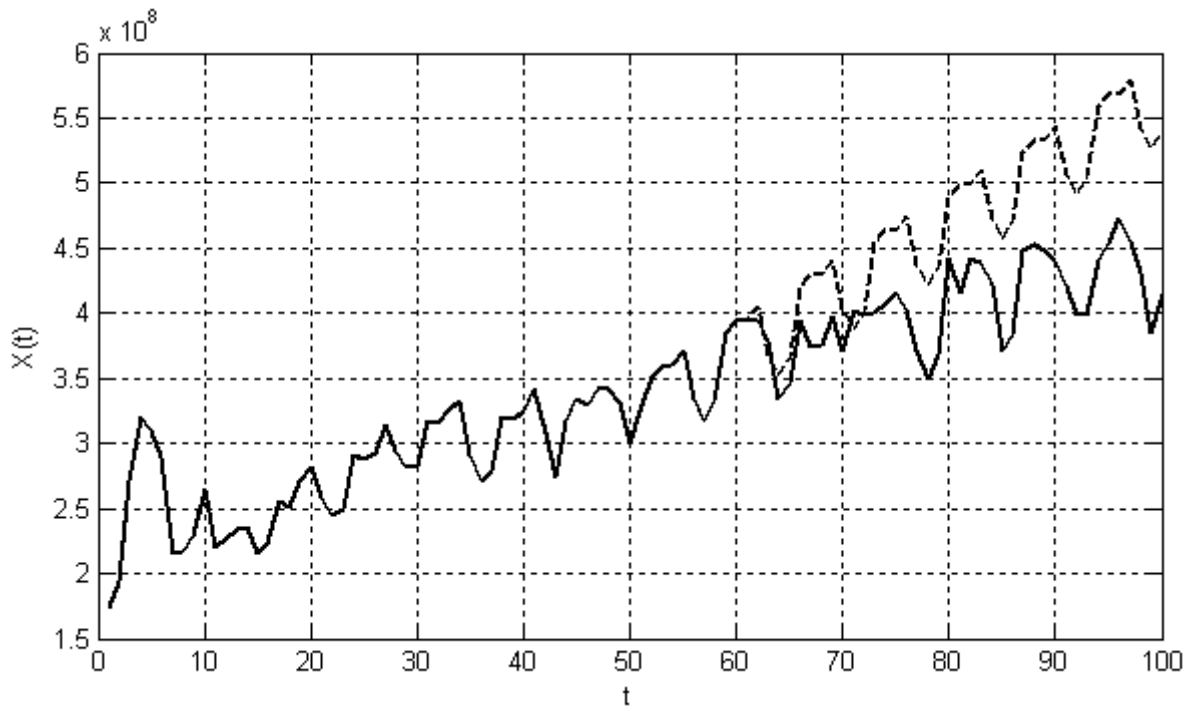


Рисунок 4.11 - Прогнозування вибірки щоденних даних трафіку передачі абонентів вгору

## ВИСНОВКИ

У кваліфікаційній роботі показано деякі проблеми, що виникають при використанні моделі АРПКС та декомпозиції моделі процесу для аналізу нестационарних даних, що містять трендову, сезонну та стаціонарну складові. Використання в моделі АРПКС операторів, що враховують ці складові, обмежує допустимі властивості сезонної складової та тренду. Вплив на модель аналізованого процесу операторів зсуву продемонстровано на завданні декомпозиції моделі нестационарного процесу. Застосування цих операторів зводиться до диференціювання вихідних даних, що спотворює характеристики вихідного процесу та призводить до неточностей. Якщо така модель може давати прийнятні результати завдання прогнозу, вона годиться для обчислення тренда. Для виділення тренда слід використовувати усереднені за певний період дані,

На прикладі прогнозу даних трафіку показано застосування моделей декомпозиції та АРПКС для вирішення важливого завдання статистичної радіотехніки – прогнозу нестационарних даних трафіку. Показано, що навіть використання простої моделі  $AR(1)$  дає позитивний результат прогнозу.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Бриллинджер Д.Р. Временные ряды. Обработка данных и теория. – М.: Мир, 1980. – 536 с.
2. Хеннан Э. Анализ временных рядов. – М.: Наука, 1964. – 216 с.
3. Марпл.–мл. С. Л. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
4. Тихонов В.А., Карташов В.М., Карташов О.В. Модель оцінювання статистичних характеристик довгострокової складової випадкового процесу на прикладі аналізу середньомісячних температур // Радіотехніка (Харків). - 2022. - № 209. - С. 239-245.
5. Тихонов В.А., Чеботарёва Д.В. Прогнозирование потока данных в сетях мобильной связи // Матеріали III 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.
6. Montgomery D.C., Johnson L. A., Gardiner J. S. Forecasting & Time Series Analysis. – Mc.Graw-Hill Inc., 1990. – P. 384.
7. Brockwell P.J., Davis R.A. Introduction to Time Series and Forecasting. – Springer, 2002. – P. 434.
8. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. – 755 с.
9. Тихонов В.И. Статистическая радиотехника. – М.: Радио и связь, 1982. – 624 с.
10. Хеннан Э. Многомерные временные ряды. – М.: Мир, 1974. – 547 с.
11. Малахов А.Н. Кумулянтный анализ случайных негауссовых процессов и их преобразований. – М.: Сов. Радио, 1978. – 376 с.
12. Бокс Дж., Дженкинс Г. Анализ временных рядов: Пер. с. англ. – М.: Мир, 1974. – Вып.2. – 197 с.
13. Бендат Д. Основы теории случайных шумов и ее приложения. – М.: Наука, 1965. – 464 с.

14. Бокс Дж., Дженкинс Г. Анализ временных рядов: Пер. с. англ. – М.: Мир, 1974. – Вып.1. – 406с.
15. Wold Н.О.А. A Study in the Analysis of Stationary Time series, dissertation, Uppsala University, 1938.
16. Сукачев З.А. Сотовые сети связи с подвижными объектами: Учебн. пособ. - Одесса: УГАС им. Попова, 2000.
17. Громаков Ю.А. Стандарты и Системы подвижной радиосвязи. ТЭК. Т.67. - М.: «Эко-Трендз», 1996.
18. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows: Учебн. пособие. – М.: Финансы и статистика, 1999. – 384 с.
19. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. – М.: Статистика, 1979. – 254 с.
20. Безрук В.М., Голиков В.С., Тихонов В.А. Распознавание случайных сигналов, описываемых авторегрессионной моделью // Радиоэлектроника. – 2004. – №4. – С.59–65. (Изв. вузов).
21. Омельченко В.А., Безрук В.М., Коваленко Н.П. Распознавание заданных радиосигналов при наличии неизвестных сигналов на авторегрессионной основе // Радиотехника. – 2001. – №123. – С. 195–199.
22. В.А. Тихонов, В.М. Безрук. Модели линейного предсказания в статистической радиотехнике. Монография. Харьков. 2020.-468 с.