

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМ. В. ДАЛЯ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ЕЛЕКТРОНІКИ
КАФЕДРА КОМП'ЮТЕРНИХ НАУК ТА ІНЖЕНЕРІЇ

До захисту допускається
Завідувач кафедри
_____ Скарга-Бандурова І.С.
« ____ » _____ 20__ р.

МАГІСТЕРСЬКА РОБОТА

НА ТЕМУ:

_____ Методи та інформаційна технологія об'єднання даних
_____ он-лайн моніторингу поверхневих вод

Освітньо-кваліфікаційний рівень “Магістр”
Спеціальність 122 – “Комп’ютерні науки”

Науковий керівник роботи:

_____ (підпис)

Кривуля Г.Ф.

_____ (ініціали, прізвище)

Консультант з охорони праці:

_____ (підпис)

Критська Я.О.

_____ (ініціали, прізвище)

Студент:

_____ (підпис)

Топчій А.О.

_____ (ініціали, прізвище)

Група:

_____ КН-17дм

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВОЛОДИМИРА ДАЛЯ

Факультет Інформаційних технологій та електроніки
Кафедра Комп'ютерних наук та інженерії
Освітньо-кваліфікаційний рівень Магістр
Напрямок підготовки _____
(шифр і назва)
Спеціальність _____
(шифр і назва)

ЗАТВЕРДЖУЮ:

Завідувач кафедри _____
І.С. Скарга-Бандурова
« ____ » _____ 20__ р.

**З А В Д А Н Н Я
НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ**

Топчій Анастасія Олександрівна

(прізвище, ім'я, по батькові)

1. Тема роботи **МЕТОДИ ТА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ОБ'ЄДНАННЯ
ДАНИХ ОН-ЛАЙН МОНІТОРИНГУ ПОВЕРХНЕВИХ ВОД**

керівник проекту (роботи) Кривуля Г.Ф., д.т.н., професор
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від " 18 " 10 2018 р. № _____

2. Строк подання студентом роботи _____
3. Вихідні дані до роботи _____

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

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) _____

АНОТАЦІЯ

Топчій А.О. Методи та інформаційна технологія об'єднання даних он-лайн моніторингу поверхневих вод.

Проведений аналітичний огляд існуючих алгоритмів, методів та програмних засобів для об'єднання даних та прогнозування випадків аномального забруднення водних об'єктів. В роботі запропоновано використання алгоритму на основі методу Демпстера-Шафера. Даний алгоритм дозволяє здійснювати об'єднання даних і максимально точно прогнозування оперуючи недостатніми вхідними даними.

Ключові слова: Метод Демпстера-Шафера, прогнозування випадків забруднення водних об'єктів, об'єднання даних.

АННОТАЦИЯ

Топчий А.А. Методы и информационная технология объединения данных он-лайн мониторинга поверхностных вод.

Проведен аналітичний огляд існуючих алгоритмів, методів та програмних засобів для об'єднання даних та прогнозування випадків аномального забруднення водних об'єктів. В роботі запропоновано використання алгоритму на основі методу Демпстера-Шафера. Даний алгоритм дозволяє здійснювати об'єднання даних і максимально точно прогнозування оперуючи недостатніми вхідними даними.

Ключевые слова: Метод Демпстера-Шафера, прогнозування забруднення водних об'єктів, об'єднання даних.

ABSTRACT

Topchy A.O. Methods and information technology for data integration on-line monitoring of surface water.

An analytical review of the existing algorithms, methods and software for data fusion and forecasting cases of abnormal pollution of water bodies was carried out. The paper proposes the use of an algorithm based on the Dempster-Schäfer method. This algorithm allows for the data fusion and the most accurate prediction in terms of insufficient input data.

Key words: Dempster-Schafer method, prediction of pollution of water bodies, data fusion.

ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ І ЗАСОБІВ ОБ'ЄДНАННЯ ДАНИХ. ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕНЬ.....	10
1.1 Аналіз вимог до актуальності питання.....	10
1.2 Аналіз програмних та інструментальних засобів	11
1.2.1 <i>Deductor</i>	11
1.2.2 <i>Flume</i>	12
1.2.3 <i>IBM Watson Analytics</i>	13
1.2.4 <i>Dell EMC Analytic Insights Module</i>	13
1.2.5 <i>Azure Machine Learning</i>	14
1.3.1 <i>Методи об'єднання даних на основі фільтра Калмана</i>	16
1.3.2 <i>Процес інтеграції даних</i>	19
1.3.3 <i>Алгоритм злиття даних на основі теорії доказів Демпстера-Шафера</i>	23
1.4 Постановка наукової задачі та обґрунтування методики досліджень	29
1.5. Висновки до першого розділу	29
РОЗДІЛ 2 МАТЕМАТИЧНА ПОСТАНОВКА ТА ОПИС АЛГОРИТМУ ПРОГНОЗУВАННЯ ВИПАДКІВ ЗАБРУДНЕННЯ ТА ОБ'ЄДНАННЯ ДАНИХ.....	30
2.1 Етапи алгоритму Демпстера Шафера	30
2.1.1 <i>Нормалізація даних</i>	30
2.1.2 <i>Модель авторегресії ARIMA</i>	33
2.1.3 <i>Байєсівський критерій</i>	38
2.1.4 <i>Метод найменших квадратів</i>	40
2.1.5 <i>Прогнозування залишків</i>	41
2.2 Висновки до другого розділу	41
РОЗДІЛ 3. РЕАЛІЗАЦІЯ МЕТОДУ ОБ'ЄДНАННЯ ТА ПРОГНОЗУВАННЯ ДАНИХ	42
3.1 Реалізація нормалізації даних	42
3.2 Прогнозування за допомогою моделі ARIMA.....	46
3.3 Розрахунок критерія Байєса.....	54
3.4 Визначення залишків	55
3.5 Висновки до третього розділу	56
РОЗДІЛ 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ. ЕКОЛОГІЯ	57

4.1 Загальні питання з охорони праці	57
4.2 Аналіз стану умов праці	58
4.2.1 <i>Вимоги до приміщень</i>	58
4.2.2 <i>Вимоги до організації місця праці</i>	59
4.2.3 <i>Навантаження та напруженість процесу праці</i>	60
4.3 Виробнича санітарія	61
4.3.1 <i>Аналіз небезпечних та шкідливих факторів при виробництві (експлуатації) виробу</i>	61
4.3.2 <i>Пожезна безпека</i>	62
4.3.3 <i>Електробезпека</i>	63
4.4 Гігієнічні вимоги до параметрів виробничого середовища	63
4.4.1 <i>Мікроклімат</i>	63
4.4.2 <i>Освітлення</i>	64
4.3 Шум та вібрація, електромагнітне випромінювання	66
4.4 Вентилювання	67
4.5 Заходи з організації виробничого середовища та попередження виникнення надзвичайних ситуацій	67
4.6 Охорона навколишнього природного середовища	70
4.6.1 <i>Загальні дані з охорони навколишнього природного середовища</i>	70
4.7 Висновки до четвертого розділу	71
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	73
ДОДАТОК А СЛАЙДИ ПРЕЗЕНТАЦІЇ	78

ВСТУП

Обґрунтування вибору теми дослідження. Тема забруднення водних середовищ є актуальною вже на протязі кількох десятків років. Вода є неодмінною складовою життя людини, і від її якості залежить якість життя і навколишнього середовища. Для виявлення якості води здійснюють он-лайн моніторинг її стану. На основі отриманих даних можливо здійснювати прогнозування випадків аномального забруднення води та запобігати їм. Дослідження якості вод проводиться в різноманітних ділянках: місцях скидання стічних і вод зливових каналізацій населених місць, сільськогосподарських комплексів; місцях скидання стічних вод окремих підприємств, електростанцій, великих промислових комплексів; місцях скидання колекторно-дренажних вод, що відводяться з зрошуваних або осушуваних земель; кінцевих створах великих і середніх річок, що впадають в моря, внутрішні водойми; межах районів, територіальних одиниць, транскордонних водних об'єктів; в районах водокористування та прилеглих територій; в місцях нересту і концентрування цінних / рідкісних видів фауни водойм; місцях розробки і транспортування корисних копалин. Основними параметрами, що вимірюються є алюміній, аміак й іони амонію, зважені речовини, гідрокарбонати, Біохімічне споживання кисню (БПК), смак і присмак, водневий показник (рН), залізо загальне, жорсткість загальна, запах (без нагрівання), кальцій, карбонати, метали важкі (мідь, свинець, цинк), нікель, пінявість, нафтопродукти, нітрати, нітрити, окислюваність біхроматна (ХПК), окислюваність перманганатна, поверхнево-активні речовини (ПАР), аніоноактивні речовини, прозорість, сухий залишок, сульфати, розчинений кисень, сірководень (сульфіди), феноли, фосфати (поліфосфати, загальний Р), хлориди, фториди, хромати (сумарно), каламутність, кольоровість.

Очевидно, що різноманітність параметрів, засобів та умов збору даних обумовлюють використання різних програмно-апаратних платформ, що потребують особливих методів збору, передачі та опрацювання даних. Більше того, в реальному житті різні засоби, середовища та платформи повинні співіснувати з іншими типами ІТ.

Тому обґрунтованою є тема магістерської роботи, у якій вирішується **науково-прикладне завдання** розроблення моделей і методу інформаційної технології об'єднання даних он-лайн моніторингу якості поверхневих вод.

Об'єкт дослідження – процеси забезпечення об'єднання даних он-лайн моніторингу поверхневих вод, .

Предмет дослідження – моделі та методи інформаційної технології оцінки і забезпечення максимальної точності результатів об'єднання даних з урахуванням

прогнозування аномальних випадків забруднення водного середовища.

Мета і завдання дослідження. Метою дослідження є підвищення якості об'єднання даних он-лайн моніторингу, та здійснення, за допомогою вдосконалених математичних моделей, досягти максимально точного прогнозування аномальних забруднень поверхневих вод.

Для досягнення мети дослідження необхідно вирішити такі **завдання**:

- проведення аналізу математичних методів і програмних засобів для проведення злиття даних та прогнозування;
- розроблення математичного алгоритму на основі теорії доказів Демпстера-Шафера, що дасть змогу робити якомога точні прогнози, щодо стану якості води;

Методи дослідження. Проведені в роботі дослідження основані на методах об'єднання та прогнозування даних, які використовувались при розробленні математичного алгоритму для об'єднання даних та прогнозування аномальних випадків забруднення водного середовища;

Особистий внесок здобувача полягає у розробленні нових моделей, методів та інструментальних засобів, що дозволяють вирішити поставлені задачі. Усі основні результати отримані автором особисто. У роботах, опублікованих у співавторстві, автору належать: опис алгоритмів об'єднання даних та методів прогнозування аномальних випадків забруднення водних середовищ.

Апробація матеріалів дисертації. Основні положення, ідеї, висновки магістерської роботи доповідалися та обговорювалися на IV Всеукраїнській науково-практичній конференції «Актуальні проблеми науково-промислового комплексу регіонів-2018» (м. Рубіжне, 23-27 квітня 2018 року).

Зв'язок з науковими програмами, планами, темами. Магістерська робота виконана у Східноукраїнському національному університеті ім. В. Даля у відповідності з державними програмами і планами науково-дослідницьких робіт кафедри комп'ютерних наук та Інженерії Східноукраїнського національного університету ім. в. Даля в межах НДР «Проектування системи моніторингу та контролю водних об'єктів на основі технології інтернет речей», реєстраційний номер 0116U005784.

Практичне значення отриманих результатів полягає в тому, що основні наукові положення магістерської роботи реалізовані у виді розрахункових моделей, які утворюють прикладну інформаційну технологію об'єднання даних та прогнозування

аномалій забруднення поверхневих вод.

Публікації. За темою магістерської роботи з викладенням її основних результатів опубліковано 3 наукових праць, серед яких 1 стаття у наукових фахових виданнях України; 2 тез доповідей.

Структура та обсяг магістерської роботи. Робота складається із вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Загальний обсяг дисертації складає 84 сторінок, з яких основний текст на 62 сторінках, список використаних джерел із 57 найменувань на 2 сторінках, додатки на 7 сторінках. Робота містить 4 таблиці, та 53 рисунків.

РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ І ЗАСОБІВ ОБ'ЄДНАННЯ ДАНИХ. ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕНЬ

1.1 Аналіз вимог до актуальності питання

За міжнародним стандартом моніторинг – це багаторазове вимірювання для спостереження за змінами будь-якого параметра в певному інтервалі часу; система довготривалих спостережень, оцінювання, контролювання і прогнозування стану й зміни об'єктів.

Нагромаджені у процесі моніторингу дані інформують про стан довкілля на певний час, основні процеси, тенденції, що відбуваються в ньому. Ці відомості допомагають спрогнозувати його розвиток, передбачити надзвичайні ситуації природного та техногенного походження, а також спланувати науково обгрунтовані природоохоронні заходи для створення безпечних умов життєдіяльності. Особливо актуальним є відстеження антропогенних змін у природі. Необхідним є дослідження середовища у динаміці, тобто оцінювання минулого, сучасного його станів, а також прогнозування змін його параметрів у майбутньому.

Об'єктом даного дослідження є дослідження математичних моделей та методів для об'єднання даних, про стан поверхневих вод, отриманих шляхом он-лайн моніторингу, та дослідження математичних алгоритмів для прогнозування аномальних випадків забруднення[24].

Мульти-сенсорне об'єднання даних (MSDF) визначається як процес об'єднання інформації з різних джерел для отримання найбільш точних та повних уніфікованих даних про об'єкт, активність або подію [14]. Загальне уявлення про з'єднання декількох датчиків представлено на рисунку 1.1. Результати отримані з кількох датчиків дають більш точні значення та є надійними, на відміну від даних, отриманих з одного датчика.

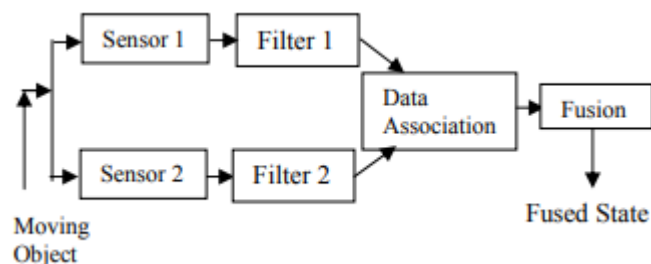


Рисунок 1.1 – Схема з'єднання даних з різних датчиків [джерело: 16]

В даній роботі проводиться огляд існуючих програмних засобів, математичних моделей та алгоритмів для злиття даних та виявлення випадків забруднення води для їх майбутнього прогнозування. На основі проведеного аналізу виявлено найбільш точний метод для проведення подальших розрахунків.

Основними проблемами проектування є пошук надійних вхідних даних для проведення розрахунків, незначна кількість програмного забезпечення для реалізацій запропонованих методів та наявність конфліктів.

Розрізняють два типи конфліктів: конфлікти, викликані невизначеними значеннями, і конфлікти, викликані суперечливими значеннями.

Невизначеність означає, що в одному джерелі даних містяться невідомі значення (nul), а в іншому - відомі. Проблема полягає в тому, що семантика невизначених значень (null) може сильно відрізнятися. Розрізняють три варіанти: невідомі значення, неіснуючі значення, приховані значення. Суперечливість значень означає поява двох різних ненульових (not null) значень. [23]

1.2 Аналіз програмних та інструментальних засобів

Окрім самих математичних методів існують вже готові програмні засоби, для вирішення питання об'єднання даних та прогнозування на основі відомих даних.

1.2.1 Deductor

Deductor [7]– аналітична платформа, розроблена компанією BaseGroup Labs. В Deductor вбудовані найбільш затребувані алгоритми аналізу (дерева рішень, нейронні мережі, які самоорганізуються карти і т. Д.), Є десятки способів візуалізації і передбачена інтеграція з безліччю джерел / приймачів даних.

В системі застосовуються технології, які на базі єдиної архітектури дозволяють пройти всі етапи побудови аналітичної платформи: від створення сховища даних до автоматичного підбору моделей і візуалізації отриманих результатів.

В Deductor реалізований сценарний підхід, що передбачає візуальне проектування логіки обробки за допомогою майстрів без програмування. Аналітики отримують всі технології аналізу: Data Warehouse, OLAP, Data Mining і можуть створювати сценарії аналізу без залучення розробників ПО. Також аналітична платформа передбачає очистку даних, зокрема, їх дедуплікація, тобто проведення оцінки схожості об'єктів, збагачення

даних і об'єднання в єдину коректну запис.

З Deductor можна:

Отримувати дані з різнорідних джерел, консолідувати їх в єдиному сховищі і відображати у вигляді звітів і OLAP-кубів.

Знаходити приховані закономірності і оцінювати якість моделей за допомогою Data Mining.

Сегментувати об'єкти аналізу, визначати цільові ринки, оптимізувати роботу зі споживачами і більш раціонально використовувати ресурси.

Платформа поставляється в 3 версіях: Academic, Professional і Enterprise. Перший варіант є абсолютно безкоштовним і призначений для освітніх цілей, другий орієнтований на професійний аналіз даних в рамках робочих груп, третій - на промислове корпоративне використання[25].

1.2.2 Flume

Flume – розподілений сервіс для збору і агрегування даних журналу подій з різних джерел в центральне сховище даних, наприклад, HDFS. Інструмент використовується переважно для передачі неструктурованих даних. Flume не обмежується тільки журналюванням. Оскільки джерела даних у Flume є налаштованим, сервіс також використовується для завантаження даних подій, даних з соціальних медіа-платформ, електронної пошти, зображень, відео та інших джерел. Завдяки підтримці декількох джерел даних і горизонтального масштабування.

Джерело Flume в зрозумілому для інструменту форматі приймає події, які передаються в систему із зовнішнього джерела, наприклад, журналу веб-сервера або даних соціальних медіа. У Flume є так звані «Колектори», які збирають дані з різних джерел і поміщають їх в централізоване сховище на кшталт HDFS або HBase. Потім дані переходять від журналів веб-серверів до каналів, які їх зберігають, поки вони не споживаються Flume Sink'ами, які потім остаточно вміщують інформацію в сховище на кшталт HDFS. Інструмент дозволяє збільшити надійність шляхом організації мульти-потоків, де дані зберігаються в кількох каналах, поки не досягнуть HDFS. А оскільки у Flume застосовується транзакційний підхід, ймовірність втрати даних перед досягненням кінцевого місця призначення виключена.

До переваг Flume відносять:

Роль посередника потоків даних між джерелом і приймачем. Коли частота

поглинання даних менше частоти їх генерації, Flume прагне врівноважити потік по своїх каналах.

Дані легко збираються з декількох джерел. Колектори Flume здатні підключатися до різних джерел з метою збору різнорідних даних і зберігати їх в централізованому сховищі.

Flume надійний. Ймовірність втрати даних перед їх надходженням до пункту призначення виключена.

Можливість відновлення. Локальна файлова система проводить резервне копіювання каналу файлу. Також в оперативній пам'яті присутній канал, який зберігає події в черзі і дуже швидко відновлює події[25].

1.2.3 IBM Watson Analytics

Програмне рішення IBM Watson Analytics[11] можна сміливо відносити до одним з найбільш потужних інструментів для аналізу великих даних. Це платформа, яка доступна для роботи на хмарі. Ввівши в систему вихідні дані, користувач отримує масив впорядкованої інформації з виявленими зв'язками між окремими елементами.

На практиці IBM Watson Analytics дозволяє компаніям з'ясувати, як сторонні чинники впливають на зміну фінансових потоків або виробничих потужностей клієнта. Система автоматично проводить складні обчислення, демонструючи користувачеві найбільш важливі для нього і його бізнесу фактори, патерни і зв'язку між окремими елементами.

Графічний інтерфейс системи зрозумілий і приємний, адже реалізована логіка «drag-and-drop». Всі необхідні дані і графіки можна розмістити на робочому просторі, просто перетягуючи їх. Візуалізація, графіки, діаграми: все це швидше допомагає розібратися в поточній ситуації. Watson Analytics швидко проводить графічну обробку даних будь-якої складності[25].

1.2.4 Dell EMC Analytic Insights Module

Analytic Insights Module від Dell EMC[8] - це інструмент, який створений з метою об'єднання обслуговуючої себе аналітики з розробкою «хмарних» додатків на єдиній платформі. Такий підхід дозволяє фахівцям з аналізу даних зосередитися на швидкому створенні (кілька днів або тижнів замість місяців) моделей з великою цінністю для

бізнесу.

Щоб реалізувати таку концепцію, Dell EMC створила відкриту платформу, призначену для відстеження повного життєвого циклу аналізу даних за допомогою декількох ключових компонентів: Data Lake, Data Curator, Data Governor, Data and Analytic Catalog. За допомогою цих складових, організації можуть збирати потрібну інформацію шляхом глибинного аналізу і формувати єдине бачення щодо всіх даних.

Data Lake відповідає за консолідацію даних в єдиному сховищі. Цей компонент нівелює складність зберігання силосу даних, пов'язаного з великою кількістю розрізної інформації.

Разом, з Dell EMC Analytic Insights Module користувач може:

досліджувати, використовувати і індексувати всі дані в єдиному форматі з допомогою Data Curator;

пізнати походження, забезпечити управління і безпеку для всіх додатків і сховищ даних за допомогою Data Governor;

перетворити всю важливу інформацію в керовані даними додатка і бізнес-моделі[25].

1.2.5 Azure Machine Learning

Azure Machine Learning[1] – це хмарний сервіс прогнозувальної аналітики, який дозволяє швидко створювати і розгортати прогнозовані моделі. Система цінується своєю простотою: для повноцінної роботи з машинним навчанням в середовищі Azure Machine Learning не потрібно бути математиком. Інтегроване середовище розробки - ML Studio - надає «drag-and-drop» інструменти і прості діаграми потоків даних, що не тільки зменшує обсяг коду, але і, завдяки вбудованій бібліотеці простих експериментів (так називаються проекти в ML Studio), зберігає користувачам багато часу.

Інструмент пропонує готовий до використання набір з бібліотек алгоритмів. Їх можна використовувати для створення інтелектуальних моделей на підключених до Інтернету комп'ютерах.[1]

В результаті огляду існуючих програмних засобів для можна зробити висновок, що дана галузь ще потребує вдосконалення і розробки нових, більш спеціалізованих програмних засобів. Існує потреба в програмному засобі, який би дав змогу здійснювати об'єднання отриманих, за допомогою он-лайн моніторингу, даних та прогнозування ймовірності забруднення поверхневих вод, з найбільш можливою точністю.

1.3 Аналіз математичних моделей і методів для вирішення задачі об'єднання та прогнозування даних

В ході дослідження було виявлено, що останнім часом методи та алгоритми об'єднання даних все більше привертають увагу науковців. Такі методи знайшли своє застосування у різних галузях і сферах діяльності. Переваги та недоліки методів об'єднання даних наводяться в таблиці 1.1.

Таблиця 1.1 – Переваги та недоліки методів об'єднання даних з датчиків[18].

Методи	Переваги	Недоліки
Ймовірнісні методи	Надання методів оцінки моделі; допускає неконтрольовану класифікацію; оцінювати стан змінних; зменшити помилки в оцінці плавного розташування; збільшити обсяг даних без зміни його структури або алгоритму; створюють плавну коваріаційну матрицю, яка краще відображає очікувану помилку розташування.	Вимагати апріорні імовірнісні знання інформації, які не завжди доступні або реалістичні; класифікація залежить від початкової точки; непридатним для великомасштабних систем; вимагає апріорного знання матриць збігів невизначеностей, пов'язаних з моделлю системи і її вимірами.
Статистичні методи	Точність поліпшується від зменшення помилки передбачення; висока точність у порівнянні з іншими місцевими оцінювачами; стійкої по відношенню до невідомої крос-коваріації.	Для отримання крос-дисперсії потрібно складне і складне обчислення; складності і більшої обчислювальної навантаження.
Методи теорії баз даних	Дозволяє включити невизначеність і неточність; легко реалізувати; Здібності до навчання; надійні і галасливі дані і здатні представляти складні функції.	Вилучення знань вимагає втручання людського досвіду (наприклад, лікарів), яке вимагає часу і / або може привести до ухилення від інтерпретації; утруднення у визначенні адекватного розміру прихованого шару; нездатність пояснити рішення; відсутність прозорості даних.
Методи доведення доказів	Призначення ступеня невизначеності для кожного джерела.	Вимагати присвоєння ступеня доказовості всім концепціям.

1.3.1 Методи об'єднання даних на основі фільтра Калмана

Серед алгоритмів Multi-Sensor Data Fusion (MSDF) розглядають методи на основі фільтра Калмана, зокрема, State Vector Fusion (SVF), Measurement Fusion (MF) та Gain fusion (GF)[14] [16].

Метод State Vector Fusion (SVF) являє собою стан злиття даних на основі фільтра Калмана (KF). KF [14] даний алгоритм застосовується незалежно для кожного датчика (даних) і генерує оцінки стану. Як показано на рис. 1.2, кожен датчик використовує оцінювач, який отримує оцінку вектора стану та пов'язаних з ним матриць коваріації від даних цього пов'язаного датчика. Тоді в центрі синтезу проводиться кореляція між доріжками та отримується вектор злитого стану [12].

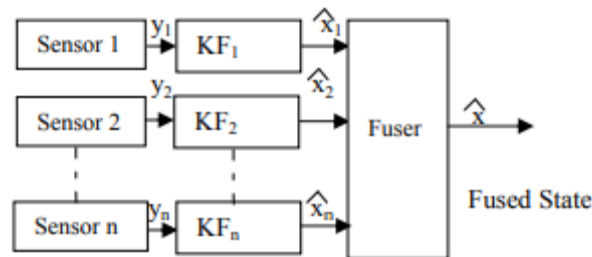


Рисунок 1.2 – Метод State Vector Fusion (SVF) [джерело: 16]

Така модель стану задається формулами:

$$X_{(k-1)} = \Phi X_{(k)} + G w_k \quad (1.1)$$

$$Z_{(k-1)} = H X_{(k)} + v_{(k)} \quad (1.2)$$

Де

X – вектор стану;

Φ – матриця переходів;

G – матриця посилення

W – процес шуму

Z – вектор вимірювання

H – матриця спостережень

V – вимірювальний шум

Рекурсивний алгоритм фільтра Калмана обчислюється наступними рівняннями.

Розповсюдження часу та коваріації задаються формулами:

$$\hat{X}_{(k+1)} = \Phi \hat{X}_{(k)} \quad (1.3)$$

$$\hat{P}_{(k+1)} = \Phi \hat{P}_{(k)} \Phi^T + G Q G^T \quad (1.4)$$

Оновлення станів та коваріантних вимірювань задаються формулами

$$K_{(k+1)} = \hat{P}_{(k+1)} \Phi^T [H \hat{P}_{(k+1)} H^T] \quad (1.5)$$

$$\hat{X}_{(k+1)} = \hat{X}_{(k+1)} + K_{(k+1)} [Z_{(k+1)} - H \hat{X}_{(k+1)}] \quad (1.6)$$

$$\hat{P}_{(k+1)} = [I - K_{(k+1)} H] \hat{P}_{(k+1)} \quad (1.7)$$

Де

$\hat{X}_{(k)}$ є вектором стану датчика в момент k

$\hat{P}_{(k)}$ це коваріація датчика, яка являє собою квадрат різниці між бажаним вектором стану та вектором стану сенсора

$K_{(k+1)}$ посилення Калмана

$\hat{X}_{(k+1)}$ це фільтрована оцінка стану

$\hat{P}_{(k+1)}$ це фільтрована оцінка коваріації

Модель Measurement Fusion (MF) з'єднує спостереження сенсора безпосередньо через вимірювальну модель і використовує фільтр Калмана [21] для оцінки вектора злитого стану. Злиття вимірювань зображено на рис. 1.3[16].

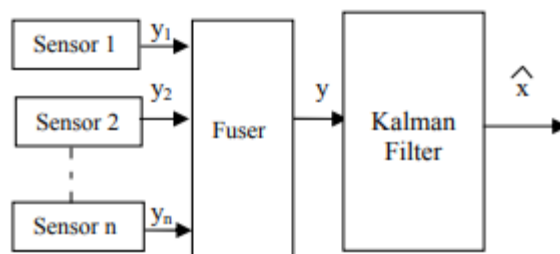


Рисунок 1.3 – Measurement Fusion [джерело: 16]

Рекурсивний алгоритм фільтра Калмана обчислюється наступними рівняннями. Розповсюдження часу та коваріації задаються формулами

$$\hat{X}_{(k+1)}^f = \Phi \hat{X}_{(k)}^f \quad (1.8)$$

$$\hat{P}_{(k+1)}^f = \Phi \hat{P}_{(k)}^f \Phi^T + G Q G^T \quad (1.9)$$

Оновлення даних з вимірювання стану та коваріації задаються формулами:

$$K_{(k+1)}^f = \hat{P}_{(k+1)}^f H^T [H \hat{P}_{(k+1)}^f H^T + R]^{-1} \quad (1.10)$$

$$\hat{X}_{(k+1)}^f = \hat{X}_{(k+1)}^f + K_{(k+1)}^f [Z_{(k+1)} - H \hat{X}_{(k)}^f] \quad (1.11)$$

$$\hat{P}_{(k+1)}^f = [I - K_{(k+1)}^f H] \quad (1.12)$$

Де

$\hat{X}_{(k+1)}^f$ це фільтрований конденсований стан

$\hat{P}_{(k+1)}^f$ представляє собою фільтровану конденсовану коваріацію

При використанні алгоритму Gain Fusion (GF) [5], глобальний процесор отримує інформацію у вигляді коефіцієнта посилення Калмана від локальних систем і формулює глобальну оцінку. Рекурсивний алгоритм фільтра Калмана для посилення злиття наведено в наступних рівняннях. Час розповсюдження глобальної оцінки задається формулами:

$$\hat{X}_{(k+1)}^f = \Phi \hat{X}_{(k)}^f \quad (1.13)$$

$$\hat{P}_{(k+1)}^f = \Phi \hat{P}_{(k)}^f \Phi^T + G Q G^T \quad (1.14)$$

Локальні фільтри позначаються як:

$$\hat{X}_{(k+1)}^m = \hat{X}_{(k+1)}^f \quad (1.15)$$

$$\hat{P}_{(k+1)}^m = \hat{P}_{(k+1)}^f \quad (1.16)$$

Де

m – кількість датчиків;

$\hat{X}_{(k)}^f$ – фільтрований конденсований стан;

$\hat{P}_{(k)}^f$ – фільтрована з'єднана коваріація;

При порівнянні цих трьох методів був зроблений висновок, виходячи з показників продуктивності, та дослідів науковців, що працювали з цими методами, найбільш точним є алгоритм State Vector Fusion. Але взагалі ці методи ще потребують подальшого вдосконалення.[16]

1.3.2 Процес інтеграції даних

Для злиття даних також застосовують процес інтеграції даних. Такий процес складається з наступних етапів: зіставлення схем, інтеграція схем, трансформація даних, дозвіл сутностей, злиття сутностей, як показано на рис. 1.4 [3][23].

Даний алгоритм розрізняє два основні підходи до злиття даних. Ці підходи засновані на операції об'єднання (union-based) або на операції з'єднання (Join-based). Розрізняють такі основні операції.

Minimum Union [2] (union-based). Операція являє собою виконання операції outer union, а потім видалення з результату всіх поглинаються (subsumed) кортежів, як показано на рис. 1.5.

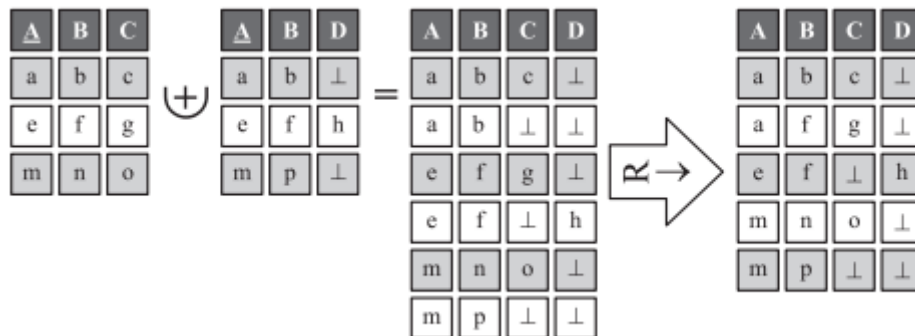


Рисунок 1.4 – Приклад операції Minimum Union [джерело: 23]

Complementation Union [2] (union-based). Операція являє собою виконання операції outer union, а потім доповнення (complementation) всіляких кортежів.

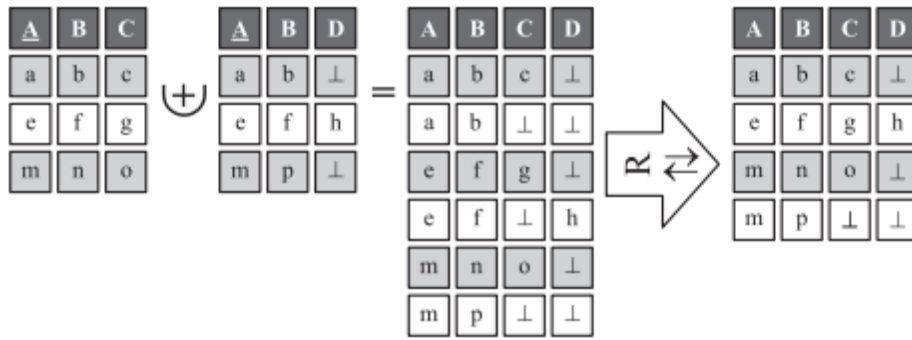


Рисунок 1.5 – Приклад операції Complemetation Union [джерело: 23]

Grouping and Aggregation [2] (union-based). Операція передбачає виконання outer union, а потім угруповання за загальним атрибутом і застосування функції агрегації до решти атрибутів. Нижче представлений приклад операції на мові SQL.

```
WITH OU AS (
  (SELECT A, B, C, NULL AS D FROM U1)
  UNION (ALL)
  (SELECT A, B, NULL AS C, D FROM U2)),
SELECT A, MAX(B), MIN(C), SUM(D)
FROM OU GROUP BY A
```

Full Disjunction [18] (join-based). Операція являє собою full outer join (стандартну реляційну операцію), після чого застосовується subsumption до результату, як показано на рис.1.6.

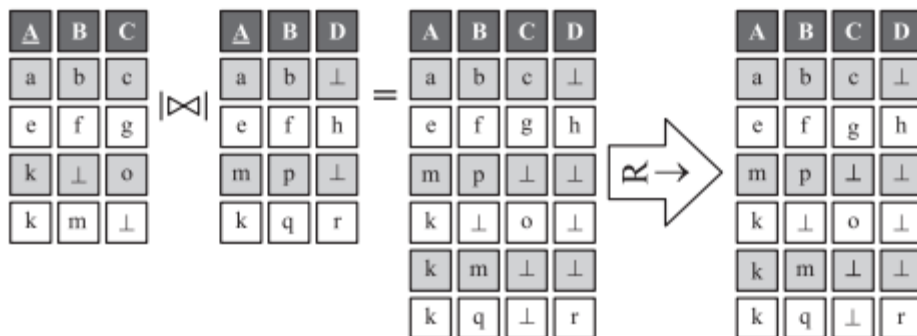


Рисунок 1.6 – Приклад операції Full Disjunction [джерело: 23]

Match Join [2] (union- + join-based). В операції вибираються всілякі комбінації значень атрибутів, після чого виконується full outer join.

Фактично реалізується outer union двох колекцій.

Потім визначається $N - 1$ допоміжних відносин, де N - число атрибутів, а кожне з відносин містить по два атрибути: один загальний і якийсь інший. Після чого відбувається full outer join ($N - 1$) -го відносини. Приклад реалізації операції на мові SQL представлений нижче [23]

```
WITH OU(A,B,C,D) AS (
  (SELECT A, B, C, NULL AS D FROM U1)
  UNION
  (SELECT A, B, NULL AS C, D FROM U2)),
// ← Outer Union
B_V(A,B) AS (SELECT DISTINCT A, B FROM OU),
// ← 1-е відношення (N = 4)
C_V(A,C) AS (SELECT DISTINCT A, C FROM OU),
// ← 2-е відношення (N = 4)
D_V(A,D) AS (SELECT DISTINCT A, D FROM OU),
// ← 3-є відношення (N = 4)
SELECT A, B, C, D
FROM B_V FULL OUTER JOIN C_V FULL OUTER
JOIN D_V // ← Full Outer Join
```

Merge (union- + join-based) [23]. Операція об'єднує операції з'єднання і об'єднання. Для кожного загального атрибута формуються дві версії значень, нульові значення видаляються функцією COALESCE (вибір першого ненульового значення).

Нехай дано два відносини: A з набором атрибутів {a, b, c} і B з набором атрибутів {a, b, d}. Нехай a - конфліктує атрибут, b - атрибут з нульовими значеннями. Приклад реалізації операції на мові SQL представлений нижче та на рис. 1.7.

```
(SELECT A.a, COALESCE(A.b, B.b), A.c, B.d
FROM A LEFT OUTER JOIN B ON A.a = B.a)
UNION
(SELECT B.a, COALESCE(B.b, A.b), A.c, B.d
FROM A RIGHT OUTER JOIN B ON A.a = B.a)
```

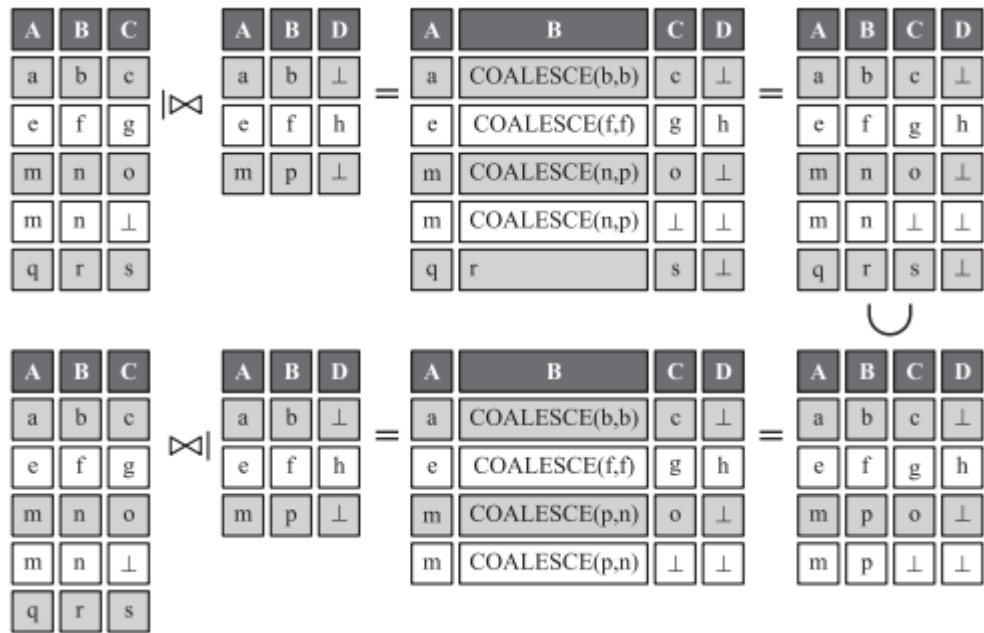


Рисунок 1.7 – Приклад операції Merge [джерело:23]

Grouping and Aggregation (union-based). Операція являє собою угруповання по деякому атрибуту, а потім використання різноманітних агрегуються функцій. Як переваги даного підходу можна виділити його реалізацію в більшості СУБД і ефективне виконання[23].

Приклад реалізації на SQL представлений нижче:

```
WITH OU AS (
  (SELECT A, B, C, NULL AS D FROM U1)
  UNION (ALL)
  (SELECT A, B, NULL AS C, D FROM U2)),
SELECT A, MAX(B), MIN(C), SUM(D)
FROM OU
GROUP BY A
```

Data Fusion оператор [4] - Fuse By (union-based).

У деяких системах пішли далі використання стандартних операцій угруповання і агрегації.

Ключове слово FUSE BY використовується замість GROUP BY, і семантика у нього аналогічна. Замість використання стандартних функцій агрегації використовується вбудована функція RESOLVE, якій параметром передається саме значення і ім'я функції вирішення конфліктів. приклад реалізації на SQL представлений нижче:

```

SELECT ID,
RESOLVE(Title, Choose(IMDB)),
RESOLVE(Year, Max),
RESOLVE(Director, COALESCE),
RESOLVE(Rating, COALESCE),
RESOLVE(Genre, Concat)
FUSE FROM IMDB, Filmdienst
FUSE BY (ID)
ON ORDER Year DESC

```

1.3.3. Алгоритм злиття даних на основі теорії доказів Демпстера-Шафера

Також, велика увага приділяється застосуванню метода Демпстера-Шафера (Dempster–Shafer). Даний метод виявлення забруднень на основі теорії доказів Демпстера-Шафера. Однією з переваг алгоритму є можливість оперувати недостатніми даними. Метод виявлення полягає в захисті систем водопостачання від випадкових або навмисних випадків забруднення та представляє собою альтернативу традиційній ймовірнісній теорії для математичного представлення невизначеностей[22]. Досягнення результату відбувається шляхом прогнозування майбутніх параметрів якості води за допомогою моделі авторегресії(AR). В свою чергу модель авторегресії, спираючись на недавні вимірювання, передбачає майбутні параметри якості води. На наступному етапі, за допомогою імовірнісного підходу визначається вірогідність часового ряду залишків, утворених шляхом порівняння прогнозованих параметрів якості води з пороговими значеннями. Останнім кроком методу є пошук аномальних імовірностей залишків і використання отриманих результатів для аналізу якості води.

Загальне комбінаторне правило теорії доказів Демпстера-Шафера задається формулами:

$$m(A) = m_1 \oplus m_2 \oplus \dots \oplus m_n \quad (1.16)$$

$$= \begin{cases} 0, & A = \Phi \\ \frac{1}{1-K} \sum_{\cap_i A_i = A} \prod_{1 \leq i \leq n} m_i(A_i), & A \neq \Phi \end{cases} \quad (1.17)$$

Де

$$K = \sum_{\cap_i A_i = \Phi} \prod_{1 \leq i \leq n} m_i(A_i), \quad (1.18)$$

Коефіцієнт узгодженості доказів K_c визначається як

$$K_c = \sum_{\cap_i A_i = \Phi} \prod_{1 \leq i \leq n} m_i(A_i), \quad (1.19)$$

Дані рівняння представляють основні рівняння злиття в теорії Демпстера-Шафера.

Алгоритм злиття даних та виявлення випадків забруднення води на основі теорії доказів Демпстера-Шафера має в основі шість етапів, як показано на рис. 1.8.

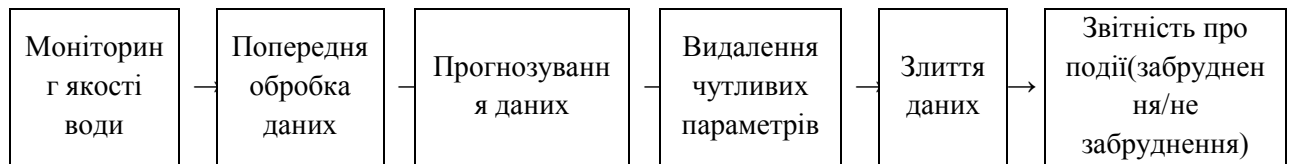


Рисунок 1.8 – Схема методу виявлення випадків забруднення води за допомогою алгоритму Демпстера-Шафера. [джерело: 9]

1. Моніторинг якості води відбувається за допомогою датчиків, які збирають та передають дані в режимі реального часу.
2. Попередня обробка даних передбачає завершення обробки даних. Для обробки відсутніх значень використовується найближче чисельне значення.
3. Прогнозування параметрів зумовлюється використанням моделі авторегресії, критерія Байеса та методу найменших квадратів. Даний підхід застосовується для калібрування параметрів моделі.
4. Вибір декількох параметрів у порядку зменшення їх залишків для подальшого вилучення чутливого параметра.
5. Проводиться обчислення результату злиття аномальних ймовірностей трьох чутливих параметрів.
6. Порівняння отриманих результатів з постійним порогом, для визначення поточного стану якості води.

Алгоритм злиття даних, на основі моделі Демпстера-Шафера має такі вимоги визначення: основи розпізнавання, основної функції розподілу ймовірностей та правил комбінування.

Основи розпізнавання подій для визначення якості води можуть бути визначені як $\Omega = \{N = \text{'нормальний'}, A = \text{'аномалія'}\}$.

Функція розподілу об'єктивних імовірностей використовується для оцінки значення $m(\{A\})$. Використовуючи комбінаційні правила теорії доказів DS, збіжність імовірнісного розподілу $m_1 \oplus m_2 \oplus \dots \oplus m_n(\{A\})$ перетворює впевненість у «аномальному» стані якості води, що виникає з різних джерел доказів, до частки в $(0, 1)$.

Різні джерела доказів можуть призвести до конфліктів доказу щодо стану якості води. Такі конфлікти повинні бути вирішені, щоб досягти визначення щодо стану якості води, яка має високу вірогідність того, що вона буде достовірною, коли відбудеться подія із забрудненням води [9].

Найважливішим кроком у теорії доказів Демпстера-Шафера є виведення основної функції розподілу (Basic Probability Assignment BPA). На цьому кроці визначається достовірність висновків, що випливають із сукупності доказів. Дана функція не має загальноприйнятого визначення. Існує три найпоширеніші способи визначення BPA: нечіткий підхід, статистичні дані та експертні знання. Найбільш важливим вважається спосіб статистичних даних, оскільки він дозволяє мінімізувати вплив суб'єктивності на висновок. [15]

Метод оцінки стану [17], який використовується для прогнозування значення параметрів якості води, має хороші показники відстеження прогнозу. Нижче наведена основна функція призначення імовірності:

$$m(\{N\}) = P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{F*\sigma^2}\right] \quad (1.20)$$

$$m(\{A\}) = 1 - m(\{N\}), \quad (1.21)$$

Де

F - постійний коефіцієнт,

μ і σ - середнє та стандартне відхилення, що розраховуються з довгострокової серії даних залишків.

Основною ідеєю середньозваженого методу є присвоєння різним вагам різні джерела доказів [20].

Міра подібності між двома множинами доказів визначається наступною формулою [6]:

$$Sim_{i,j} = \frac{1}{2} \left[\cos\left(\pi d_{BPA}(m_i, m_j)\right) + 1 \right] \quad i, j = 1, 2, \dots, n, \quad (1.22)$$

Де

$d_{BPA}(m_i, m_j)$ відстань між m_i, m_j [13].

Звідси виходить, чим менша відстань $d_{BPA}(m_i, m_j)$, тим більше двох подібних наборів доказів. Кожен набір доказів отримує рейтинг надійності:

$$Crd_i = \frac{Sup(m_i)}{\sum_{i=1}^n Sup(m_i)}, \quad i = 1, 2, \dots, n \quad (1.23)$$

Де

$$Sup(m_i) = \sum_{j=1, j \neq i}^n Sim_{i,j}, \quad i = 1, 2, \dots, n \quad (1.24)$$

Багатозначне комбінаторне правило має вигляд:

$$= \begin{cases} 0, & A = A \\ \frac{1}{1-K^*} \sum_{\cap_i A_i = A} \prod_{1 \leq i \leq n} m_i^*(A_i), & A \neq A \end{cases} \quad (1.25)$$

Де

$$K^* = \sum_{\cap_i A_i = A} \prod_{1 \leq i \leq n} m_i^*(A_i), \quad (1.26)$$

Дані рівняння представляють зважені середні рівняння злиття. При виникненні проблеми конфлікту доказів застосовується аналіз часу.

Для того, щоб виявляти події забруднення води в режимі реального часу проводиться послідовна серія розрахунків (рис 1.9), на етапі k-часу. До цих розрахунків належить моніторинг якості води, попередня обробка даних, прогнозування, розрахунок залишків, вилучення чутливих параметрів, злиття даних та аналіз аномалій.

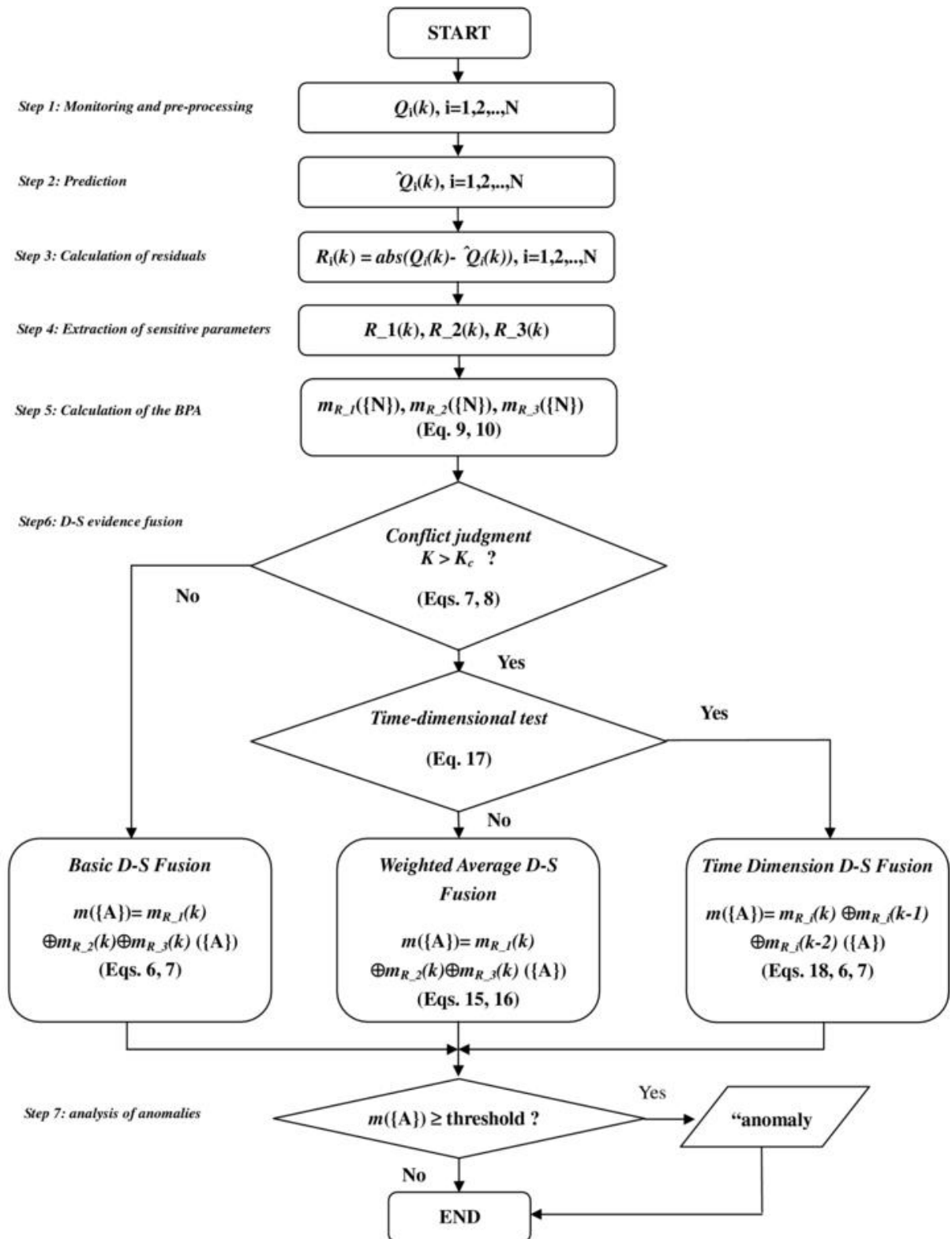


Рисунок 1.9 – Етапи проведення розрахунків [джерело:9]

Якщо залишки з трьох чутливих параметрів не конфліктують один з одним, для отримання остаточної (водяного забруднення) імовірності події використовується основний метод злиття Демпстера-Шафера на основі узагальненого комбінаторного

правила теорії Демпстера-Шафера. В іншому випадку вводиться додатковий тест, який може бути виражений наступним чином:

$$m_{R-i}(k) \oplus m_{R-i}(k-1)(\{A\}) > P \text{ AND} \quad (1.27)$$

$$m_{R-i}(k-1) \oplus m_{R-i}(k-2)(\{A\}) > P \text{ AND} \quad (1.28)$$

$$m_{R-i}(k-2) \oplus m_{R-i}(k-3)(\{A\}) > P \quad (1.29)$$

Де

$m_{R-i}(k_j)$ позначає вірогідність, що присвоєна i -му залишку на етапі часу k_j , $j=0,1,2$;

P – постійний поріг, зазвичай встановлений рівним 0,8 або 0,9.

Якщо існує чутливий параметр, який вище попередньої нерівності, то наступне рівняння, яке можна обчислити за методом Демпстера-Шафера, – це ймовірність події, пов'язаної із забрудненням води:

$$m_{R-i}(k) \oplus m_{R-i}(k-1) \oplus m_{R-i}(k-2)(\{A\}) \quad (1.30)$$

Середньозважений підхід, що базується на рівняннях багатозначного комбінаторного правила, використовується для синтезу Демпстера-Шафера та отримання остаточної ймовірності подій. Очевидно, що залишки трьох вибраних чутливих параметрів використовуються як джерела доказів для синтезу Демпстера-Шафера, і результатом синтезу є ймовірність події, пов'язаної з забрудненням води. Науковцями доведено, що метод Демпстера-Шафера ефективно застосовувати для виявлення подій пов'язаних із забрудненням води.[10]

Таким чином, для вирішення питання об'єднання даних та прогнозування випадків забруднення води, з розглянутих моделей, найбільш відповідним з точки зору точності та доцільності використання є модель теорії доказів Демпстер-Шафера. Також актуальність вибору даної моделі зумовлює можливість оперувати відсутніми даними.

1.4 Постановка наукової задачі та обґрунтування методики досліджень

Результати проведеного аналізу моделей, методів й інструментальних засобів для проведення об'єднання даних та прогнозування випадків аномалій забруднення поверхневих вод, показали, що у відомих публікаціях не вирішеною є задача злиття отриманих даних та прогнозування випадків аномальних забруднень безпосередньо річкових вод. Тому планується провести покращення алгоритму Демпстера-Шафера, що дасть змогу робити прогнози щодо якості води згідно отриманих даних, з різних моніторингових точок, розташованих на певній відстані одна від одної.

Далі планується за допомогою отриманого алгоритму провести розрахунки та виявити випадки аномальних забруднень поверхневих вод на території України.

1.5. Висновки до першого розділу

Таким чином, в результаті дослідження було виявлено, що питання злиття даних, на наш час, є досить актуальними та потребують подальшого вдосконалення та вирішення. Зокрема об'єднання даних он-лайн моніторингу поверхневих вод досить мало розглядалось науковцями. Для вирішення цього питання було вирішено проаналізувати існуючі програмні засоби та математичні методи, і в результаті обрати алгоритм, що дозволяє отримувати найбільш точні кінцеві данні. Тому був обраний метод Демпстера-Шафера, оскільки саме цей метод дозволяє не тільки робити злиття даних, але й на основі отриманих даних робити прогнози щодо випадків аномального забруднення води.

РОЗДІЛ 2 МАТЕМАТИЧНА ПОСТАНОВКА ТА ОПИС АЛГОРИТМУ ПРОГНОЗУВАННЯ ВИПАДКІВ ЗАБРУДНЕННЯ ТА ОБ'ЄДНАННЯ ДАНИХ

Для проведення розрахунків, прогнозування випадків забруднення та об'єднання даних, було вирішено використовувати готовий набір даних зібраних завдяки моніторингу поверхневих вод річок України. Отримані дані включають в себе наступні параметри: назва станції з показником кілометражу, біологічне споживання кисню, кількість кисню, сульфатів, хлориду, нітрату, нітриту та фосфатів за певний проміжок часу. Для здійснення розрахунків був розроблений алгоритм дій на основі методу Демпстера-Шафера. Першим кроком алгоритму є нормалізація даних, далі виконується прогнозування даних завдяки інтегрованій моделі авторегресії ARIMA. На наступному кроці розраховується Байесовський критерій. І на останньому етапі, використовуючи нормалізовані дані, знаходиться прогнозування залишів .

2.1 Етапи алгоритму Демпстера Шафера

2.1.1 Нормалізація даних

Необхідність проведення нормалізації вхідних даних обумовлюється тим, що різні дані можуть відрізнятися між собою за абсолютними величинами. Нормалізація дозволяє привести всі використовувані числові значення змінних до однакової області їх змінення, завдяки чому з'являється можливість зведення даних, за допомогою відповідних методів. Для проведення нормалізації даних, необхідно точно знати границі змінення значень відповідних змінних. Тоді їм будуть відповідати границі інтервалу нормалізації [30].

Нормалізація дозволяє перетворити діапазон зміни значень числових ознак в інший діапазон, більш зручний для застосування до даних тих чи інших аналітичних алгоритмів, а також узгодити діапазони змін різних ознак. Часто використовується приведення до одиниці, коли весь наявний діапазон даних «стискається» в інтервал $[0; 1]$ або $[1; 1]$. Особливо важливо провести правильну нормалізацію даних в алгоритмах Data Mining, які засновані на вимірі відстані між векторами об'єктів в багатовимірному просторі ознак. Серед основних методів нормалізації виділяють: десятичне масштабування, мінімаксу нормалізацію, нормалізацію за допомогою стандартного відхилення та нормалізацію за допомогою поелементних перетворень.

Оцінюється стандартне відхилення за вибіркою. Стандартне відхилення (рис. 2.2) - це міра того, наскільки широко розкидані точки даних відносно свого середнього (рис 2.1).

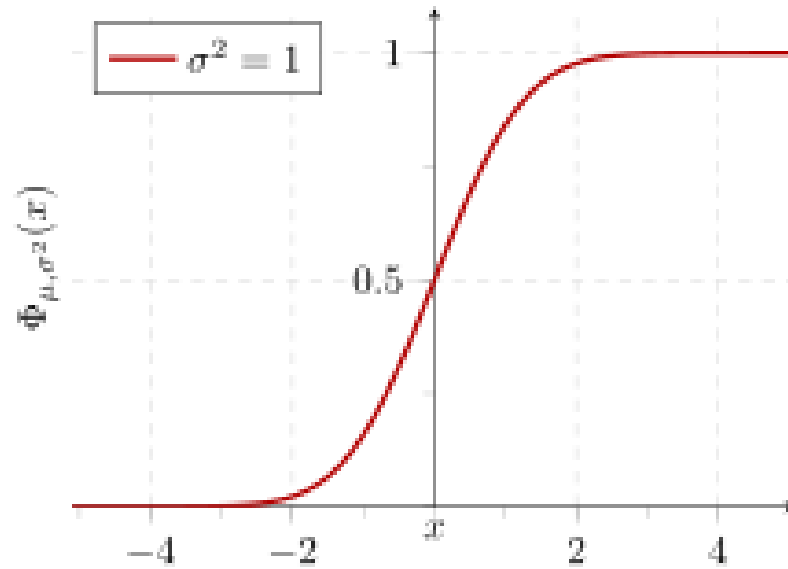


Рисунок 2.1 – Стандартне відхилення [джерело:]

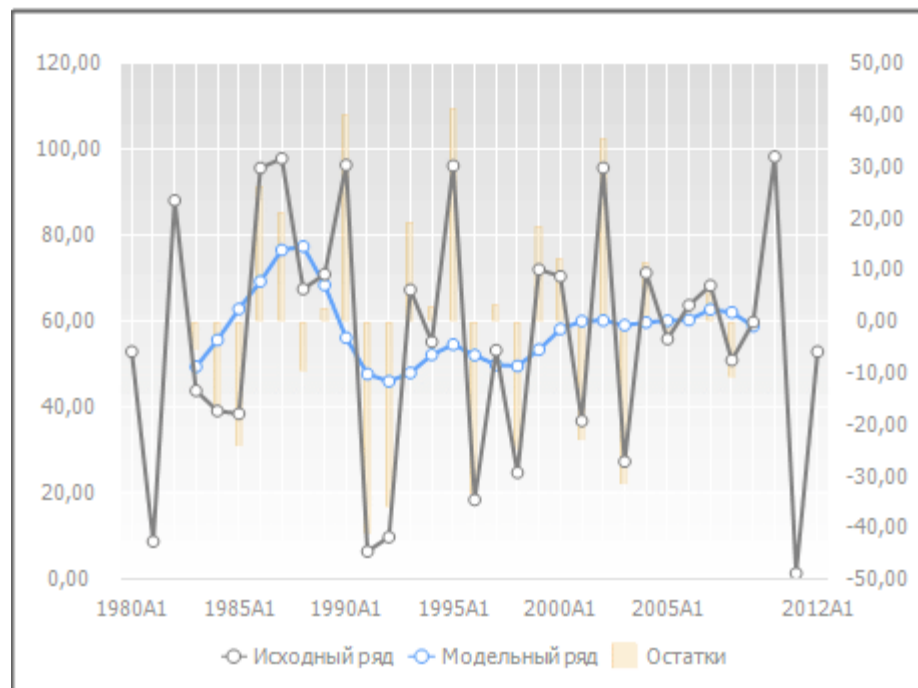


Рисунок 2.2 – Нормалізація за допомогою стандартного відхилення [джерело:26]

Десятичне масштабування Проводиться шляхом переміщення десяткового дробу на кількість цифр в числі, яке визначається виходячи з максимального значення ознаки. При цьому перетворення кожного вихідного значення ознаки $V(i)$ в нормалізоване значення $V'(i)$ проводиться за допомогою виразу:

$$V'(i) = V(i)/10^k \quad (2.1)$$

Мінімаксна нормалізація використовується для кращого розподілення значень в межах інтервалу $[0;1]$, якщо значення сконцентровані в дуже невеликому діапазоні. Мінімаксна нормалізація оптимальна, коли значення ознаки V щільно заповнюють певний інтервал. Для обчислення використовують наступну мінімаксну формулу:

$$V'(i) = \frac{V(i) - \min(V(i))}{\max(V(i)) - \min(V(i))} \quad (2.2)$$

Нормалізація за допомогою стандартного відхилення використовується, якщо в даних наявні відносно рідкі викиди, набагато перевищуючі типовий розкид, саме вони визначають масштаб нормалізації. Це призводить до того, що основна маса значень нормованої змінної $V'(i)$ зосередиться коло нуля. В такому випадку більш надійною з орієнтування не на екстримальні значення, а на типові і використовувати для нормалізації статистичні характеристики даних: середнє значення і дисперсію, тобто стандартне відхилення. Перетворення відбувається за формулою:

$$V'(i) = \frac{V(i) - \bar{V}}{\sigma_V} \quad (2.3)$$

Нормалізація за допомогою поелементних перетворень – поелементне перетворення членів ряду за допомогою різних нелінійних функцій, які здатні відображати вихідний діапазон значень в діапазон, відповідний параметрам функції перетворення.

Початковий діапазон $[v1;v2]$ перетворюється на більш вузький $[v1';v2']$ [31].

Найбільш розповсюджений спосіб нормалізації – лінійна нормалізація. [30]. В табличному процесорі Microsoft Excel серед вбудованих функцій є функція нормалізації яка повертає нормалізоване значення для розподілу, що характеризується середнім і стандартним відхиленням.

2.1.2 Модель авторегресії ARIMA

В даний час відомо велика кількість різних методів прогнозування, що ґрунтуються тільки на аналізі минулих значень часовій послідовності, тобто методів, які використовують принципи, прийняті в технічному аналізі. Основним інструментом цих методів є схема екстраполяції, коли властивості послідовності, виявлені на даному інтервалі часу поширюються за його межі.

При цьому передбачається, що властивості послідовності в майбутньому будуть такими ж, як в минулому і сьогодні. Рідше в процесі прогнозування використовується більш складна схема екстраполяції, що припускає дослідження динаміки зміни властивостей послідовності і облік цієї динаміки на інтервалі прогнозу.[29]

Напевно, найбільш відомими методами прогнозування, заснованими на екстраполяції, є методи, які використовують модель авторегресії (рис. 2.3) і ковзного середнього (ARIMA). Своєю популярністю ці методи, в першу чергу, зобов'язані робіт Боксу і Дженкінса (Boks Dzh., Dzhenkins G.), який запропонував і розвинув узагальнену модель ARIMA. Але крім представлених Боксом і Дженкінсом моделей, звичайно, існують і інші моделі і методи прогнозування. Модель ARIMA - одна з найбільш популярних моделей для побудови короткострокових прогнозів. Для опису даної моделі використовуються три групи параметрів. Авторегресійне інтегроване ковзне середнє (autoregressive integrated moving average, ARIMA) є узагальненням моделі авторегресійного змінного середнього. Ці моделі використовуються при роботі з тимчасовими рядами для більш глибокого розуміння даних або передбачення майбутніх точок ряду. Зазвичай модель згадується, як ARIMA (p, d, q), де p, d і q - цілі невід'ємні числа, що характеризують порядок для частин моделі [32].

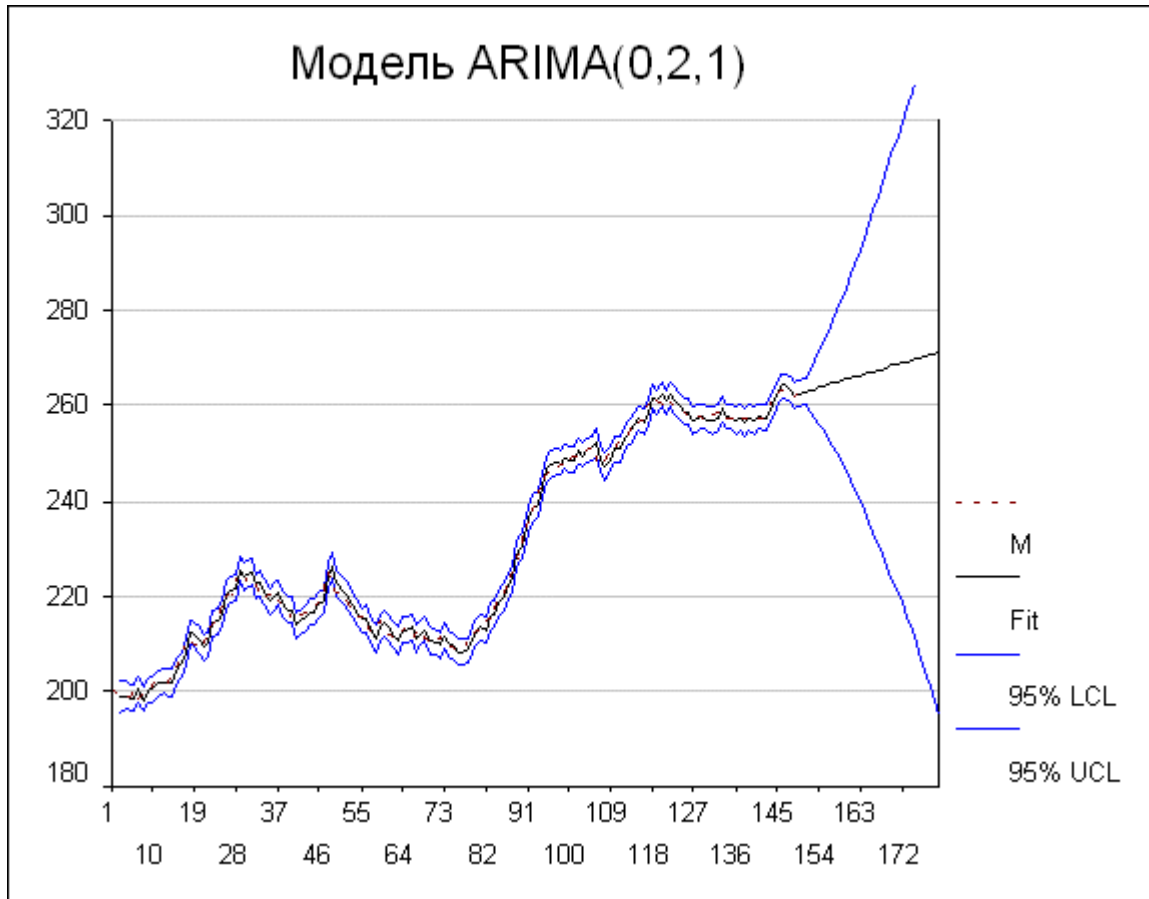


Рисунок 2.3 – Графік побудови прогнозування на основі моделі авторегресії
[джерело: 27]

До параметрів моделі належать: параметри несезонної авторегресії AR (p) і змінного середнього MA (q), параметри сезонної авторегресії SAR (P) і сезонного змінного середнього SMA (Q) та параметри диференціювання вихідного ряду: d , D , s .

Модель ARIMA є узагальненням моделі ARMA і будується в два етапи:

- Диференціювання вихідного ряду (якщо це необхідно).
- Опис продиференційованого ряду за допомогою моделі ARMA.

Необхідність диференціювання вихідного ряду виникає при відсутності стаціонарності. Наприклад, при наявності яскраво вираженої тенденції до зростання. В даному випадку більш правильно застосовувати модель ARMA до ряду, який містить значення приросту вихідного ряду. Для отримання моделі ARMA при заданому часовому ряді X_t , де t – цілий індекс і X_t – дійсні числа:

$$(1 - \sum_{i=1}^p \varphi_i L^i) X_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (2.4)$$

Де

L – оператор затримки, φ_i – параметри авторегресійної частини моделі, θ_i – параметри ковзного середнього, а ϵ_t – значення помилки. Зазвичай передбачають, що помилки ϵ_t є незалежними однаково розподіленими випадковими величинами з нормального розподілу з нульовим середнім.

ARIMA (p, d, q) виходить інтегруванням ARMA (p, q).

$$(1 - \sum_{i=1}^p \varphi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (2.5)$$

Де

d – додатне число, яке задає рівень диференціювання.

Процес авторегресії p -го порядку в його класичному розумінні може бути представлений у формі:

$$y_t = p_1 y_{t-1} + p_2 y_{t-2} + \dots + p_p y_{t-p} + \epsilon_t \quad (2.6)$$

$$(1 - p_1 L - p_2 L^2 - \dots - p_p L^p) y_t = \epsilon_t \quad (2.7)$$

Де

L – оператор затримки: $L^n x_t = x_{t-n}$;

$\{p_1; p_2; \dots; p_p\}$ коефіцієнти авторегресії

Модель ARIMA може враховувати і сезонну авторегресію. Так для квартальної сезонності модель з авторегресії p -го порядку з урахуванням сезонної авторегресії першого порядку буде наступна:

$$(1 - p_1 L - p_2 L^2 - \dots - p_p L^p)(1 - \beta L^4) y_t = \epsilon_t \quad (2.8)$$

Де

β коефіцієнт сезонної авторегресії першого порядку.

Параметри диференціювання вихідного ряду: d, D, s . Звичайний оператор диференціювання виглядає наступним чином:

$$\Delta(y, d) = (1 - L)^d y \quad (2.9)$$

Де

d – порядок диференціювання. При диференціюванні даних можна враховувати сезонність:

$$\Delta(y, d, D, s) = (1 - L)^d (1 - L^s)^D y \quad (2.10)$$

Де

d – порядок несезонного диференціювання

D – порядок сезонного диференціювання

S – період сезонності

Сезонні і несезонні параметри повинні задовольняти нерівності:

$$C - (p + P + d + D * s) > 1 \quad (2.11)$$

Де

C – довжина початкового ряду;

p – максимальний порядок авторегресії;

P – максимальний порядок сезонної авторегресії;

d – Різність;

D – Сезонна різність;

s – період сезонності;

При наявності пропусків всередині початкового ряду кількість використаних для розрахунків значень зменшується на $n * (p + P + d + D * s)$, де n – кількість пропусків в початковому ряду. [33]

В програмному середовищі Microsoft Excel для проведення прогнозування існує спеціальна функція, що називається «Лист прогнозу». У цій функції Excel використовується алгоритм експоненціальне згладжування (ETS). Це один з найпопулярніших способів прогнозування, який вже став галузевим стандартом.

Одне з головних переваг методу ETS - можливість виявляти сезонні закономірності і довірчі інтервали.

У багатьох бізнес-сценаріях існує сезонна закономірність, яку бажано враховувати при прогнозі [28].

У прикладі нижче (рис. 2.4) видно, як щорічна сезонність виявляється і застосовується до прогнозу. Оскільки дані наведені по місяцях і повторюються кожні 12 точок, виявлена сезонність дорівнює 12.

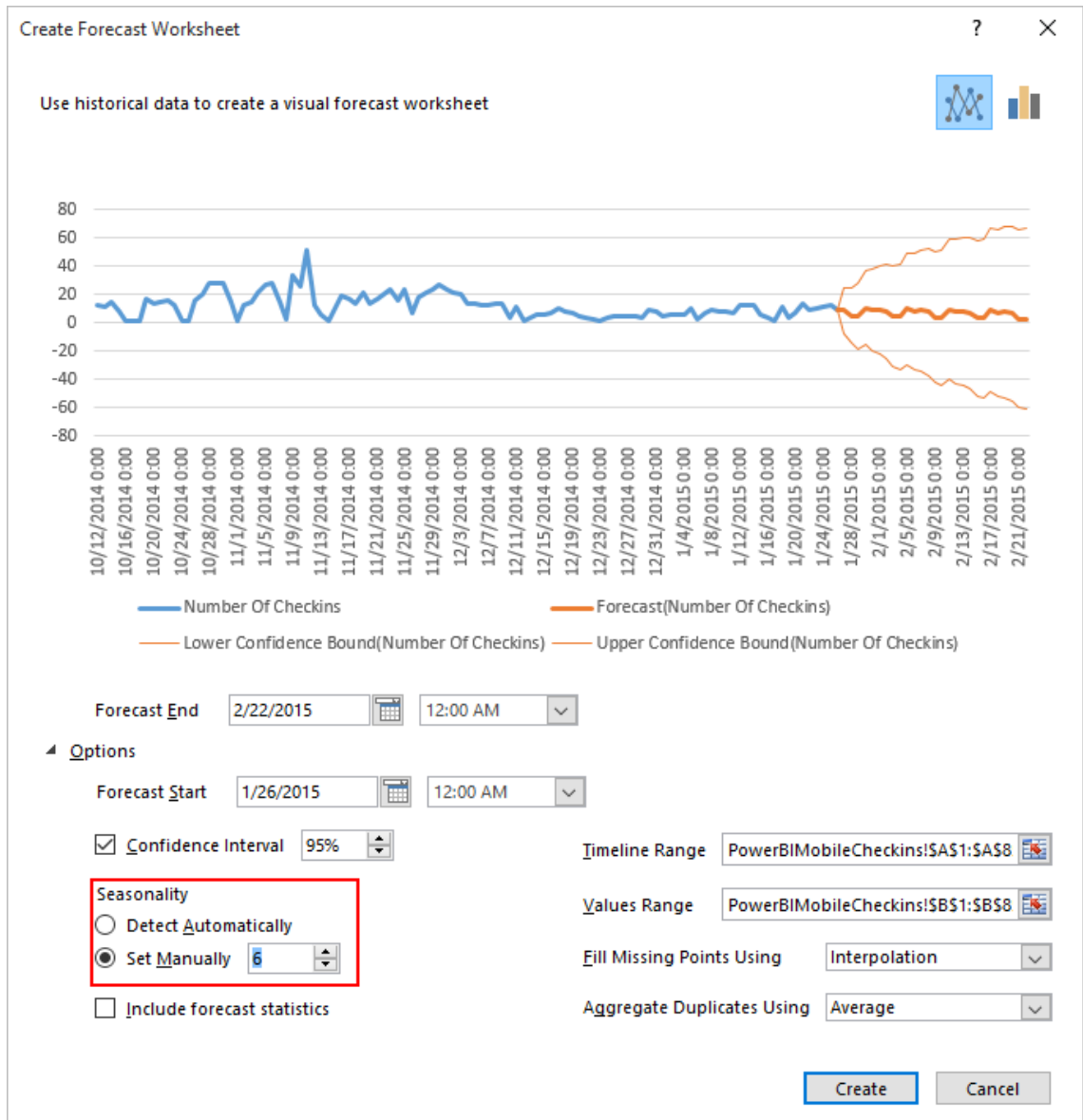


Рисунок 2.4 – Приклад щорічної сезонності. [джерело: 28]

Тривалість сезонності відома не завжди. Цей метод прогнозу виявляє сезонні закономірності в хронологічних даних, визначаючи ту, яка найбільш точно відповідає даним. Для найбільш точного прогнозу хронологічні дані повинні містити якомога більше повторюваних циклів. Бажано, щоб повних сезонних циклів було хоча б 2-3. Автоматично виявляється значення в розділі сезонності можна знайти в діалоговому вікні Створення прогнозу в розділі Параметри. Якщо сезонні дані виявити не вдалося або

сезонність відома заздалегідь, то це значення можна перевизначити, вибравши параметр Задати вручну.

Крім прогнозування майбутніх значень для введеного часового ряду, прогноз ETS також може визначати довірчий інтервал.

Довірчий інтервал - це діапазон, навколишній кожне прогнозоване значення, в який, згідно з прогнозом (з нормальним розподілом), має потрапити 95% майбутніх точок. Довірчий інтервал допомагає з'ясувати точність прогнозу. Чим менше інтервал, тим вище довіра до прогнозу для тієї чи іншої точки. Значення коефіцієнта довіри, за замовчуванням рівне 95%, можна змінити за допомогою стрілок вгору і вниз в наступних цілях:

На основі ширини довірчого інтервалу можна визначити точність прогнозу.

Експериментуючи з розширеними функціями (облік відсутніх точок, сезонності і ін.), Можна відстежувати, як звужується і розширюється попередній довірчий інтервал. Це дозволяє визначити, наскільки отримана модель відповідає хронологічним даними.

Щоб створити лист прогнозу, спочатку підготуйте набір даних часового ряду (з тимчасовим поруч і поруч значень). Потім на вкладці «Дані» натисніть на кнопку Лист прогнозу. Запуститься діалогове вікно створення прогнозу з покроковою інструкцією.[28]

2.1.3 Байєсівський критерій

Байєсівський підхід зазвичай розглядають як спосіб переоцінки наукових уявлень з допомогою знову отриманих даних.

Мінімальний Байєсівський критерій являє собою об'єктивний показник, здатний з успіхом замінити величину p . На відміну від величини p теоретичне обґрунтування та інтерпретація Байєсівського критерію дозволяють використовувати його як в процесі перевірки гіпотез, так і в процесі прийняття рішень. Байєсівський фактор показує, що оцінка даних за допомогою величини p перебільшує докази, які спростовують нульову гіпотезу. І найголовніше, Байєсівський критерій має на увазі включення в аналіз минулого досвіду у вигляді ймовірності, що той чи інший висновок правильний або помилковий. Байєсівський критерій відокремлює логічне умовивід від даних досвіду і в той же час дає досліднику можливість комбінувати стару і нову інформацію.[34]

У порівнянні зі стандартними статистичними методами Байєсівський підхід, що припускає, що результати кожного дослідження представляють собою більш-менш переконливий доказ на користь кожної з висунутих гіпотез, полегшує мета-аналіз і

піднімає його на принципово новий рівень. Завдання дослідника полягає в об'єднанні доказів, отриманих у всіх включених в мета-аналіз дослідженнях, для кожної з гіпотез. При використанні логарифма Байєсівського критерію (логарифма відношення правдоподібності) це дія зводиться до простого додавання [35],[36],[37].

Критерій Байєса – правило, згідно з яким стратегія рішень обирається таким чином, щоб забезпечити мінімум середнього ризику. Застосування критерія Байєса доцільно у випадку, коли система розпізнавання багаторазово здійснює розпізнавання невідомих об'єктів або явищ в умовах незмінного ознакового простору. [38]

Введення критерія Байєса необхідне для визначення порядку моделі прогнозування авторегресії для параметрів якості води. Моделі з мінімальним значенням Байєсівського критерія є оптимальними.

Критерій Байєса - це числова характеристика стратегій в іграх з природою.

Значення критерію Байєса - це найбільше значення математичного очікування виграшу.

Обчислюється за формулою:

$$B(X^*) = \max_{i \in N_m} \left\{ \sum_{j=1}^n p_j a_{ij} \right\}$$

Де

m – число стратегій

n – число станів природи

a_{ij} – виграш при i -ї стратегії при j -м стані природи

p_j – ймовірність j -го стану природи

$B(X^*)$ – критерій Байєса

[39]

В програмному середовищі Excel критерій Байєса можна визначити за допомогою функції Сума Добутку де в якості параметрів задається діапазон ряду параметрів критеріїв та діапазон ряду ймовірностей.

=СУММПРОИЗВ(B3:H3;\$B\$1558:\$H\$1558)							
В	С	Д	Е	Ф	Г	Н	І
Нормалізація БСК	Нормалізація оксиген	Нормалізація сульфат	Нормалізація хлорид	Нормалізація Нітрат	Нормалізація Нітрит	Нормалізація Фосфат	Середнє
-6,91975	0,632698	0,315891	1,861201	0,400566	0,333483	-0,89536	-1,74781
0,231034	0,988054	0,315891	1,448015	0,210121	0,333483	-1,17622	0,38005
0,332607	0,810376	0,315891	1,510569	0,362477	0,481492	-1,35175	0,393725
-0,27683	0,988054	0,794117	0,575551	0,337084	0,259478	-1,28154	0,18303
0,149775	0,401717	1,377833	1,04306	-1,0976	-0,25855	-1,19377	0,112372
-0,03306	0,401717	0,122993	1,978078	-1,08491	-0,33256	-0,57939	0,080847
0,231034	0,25069	0,578111	3,105697	-0,77067	-0,18455	-0,29854	0,362453
0,759217	-0,78873	0,592177	1,638969	-0,64688	0,111469	-0,28098	0,211495
-0,01274	1,698766	0,894584	2,679342	-0,71036	0,777511	-1,36931	0,563107
-0,17526	0,45502	0,605238	2,318833	0,660841	0,407487	0,421163	0,479782
0,820161	0,45502	1,184935	2,546003	0,114898	0,481492	-0,19321	0,750464

Рисунок 2.5 – Формула для розрахунку Байєсівського критерію

2.1.4 Метод найменших квадратів

Метод найменших квадратів (МНК, англ. Ordinary Least Squares, OLS) - математичний метод, застосований для вирішення різних завдань, заснований на мінімізації суми квадратів відхилень деяких функцій від шуканих змінних. Він може використовуватися для «вирішення» перевизначених систем рівнянь (коли кількість рівнянь перевищує кількість невідомих), для пошуку рішення в разі звичайних (НЕ перевизначених) нелінійних систем рівнянь, для апроксимації точкових значень деякої функції. МНК є одним з базових методів регресійного аналізу для оцінки невідомих параметрів регресійних моделей за вибірковими даними.

Метод найменших квадратів (МНК) є одним із методів регресійного аналізу, який використовується для статистичного оцінювання параметрів регресійної моделі за емпіричними даними. Згідно з цим методом параметри моделі повинні відповідати такому рівнянню регресії, що забезпечує найменшого значення суми квадратів відхилень емпіричніх даних від тих, що обчислені за рівнянням регресії. Так, з двох різних наближень тієї ж самої емпіричної функції, що задана у вигляді таблиці, кращим вважаються ті, для яких сума квадратів відхилення має найменше значення[40]. Для ілюстрації суті методу найменших квадратів можна розглянути рис. 2.6.

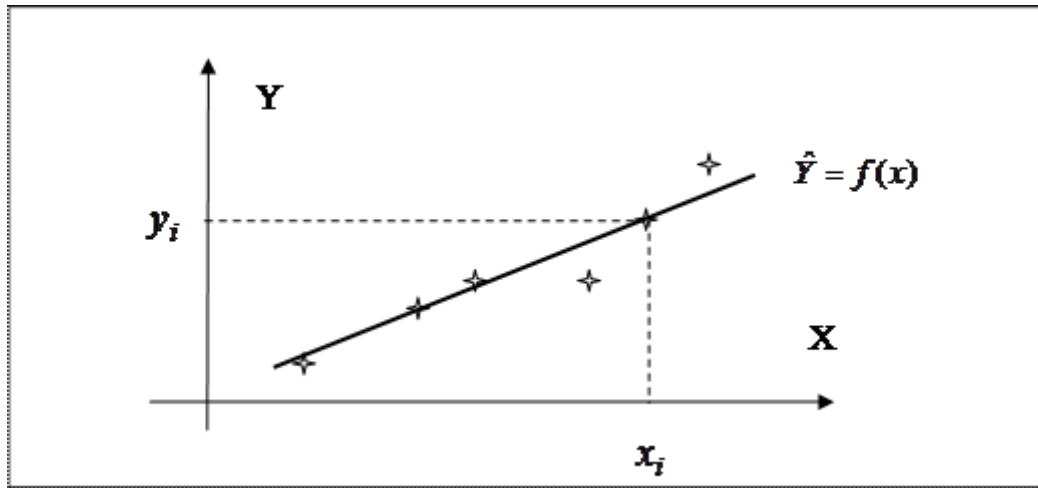


Рисунок 2.6 – Візуалізація методу найменших квадратів [джерело:40]

2.1.5 Прогнозування залишків

2.2 Висновки до другого розділу

Отже, в результаті дослідження, в другому розділі було розглянуто основні кроки алгоритму об'єднання даних та прогнозування випадків аномального забруднення поверхневих вод, на основі методу Демпстера-Шафера. Для проведення прогнозування запропоновано використовувати наступні математичні моделі: нормалізації даних, прогнозування авторегресії ARIMA, розрахування Байєсівського критерію та виявлення залишків. Нормалізація виконується за допомогою середнього на стандартного відхилення і необхідна для приведення даних до загального вигляду, прогнозування за допомогою математичного методу авторегресії ARIMA дозволяє робити короткострокові прогнози на основі існуючих даних. Наступним кроком алгоритму є виявлення Байєсівського критерію та виявлення залишків за допомогою математичних формул.

РОЗДІЛ 3. РЕАЛІЗАЦІЯ МЕТОДУ ОБ'ЄДНАННЯ ТА ПРОГНОЗУВАННЯ ДАНИХ

З кожним роком проблема забруднення водних середовищ стає більш актуальною. Тому виникає потреба в методах, які дозволять робити якомога точні прогнози, щодо аномальних випадків забруднення води. В свою чергу застосування можливості об'єднання отриманих, під час он-лайн моніторингу, даних дозволить В роботі пропонується поліпшення методу прогнозування та об'єднання даних стосовно проточних поверхневих вод, таких як річки.

Для дослідження були обрані готові дані отримані з моніторингу річки: Сіверський Дінець з . Дані, над якими були проведені розрахунки включають в себе відомості про кількість кисню, біологічне споживання кисню, наявність сульфатів, хлоридів, нітратів, нітритів та фосфатів у водному середовищі, за визначений проміжок часу на певному відрізку русла річки.

Для проведення розрахунків, через свою багатofункціональність та можливість наочно представляти дані, був обраний програмний засіб Microsoft Office Excel.

За допомогою даного програмного засобу були проведені наступні етапи розрахунків, роль яких детально описана в попередньому розділі:

- Нормалізація даних
- Побудова моделі авторегресії ARIMA
- Визначення Байєсівського критерію
- Прогнозування залишків

3.1 Реалізація нормалізації даних

На цьому етапі готові дані були приведені до зручного застосування при проведенні наступних розрахунків що надає можливість у подальшому проводити зведення даних. Серед вбудованих функцій Microsoft Excel є функція нормалізації (рис 3.1–3.2). Для розрахунку нормалізації потрібно заздалегідь порахувати середнє та стандартне відхилення для кожного критерія. Середнє та стандартне відхилення було реалізовано за допомогою вбудованих функцій Microsoft Excel, таких як: СРЗНАЧ та СТАНДОТКЛОН, як наведено на рис. 3.3–3.6 [41].

B1558									
=CP3HACH(B2:B1557)									
	A	B	C	D	E	F	G	H	I
717	р. Сіверський Донець, 522 км	3,58	9,47	269	71	7,36	0,08	1,48	04.09.2017
718	р. Сіверський Донець, 522 км	3,56	9,57	273,8	71	7,21	0,076	1,52	10.10.2017
719	р. Сіверський Донець, 522 км	3,6	9,62	278,6	78,1	7,53	0,082	1,49	08.11.2017
720	р. Сіверський Донець, 522 км	3,56	9,69	288,2	78,1	7,49	0,076	1,51	12.12.2017
721	р. Сіверський Донець, 573 км, с.Єремі		9,6	288,2	78,1	4,9	0,06	1,24	15.01.2008
722	р. Сіверський Донець, 573 км, с.Єремі		9,6	317	78,1	10,61	0,07	1,35	12.02.2008
723	р. Сіверський Донець, 573 км	3,1	9,6	297,8	78,1	9,51	0,09	1,1	06.03.2008
724	р. Сіверський Донець, 573 км	3,4	9,3	317	62,1	6,9	0,07	1,08	13.03.2008
725	р. Сіверський Донець, 573 км	3,49	9,5	307,4	85,2	9,7	0,03	1,2	18.04.2008
726	р. Сіверський Донець, 573 км	3,56	9,57	163,3	78,1	4,35	0,063	0,94	12.05.2008
727	р. Сіверський Донець, 573 км, с.Єремі		9,65	259,4	76,6	4,28	0,084	0,79	09.06.2008
728	р. Сіверський Донець, 573 км	3,55	9,24	215,6	83,5	6,98	0,13	0,89	14.07.2008
729	р. Сіверський Донець, 573 км	3,55	9,24	215,6	83,5	6,98	0,13	0,89	14.07.2008
1547	р. Сіверський Донець, 944 км	2,56	7,55	134,42	35,25	3,42	0,074	1,08	14.02.2017
1548	р. Сіверський Донець, 944 км	2,55	7,52	139,77	34,76	3,5	0,072	1,1	15.03.2017
1549	р. Сіверський Донець, 944 км	3,05	7,28	130,45	37,68	4,29	0,051	0,86	10.04.2017
1550	р. Сіверський Донець, 944 км	3,12	7,32	132,18	38,15	3,85	0,07	1,05	15.05.2017
1551	р. Сіверський Донець, 944 км	3,06	7,49	128,64	37,88	3,38	0,068	1,03	08.06.2017
1552	р. Сіверський Донець, 944 км	2,5	8,32	111,81	36,5	3,79	0,07	1,14	04.07.2017
1553	р. Сіверський Донець, 944 км	2,58	8,08	121,13	36,62	3,52	0,072	1,18	07.08.2017
1554	р. Сіверський Донець, 944 км	2,61	7,94	130,45	35,62	3,46	0,078	1,16	04.09.2017
1555	р. Сіверський Донець, 944 км	2,58	7,35	130,45	38,33	3,68	0,073	1,12	10.10.2017
1556	р. Сіверський Донець, 944 км	2,51	7,6	139,77	37,98	3,62	0,07	1,08	14.11.2017
1557	р. Сіверський Донець, 944 км	2,54	7,54	139,77	36,26	3,57	0,072	1,06	05.12.2017
1558	Середнє	3,40627	8,687815	295,158	107,037	8,43801	0,13494	1,51007	
1559	Стандартне відхилення	0,49225	1,125632	99,5346	60,7475	3,15051	0,13513	0,56968	
1560									

Рисунок 3.1 – Формула та розрахунок середнього значення для кожного критерію

B1559									
=СТАНДОТКЛОН.В(B2:B1557)									
	A	B	C	D	E	F	G	H	I
717	р. Сіверський Донець, 522 км	3,58	9,47	269	71	7,36	0,08	1,48	04.09.2017
718	р. Сіверський Донець, 522 км	3,56	9,57	273,8	71	7,21	0,076	1,52	10.10.2017
719	р. Сіверський Донець, 522 км	3,6	9,62	278,6	78,1	7,53	0,082	1,49	08.11.2017
720	р. Сіверський Донець, 522 км	3,56	9,69	288,2	78,1	7,49	0,076	1,51	12.12.2017
721	р. Сіверський Донець, 573 км, с.Єремі		9,6	288,2	78,1	4,9	0,06	1,24	15.01.2008
722	р. Сіверський Донець, 573 км, с.Єремі		9,6	317	78,1	10,61	0,07	1,35	12.02.2008
723	р. Сіверський Донець, 573 км	3,1	9,6	297,8	78,1	9,51	0,09	1,1	06.03.2008
724	р. Сіверський Донець, 573 км	3,4	9,3	317	62,1	6,9	0,07	1,08	13.03.2008
725	р. Сіверський Донець, 573 км	3,49	9,5	307,4	85,2	9,7	0,03	1,2	18.04.2008
726	р. Сіверський Донець, 573 км	3,56	9,57	163,3	78,1	4,35	0,063	0,94	12.05.2008
727	р. Сіверський Донець, 573 км, с.Єремі		9,65	259,4	76,6	4,28	0,084	0,79	09.06.2008
728	р. Сіверський Донець, 573 км	3,55	9,24	215,6	83,5	6,98	0,13	0,89	14.07.2008
729	р. Сіверський Донець, 573 км	3,55	9,24	215,6	83,5	6,98	0,13	0,89	14.07.2008
1547	р. Сіверський Донець, 944 км	2,56	7,55	134,42	35,25	3,42	0,074	1,08	14.02.2017
1548	р. Сіверський Донець, 944 км	2,55	7,52	139,77	34,76	3,5	0,072	1,1	15.03.2017
1549	р. Сіверський Донець, 944 км	3,05	7,28	130,45	37,68	4,29	0,051	0,86	10.04.2017
1550	р. Сіверський Донець, 944 км	3,12	7,32	132,18	38,15	3,85	0,07	1,05	15.05.2017
1551	р. Сіверський Донець, 944 км	3,06	7,49	128,64	37,88	3,38	0,068	1,03	08.06.2017
1552	р. Сіверський Донець, 944 км	2,5	8,32	111,81	36,5	3,79	0,07	1,14	04.07.2017
1553	р. Сіверський Донець, 944 км	2,58	8,08	121,13	36,62	3,52	0,072	1,18	07.08.2017
1554	р. Сіверський Донець, 944 км	2,61	7,94	130,45	35,62	3,46	0,078	1,16	04.09.2017
1555	р. Сіверський Донець, 944 км	2,58	7,35	130,45	38,33	3,68	0,073	1,12	10.10.2017
1556	р. Сіверський Донець, 944 км	2,51	7,6	139,77	37,98	3,62	0,07	1,08	14.11.2017
1557	р. Сіверський Донець, 944 км	2,54	7,54	139,77	36,26	3,57	0,072	1,06	05.12.2017
1558	Середнє	3,40627	8,687815	295,158	107,037	8,43801	0,13494	1,51007	
1559	Стандартне відхилення	0,49225	1,125632	99,5346	60,7475	3,15051	0,13513	0,56968	
1560									

Рисунок 3.2 – Формула та розрахунок стандартного відхилення для кожного критерія

СЕРВЕРСКИЙ ДОНЕЦ(расчеты) - Excel

Анастасия Топчий

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Разработчик Справка Что вы хотите сделать? Общий доступ

Вставить Буфер обмена Шрифт Выравнивание Число Стили

Условное форматирование * Форматировать как таблицу * ячеек * Стили Удалить * Формат * Ячейки

Вставить * Удалить * Формат * Сортировка и фильтр * Найти и выделить * Редактирование

station_name	bsk5	oxygen_in_water	sulfat	chloride	nitrat	nitryt	phoshat	date	Нормализация БСК	Нормализация кислорода	Нормализация сульфат	Нормализация хлорид	Нормализация нитрат
р. Сиверський Донець, 406 км, с.Світлич		9,4	326,6	220,1	9,7	0,18	1	22.01.2008					
р. Сиверський Донець, 406 км,	3,52	9,8	326,6	195	9,1	0,18	0,84	18.02.2008					
р. Сиверський Донець, 406 км,	3,57	9,6	326,6	198,8	9,58	0,2	0,74	17.03.2008					
р. Сиверський Донець, 406 км,	3,27	9,8	374,2	142	9,5	0,17	0,78	21.04.2008					
р. Сиверський Донець, 406 км,	3,48	9,14	432,3	170,4	4,98	0,1	0,83	21.05.2008					
р. Сиверський Донець, 406 км,	3,39	9,14	307,4	227,2	5,02	0,09	1,18	17.06.2008					
р. Сиверський Донець, 406 км,	3,52	8,97	352,7	295,7	6,01	0,11	1,34	21.07.2008					
р. Сиверський Донець, 406 км,	3,78	7,8	354,1	206,6	6,4	0,15	1,35	19.08.2008					
р. Сиверський Донець, 944 км,	2,61	7,94	130,45	35,62	3,46	0,078	1,16	04.09.2017					
р. Сиверський Донець, 944 км,	2,58	7,35	130,45	38,33	3,68	0,073	1,12	10.10.2017					
р. Сиверський Донець, 944 км,	2,51	7,6	139,77	37,98	3,62	0,07	1,08	14.11.2017					
р. Сиверський Донець, 944 км,	2,54	7,54	139,77	36,26	3,57	0,072	1,06	05.12.2017					
Среднее	=Ср												
Стандартное відхилення	=СТ												

Возвращает среднее арифметическое своих аргументов, которые могут быть числами, именами, массивами или ссылками на ячейки с числами

Лист1 Зависимость БПК и кислорода Корреляция

Ввод Scroll Lock

Рисунок 3.3– Розрахунок середнього значення для кожного з критеріїв

СЕРВЕРСКИЙ ДОНЕЦ(расчеты) - Excel

Анастасия Топчий

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Разработчик Справка Что вы хотите сделать? Общий доступ

Вставить Буфер обмена Шрифт Выравнивание Число Стили

Условное форматирование * Форматировать как таблицу * ячеек * Стили Удалить * Формат * Ячейки

Вставить * Удалить * Формат * Сортировка и фильтр * Найти и выделить * Редактирование

station_name	bsk5	oxygen_in_water	sulfat	chloride	nitrat	nitryt	phoshat	date	Нормализация БСК	Нормализация кислорода	Нормализация сульфат	Нормализация хлорид	Нормализация нитрат
р. Сиверський Донець, 406 км, с.Світлич		9,4	326,6	220,1	9,7	0,18	1	22.01.2008					
р. Сиверський Донець, 406 км,	3,52	9,8	326,6	195	9,1	0,18	0,84	18.02.2008					
р. Сиверський Донець, 406 км,	3,57	9,6	326,6	198,8	9,58	0,2	0,74	17.03.2008					
р. Сиверський Донець, 406 км,	3,27	9,8	374,2	142	9,5	0,17	0,78	21.04.2008					
р. Сиверський Донець, 406 км,	3,48	9,14	432,3	170,4	4,98	0,1	0,83	21.05.2008					
р. Сиверський Донець, 406 км,	3,39	9,14	307,4	227,2	5,02	0,09	1,18	17.06.2008					
р. Сиверський Донець, 406 км,	3,52	8,97	352,7	295,7	6,01	0,11	1,34	21.07.2008					
р. Сиверський Донець, 406 км,	3,78	7,8	354,1	206,6	6,4	0,15	1,35	19.08.2008					
р. Сиверський Донець, 944 км,	2,61	7,94	130,45	35,62	3,46	0,078	1,16	04.09.2017					
р. Сиверський Донець, 944 км,	2,58	7,35	130,45	38,33	3,68	0,073	1,12	10.10.2017					
р. Сиверський Донець, 944 км,	2,51	7,6	139,77	37,98	3,62	0,07	1,08	14.11.2017					
р. Сиверський Донець, 944 км,	2,54	7,54	139,77	36,26	3,57	0,072	1,06	05.12.2017					
Среднее	3,406273	8,6878152	295,1579	107,0367	8,438012	0,134938	1,510071						
Стандартное відхилення	=СТ												

Оценивает стандартное отклонение по выборке. Логические и текстовые значения игнорируются

Лист1 Зависимость БПК и кислорода Корреляция

Ввод Scroll Lock

Рисунок 3.4 – Розрахунок стандартного відхилення для кожного з критеріїв

station_name	bsk5	oxygen_in_water	sulfat	chloride	nitrat	nitryt	phoshat	date	НОРМАЛІЗАЦІЯ БСК	НОРМАЛІЗАЦІЯ оксиген	НОРМАЛІЗАЦІЯ сульфат	НОРМАЛІЗАЦІЯ хлорид	НОРМАЛІЗАЦІЯ Нітрат
р. Сіверський Донець, 406 км, с.Світличн	9,4	326,6	220,1	9,7	0,18	1	22.01.2008						
р. Сіверський Донець, 406 км,	3,52	9,8	326,6	195	9,1	0,18	0,84	18.02.2008					
р. Сіверський Донець, 406 км,	3,57	9,6	326,6	198,8	9,58	0,2	0,74	17.03.2008					
р. Сіверський Донець, 406 км,	3,27	9,8	374,2	142	9,5	0,17	0,78	21.04.2008					
р. Сіверський Донець, 406 км,	3,48	9,14	432,3	170,4	4,98	0,1	0,83	21.05.2008					
р. Сіверський Донець, 406 км,	3,39	9,14	307,4	227,2	5,02	0,09	1,18	17.06.2008					
р. Сіверський Донець, 406 км,	3,52	8,97	352,7	295,7	6,01	0,11	1,34	21.07.2008					
р. Сіверський Донець, 406 км,	3,78	7,8	354,1	206,6	6,4	0,15	1,35	19.08.2008					
1554 р. Сіверський Донець, 944 км,	2,61	7,94	130,45	35,62	3,46	0,078	1,16	04.09.2017					
1555 р. Сіверський Донець, 944 км,	2,58	7,35	130,45	38,33	3,68	0,073	1,12	10.10.2017					
1556 р. Сіверський Донець, 944 км,	2,51	7,6	139,77	37,98	3,62	0,07	1,08	14.11.2017					
1557 р. Сіверський Донець, 944 км,	2,54	7,54	139,77	36,26	3,57	0,072	1,06	05.12.2017					
1558 Середнє	3,406273	8,6878152	295,1579	107,0367	8,438012	0,134938	1,510071						
1559 Стандартне відхилення	0,492254	1,1256315	99,53462	60,74748	3,150514	0,135127	0,569683						

Рисунок3.5 – Розрахунок нормалізованого значення для кожного з критеріїв

station_name	bsk5	oxygen_in_water	sulfat	chloride	nitrat	nitryt	phoshat	date	НОРМАЛІЗАЦІЯ БСК	НОРМАЛІЗАЦІЯ оксиген	НОРМАЛІЗАЦІЯ сульфат	НОРМАЛІЗАЦІЯ хлорид	НОРМАЛІЗАЦІЯ Нітрат	НОРМАЛІЗАЦІЯ Нітрит	НОРМАЛІЗАЦІЯ фосфат
р. Сіверський Донець, 406 км, с.Світличн	9,4	326,6	220,1	9,7	0,18	1	22.01.2008	-6,919751223	0,632698016	0,315890993	1,861200758	0,400565715	0,333482734	-0,895358651	
р. Сіверський Донець, 406 км,	3,52	9,8	326,6	195	9,1	0,18	0,84	18.02.2008	0,23103364	0,988054097	0,315890993	1,448014927	0,210120585	0,333482734	
р. Сіверський Донець, 406 км,	3,57	9,6	326,6	198,8	9,58	0,2	0,74	17.03.2008	0,332607288	0,810376057	0,315890993	1,510568957	0,362476689	0,48149211	
р. Сіверський Донець, 406 км,	3,27	9,8	374,2	142	9,5	0,17	0,78	21.04.2008	-0,276834603	0,988054097	0,794116577	0,575550821	0,337084005	0,259478046	
р. Сіверський Донець, 406 км,	3,48	9,14	432,3	170,4	4,98	0,1	0,83	21.05.2008	0,149774721	0,401716563	1,3778331	1,043059889	-1,097602644	-0,25854772	
р. Сіверський Донець, 406 км,	3,39	9,14	307,4	227,2	5,02	0,09	1,18	17.06.2008	-0,033057847	0,401716563	0,122993278	1,978078025	-1,084906302	-0,33255946	
р. Сіверський Донець, 406 км,	3,52	8,97	352,7	295,7	6,01	0,11	1,34	21.07.2008	0,23103364	0,250690228	0,578111324	3,105696728	-0,770671837	-0,184550084	
р. Сіверський Донець, 406 км,	3,78	7,8	354,1	206,6	6,4	0,15	1,35	19.08.2008	0,759216613	-0,78872631	0,592176782	1,638969335	-0,646882502	0,111468669	
р. Сіверський Донець, 406 км,	3,4	10,6	384,2	269,8	6,2	0,24	0,73	16.09.2008	-0,012743117	1,69876626	0,894584137	2,679341627	-0,710364212	0,777510863	
1547 р. Сіверський Донець, 944 км	2,56	7,55	134,42	35,25	3,42	0,074	1,08	14.02.2017	-1,719180414	-0,1010823861	-1,614894572	-1,18172358	-1,592759983	-0,450966961	
1548 р. Сіверський Донець, 944 км	2,55	7,52	139,77	34,76	3,5	0,072	1,1	15.03.2017	-1,739490144	-1,037475567	-1,561144427	-1,189789758	-1,567367299	-0,465767899	
1549 р. Сіверський Донець, 944 км	3,05	7,28	130,45	37,68	4,29	0,051	0,86	10.04.2017	-0,723758657	-1,250689216	-1,654780193	-1,141721924	-1,316614544	-0,621177744	
1550 р. Сіверський Донець, 944 км	3,12	7,32	132,18	38,15	3,85	0,07	1,05	15.05.2017	-0,581555549	-1,215153608	-1,637999305	-1,133984978	-1,456274306	-0,480568836	
1551 р. Сіверський Донець, 944 км	3,06	7,49	128,64	37,88	3,38	0,068	1,03	08.06.2017	-0,703443928	-1,064127273	-1,672964821	-1,138429606	-1,605456325	-0,495369774	
1552 р. Сіверський Донець, 944 км	2,5	8,32	111,81	36,5	3,79	0,07	1,14	04.07.2017	-1,841068792	-0,326763404	-1,842051724	-1,161146596	-1,475318819	-0,480568836	
1553 р. Сіверський Донець, 944 км	2,58	8,08	121,13	36,62	3,52	0,072	1,18	07.08.2017	-1,678530954	-0,539977053	-1,748415959	-1,159171206	-1,561019128	-0,465767899	
1554 р. Сіверський Донець, 944 км	2,61	7,94	130,45	35,62	3,46	0,078	1,16	04.09.2017	-1,617606765	-0,664351681	-1,654780193	-1,175632793	-1,580063641	-0,421365086	
1555 р. Сіверський Донець, 944 км	2,58	7,35	130,45	38,33	3,68	0,073	1,12	10.10.2017	-1,678550954	-1,188501901	-1,654780193	-1,131021892	-1,51023376	-0,45836743	
1556 р. Сіверський Донець, 944 км	2,51	7,6	139,77	37,98	3,62	0,07	1,08	14.11.2017	-1,820754062	-0,966404351	-1,561144427	-1,136783448	-1,529278273	-0,480568836	
1557 р. Сіверський Донець, 944 км	2,54	7,54	139,77	36,26	3,57	0,072	1,06	05.12.2017	-1,758089873	-1,019707763	-1,561144427	-1,165097377	-1,5451487	-0,465767899	
1558 Середнє	3,406273	8,687815	295,158	107,037	8,43801	0,13494	1,51007								
1559 Стандартне відхилення	0,49225	1,125632	99,5346	60,7475	3,15051	0,13513	0,56968								

Рисунок 3.6 – Результат розрахунків нормалізації даних

3.2 Прогнозування за допомогою моделі ARIMA

Модель авторегресії ARIMA дозволяє будувати короткострокові прогнози на основі отриманих даних.

За початковий діапазон обрано проміжок часу з 12.01.2016 по 05.12.2017 та відповідні діапазони даних біологічного споживання кисню, рівень кисню, рівень нітратів, рівень нітритів, рівень сульфатів, та рівень фосфатів за цей період. Отримання даних прогнозування та побудова графіків відбувається за допомогою інструмента Лист прогнозу групи Прогноз, на вкладці Данні прогнозуємо рівень біологічного споживання кисню, рівень кисню, рівень нітратів, рівень нітритів, рівень сульфатів та рівень фосфатів до 1 січня 2019 року.

Необхідні налаштування задаються за допомогою форми «Создание листа прогноза» (рис. 3.7), де вказується початковий період, період прогнозування, діапазон даних для прогнозування, та обираються налаштування сезонності. На отриманих табличних даних та графіках можна спостерігати як зміниться рівень речовин на протязі обраного періоду прогнозування для кожного з обраних критеріїв [28].

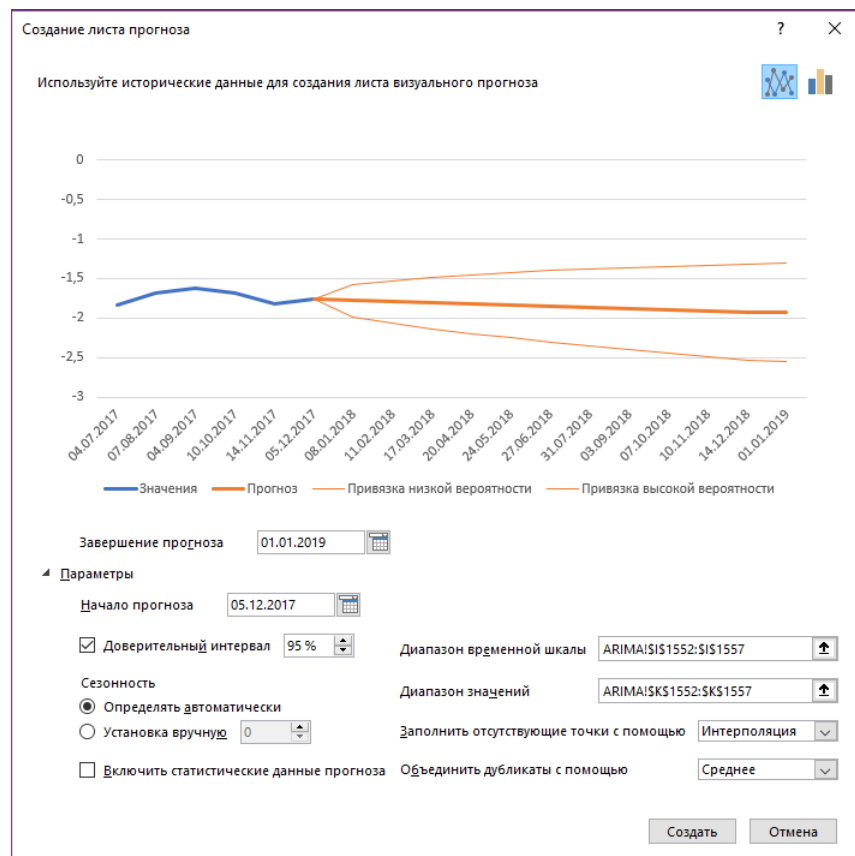


Рисунок 3.7– Додавання даних та налаштування параметрів прогнозування

	A	B	C	D	E
1	Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
2	04.07.2017	-1,84106879			
3	07.08.2017	-1,67855095			
4	04.09.2017	-1,61760677			
5	10.10.2017	-1,67855095			
6	14.11.2017	-1,82075406			
7	05.12.2017	-1,75980987	-1,7598099	-1,76	-1,76
8	08.01.2018		-1,7804218	-1,98	-1,58
9	11.02.2018		-1,7948104	-2,06	-1,53
10	17.03.2018		-1,8091991	-2,13	-1,48
11	20.04.2018		-1,8235877	-2,19	-1,45
12	24.05.2018		-1,8379763	-2,25	-1,42
13	27.06.2018		-1,852365	-2,30	-1,40
14	31.07.2018		-1,8667536	-2,35	-1,38
15	03.09.2018		-1,8811422	-2,40	-1,36
16	07.10.2018		-1,8955309	-2,45	-1,35
17	10.11.2018		-1,9099195	-2,49	-1,33
18	14.12.2018		-1,9243081	-2,53	-1,32
19	01.01.2019		-1,9319256	-2,55	-1,31

Рисунок 3.8 – Дані результатів прогнозування рівня БСК у водному середовищі

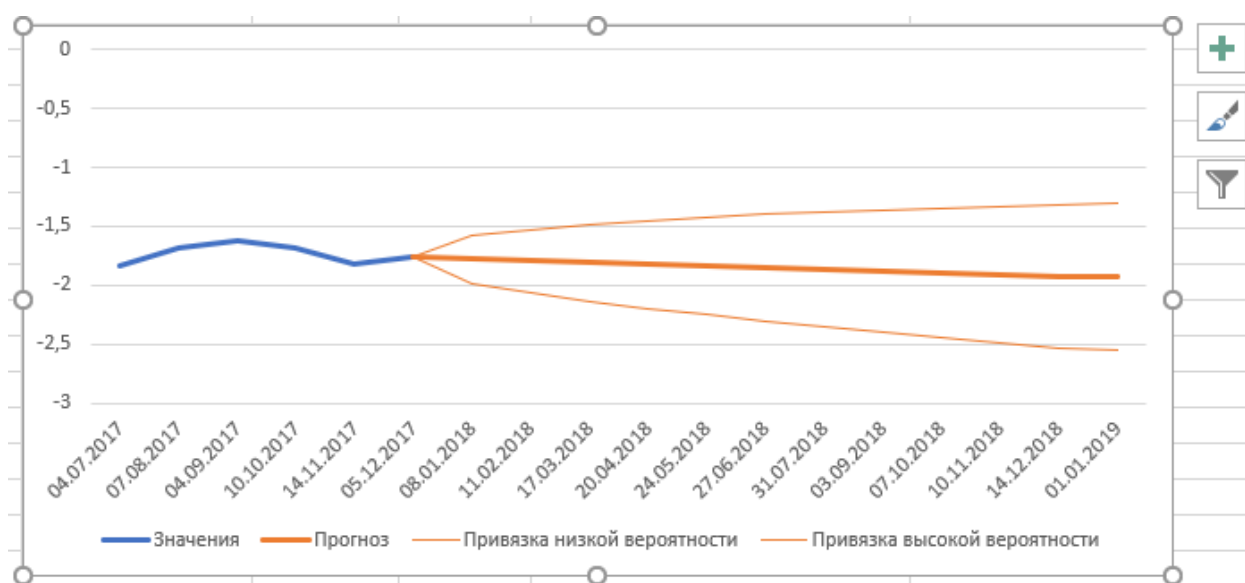


Рисунок 3.9– Графік прогнозування рівня БСК (Біологічного споживання кисню)

Прогнозований рівень БСК залежить від багатьох чинників, серед них рівень кисню у водному середовищі, перепад температур у зв'язку з сезонними змінами, швидкість течії, кількість організмів та мікроорганізмів, що перебувають у водному середовищі. Не менш важливим чинником є наявність органічних відходів, що можуть викликати аномальні перепади рівня БСК у водному середовищі, що в свою чергу буде вказувати на забруднення.

Далі наведені прогнозовані дані критеріїв моніторингу Сіверського Дінця, в часовому діапазоні з 05.12.2017 по 01.01.2019, а також для кожного параметру наведено графік, що візуально відображає прогнозування даних рис. 3.7 – 3.18.

	A	B	C	D	E
1	Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
2	04.07.2017	-0,3267634			
3	07.08.2017	-0,53997705			
4	04.09.2017	-0,66435168			
5	10.10.2017	-1,1885019			
6	14.11.2017	-0,96640435			
7	05.12.2017	-1,01970776	-1,0197078	-1,02	-1,02
8	08.01.2018		-1,236837	-1,59	-0,89
9	11.02.2018		-1,379109	-1,73	-1,03
10	17.03.2018		-1,5213811	-1,88	-1,17
11	20.04.2018		-1,6636532	-2,02	-1,30
12	24.05.2018		-1,8059252	-2,17	-1,44
13	27.06.2018		-1,9481973	-2,31	-1,58
14	31.07.2018		-2,0904694	-2,46	-1,72
15	03.09.2018		-2,2327414	-2,60	-1,86
16	07.10.2018		-2,3750135	-2,75	-2,00
17	10.11.2018		-2,5172856	-2,89	-2,14
18	14.12.2018		-2,6595576	-3,04	-2,28
19	01.01.2019		-2,7348782	-3,12	-2,35

Рисунок 3.10– Дані прогнозування рівня кисню у водному середовищі



Рисунок 3.11 – Графік прогнозування рівня кисню

Прогнозування рівня кисню. При побудові даного прогнозу не бралось до уваги залежність рівня кисню та біологічного споживання кисню мікроорганізмами.

	A	B	C	D	E
1	Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
2	04.07.2017	-1,84205172			
3	07.08.2017	-1,74841596			
4	04.09.2017	-1,65478019			
5	10.10.2017	-1,65478019			
6	14.11.2017	-1,56114443			
7	05.12.2017	-1,56114443	-1,5611444	-1,56	-1,56
8	08.01.2018		-1,4891475	-1,56	-1,42
9	11.02.2018		-1,4362392	-1,52	-1,36
10	17.03.2018		-1,383331	-1,47	-1,29
11	20.04.2018		-1,3304227	-1,43	-1,23
12	24.05.2018		-1,2775144	-1,38	-1,17
13	27.06.2018		-1,2246061	-1,33	-1,12
14	31.07.2018		-1,1716978	-1,29	-1,06
15	03.09.2018		-1,1187896	-1,24	-1,00
16	07.10.2018		-1,0658813	-1,19	-0,94
17	10.11.2018		-1,012973	-1,14	-0,88
18	14.12.2018		-0,9600647	-1,10	-0,82
19	01.01.2019		-0,9320545	-1,07	-0,79

Рисунок 3.12– Дані прогнозування рівня сульфатів у водному середовищі

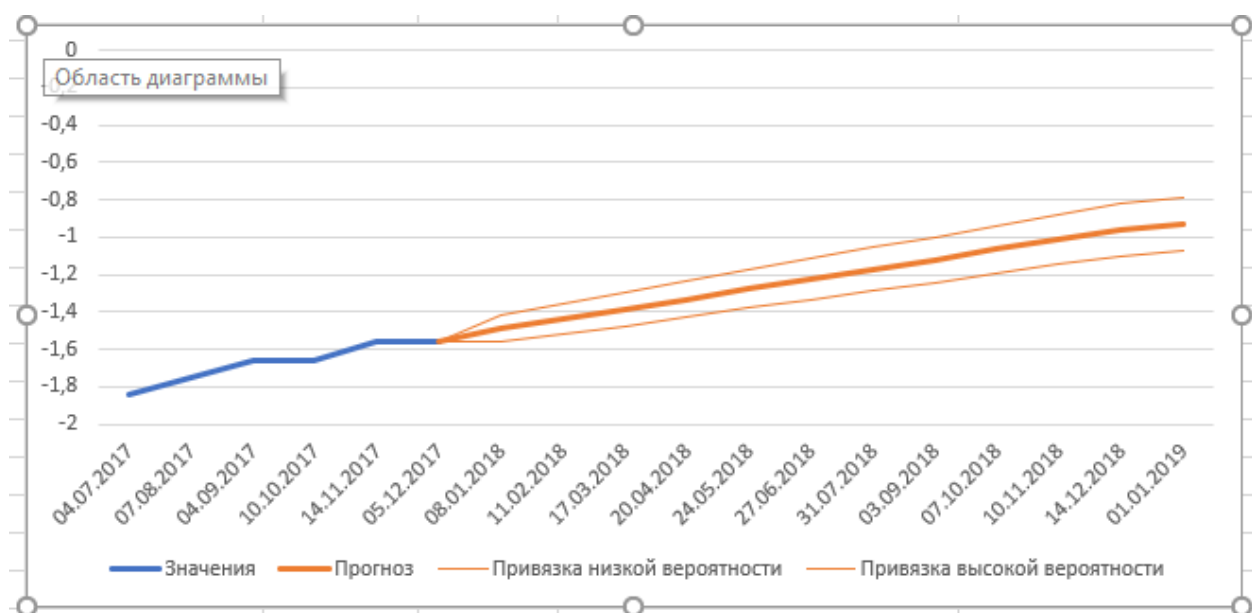


Рисунок 3.13– Графік прогнозування рівня сульфатів у водному середовищі

	A	B	C	D	E
1	Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
2	04.07.2017	-1,1611466			
3	07.08.2017	-1,15917121			
4	04.09.2017	-1,17563279			
5	10.10.2017	-1,13102189			
6	14.11.2017	-1,13678345			
7	05.12.2017	-1,16509738	-1,1650974	-1,17	-1,17
8	08.01.2018		-1,1452296	-1,17	-1,12
9	11.02.2018		-1,142577	-1,17	-1,11
10	17.03.2018		-1,1399244	-1,17	-1,11
11	20.04.2018		-1,1372718	-1,17	-1,11
12	24.05.2018		-1,1346192	-1,16	-1,10
13	27.06.2018		-1,1319666	-1,16	-1,10
14	31.07.2018		-1,1293139	-1,16	-1,10
15	03.09.2018		-1,1266613	-1,16	-1,10
16	07.10.2018		-1,1240087	-1,15	-1,09
17	10.11.2018		-1,1213561	-1,15	-1,09
18	14.12.2018		-1,1187035	-1,15	-1,09
19	01.01.2019		-1,1172992	-1,15	-1,09

Рисунок 3.14– Дані прогнозування рівня хлоридів у водному середовищі

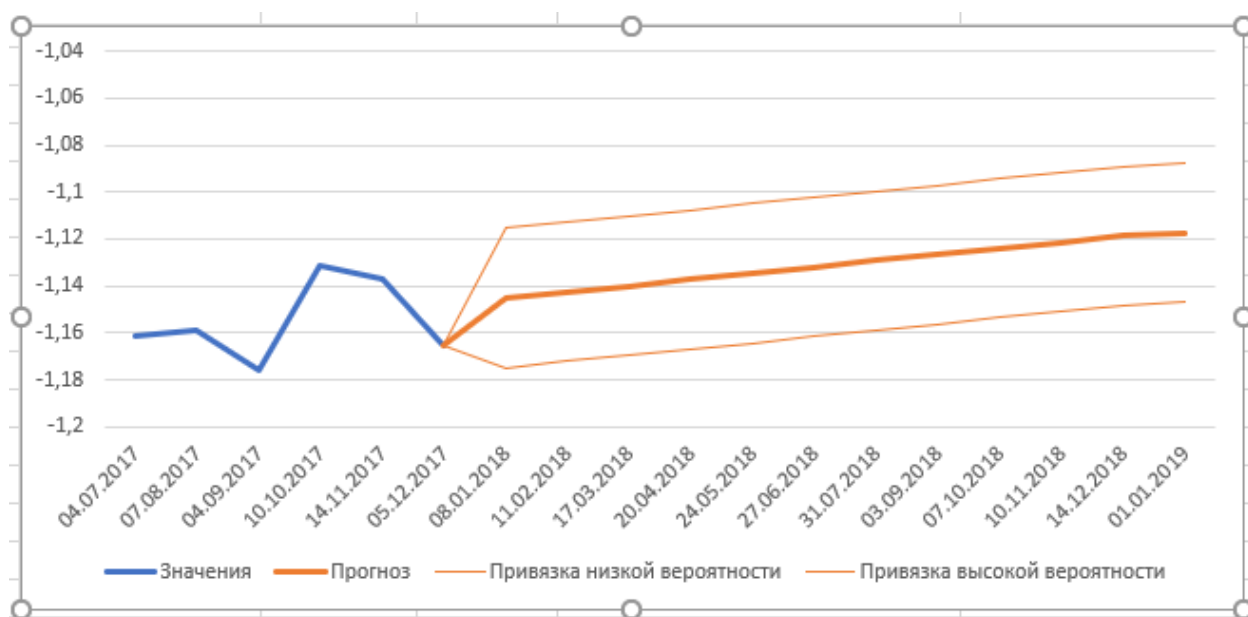


Рисунок 3.15 – Графік прогнозування рівня хлоридів у водному середовищі

	A	B	C	D	E
	Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
	04.07.2017	-1,47531882			
	07.08.2017	-1,56101913			
	04.09.2017	-1,58006364			
	10.10.2017	-1,51023376			
	14.11.2017	-1,52927827			
	05.12.2017	-1,5451487	-1,5451487	-1,55	-1,55
	08.01.2018		-1,5383898	-1,63	-1,45
	11.02.2018		-1,5389436	-1,64	-1,44
	17.03.2018		-1,5394974	-1,65	-1,43
	20.04.2018		-1,5400513	-1,66	-1,42
	24.05.2018		-1,5406051	-1,67	-1,41
	27.06.2018		-1,5411589	-1,68	-1,40
	31.07.2018		-1,5417127	-1,69	-1,40
	03.09.2018		-1,5422665	-1,69	-1,39
	07.10.2018		-1,5428203	-1,70	-1,38
	10.11.2018		-1,5433741	-1,71	-1,38
	14.12.2018		-1,543928	-1,71	-1,37
	01.01.2019		-1,5442212	-1,72	-1,37

Рисунок3.16 – Дані прогнозування рівня нітратів у водному середовищі

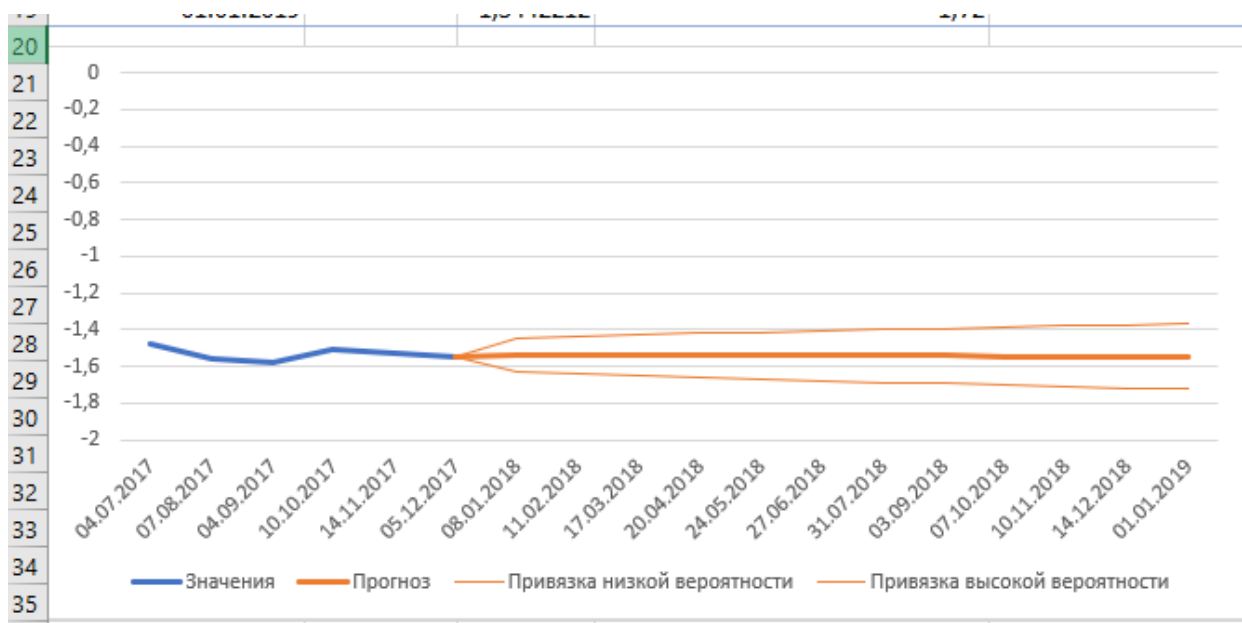


Рисунок3.17 – Графік прогнозування рівня нітратів у водному середовищі

Прогнозування рівня нітратів у водному середовищі за проміжок часу з 04.07.2017 по 01.01.2019, на основі даних спостереження з 04.07.2017 по 05.12.2017

Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
04.07.2017	-0,48056884			
07.08.2017	-0,4657679			
04.09.2017	-0,42136509			
10.10.2017	-0,45836743			
14.11.2017	-0,48056884			
05.12.2017	-0,4657679	-0,4657679	-0,47	-0,47
08.01.2018		-0,4705673	-0,52	-0,42
11.02.2018		-0,4727711	-0,53	-0,42
17.03.2018		-0,4749749	-0,54	-0,41
20.04.2018		-0,4771787	-0,54	-0,41
24.05.2018		-0,4793826	-0,55	-0,41
27.06.2018		-0,4815864	-0,56	-0,41
31.07.2018		-0,4837902	-0,56	-0,40
03.09.2018		-0,485994	-0,57	-0,40
07.10.2018		-0,4881979	-0,58	-0,40
10.11.2018		-0,4904017	-0,58	-0,40
14.12.2018		-0,4926055	-0,59	-0,40
01.01.2019		-0,4937722	-0,59	-0,40

Рисунок3.18 – Дані прогнозування рівня нітритів у водному середовищі

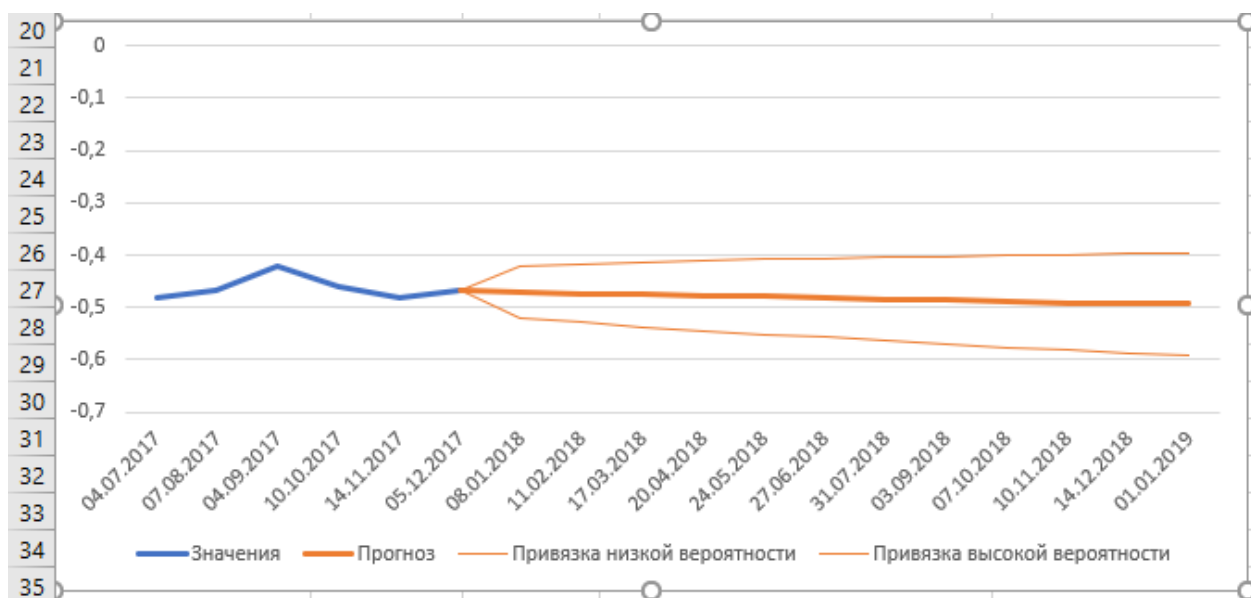


Рисунок3.19 – Графік прогнозування рівня нітритів у водному середовищі

	A	B	C	D	E
1	Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
2	04.07.2017	-0,64960799			
3	07.08.2017	-0,57939351			
4	04.09.2017	-0,61450075			
5	10.10.2017	-0,68471523			
6	14.11.2017	-0,7549297			
7	05.12.2017	-0,79003694	-0,7900369	-0,79	-0,79
8	08.01.2018		-0,8344547	-0,93	-0,74
9	11.02.2018		-0,8782877	-1,01	-0,75
10	17.03.2018		-0,9221208	-1,08	-0,76
11	20.04.2018		-0,9659538	-1,15	-0,78
12	24.05.2018		-1,0097868	-1,21	-0,81
13	27.06.2018		-1,0536199	-1,27	-0,83
14	31.07.2018		-1,0974529	-1,33	-0,86
15	03.09.2018		-1,141286	-1,39	-0,89
16	07.10.2018		-1,185119	-1,45	-0,92
17	10.11.2018		-1,2289521	-1,51	-0,95
18	14.12.2018		-1,2727851	-1,57	-0,98
19	01.01.2019		-1,2959908	-1,60	-0,99

Рисунок 3.20– Дані прогнозування рівня фосфатів у водному середовищі



Рисунок 3.21– Графік прогнозування рівня фосфатів у водному середовищі.

Як видно, з наведених вище результатів, прогнозування забруднення водного середовища залежить від рівня наявності різних речовин у водному середовищі. В свою чергу наявність та кількість речовин обумовлюють різні факторів. Органічні і не органічні речовини, що впливають на забруднення води, можуть потрапляти до водного

середовища у результаті промислової діяльності, людського фактору, кліматичних умов та інших чинників.

3.3 Розрахунок критерія Байєса

Для визначення мінімального Байєсівського критерію для кожного критерію був обраний показник ймовірності, та розрахований критерій Байєса, користуючись вбудованою функцією добутку ряду числових показників та показників ймовірності .

	A	B	C	D	E	F	G	H	I
		Нормалізація БСК	Нормалізація оксиген	Нормалізація сульфат	Нормалізація хлорид	Нормалізація Нітрат	Нормалізація Нітрит	Нормалізація Фосфат	Середнє
1									
2		-6,91975	0,632698	0,315891	1,861201	0,400566	0,333483	-0,89536	-1,74781
3		0,231034	0,988054	0,315891	1,448015	0,210121	0,333483	-1,17622	0,38005
4		0,332607	0,810376	0,315891	1,510569	0,362477	0,481492	-1,35175	0,393725
5		-0,27683	0,988054	0,794117	0,575551	0,337084	0,259478	-1,28154	0,18303
6		0,149775	0,401717	1,377833	1,04306	-1,0976	-0,25855	-1,19377	0,112372
7		-0,03306	0,401717	0,122993	1,978078	-1,08491	-0,33256	-0,57939	0,080847
8		0,231034	0,25069	0,578111	3,105697	-0,77067	-0,18455	-0,29854	0,362453
9		0,759217	-0,78873	0,592177	1,638969	-0,64688	0,111469	-0,28098	0,211495
10		-0,01274	1,698766	0,894584	2,679342	-0,71036	0,777511	-1,36931	0,563107
11		-0,17526	0,45502	0,605238	2,318833	0,660841	0,407487	0,421163	0,479782
12		0,820161	0,45502	1,184935	2,546003	0,114898	0,481492	-0,19321	0,750464
13		0,393551	0,366181	0,912668	1,904001	1,130605	0,92552	-0,3512	0,643461
551		-0,70344	-1,06413	-1,67296	-1,13843	-1,60546	-0,49537	-0,8427	-0,99935
552		-1,84107	-0,32676	-1,84205	-1,16115	-1,47532	-0,48057	-0,64961	-1,17854
553		-1,67855	-0,53998	-1,74842	-1,15917	-1,56102	-0,46577	-0,57939	-1,16294
554		-1,61761	-0,66435	-1,65478	-1,17563	-1,58006	-0,42137	-0,6145	-1,16279
555		-1,67855	-1,1885	-1,65478	-1,13102	-1,51023	-0,45837	-0,68472	-1,28518
556		-1,82075	-0,9664	-1,56114	-1,13678	-1,52928	-0,48057	-0,75493	-1,28578
557		-1,75981	-1,01971	-1,56114	-1,1651	-1,54515	-0,46577	-0,79004	-1,2846
558	Ймовірність	0,30	0,20	0,10	0,10	0,10	0,10	0,10	-4,36834

Рисунок 3.22 – Розрахування критерію Байєса

3.4 Визначення залишків

Для визначення залишків прогнозування серед нормалізованих даних для кожного критерію було визначено середнє значення та стандартне відхилення. Розрахунки були проведені з використанням влаштованих функцій Microsoft Excel, для розрахунку середнього і стандартного відхилення нормалізованих даних, а саме функції СРЗНАЧ і СТАНДОТКЛОН, результати розрахунків за якими зображені на рис. 3.33-3.34 [42],[].

B1559 =СРЗНАЧ(B2:B1557)									
	A	B	C	D	E	F	G	H	I
		Нормалізація БСК	Нормалізація окисген	Нормалізація сульфат	Нормалізація хлорид	Нормалізація Нітрат	Нормалізація Нітрит	Нормалізація Фосфат	Середнє
1									
2		-6,91975	0,632698	0,315891	1,861201	0,400566	0,333483	-0,89536	-1,74781
3		0,231034	0,988054	0,315891	1,448015	0,210121	0,333483	-1,17622	0,38005
4		0,332607	0,810376	0,315891	1,510569	0,362477	0,481492	-1,35175	0,393725
5		-0,27683	0,988054	0,794117	0,575551	0,337084	0,259478	-1,28154	0,18303
6		0,149775	0,401717	1,377833	1,04306	-1,0976	-0,25855	-1,19377	0,112372
7		-0,03306	0,401717	0,122993	1,978078	-1,08491	-0,33256	-0,57939	0,080847
8		0,231034	0,25069	0,578111	3,105697	-0,77067	-0,18455	-0,29854	0,362453
9		0,759217	-0,78873	0,592177	1,638969	-0,64688	0,111469	-0,28098	0,211495
10		-0,01274	1,698766	0,894584	2,679342	-0,71036	0,777511	-1,36931	0,563107
11		-0,17526	0,45502	0,605238	2,318833	0,660841	0,407487	0,421163	0,479782
12		0,820161	0,45502	1,184935	2,546003	0,114898	0,481492	-0,19321	0,750464
13		0,393551	0,366181	0,912668	1,904001	1,130605	0,92552	-0,3512	0,643461
1556		-1,82075	-0,9664	-1,56114	-1,13678	-1,52928	-0,48057	-0,75493	-1,28578
1557		-1,75981	-1,01971	-1,56114	-1,1651	-1,54515	-0,46577	-0,79004	-1,2846
1558	Ймовірність	0,30	0,20	0,10	0,10	0,10	0,10	0,10	-4,36834
1559	Середнє	-0,10673	-0,20337	-4,8E-16	-8,5E-15	3,69E-15	-0,00064	4,79E-15	
1560	Стандартне відхилення	1,308509	1,582058	1	1	1	0,999999	1	

Рисунок 3.33 – Формула середнього значення за вибіркою даних

	A	B	C	D	E	F	G	H	I
		Нормалізація БСК	Нормалізація кисень	Нормалізація сульфат	Нормалізація хлорид	Нормалізація нітрат	Нормалізація нітрит	Нормалізація фосфат	Середнє
1									
2		-6,91975	0,632698	0,315891	1,861201	0,400566	0,333483	-0,89536	-1,74781
3		0,231034	0,988054	0,315891	1,448015	0,210121	0,333483	-1,17622	0,38005
4		0,332607	0,810376	0,315891	1,510569	0,362477	0,481492	-1,35175	0,393725
5		-0,27683	0,988054	0,794117	0,575551	0,337084	0,259478	-1,28154	0,18303
6		0,149775	0,401717	1,377833	1,04306	-1,0976	-0,25855	-1,19377	0,112372
7		-0,03306	0,401717	0,122993	1,978078	-1,08491	-0,33256	-0,57939	0,080847
8		0,231034	0,25069	0,578111	3,105697	-0,77067	-0,18455	-0,29854	0,362453
9		0,759217	-0,78873	0,592177	1,638969	-0,64688	0,111469	-0,28098	0,211495
10		-0,01274	1,698766	0,894584	2,679342	-0,71036	0,777511	-1,36931	0,563107
11		-0,17526	0,45502	0,605238	2,318833	0,660841	0,407487	0,421163	0,479782
12		0,820161	0,45502	1,184935	2,546003	0,114898	0,481492	-0,19321	0,750464
13		0,393551	0,366181	0,912668	1,904001	1,130605	0,92552	-0,3512	0,643461
556		-1,82075	-0,9664	-1,56114	-1,13678	-1,52928	-0,48057	-0,75493	-1,28578
557		-1,75981	-1,01971	-1,56114	-1,1651	-1,54515	-0,46577	-0,79004	-1,2846
558	Ймовірність	0,30	0,20	0,10	0,10	0,10	0,10	0,10	-4,36834
559	Середнє	-0,10673	-0,20337	-4,8E-16	-8,5E-15	3,69E-15	-0,00064	4,79E-15	
560	Стандарт відхилення	1,308509	1,582058	1	1	1	0,999999	1	

Рисунок 3.34 – Формула стандартного відхилення за нормалізованою вибіркою даних

Залишком спостережуваної змінної є різниця між цим спостережуваним значенням та оцінкою значення досліджуваної величини.

3.5 Висновки до третього розділу

В результаті проведення розрахунків було використано програмний засіб Microsoft Excel. За допомогою даного програмного засобу проведено нормалізацію даних, для приведення їх до загального вигляду з використання середнього та стандартного відхилення. Також було проведено прогнозування даних для кожного з критеріїв, за допомогою відповідного інструмента прогнозування. На основі отриманих даних нормалізації проведено розрахунок Байєсівського критерію та виявлення залишків, з використанням середнього та стандартного відхилення нормалізованих даних.

РОЗДІЛ 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ. ЕКОЛОГІЯ

В даному розділі проведено аналіз потенційних небезпечних та шкідливих виробничих факторів, причин пожеж. Розглянуті заходи, які дозволяють забезпечити гігієну праці і виробничу санітарію. На підставі аналізу розроблені заходи з техніки безпеки та рекомендації з пожежної профілактики.

Завданням даної магістерської роботи було дослідити існуючі алгоритми та методи для об'єднання даних та прогнозування випадків аномального забруднення водних об'єктів та обрати найбільш оптимальний алгоритм, і як результат був проведений аналіз методів об'єднання даних та обраний метод Демпстера-Шафера, що дозволяє здійснювати злиття даних для подальшого прогнозування випадків забруднення водних середовищ, навіть в умовах наявності неповних даних. А також були проведені розрахунки, з використанням даної математичної моделі для виявлення випадків аномального забруднення в річці Сіверський Дінець, з використанням готових даних, проведеного моніторингу в період з 22 січня 2008 року по 5 грудня 2017 року. За цим для вдосконалення точності роботи даного методу в подальшому розроблятиметься реальна система, яка значно полегшить процес прогнозування та об'єднання даних, отриманих з річок, що дозволить прогнозувати випадки аномального забруднення води з урахуванням таких критеріїв як течія, наявність потенційно забруднюючих об'єктів та інших географічних, біологічних і хімічних чинників. В процесі проектування не використовувалося спеціалізоване обладнання, то аналіз потенційно небезпечних і шкідливих виробничих чинників виконується для робочого місця, з використанням персонального комп'ютера, на якому розроблялася математична модель.

4.1 Загальні питання з охорони праці

Умови праці на робочому місці, безпека технологічних процесів, машин, механізмів, устаткування та інших засобів виробництва, стан засобів колективного та індивідуального захисту, що використовуються працівником, а також санітарно-побутові умови повинні відповідати вимогам нормативних актів про охорону праці. В законі України «Про охорону праці» [1] визначається, що охорона праці - це система правових, соціально-економічних, організаційно-технічних, санітарно-гігієнічних і лікувально-

профілактичних заходів та засобів, спрямованих на збереження життя, здоров'я і працездатності людини у процесі трудової діяльності.

При роботі з обчислювальною технікою змінюються фізичні і хімічні фактори навколишнього середовища: виникає статична електрика, електромагнітне випромінювання, змінюється температура і вологість, рівень вміст кисню і озону в повітрі. Умови праці на робочому місці, безпека технологічних процесів, машин, механізмів, устаткування та інших засобів виробництва, стан засобів колективного та індивідуального захисту, що використовуються працівником, а також санітарно-побутові умови повинні відповідати вимогам нормативних актів про охорону праці.

4.2 Аналіз стану умов праці

Робота над створенням вдосконаленої математичної моделі для об'єднання даних і прогнозування аномальних випадків забруднення водних об'єктів проходитиме в приміщенні жилої кімнати. Для даної роботи достатньо однієї людини, для якої надано робоче місце зі стаціонарним комп'ютером.

4.2.1 Вимоги до приміщень

Геометричні розміри приміщення зазначені в табл. 4.1.

Таблиця 4.1 – Розміри приміщення

Найменування	Значення
Довжина, м	4,53
Ширина, м	2,36
Висота, м	2,50
Площа, м ²	10,69
Об'єм, м ³	26,72

Згідно з ДСН 3.3.6.042-99 «Санітарні норми мікроклімату виробничих приміщень» [45] розмір площі для одного робочого місця оператора персонального комп'ютера має бути не менше 6 кв. м, а об'єм — не менше 20 куб. м. Отже, дане

приміщення цілком відповідає зазначеним нормам.

Для зручності роботи в кімнаті є робоче місце а саме письмовий стіл, з необхідними, для роботи, матеріалами. Також робочий процес пов'язаний з багатьма електронними документами, для чого приміщення облаштоване принтером. Задля дотримання визначеного рівня мікроклімату в будівлі встановлено систему опалення.

Для забезпечення потрібного рівня освітленості кімната має вікно та систему загального рівномірного освітлення, що встановлена на стелі.

4.2.2 Вимоги до організації місця праці

При порівнянні відповідності характеристик робочого місця нормативним основні вимоги до організації робочого місця за ДСанПіН 3.3.2.007-98 «Правила і норми роботи з візуальними дисплейними терміналами електронно-обчислювальних машин» [46] і відповідними фактичними значеннями для робочого місця, констатуємо повну відповідність.

Таблиця 4.2 - Характеристики робочого місця

Найменування параметра	Фактичне значення	Нормативне значення
Висота робочої поверхні, мм	750	680 ÷ 800
Висота простору для ніг, мм	730	не менше 600
Ширина простору для ніг, мм	660	не менше 500
Глибина простору для ніг, мм	700	не менше 650
Висота поверхні сидіння, мм	470	400 ÷ 500
Ширина сидіння, мм	400	не менше 400
Глибина сидіння, мм	400	не менше 400
Висота поверхні спинки, мм	600	не менше 300
Ширина опорної поверхні спинки, мм	500	не менше 380
Радіус кривини спинки в горизонтальній площині, мм	400	400
Відстань від очей до екрану дисплея, мм	800	700 ÷ 800

Робочий стіл на досліджуваному місці також містить достатньо простору для ніг. Крісло, що використовується в якості робочого сидіння, дає можливість працювати у комфорті. Екран монітору знаходиться на відстані 0.8 м. Отже, за всіма параметрами робоче місце відповідає нормативним вимогам.

Приміщення кімнати знаходиться на четвертому поверсі чотирьох поверхової будівлі і має об'єм 26,72 м³, площу – 10,69 м². У цій кімнаті обладнано одне місце праці з можливістю користуватися ноутбуком.

Температура в приміщенні протягом року коливається у межах 18–25°C, відносна вологість — близько 50%. Швидкість руху повітря не перевищує 0,2 м/с. Система вентилявання приміщення — природна неорганізована, а опалення — індивідуальне, електричне.

Розміщення вікон забезпечує природне освітлення з коефіцієнтом природного освітлення не менше 1,5%, а загальне штучне освітлення, яке здійснюється за допомогою однієї енергозберігаючої лампи, забезпечує достатній рівень освітленості.

У кімнаті є електрична мережа з напругою 220 В, яка створює небезпеку ураження електричним струмом. Ноутбук та периферійні пристрої можуть бути джерелами електромагнітних випромінювань, аерозолів та шкідливих речовин.

4.2.3 Навантаження та напруженість процесу праці

Під час виконання магістерської роботи: за фізичним навантаженням робота відноситься до категорії легкі роботи (Ia), її виконують сидячи з періодичним ходінням. Щодо характеру організування виконання дипломної роботи, то він підпадає під нав'язаний режим, оскільки певні розділи роботи необхідно виконати у встановлені конкретні терміни. За ступенем нервово-психічної напруги виконання роботи можна віднести до II – III ступеня і кваліфікувати як помірно напружений – напружений за умови успішного виконання поставлених завдань.

Під час виконання робіт використовують ПК та периферійні пристрої (струменеві), що призводить до навантаження на окремі системи організму. Такі перекося у напруженні різних систем організму, що трапляються під час роботи з ПК, зокрема, значна напруженість зорового аналізатора і довготривале малорухоме положення перед екраном, не тільки не зменшують загального напруження, а навпаки, призводять до його посилення і появи стресових реакцій.

Найбільшому ризику виникнення різноманітних порушень піддаються: органи

зору, м'язово скелетна система, нервово-психічна діяльність, репродуктивна функція у жінок.

Роботу за дипломним проектом визнано, таку, що займає 50% часу робочого дня та за восьмигодинної робочої зміни рекомендовано встановити додаткові регламентовані перерви:

- для операторів персональних комп'ютерів тривалістю 15 хв через дві години роботи;
- для операторів комп'ютерного набору тривалістю 10 хв через кожен годину роботи.

4.3 Виробнича санітарія

На підставі аналізу небезпечних та шкідливих факторів при виробництві (експлуатації), пожежної безпеки можуть бути надалі вирішені питання необхідності забезпечення працюючих достатньою кількістю освітлення, вентиляції повітря, організації заземлення, тощо.

4.3.1 Аналіз небезпечних та шкідливих факторів при виробництві (експлуатації) виробу

Роботу, пов'язану з ЕОП з ВДТ, у тому числі на тих, які мають робочі місця, обладнані ЕОМ з ВДТ і ПП, виконують із забезпеченням виконання НПАОП 0.00.-1.28-10 «Правил охорони праці під час експлуатації електронно-обчислювальних машин» [47], які встановлюють вимоги безпеки до обладнання робочих місць, до роботи із застосуванням ЕОМ з ВДТ і ПП. Переважно роботи за проектами виконують у кабінетах чи інших приміщеннях, де використовують різноманітне електрообладнання, зокрема персональні комп'ютери (ПК) та периферійні пристрої.

Основними робочими характеристиками персонального комп'ютера є наступні:

- робоча напруга $U = +220\text{В} \pm 5\%$;
- робочий струм $I = 2\text{А}$;
- споживана потужність $P = 350\text{Вт}$.

4.3.2 Пожежна безпека

Небезпека розвитку пожежі на обчислювальному центрі обумовлюється застосуванням розгалужених систем електроживлення ЕОМ, вентиляції і кондиціонування. Небезпека загоряння пов'язана з особливістю комп'ютерів - із значною кількістю щільно розташованих на монтажній платі і блоках електронних вузлів і схем, електричних і комутаційних кабелів, резисторів, конденсаторів, напівпровідникових діодів і транзисторів. Надійна робота окремих елементів і мікросхем в цілому забезпечується тільки в певних інтервалах температури, вологості і при заданих електричних параметрах. При відхиленні реальних умов експлуатації від розрахункових можуть виникнути пожежонебезпечні ситуації.

Висока щільність елементів в електронних схемах призводить до значного підвищення температури окремих вузлів (80...100 °С). При проходженні електричного струму по провідниках і деталях виділяється тепло, що в умовах їх високої щільності може привести до перегріву, і може служити причиною запалювання ізоляційних матеріалів. Слабкий опір ізоляційних матеріалів дії температури може викликати порушення ізоляції і привести до короткого замикання між струмоведучими частинами обладнання (шини, електроди). Для зниження займистості і здатності поширювати полум'я кабелі покривають вогнезахисними покриттями. Проектом передбачено прокладати проводку: приховано, під знімною підлогою розділяючи негорючими діафрагмами, в малодоступних місцях.

Для гасіння пожеж в приміщенні пропонується використовувати кошму. Заземлені конструкції, що знаходяться в приміщеннях, де розміщені робочі місця (батареї опалення, водопровідні труби, кабелі із заземленим відкритим екраном), надійно захищені діелектричними щитками та/або сітками з метою недопущення потрапляння працівника під напругу.

Згідно НАПБ Б.03.002-2007 «Норми визначення категорій приміщень, будинків та зовнішніх установок за вибухопожежною та пожежною безпекою» [10], таке приміщення, площею 10,69 м², відноситься до категорії "В" (пожежонебезпечної).

Продуктами згорання, що виділяються на пожежі, є: окис вуглецю; сірчистий газ; окис азоту; синильна кислота; аромін; фосген; хлор і ін. При горінні пластмас, окрім звичних продуктів згорання, виділяються різні продукти термічного розкладання: хлорангідридні кислоти, формальдегіди, хлористий водень, фосген, синильна кислота, аміак, фенол, ацетон, стирол (ГОСТ 12.1.044-89 [49]).

4.3.3 Електробезпека

На робочому місці виконуються наступні вимоги електробезпеки: ПК, периферійні пристрої та устаткування для обслуговування, електропроводи і кабелі за виконанням та ступенем захисту відповідають класу зони за ПУЕ (правила улаштування електроустановок), мають апаратуру захисту від струму короткого замикання та інших аварійних режимів. Лінія електромережі для живлення ПК, периферійних пристроїв і устаткування для обслуговування, виконана як окрема групова три-провідна мережа, шляхом прокладання фазового, нульового робочого та нульового захисного провідників. Нульовий захисний провідник використовується для заземлення (занулення) електроприймачів. Електромережа штепсельних розеток для живлення персональних ПК, укладено по підлозі поруч зі стінами відповідно до затвердженого плану розміщення обладнання та технічних характеристик обладнання. Металеві труби та гнучкі металеві рукави заземлені. Захисне заземлення включає в себе заземлюючих пристроїв і провідник, який з'єднує заземлюючий пристрій з обладнанням, яке заземлюється - заземлюючий провідник.

4.4 Гігієнічні вимоги до параметрів виробничого середовища

4.4.1 Мікроклімат

Мікроклімат робочих приміщень – це клімат внутрішнього середовища цих приміщень, що визначається діючої на організм людини з'єднанням температури, вологості, швидкості переміщення повітря. В даному приміщенні проводяться роботи, що виконуються сидячи і не потребують динамічного фізичного напруження, то для нього відповідає категорія робіт Ia. Отже оптимальні значення для температури, відносної вологості й рухливості повітря для зазначеного робочого місця відповідають ДСН 3.3.6.042-99 «Санітарні норми мікроклімату виробничих приміщень» [45] і наведені в табл. 4.4:

Таблиця 4.4 – Норми мікроклімату робочої зони об'єкту

Період року	Категорія робіт	Температура С ⁰	Відносна вологість %	Швидкість руху повітря, м/с
Холодна	легка-1 а	22 - 24	40 – 60	0,1
Тепла	легка-1 а	23 - 25	40 – 60	0,1

Дане приміщення обладнане системами опалення, кондиціонування повітря. У приміщенні на робочому місці забезпечуються оптимальні значення параметрів мікроклімату: температури, відносної вологості й рухливості повітря у відповідності до ДСН 3.3.6.042-99 [45]. Рівні позитивних і негативних іонів у повітрі мають відповідати ДСН 3.3.6.042-99 [45]. Для забезпечення оптимальних параметрів мікроклімату в приміщенні проводяться перерви в роботі, з метою його провітрювання. Існують спеціальні системи кондиціонування, які забезпечують підтримання в приміщенні балансу оптимальних параметрів мікроклімату. Контроль параметрів мікроклімату в холодний і теплий період року здійснюється не менше 3-х разів на зміну (на початку, середині, в кінці).

4.4.2 Освітлення

Світло є природною умовою існування людини. Воно впливає на стан вищих психічних функцій і фізіологічні процеси в організмі. Хороше освітлення діє тонізуюче, створює гарний настрій, покращує протікання основних процесів вищої нервової діяльності.

Збільшення освітленості сприяє поліпшенню працездатності навіть в тих випадках, коли процес праці практично не залежить від зорового сприйняття. При поганому освітленні людина швидко втомлюється, працює менш продуктивно, виникає потенційна небезпека помилкових дій і нещасних випадків.

Освітленість приміщення має велике значення при роботі на ПЕОМ. Вона багато в чому визначається колірною і мережевий обстановкою. Для зменшеного поглинання світла стеля і стіни вище панелей (1,5-1,7м.). Якщо вони не облицьовані звукопоглинальним матеріалом, фарбуються білою водоемульсійною фарбою (коефіцієнт відбиття повинен бути не менше 0,7). Для забарвлення стіни панелей рекомендується віддавати перевагу світлим фарбам.

При експлуатації ЕОМ виконується зорова робота IV в розряді точності (середня точність). При цьому нормована освітленість на робочому місці (E_n) рівна 200 лк. Джерелом природного освітлення є сонячне світло.

У приміщенні, де розташовані ЕОМ передбачається природне бічне освітлення, рівень якого відповідає ДБН В.2.5-28:2015 [50]. Джерелом природного освітлення є сонячне світло. Регулярно повинен проводитися контроль освітленості, який підтверджує, що рівень освітленості задовольняє ДБН і для даного приміщення в світлий час доби достатньо природного освітлення.

Розрахунок освітлення.

Для будівель виробництв світловий коефіцієнт приймається в межах 1/6 - 1/10:

$$\sqrt{a^2 + b^2} \cdot S_b = (1/8 \div 1/10) \cdot S_n \quad (4.1)$$

де S_b – площа віконних прорізів, м²;

S_n – площа підлоги, м².

$$S_n = a \cdot b = 4,53 \cdot 2,36 = 10,69 \text{ м}^2$$

$$S_{\text{вік}} = 1,30 \cdot 1,70 = 2,21 \text{ м}^2$$

Приймаємо 1 вікно площею $S = 2,21 \text{ м}^2$.

Світильник загального освітлення розташовується в центрі приміщення. Для організації освітлення в темний час доби передбачається обладнати приміщення, довжина якого складає 4,53 м, ширина 2,36 м, світильником, оснащеним енергозберігаючою лампою (30 Вт) з світловим потоком 4000 лм.

Розрахунок штучного освітлення виробляється по коефіцієнтах використання світлового потоку, яким визначається потік, необхідний для створення заданої освітленості при загальному рівномірному освітленні. Розрахунок кількості світильників n виробляється по формулі (4.2):

$$n = \frac{E \cdot S \cdot Z \cdot K}{F \cdot U \cdot M} \quad (4.2)$$

де E – нормована освітленість робочої поверхні, визначається нормами – 150 лк для житлових приміщень;

S – освітлювана площа, м²; $S = 10,69 \text{ м}^2$;

Z – поправочний коефіцієнт світильника ($Z = 1,15$ для ламп розжарювання та ДРЛ; $Z = 1,1$ для люмінесцентних та енергоощадних ламп) приймаємо рівним 1,1;

K – коефіцієнт запасу, що враховує зниження освітленості в процесі експлуатації – 1,5;

U – коефіцієнт використання, залежний від типу світильника, показника індексу приміщення і т.п. – 0,575

M – число енергозберігаючих ламп в світильнику – 1;

F – світловий потік лампи – 4000лм (для 30 Вт).

Підставивши числові значення у формулу (4.2), отримуємо:

$$n = \frac{150 \cdot 10,69 \cdot 1,1 \cdot 1,5}{4000 \cdot 0,575 \cdot 1} \approx 1$$

Приймаємо освітлювальну установку, яка складається з 1 світильника, що складається з однієї енергозберігаючої лампи загальною потужністю 30 Вт, напругою – 220 В.

4.3 Шум та вібрація, електромагнітне випромінювання

Рівень шуму, що супроводжує роботу користувачів персональних комп'ютерів (зумовлений як роботою системних блоків, клавіатури, так і друкуванням на принтерах, а також зовнішніми чинниками), коливається у межах 50–65 дБА згідно ДСН 3.3.6.037-99 [51].

В роботі використовувався ноутбук LENOVO IdeaPad 310, тому рівень шуму та вібрації відповідав затвердженим нормативам.

4.4 Вентилювання

У приміщенні, де знаходяться ЕОМ, повітрообмін реалізується за допомогою природної організованої вентиляції (вентиляційні шахти в віконному отворі. Цей метод забезпечує приток потрібної кількості свіжого повітря, що визначається в СНіП (30 м³ на годину на одного працюючого).

Також має здійснюватися провітрювання приміщення, в залежності від погодних умов, тривалість повинна бути не менше 10 хв. Найкращий обмін повітря здійснюється при наскрізному провітрюванні.

4.5 Заходи з організації виробничого середовища та попередження виникнення надзвичайних ситуацій

Відповідно до санітарно-гігієнічних нормативів та правил експлуатації обладнання наводимо приклади деяких заходів безпеки.

1) Заходи безпеки під час експлуатації персонального комп'ютера та периферійних пристроїв передбачають:

- правильне організування місця праці та дотримання оптимальних режимів праці та відпочинку під час роботи з ПК;
- експлуатацію сертифікованого обладнання;
- дотримання заходів електробезпеки;
- забезпечення оптимальних параметрів мікроклімату;
- забезпечення раціонального освітлення місця праці (освітленість робочого місця не перевищувала 2/3 нормальної освітленості приміщення);

2) Заходи безпеки під час експлуатації інших електричних приладів передбачають дотримання таких правил:

- постійно стежити за справним станом електромережі, розподільних щитків, вимикачів, штепсельних розеток, лампових патронів, а також мережевих кабелів живлення, за допомогою яких електроприлади під'єднують до електромережі;
- постійно стежити за справністю ізоляції електромережі та мережевих кабелів, не допускаючи їхньої експлуатації з пошкодженою ізоляцією;
- не тягнути за мережевий кабель, щоб витягти вилку з розетки;
- не закривати меблями, різноманітним інвентарем вимикачі, штепсельні розетки;

- не підключати одночасно декілька потужних електропристроїв до однієї розетки, що може викликати надмірне нагрівання провідників, руйнування їхньої ізоляції, розплавлення і загоряння полімерних матеріалів;

- не залишати включені електроприлади без нагляду;

- не допускати потрапляння всередину електроприладів крізь вентиляційні отвори рідин або металевих предметів, а також не закривати їх та підтримувати в належній чистоті, щоб уникнути перегрівання та займання приладу;

- не ставити на електроприлади матеріали, які можуть під дією теплоти, що виділяється, загорітися (канцелярські товари, сувенірну продукцію тощо).

Розрахунок захисного заземлення (забезпечення електробезпеки будівлі).

Загальний опір захисного заземлення визначається за формулою:

$$R_{ззп} = \frac{R_з \cdot R_n}{R_n \cdot n \cdot \eta_з + R_з \cdot \eta_n}, \quad (4.3)$$

де $R_з$ - опір заземлення, якими когут бать труби, опори, кути і т.п., Ом;

R_n - опір опори, яке з'єднує заземлювачі, Ом;

n - кількість заземлювачів;

$\eta_з$ - коефіцієнт екранування заземлювача; приймається в межах $0,2 \div 0,9$; $\eta_з = 0,7$

η_n - коефіцієнт екранування сполучної стійки; приймається в межах $0,1 \div 0,7$; $\eta_n = 0,5$;

Опір заземлення визначається за формулою:

$$R_з = \frac{\rho}{2\pi \cdot l} \cdot \left(\ln \frac{2 \cdot l}{d} + \frac{1}{2} \ln \frac{4 \cdot t + l}{4 \cdot t - l} \right), \quad (4.4)$$

де ρ - питомий опір ґрунту, залежить від типу ґрунту, Ом·м;

для піску - $400 \div 700$ Ом·м; приймаємо $\rho = 400$ Ом·м;

l - довжина заземлювача, м; для труб - 2-3 м; $l = 3$ м;

d - діаметр заземлювача, м; для труб - 0,03-0,05 м; $d = 0,05$ м;

t - відстань від середини забитого в ґрунт заземлювача до рівня землі, м; $t = 2$ м.

$$R_з = \frac{400}{2 \cdot 3,14 \cdot 3} \left(\ln \frac{2 \cdot 3}{0,05} + \frac{1}{2} \ln \frac{4 \cdot 2 + 3}{4 \cdot 2 - 3} \right) = 110, \text{ Ом}$$

Опір смуги, що з'єднує заземлювачі, визначається за формулою:

$$R_u = \frac{\rho}{2\pi \cdot L} \cdot \ln \frac{2 \cdot L^2}{b \cdot t^1}, \quad (4.5)$$

де L - довжина смуги, що з'єднує заземлювачі (м) і приблизно дорівнює периметру будівлі: $P_{\text{буд.}} = 42 \cdot 2 + 38 \cdot 2 = 160$ м; $L = 160$ м;

b - ширина смуги, м; $b = 0,03$ м;

t_1 - глибина заземлення від рівня землі, м; $t_1 = 0,5$ м.

$$R_u = \frac{400}{2 \cdot 3,14 \cdot 160} \cdot \ln \frac{2 \cdot 160^2}{0,03 \cdot 0,5} = 5,99, \text{ Ом}$$

Кількість заземлювачів захисного заземлення визначається за формулою:

$$n = \frac{2 \cdot R_s}{4 \cdot \eta_s}, \quad (4.6)$$

де 4 - допустимий загальний опір, Ом;

2 - коефіцієнт сезонності.

Визначаємо загальний опір захисного заземлення:

$$R_{\text{ззп}} = \frac{110 \cdot 5,99}{5,99 \cdot 79 \cdot 0,7 + 110 \cdot 0,5} = 1,7 \text{ Ом}$$

Висновок: дане захисне заземлення буде забезпечувати електробезпеку будівлі, так як виконується умова: $R_{\text{ззп}} < 4$ Ом.

3) При виникненню пожеж при роботі на ПЕОМ від таких можливими джерел запалювання як:

- іскри і дуги коротких замикань;
- перегрів провідників, резисторів та інших радіодеталей ПЕОМ, від тривалої перевантаження та наявність перехідного опору;
- іскри при розмиканні і розмиканні ланцюгів;
- розряди статичної електрики;

– необережному поводженню з вогнем, а також вибухи газо-повітряних і пароповітряних сумішей.

Важливу увагу слід звернути на пожежну безпеку підприємства в цілому і окремих його приміщень. В приміщеннях не повинно накопичуватися сміття, непотрібний папір, мотлох та ін. речі, які не використовуються у виробничому процесі. Наявний вільний аварійний вихід за межі приміщення в разі пожежі, бути передбачені вогнегасники. Вони повинні бути в робочому стані і перевірятися згідно з нормами. У приміщеннях повинна бути пожежна сигналізація, вогнегасник. У разі виникнення пожежі необхідно повідомити в найближчу пожежну частину, убезпечити інших працівників і по можливості прийняти кроки по запобіганню можливих наслідків та усуненню пожежі.

4.6 Охорона навколишнього природного середовища

4.6.1 Загальні дані з охорони навколишнього природного середовища

Діяльність за темою магістерської роботи, а саме: Методи та інформаційна технологія об'єднання даних он-лайн моніторингу поверхневих вод в процесі її виконання впливає на навколишнє природне середовище і регламентується нормами діючого законодавства: Законом України «Про охорону навколишнього природного середовища» [52], Законом України «Про забезпечення санітарного та епідемічного благополуччя населення» [53], Законом України «Про відходи» [54], Законом України «Про охорону атмосферного повітря» [55], Законом України «Про захист населення і територій від надзвичайних ситуацій техногенного та природного характеру» [56], Водний кодекс України [57].

Основним екологічним аспектом в процесі діяльності за даними спеціальностями є процеси впливу на атмосферне повітря та процеси поводження з відходами, які утворюються, збираються, розміщуються, передаються на віддалення (знешкодження), утилізацію, тощо в ІТ галузі.

Вплив на атмосферне повітря при нормальних умовах праці не оказує, бо не має в приміщенні сканерів, принтерів та інших джерел викиду забруднюючих речовин в повітря робочої зони.

4.7 Висновки до четвертого розділу

В результаті проведеної роботи було зроблено аналіз умов праці, шкідливих та небезпечних чинників, з якими стикається робітник. Було визначено параметри і певні характеристики приміщення для роботи над запропонованим проектом написаному в дипломній роботі, описано, які заходи потрібно зробити для того, щоб дане приміщення відповідало необхідним нормам і було комфортним і безпечним для робітника. Приведені рекомендації щодо організації робочого місця, а також важливу інформацію щодо пожежної та електробезпеки. Були наведені розміри приміщення та наведено значення температури, вологості й рухливості повітря, необхідна кількість і потужність ламп та інші параметри, значення яких впливає на умови праці робітника, а також – наведені інструкції з охорони праці, техніки безпеки при роботі на комп'ютері.

А також визначені основні екологічні аспекти впливу на навколишнє природне середовище та зазначені заходи щодо поводження з ними.

ВИСНОВКИ

В результаті роботи був проведений аналітичний огляд існуючих інструментальних засобів для проведення прогнозування та об'єднання даних. Також було розглянуто найбільш відомі математичні методи та алгоритми для прогнозування та злиття даних. Серед розглянутих методів був процес інтеграції даних, методи об'єднання даних на основі фільтра Калмана та алгоритм злиття даних на основі теорії-доказів Демпстера-Шафера. Найбільш продуктивним виявився метод на основі теорії доказів Демпстера-Шафера. Перевагою даного методу є можливість оперувати неповними даними та найбільш точні результати. Для використання методу на основі теорії доказів Демпстера-Шафера був запропонований алгоритм з використанням, в якості даних, готового набору даних, моніторингу річки Сіверський Дінець. Для проведення розрахунків використовувалось програмне середовище Microsoft Excel, через можливість використання необхідних для роботи математичних та статистичних функцій, і здатність програмного засобу, дозволяти створювати графіки та прогнози. Для роботи була запропонована та проведена нормалізація на основі середнього на стандартного відхилення. Також було проведено прогнозування за допомогою інструмента авторегресії ARIMA. Наступним етапом запропонованого алгоритму було знаходження Байєсівського критерію і залишків прогнозування. В роботі наведені таблиці та графіки проведених розрахунків. В розділі охорони праці та безпеки в надзвичайних ситуаціях наведені розрахунки для приміщення, в якому виконувалась робота, а також розроблені заходи з техніки безпеки та рекомендації з пожежної профілактики під час написання магістерської роботи.

Як результат роботи, запропонований математичний алгоритм, що дозволяє проводити об'єднання даних та прогнозування випадків аномального забруднення водного середовища.

В подальшому можливе вдосконалення алгоритму, шляхом урахування впливу сезонних факторів, швидкості течії, погодних умов, виявлення залежності між факторами та інших критерій для отриманні більш точних результатів прогнозування.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Azure Machine Learning Studio [Електронний ресурс] – Режим доступу: <https://azure.microsoft.com/en-us/services/machine-learning-studio/> (дата звернення 29.10.2018р.) – Назва з екрана.
2. Bleiholder J. Data fusion and conflict resolution in integrated information systems. — Potsdam: Hasso-PlattnerInstitut, 2010. D.Sc. Diss. 184 p.
3. Bleiholder J., Naumann F. Data fusion // ACM Computing Surveys (CSUR), 2009. Vol. 41. Iss. 1. Article No. 1. doi: 10.1145/1456650.1456651.
4. Bleiholder J., Naumann F. Declarative data fusion — syntax, semantics, and implementation // East European Conference on Advances in Databases and Information Systems (ADBIS) Proceedings, 2005. P. 58–73.
5. B.S.Paik, J.H.Oh, “Gain fusion algorithm for decentralized parallel Kalman filters,” Control Theory and Applications IEE Proceedings, vol. 147, I ISSN: 1350-2379.
6. Chen L Z, Shi W K, Deng Y and Zhu Z F 2005 A new fusion approach based on distance of evidences J. Zhejiang Univ. Sci. 6 476–82
7. Deductor [Електронний ресурс] – Режим доступу: <https://basegroup.ru/deductor/description> (дата звернення 29.10.2018р.)
8. Dellemc. Большие данные[Електронний ресурс] – Режим доступу: <https://www.dellemc.com/uk-ua/big-data/index.htm> (дата звернення 29.10.2018р.) – Назва з екрана.
9. Hall J and Szabo J 2005 WaterSentinel Online Water Quality Monitoring as an Indicator of Drinking Water Contamination (Washington, DC: Environmental Protection Agency)
10. Hou, Dibo & He, Huimei & Huang, Pingjie & Zhang, Guangxin & Loaiciga, Hugo. (2013). Detection of water-quality contamination events based on multi-sensor fusion using an extended Dempster–Shafer method. Measurement Science and Technology. 24. 055801. 10.1088/0957-0233/24/5/055801.
11. IBM Watson Analytics [Електронний ресурс] – Режим доступу: <https://www.ibm.com/watson-analytics> (дата звернення 29.10.2018р.)
12. Jitendra R. Raol, “Multi-Sensor Data Fusion with Matlab,”CRC Press, Taylor & Francis Group, ISBN 978-1- 4398-0003-4, 2010.
13. Jusselme A L, Grenier D and Bosse E 2001 A new distance between two bodies of evidence Inf. Fusion 2 91–101
14. Khaleghi, Bahador & Khamis, Alaa & Karray, Fakhri & Razavi, Saiedeh. (2013). Multisensor Data Fusion: A Review of the State-of-the-art. Information Fusion - INFFUS. 14. 10.1016/j.inffus.2011.08.001.

15. Luan C J, Wang X F, Zhang H Z, Jin S S and Li F 2012 Approach for constructing basic probability assignment for evidence theory in problems of ordered propositions *Comput. Eng. Appl.* 48 217–21
16. Multi sensor data fusion algorithms for target tracking using multiple measurements. Anitha, R. /Renuka, S. / Abudhahir, A. | 2013. *Elektronische Ausgabe.* 1.
17. Murray R et al 2010 Water quality event detection systems for drinking water contamination warning systems Development, Testing, and Application of Canary (Washington, DC: Environmental Protection Agency)
18. Pires, IM, Garcia, NM, Pombo, N., & Flórez-Revuelta, F. (2016). От сбора данных до сбора данных: всесторонний обзор и дорожная карта для определения активности повседневной жизни с использованием мобильных устройств. *Датчики (Базель, Швейцария)*, 16 (2), 184. doi: 10.3390 / s16020184
19. Rajaraman A., Ullman J. D. Integrating information by outerjoins and full disjunctions // PODS'96: 15th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems Proceedings, 1996. P. 238–248.
20. Sentz K and Ferson S 2002 *Combination of Evidence in Dempster-Shafer Theory* (Albuquerque, NM: Sandia National Laboratories)
21. X.R.Li, “Comparison of Two Measurement Fusion Methods for Kalman-Filter-Based Multisensor Data Fusion,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 273-280, 2001.
22. ZhuGe J W, Wang D W, Chen Y, Ye Z Y and Zou W 2006 A network anomaly detector based on the D-S evidence theory *J. Softw.* 17 463–71
23. А. Е. Вовченко, Л. А. Калиниченко, Д. Ю. Ковалев, “Методы разрешения сущностей и слияния данных в ETL-процессе и их реализация в среде Hadoop”, *Информ. и е ё примен.*, 8:4 (2014), 94–109
24. Навчальний посібник/ Степова О.В., Рома В.В. // Моніторинг поверхневих вод 2017 .–82с.
25. Наиболее полный список инструментов для анализа данных и машинного обучения. February 5, 2017 [Электронный ресурс] – Режим доступа: <http://ru.datasides.com/big-data-analytic-tools/#Flume> (дата звернення 29.10.2017р.) – Назва з екрана.
26. Режим аналізу часових рядів [Электронный ресурс] – режим доступа: http://help.prognoz.com/ru/mergedProjects/ExcelAddin/work/timeseries_mode/timeseries_work/chart_work.htm (дата звернення 11.01.2019р.) – Назва з екрана.
27. Правила построения АРПСС-моделей [Электронный ресурс] – Режим доступа: <https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUKEwiv6cbD0OvfAhVutIsKHdLyByUQjhx6BAgBEAM&url=http%3A%2F%2Fwww.geocities.ws%2Fflukistrela%2FStatistics%2Ftrends%2FARIMA%2FModellingRules%2Findex.html&psig=AOvVaw2f17KvsDMG4lcFKm7uBiN3&ust=1547498829304319> (дата звернення 11.01.2019 р.) – Назва з екрану.

28. Прогнозирование в один клик в Excel [Электронный ресурс] – Режим доступа: <https://blogs.technet.microsoft.com/tasush/2015/11/16/excel-2016/> дата звернення 11.01.2019 р.) – Назва з екрану.
29. Прогнозирование временных рядов при помощи экспоненциального сглаживания [Электронный ресурс] – Режим доступа: <https://www.mql5.com/ru/articles/318> (дата звернення 11.01.2019 р.) – Назва з екрану.
30. Засоби нормалізації змінних [Электронный ресурс] – Режим доступа: <https://neuronus.com/theory/nn/925-sposoby-normalizatsii-peremennykh.html> (дата звернення 11.01.2019 р.) – Назва з екрану.
31. Трансформація даних [Электронный ресурс] – Режим доступа: <http://webcache.googleusercontent.com/search?q=cache:D61dIR6U-2cJ:kek.ksu.ru/eos/dw/transform.doc+&cd=1&hl=ru&ct=clnk&gl=ua> (дата звернення 11.01.2019 р.) – Назва з екрану.
32. ARIMA [Электронный ресурс] – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%B2%D1%82%D0%BE%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D0%B8%D0%BD%D1%82%D0%B5%D0%B3%D1%80%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D0%BE%D0%B5_%D1%81%D0%BA%D0%BE%D0%BB%D1%8C%D0%B7%D1%8F%D1%89%D0%B5%D0%B5_%D1%81%D1%80%D0%B5%D0%B4%D0%BD%D0%B5%D0%B5 (дата звернення 11.01.2019 р.) – Назва з екрану.
33. Arima. [Электронный ресурс] – Режим доступа: http://help.prognoz.com/ru/mergedProjects/Lib/02_time_series_analysis/uimodelling_arima.htm (дата звернення 11.01.2019 р.) – Назва з екрану.
34. Translated, with permission of the ACP — ASIM, from: Goodman S.N. Towards evidence-based medical statistics: 1: The Bayes factor. *Ann Intern Med* 1999;130:1005—13.
35. Goodman S.N. Meta-analysis and evidence. *Control Clin Trials* 1989;10:188—204, 435.
36. Efron B. Empirical Bayes methods for combining likelihoods. *Journal of the American Statistical Association* 1996;91:538—50.
37. Hardy R.J., Thompson S.G. A likelihood approach to meta-analysis with random effects. *Stat Med* 1996;15:619—29.
38. Циганов О.В. Основи проектування систем штучного інтелекту. Конспект лекцій для студентів спеціальності 7.080402 – інформаційні технології проектування. – Одеса: Наука і техніка, 2006 – 196 с.
39. Кузнецов Ю. Н., Кузубов В. И., Волощенко А. Б. Математическое программирование — М.: «Высшая школа», 1980, стр.291.
40. Метод наименьших квадратов [Электронный ресурс] – Режим доступа: https://studopedia.su/5_28356_metod-naumenshih-kvadrativ.html (дата звернення 12.01.2019 р.) – Назва з екрану.

41. Нормалізація (функція НОРМАЛІЗАЦІЯ) [Електронний ресурс] – Режим доступу: <https://support.office.com/ru-ru/article/%D0%9D%D0%9E%D0%A0%D0%9C%D0%90%D0%9B%D0%98%D0%97%D0%90%D0%A6%D0%98%D0%AF-%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F-%D0%9D%D0%9E%D0%A0%D0%9C%D0%90%D0%9B%D0%98%D0%97%D0%90%D0%A6%D0%98%D0%AF-35dc2f30-b032-4c63-8f45-308f025a7da4> (дата звернення 12.01.2019 р.) – Назва з екрану.
42. Стандарт отклоня (Функція СТАНДАРТОТКЛОНА) [Електронний ресурс] – Режим доступу: <https://support.office.com/ru-ru/article/%D0%A1%D0%A2%D0%90%D0%9D%D0%94%D0%9E%D0%A2%D0%9A%D0%9B%D0%9E%D0%9D%D0%90-%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F-%D0%A1%D0%A2%D0%90%D0%9D%D0%94%D0%9E%D0%A2%D0%9A%D0%9B%D0%9E%D0%9D%D0%90-5ff38888-7ea5-48de-9a6d-11ed73b29e9d> (дата звернення 12.01.2019 р.) – Назва з екрану.
43. Срзнач (Функція СРЗНАЧ) [Електронний ресурс] – Режим доступу: <https://support.office.com/ru-ru/article/%D0%A1%D0%A0%D0%97%D0%9D%D0%90%D0%A7-%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F-%D0%A1%D0%A0%D0%97%D0%9D%D0%90%D0%A7-047bac88-d466-426c-a32b-8f33eb960cf6> (дата звернення 12.01.2019 р.) – Назва з екрану
44. Закон України Про охорону праці
45. ДСН 3.3.6.042-99 Санітарні норми мікроклімату виробничих приміщень
46. ДСанПіН 3.3.2.007-98 Гігієнічні вимоги до організації роботи з візуальними дисплейними терміналами електронно-обчислювальних машин
47. НПАОП 0.00-1.28-10 Правила охорони праці під час експлуатації електронно-обчислювальних машин
48. НАПБ Б.03.002-2007 Норми визначення категорій приміщень, будинків та зовнішніх установок за вибухопожежною та пожежною небезпекою
49. ГОСТ 12.1.044-89 ССБТ. Пожаровзрывоопасность веществ и материалов. Номенклатура показателей и методы их определения
50. ДБН В.2.5-28:2015. Природне і штучне освітлення
51. ДСН 3.3.6.037-99 Санітарні норми виробничого шуму, ультразвуку та інфразвуку
52. Закон України Про охорону навколишнього природного середовища
53. Закон України Про забезпечення санітарного та епідемічного благополуччя населення
54. Закон України Про відходи
55. Закон України Про охорону атмосферного повітря
56. Закон України „Про захист населення і територій від надзвичайних ситуацій техногенного та природного характеру”

57. Кодекс Водний кодекс України

ДОДАТОК А СЛАЙДИ ПРЕЗЕНТАЦІЇ

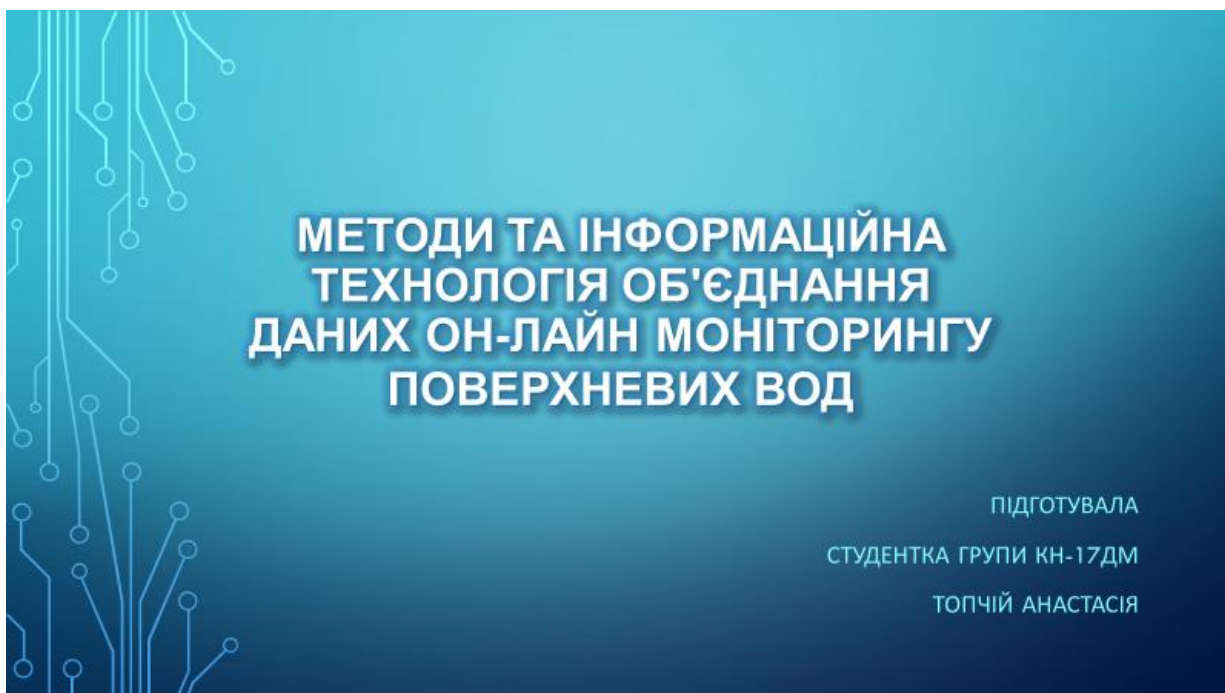


Рисунок А.1 – Титульний слайд



Рисунок А.2 – Актуальність

МЕТА І ЗАВДАННЯ ДОСЛІДЖЕННЯ

Мета: підвищення якості об'єднання даних он-лайн моніторингу, та досягнення, за допомогою вдосконалених математичних моделей, максимально точного прогнозування аномальних забруднень поверхневих вод.

Завдання:

- аналіз математичних методів і програмних засобів для проведення злиття даних та прогнозування;
- розроблення математичного алгоритму на основі теорії доказів Демпстера-Шафера, що дає змогу робити якомога точніші прогнози, щодо якості води;

Рисунок А.3 – Мета і завдання

МЕТОДИ ДОСЛІДЖЕННЯ

- Проведені в роботі дослідження засновані на методах об'єднання та прогнозування даних, які використовувались при розробленні математичного алгоритму для об'єднання даних та прогнозування аномальних випадків забруднення водного середовища;

Рисунок А.4 – Методи дослідження

ПРАКТИЧНЕ ЗНАЧЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Основні наукові положення роботи реалізовані у виді розрахункових моделей, які утворюють прикладну інформаційну технологію об'єднання даних та прогнозування аномалій забруднення поверхневих вод.

Рисунок А.5 – Практичне значення отриманих результатів

- В результаті дослідження було виявлено, що питання злиття даних, на наш час, є досить актуальними та потребують подальшого вдосконалення та вирішення. Зокрема об'єднання даних он-лайн моніторингу поверхневих вод досить мало розглядалось науковцями.
- Для вирішення цього питання було вирішено проаналізувати існуючі програмні засоби та математичні методи, і в результаті обрати алгоритм, що дозволяє отримувати найбільш точні кінцеві данні.

Рисунок А.6 – Результати аналізу

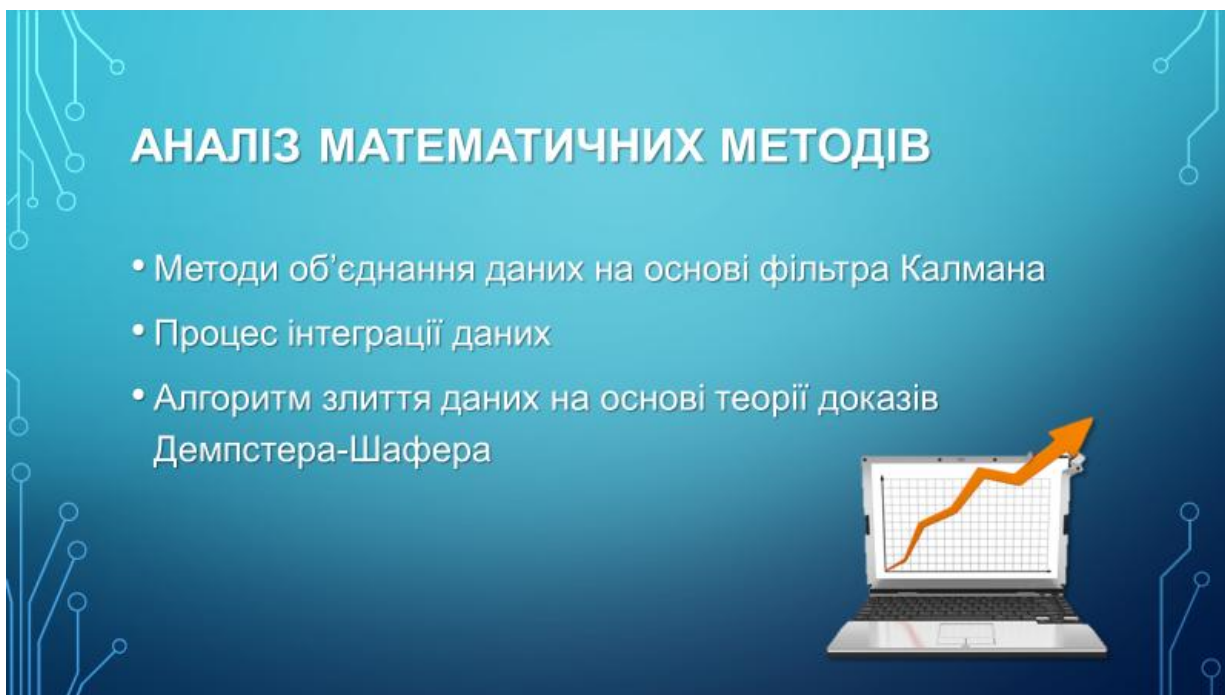


Рисунок А.7 – Аналіз математичних методів



Рисунок А.8 – Алгоритм математичного методу на основі теорії доказів Демпстера-Шафера

ДАНІ ДОСЛІДЖЕННЯ

- Були обрані готові дані отримані завдяки моніторингу річки Сіверський Дінець з 22.01.2008 по 05.12.2017 р.
- Дані, над якими були проведені розрахунки включають в себе відомості про кількість кисню, біологічне споживання кисню, наявність сульфатів, хлоридів, нітратів, нітритів та фосфатів у водному середовищі, за визначений проміжок часу на певному відрізку русла річки.

Рисунок А.9 – Дані дослідження

- В результаті проведення розрахунків було використано програмний засіб Microsoft Excel. За допомогою даного програмного засобу проведено нормалізацію даних, для приведення їх до загального вигляду з використання середнього та стандартного відхилення. Також було проведено прогнозування даних для кожного з критеріїв, за допомогою відповідного інструмента прогнозування. На основі отриманих даних нормалізації проведено розрахунок критерія Байєса та виявлення залишків, з використанням середнього та стандартного відхилення нормалізованих даних.

Рисунок А.10 – Проведення розрахунків

МОДЕЛЬ АВТОРЕГРЕСІЇ ARIMA

- Дозволяє будувати короткострокові прогнози на основі отриманих даних.
- За початковий діапазон обрано проміжок часу з 12.01.2016 по 05.12.2017 та відповідні діапазони даних біологічного споживання кисню, рівень кисню, рівень нітратів, рівень нітритів, рівень сульфатів, та рівень фосфатів за цей період.

Рисунок А.11 – Модель авторегресії ARIMA

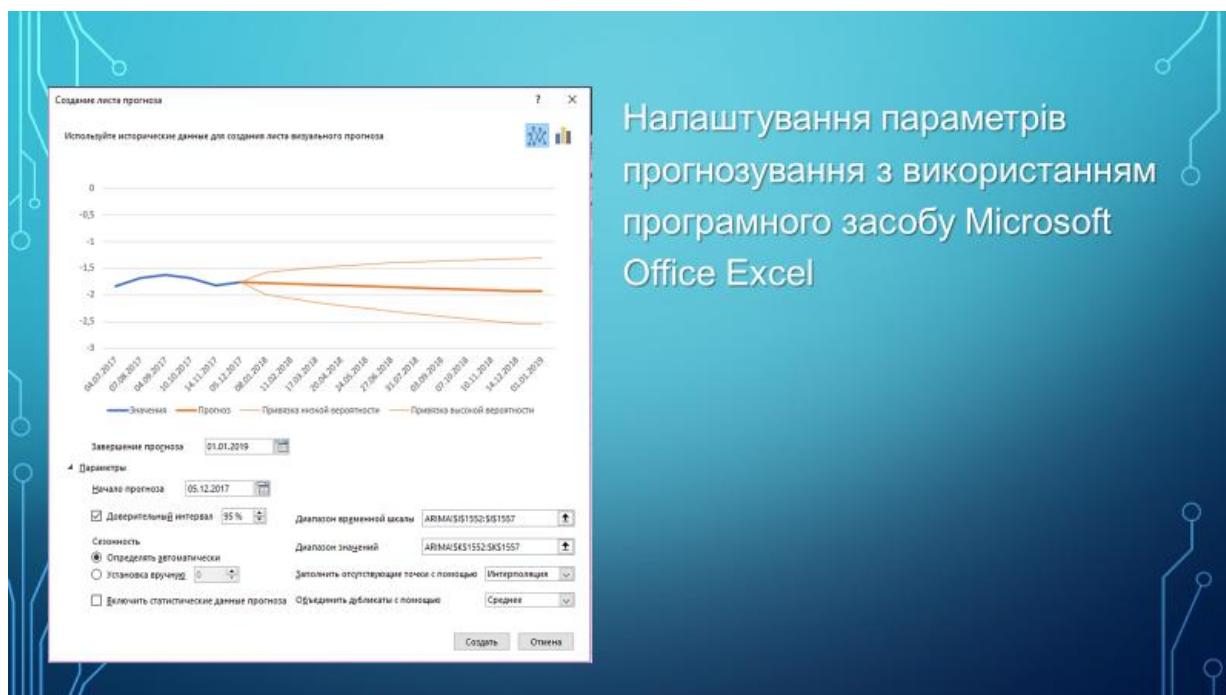


Рисунок А.12 – Параметри прогнозування

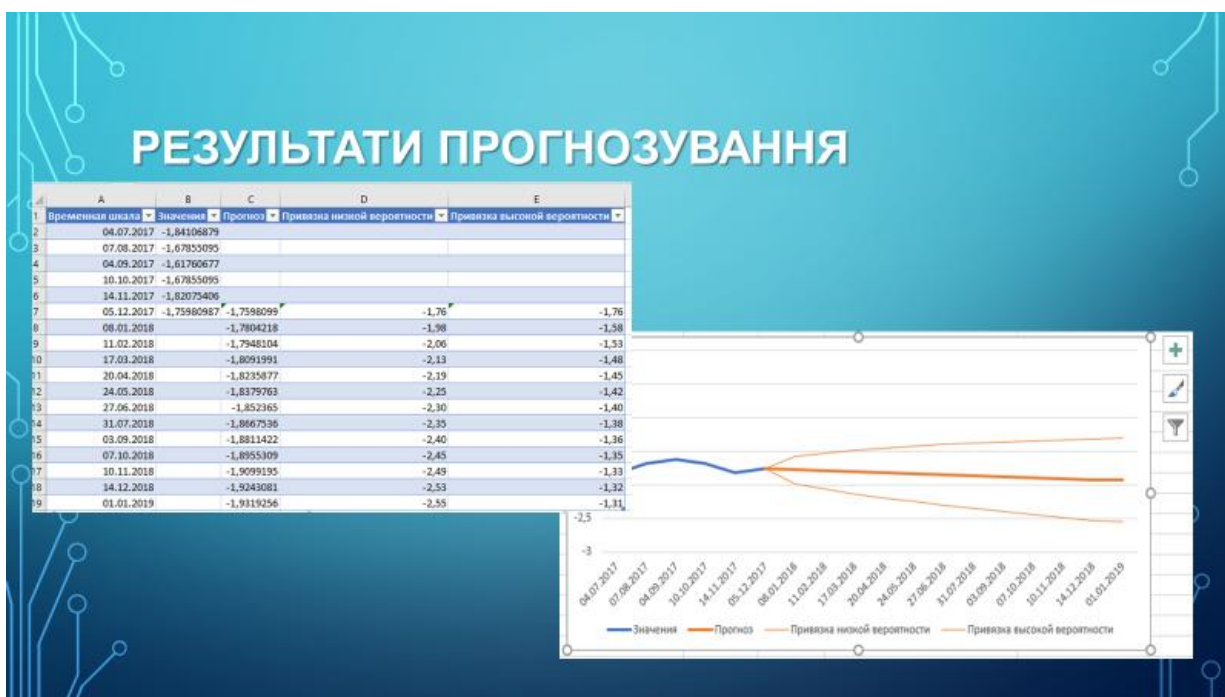


Рисунок А.13 – Результати прогнозування

ВИСНОВКИ

- В результаті роботи було:
 - Здійснено аналітичний огляд математичних алгоритмів та програмних засобів для об'єднання та прогнозування даних;
 - Розроблено математичний алгоритм для здійснення прогнозування та злиття даних;
 - Проведено розрахунки, щодо прогнозування випадків забруднення, використовуючи дані моніторингу річки Сіверський Дінець;
- В подальшому можливе вдосконалення алгоритму, шляхом урахування впливу сезонних факторів, швидкості течії, погодних умов, виявлення залежності між факторами та інших критерій для отриманні більш точних результатів прогнозування.

Рисунок А.14 Висновки