

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

Факультет інформатики та обчислювальної техніки

Кафедра інформаційних систем та технологій

«На правах рукопису»
УДК _____

До захисту допущено:
Завідувач кафедри
_____ Олександр РОЛІК
«__» _____ 20__ р.

Магістерська дисертація

на здобуття ступеня магістра

**за освітньо-науковою програмою «Інтегровані інформаційні
системи»**

зі спеціальності 126 «Інформаційні системи та технології»

**на тему: «Моделі дослідження розповсюдження
епідемічних захворювань»**

Виконав (-ла):
студент (-ка) VI курсу, групи ІА-01мн
Капченко Поліна Володимирівна _____

Керівник:
доцент каф. ІСТ, к.т.н., доцент
Репнікова Наталія Борисівна _____

Рецензент:
Комерційний директор ТОВ НВП «Мадек», к.т.н.
Коломейко Дмитро Анатолійович _____

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів без
відповідних посилань.
Студент (-ка) _____

Київ – 2022 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій
Рівень вищої освіти – другий (магістерський)
Спеціальність – 126 «Інформаційні системи та технології»
Освітньо-професійна програма «Інтегровані інформаційні системи»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олександр РОЛІК

«__» _____ 2022 р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Капченко Поліні Володимирівні

1. Тема дисертації «Модель дослідження розповсюдження епідемічних захворювань», науковий керівник дисертації Репнікова Наталія Борисівна к.т.н, доцент затверджені наказом по університету НС/88/2022 від « 26 » 04 2022р.
2. Термін подання студентом дисертації 08.06.2022
3. Об'єкт дослідження епідемічне захворювання
4. Предмет дослідження моделі дослідження розповсюдження епідемічних захворювань.
5. Перелік завдань, які необхідно розробити аналіз предметної області, огляд та аналіз існуючих рішень для моделей дослідження розповсюдження епідемічних захворювань, розроблення базової моделі, розроблення авторської моделі, розроблення автроської моделі з температурним параметром, програмна реалізація моделей у програмному пакеті
6. Орієнтовний перелік ілюстративного (графічного) матеріалу структурна схема базової моделі, графіки базової моделі з реальними даними,

порівняльна таблиця базової моделі та з інкубаційним періодом, графіки з моделлю з оптимізованими параметрами, діаграми середньомісячної температури повітря, графіки з даними інфікованих та температурні проміжки, графіки моделі з температурним параметром, результуюча таблиця методів

7. Орієнтовний перелік публікацій Репнікова Н. Б Вдосконалення функціональних можливостей моделі прогнозування епідемічних захворювань / Н. Б. Репнікова, П. В. Капченко - <https://www.inter-nauka.com/issues/2022/4/7973>

8. Консультанти розділів дисертації*

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

9. Дата видачі завдання 31.01.22

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Аналіз предметної області	01.02.22 – 04.02.22	
2	Огляд і аналіз існуючих рішень	07.02.22- 18.02.22	
3	Розроблення математичної моделі для базової моделі	21.02.22-11.03.22	
4	Розроблення вдосконаленої моделі	14.03.22 – 18.03.22	
5	Розроблення вдосконаленої моделі з температурним параметром	21.03.22-01.04.22	
6	Моделювання в MATLAB	04.04.22-22.04.22	
7	Оформлення текстової документації	28.04.22-24.05.22	
8	Подання дисертації до попереднього захисту	25.05.2022	

Студент

Поліна КАПЧЕНКО

Науковий керівник

Наталія РЕПНІКОВА

* Якщо визначені консультанти. Консультантом не може бути зазначено наукового керівника магістерської дисертації.

РЕФЕРАТ

Магістерська дисертація на здобуття ступеня «магістр» за освітньо-науковою програмою підготовки «Інтегровані інформаційні системи» на тему «Моделі дослідження розповсюдження епідемічних захворювань». Дисертація містить 105 сторінок, 45 рисунків, 19 таблиць, 1 додаток, 41 джерел.

Актуальність. Людство стикається з новими захворюваннями все частіше і частіше, останні роки трапляються часті епідемії або ж навіть пандемії. Зважаючи на нові віруси, моделі з вже відомими підходами не є досить точними. Тому задача вдосконалення точності моделі дослідження розповсюдження епідемічних захворювань є цікавою, а головне корисною.

Метою магістерської дисертації є підвищення точності базової моделі дослідження розповсюдження епідемічних захворювань.

Об'єкт дослідження: епідемічне захворювання.

Предмет дослідження: моделі дослідження розповсюдження епідемічних захворювань.

Наукова новизна одержаних у магістерській дисертації результатів полягає у вдосконаленні функціональних можливостей існуючої моделі дослідження розповсюдження захворювань, а саме – у реалізації логіки з оптимізованими параметрами та використанням температури як фактору впливу на хворобу.

Публікація результатів дисертації. За результатами роботи було опубліковано статтю: Репнікова Н. Б Вдосконалення функціональних можливостей моделі прогнозування епідемічних захворювань / Н. Б. Репнікова, П. В. Капченко - Міжнародний науковий журнал "Інтернаука". — 2022. — №4

Ключові слова: ВІРУСИ, МОДЕЛІ ДОСЛІДЖЕННЯ, ЕПІДЕМІЯ, МЕТОД НАЙМЕНШИХ КВАДРАТІВ, ПРОГНОЗУВАННЯ, MATLAB

ABSTRACT

Master's on the educational-scientific level training program “Master” and professional training program "Integrated Information Systems" on "Models for studying epidemic spread." The dissertation contains 105 pages, 45 figures, 19 tables, 1 appendices, 41 sources.

The topicality. Mankind is facing new diseases more and more often, in recent years there have been frequent epidemics or even pandemics. Due to new viruses, models with already known approaches are not accurate enough. Therefore, the task of improving the accuracy of the model of research on the spread of epidemic diseases is interesting and, most importantly, useful.

The aim of the master's dissertation is to increase the accuracy of the basic model of research on the spread of epidemic diseases.

The object of study: epidemic disease.

The subject of study: models of research on the spread of epidemic diseases.

The scientific originality of the results obtained in the master's dissertation is to improve the functionality of the existing model for studying the spread of diseases, namely - the implementation of logic with optimized parameters and the use of temperature as a factor influencing the disease.

Publication of dissertation results. Based on the results of the work,, an article was published: Repnikova NB Improving the functional capabilities of the model of forecasting epidemic diseases / NB Repnikova, PV Kapchenko - International scientific journal "Internauka". - 2022. - №4

Keywords: VIRUSES, RESEARCH MODELS, EPIDEMIC, LEAST SQUARES METHOD, FORECASTING, MATLAB

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	6
ВСТУП.....	7
1 АНАЛІЗ ПЕРДМЕТНОЇ ОБЛАСТІ.....	10
1.1 Загальні відомості про епідемію.....	10
1.2 Загальні відомості про COVID-19	12
2 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ДЛЯ МОДЕЛЕЙ ДОСЛІДЖЕННЯ РОЗПОВСЮДЖЕННЯ ЕПІДЕМІЧНИХ ЗАХВОРЮВАНЬ	18
2.1 Регресійні моделі	19
2.2 Байєсова мережа	20
2.3 Міркування на основі прецедентів	21
2.4 Штучні нейронні мережі.....	23
2.4 Аналітичні моделі розповсюдження	25
3 РОЗРОБЛЕННЯ МАТЕМАТИЧНОЇ МОДЕЛІ SIR та SEIR	29
3.1 Математична модель SIR.....	29
3.2 Реалізація SIR моделі	36
3.3 Моделювання MATLAB	38
3.3.1. Вибір програмного пакету для моделювання	38
3.4 Моделювання та аналіз	40
3.5 Модифікація моделі SIR - SEIR модель	50
3.6 Висновки до розділу.....	58
4 РОЗРОБЛЕННЯ МОДЕЛІ ПРОГНОЗУВАННЯ З ОПТИМІЗОВАНИМИ ПАРАМЕТРАМИ	59
4.1 Метод найменших квадратів.....	60

4.2 Метод Левенберга-Марквардта	63
4.3 Застосування методу Левенберга-Марквардта для SIR моделі	65
4.4 Вибір програмного забезпечення для розробки моделі	67
4.5 Програмна реалізація моделі з використанням методу Левенберг-Марквардта	70
4.6 Вплив температури на перебіг захворюваності	76
4.7 Математична модель для моделі розповсюдження з врахуванням температури	83
4.8 Моделювання моделі в програмному пакеті MATLAB	84
4.9 Висновки до розділу	90
5 ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛІ З ВРАХУВАННЯМ ТЕМПЕРАТУРИ ВІДНОСНО ВЖЕ ІСНУЮЧИХ МОДЕЛЕЙ	91
ВИСНОВКИ	99
ДОДАТОК А	106

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

COVID-19 – Corona Virus Disease 19 (короновірусна хвороба)

ПЛР – метод полімеразної ланцюгової реакції

ВООЗ – Всесвітня організація охорони здоров'я

SIR – «Susceptible-Infected-Recovered»

SEIR – «Susceptible-Exposed- Infected-Recovered»

НМ – нейронна мережа

США – Сполучені Штати Америки

CDC – Centers for Disease Control and Prevention

ВСТУП

За останні декілька років з новим вірусом тема епідемій стала широко обговорюваною серед всього світу. Адже багато людей стикнулись з цим лице в лице. Майже кожного року епідемії грипу з новими штамми поглинали світ, але з часом всі звикли до такого перебігу подій, а от з появою коронавірусної хвороби (COVID-19) почалась нова хвиля, більшість галузей мало адаптовуватись до нових реалій, так само науковці почали робити як можна більше всього аби можна було якісно оцінити процес розвитку розповсюдження вірусу та прогнозувати подальший перебіг подій. Відомі моделі дослідження розвитку епідемічних захворювань почали активно досліджуватись і до тепер продовжуються неодноразові спроби покращити їх.

З розвитком технології з'явилась можливість збирати дані щодо перебігу захворювань та проводити дослідження, щоб можна було забезпечити відповідним лікуванням, а також передбачити спалахи на основі зібраної інформації. Епідемічні прогнози виконуються для різних потреб і на різні часові проміжки, також і в залежності від цього служать різним цілям.

Для того, щоб успішно протистояти поширенню інфекції, необхідно аналізувати динаміку зростання захворюваності, розраховувати навантаження на охорону здоров'я та робити прогнози, щоб запобігти жахливим наслідкам захворювань. Все це потребує розробки реалістичної математичної моделі поширення епідемії.

Найбільш актуальними моделі дослідження розповсюдження вірусу є під час того, як лікування і вакцинація ще не винайдені для новітньої хвороби або певних нових штамів, адже фактично, єдиним способом врегулювання розповсюдження захворювань може відбуватись тільки на основі моделей прогнозування. Якщо ж моделі будуть мати велику похибку, це може призвести до неправильного врегулювання ситуації

Актуальність дослідження моделей розповсюдження епідемічних захворювань є гострою в наші дні, адже нова хвороба така як COVID-19 нікуди не зникла і епідемія далі тримає, також перебіг даної хвороби є незвичним і не схожим на вже відомі віруси, тому актуалізація моделей для дослідження розповсюдження епідемічних захворювань є досить перспективною і може допомогти багатьом сферам не тільки в подоланні захворювання, а й в економічних планах та іншою

Мета: підвищення точності базової моделі дослідження розповсюдження епідемічних захворювань

Для досягнення поставленої мети вирішені наступні **задачі:**

- огляд і аналіз предметної області та перебіг безпосередньо хвороби COVID-19;
- огляд і аналіз існуючих рішень для моделей дослідження розповсюдження захворювань;
- дослідження базової моделі та модифікованої моделі задля вибору початкової моделі для подальшої розробки;
- реалізація базової моделі за допомогою програмного пакету;
- виконання модифікації базової моделі для подальшого застосування для оптимізації параметрів;
- розроблення задачі найменших квадратів за допомогою методу Левенберг-Марквардта для моделі дослідження розповсюдження COVID-19;
- оптимізація параметрів за допомогою реалізації в програмному пакеті MATLAB
- дослідження моделі з оптимізованими параметрами
- розроблення та реалізація моделі з оптимізованими параметрами та залежність від температурного режиму
- дослідження моделі з застосуванням температурного режиму.

Для вирішення задач магістерської дисертації використані наступні методи:

- метод комп'ютерного моделювання;
- моделі дослідження розповсюдження захворювань;

- метод Левенберг-Марквардта;
- алгоритми та конструкції Matlab.

Об'єктом дослідження є епідемічне захворювання, а **предметом** - моделі дослідження розповсюдження епідемічних захворювань.

Наукова новизна одержаних у магістерській дисертації результатів полягає у вдосконаленні функціональних можливостей існуючої базової моделі дослідження розповсюдження захворювань, а саме – у реалізації логіки з оптимізованими параметрами та використанням температури як фактору впливу на хворобу.

Публікація результатів дисертації. За результатами роботи було опубліковано статтю: Репнікова Н. Б. Вдосконалення функціональних можливостей моделі прогнозування епідемічних захворювань / Н. Б. Репнікова, П. В. Капченко // Інтернаука – 2022. – Т

1 АНАЛІЗ ПЕРДМЕТНОЇ ОБЛАСТІ

1.1 Загальні відомості про епідемію

В наші дні, мабуть, всі хоч раз чули про епідемію, а також пандемію, особливо, що стосується COVID-19 викликаний новим вірусом SARS-CoV-2. Терміни «епідемія», «пандемія» у всіх на слуху, але що ж саме вони означають і в чому основна різниця, і чи взагалі є якась відмінність у цих поняттях.

Доцільно було б розпочати з тлумачення кожного з понять та ввести ще декілька нових понять, для краще заглиблення у предметну область.

Спалах — це коли хвороба виникає у несподівано великих кількостях. Він може залишатися в одній області або поширюватися ширше. Спалах може тривати дні або роки. Іноді фахівці вважають спалахом одиничний випадок інфекційного захворювання. Це може бути правдою, якщо це невідоме захворювання, якщо воно є новим для спільноти або якщо воно відсутнє в популяції протягом тривалого часу. Закінчення спалаху оголошується за умови, якщо пройшло два інкубаційних періодів і протягом цих днів, не було зафіксовано ні одного нового випадку. Якщо ж хвороба має нетривалий інкубаційний період, менше десяти днів, то закінчення спалаху визначають по закінченню трьох інкубаційних періодів.[1]

Епідемія – це коли інфекційне захворювання швидко поширюється на більше людей, ніж очікували експерти. Зазвичай воно вражає більшу площу, ніж спалах. До прикладу, у 2003 році епідемія важкого гострого респіраторного синдрому (ГРВІ) забрала життя майже 800 людей у всьому світі. [2] Тип захворювання, який є новим для даної популяції впродовж періоду збереження імунної «пам'яті» та поширюється зі швидкістю, що значно перевищує очікувану, ґрунтуючись на попередньому досвіді (тобто, числі нових випадків за одиницю часу). Українське законодавство визначає: «епідемія — масове поширення інфекційного захворювання серед населення відповідної території за короткий проміжок часу». [3], що не відповідає світовому визначенню, яке поширює поняття епідемії на усі випадки масового ураження, в тому числі й отруєння.

Пандемія — це глобальний спалах хвороби, що поширюється на країни чи континенти. Це вражає більше людей і забирає більше життів, ніж епідемія. Всесвітня організація охорони здоров'я (ВООЗ) несе відповідальність за оголошення про глобальну пандемію. ВООЗ робить це, відстежуючи спалахи захворювання та користуючись порадами міжнародних експертів у сфері охорони здоров'я. Однак Австралія та інші країни, ймовірно, зроблять кроки для зменшення впливу пандемії до того, як ВООЗ зробить офіційну заяву[4].

Він відрізняється від спалаху чи епідемії тим, що:

- впливає на більш широку географічну територію, часто по всьому світу;
- заражає більшу кількість людей, ніж епідемія;
- часто викликається новим вірусом або штамом вірусу, який тривалий час не циркулював серед людей, зазвичай немає імунітету проти нього, а сам вірус швидко поширюється від людини до людини по всьому світу;
- викликає набагато більшу кількість смертей, ніж епідемія;
- часто створює соціальні руйнування, економічні втрати та загальні труднощі.

Пандемія грипу 1918-1919 років забрала життя від 20 до 40 мільйонів людей. Це одна з найбільш руйнівних пандемій у зареєстрованій світовій історії. [5] Грип H1N1 у 2009 році став новою глобальною пандемією.

Кількість втрат протягом пандемії залежить від багатьох чинників, а саме :

- як швидко відбувається інфікування населення;
- які групи населення є в групі ризику;
- наскільки складно протікає захворювання;
- кількість заходів вжитих органами влади та їх ефективність.

В пандемії є певна кількість фаз, які визначенні ВООЗ , варіюється від фази 1 (низький ризик) до фази 6 (повна пандемія) :

- фаза 1: вірус у тварин не викликав жодних відомих інфекцій у людей;
- фаза 2: вірус тварин спричинив інфекцію у людей;
- фаза 3: є окремі випадки або невеликі групи захворювань у людей. Якщо хвороба поширюється від людини до людини, вона недостатньо широка, щоб викликати спалахи на рівні громад;

- фаза 4: хвороба поширюється від людини до людини з підтвердженими спалахами на рівні громади;
- фаза 5: хвороба поширюється між людьми в більш ніж одній країні одного з регіонів ВООЗ;
- фаза 6: принаймні ще в одній країні, що відрізняється від Фази 5, є спалахи на рівні громад.[6]

1.2 Загальні відомості про COVID-19

Коронавірусна хвороба (COVID-19) – це інфекційне захворювання, спричинене вірусом SARS-CoV-2. Перший випадок COVID-19 був зареєстрований 1 грудня 2019 року, а причиною став новий коронавірус, пізніше названий SARS-CoV-2.[7] Є припущення, що SARS-CoV-2 виник у тварини та змінився (мутував) та зміг викликати захворювання у людей. Раніше вже спостерігались кілька спалахів інфекційних захворювань, які були матованими і походили ці віруси, саме від птахів, свиней, кажанів та інших тварин, щоб стати небезпечними для людей. Дослідження тривають, і нові дослідження можуть виявити, як і чому коронавірус еволюціонував, щоб викликати пандемію.

Коронавіруси – це велике сімейство вірусів, які можуть викликати захворювання у тварин або людей. У людей відомі кілька коронавірусів, які викликали респіраторні інфекції, від звичайної застуди до інших важких захворювань, такі як близькосхідний респіраторний синдром (MERS-CoV) і важкий гострий респіраторний синдром (ГРВІ). Останній виявлений коронавірус викликає коронавірусну хворобу COVID-19. [7]

Більшість людей, інфікованих вірусом, відчувають респіраторне захворювання легкого та середнього ступеня тяжкості та одужують без спеціального лікування. Однак деякі серйозно захворіють і потребують медичної допомоги. Люди похилого віку та ті, хто має супутні захворювання, такі як серцево-судинні захворювання, діабет, хронічні респіраторні захворювання або рак, мають

більшу ймовірність розвитку серйозних захворювань. Захворіти на COVID-19 і серйозно захворіти чи померти може будь-хто в будь-якому віці.

Визначити чи це захворювання грип чи COVID-19 або ж застуда без спеціального тестування неможливо, адже симптоми є схожими (рисунок 1.1).

СИМПТОМИ	КОРОНАВІРУС	ЗАСТУДА	ГРИП
 Лихоманка	поширено	рідко	поширено
 Втома	іноді	іноді	поширено
 Кашель	часто (сухий)	помірний	часто (сухий)
 Чихання	ні	поширено	ні
 Ломота	іноді	поширено	поширено
 Нежить	зрідка	поширено	іноді
 Біль у горлі	іноді	поширено	іноді
 Діарея	рідко	ні	іноді у дітей
 Головний біль	іноді	рідко	поширено
 Задишка	іноді	ні	ні

Джерело: Всесвітня організація охорони здоров'я, центри контролю та профілактики хвороб

Рисунок 1.1 – Відмінність між COVID-19, застудою та грипом [8]

Існує декілька видів тестування, щоб виявити чи хворіє людина саме на COVID-19:

- полімеразна ланцюгова реакція (ПЛР) - це метод, який використовується для багатьох видів хвороб, крім COVID-19, він дозволяє знайти ДНК вірус

в генетичному матеріалі людини, найчастіше це слюна, також можуть брати мазок з носу;

- експрес тест – виявляє не вірус, а антитіла до нього, які виробляються аби протистояти цьому вірусу, проводиться на основі взяття крові з вени або ж пальця.

Експрес-тести виявляють в організмі людини не сам вірус, а антитіла, які можуть з'явитися в крові лише на 5-7 день з моменту прояву перших клінічних проявів, тож робити швидкий тест раніше немає сенсу.

ПЛР тести може показати чи людина має коронавірус тільки в перші дні. Основним недоліком ПЛР тесту також є те, що він може показати невірний результат, якщо не були дотримані спеціальні умови перед виконанням тесту (не можна пити ніяку рідину за дві години до тесту, також не можна їсти, палити та використовувати будь-які спреї та інше.) Перевагою даних тестів є те, що навіть безсимптомний пацієнт може зробити тест даним методом і отримати результат чи він хворіє, це буде доцільним, якщо людина контактувала з хворими і не може бути впевненою чи вона не є переносником хвороби (рисунок 1.2) .

Великий недолік є в тому, що можна зробити один з тестів не в коректний день і результати будуть невірними. Також, якщо людина вже перехворіла і зробила ПЛР тестування результат на коронавірус буде негативним. Результати ПЛР тестування потрібно чекати більше 12 годин, а ось експрес тест результат можна отримати через 10-15 хвилин. На початку епідемії COVID-19 тестів не було, також було досить багато бракованих тестів, які показували невірний результат, що сприяло швидкому зараженню.

Враховуючи неточність методів тестування основним фактором в запобіганні захворювання та уповільнити передачу – це бути добре проінформованим про хворобу та способи поширення вірусу. Захистити себе інших від зараження, тримаючись на відстані приблизно одного метра від інших, обов'язкового одягаючи належно підібрану маску та часто мити руки або використовуйте спиртовмісні засоби. Зробіть щеплення, коли настане ваша черга, і дотримуйтесь місцевих інструкцій.


**МІНІСТЕРСТВО
ОХОРОНИ
ЗДОРОВ'Я
УКРАЇНИ**

Різновиди тестування на COVID-19

	ПЛР-тест	Антиген-тест*	ІФА-тест
Визначає активну фазу COVID-19	✓	✓	✗
Скільки часу чинний	72 години після збору матеріалу	72 години після збору матеріалу	✗
Зразок матеріалу	Мазок із носа та рота	Мазок із носа або рота	Кров
Виявляє антитіла до COVID-19	✗	✗	✓
Дає підстави встановити діагноз «COVID-19»	✓	✓	✓
Підходить для роботи під час карантину	✓	✓	✗

* Є валідним для роботи в умовах карантину, поїздок у транспорті.

Рисунок 1.2– Різновиди тестування на COVID-19 [9]

Вірус може поширюватися з рота або носа інфікованої людини дрібними частинками рідини, тобто повітряно-крапельним шляхом інфікованої, коли вони кашляють, чхають, говорять, співають або дихають. Ці частинки варіюються від більших дихальних крапель до менших аерозолів. Важливо дотримуватися респіраторного етикету, наприклад, кашляти в зігнутий лікоть, а також залишатися вдома та самоізолюватися, поки не одужаєш, якщо відчуваєшся погано.

Ступінь тяжкості симптомів COVID-19 може варіюватися від дуже легкого до важкого. Деякі люди можуть мати лише кілька симптомів. Особливо небезпечною хвороба є через те, що частина людей може взагалі не мати симптомів, але все ще можуть поширювати їх (безсимптомна передача). Деякі люди можуть

відчувати погіршення симптомів, наприклад, посилення задишки та пневмонії, приблизно через тиждень після появи симптомів.

Деякі люди відчують симптоми COVID-19 більше чотирьох тижнів після того, як їм поставили діагноз. Ці проблеми зі здоров'ям іноді називають станами після COVID-19. Деякі діти відчують мультисистемний запальний синдром, синдром, який може вражати деякі органи і тканини, через кілька тижнів після захворювання на COVID-19. Рідко деякі дорослі також відчують синдром.

Люди старшого віку мають вищий ризик серйозних захворювань від COVID-19, і ризик зростає з віком. Люди з наявними захворюваннями також можуть мати більший ризик серйозних захворювань.

Важливою особливістю нового захворювання є інкубаційний період (рисунк 1.3), який може тривати від 2 до 14 днів. Під час інкубаційного періоду людина може вже заражати, але ще не відчувати симптомів, що має великий вплив на розповсюдження вірусу, адже під час інкубаційного періоду люди найчастіше не починають дотримуватись самоізоляції терміном на 14 днів, щоб убезпечити інших від зараження.

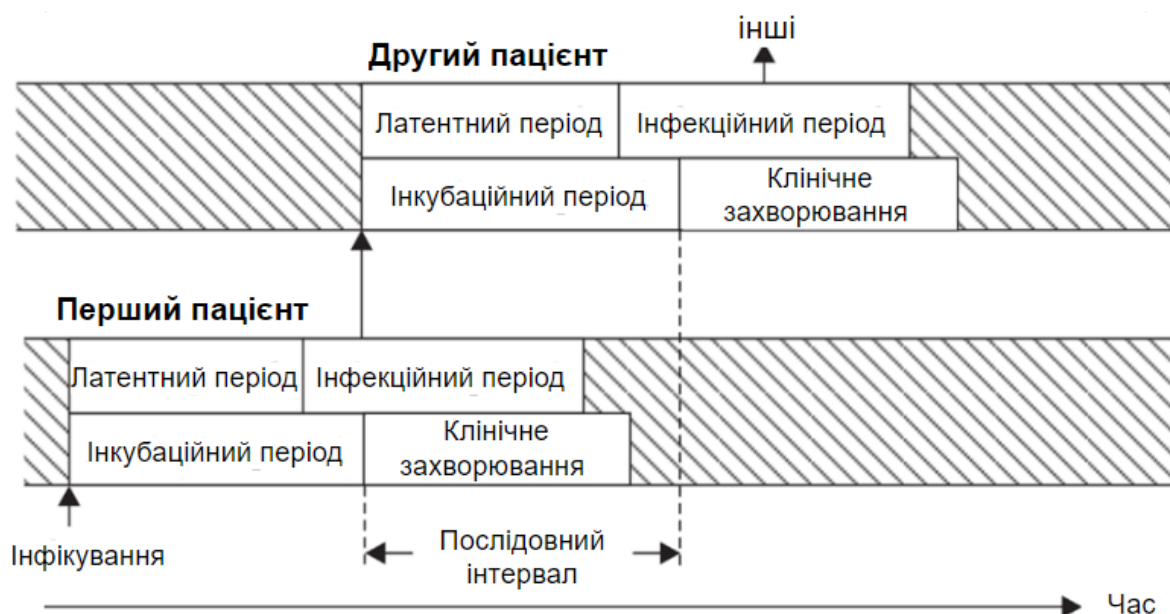


Рисунок 1.3 – Принцип зараження захворюванням COVID-19

Також, є інша особливість захворювання – це безсимптомне перенесення хвороби. Під час безсимптомного перебігу хвороби заражена особина може передавати вірус і не відчувати симптомів, натомість заражена людина від безсимптомного пацієнту буде не обов'язково також безсимптомною, тому що перебіг хвороби залежить від сукупності факторів, а особливо від імунітету пацієнта. То ж основне, якщо люди мали будь який контакт з захворівшими на COVID-19 і не відчують ще симптомів, то краще витримати карантин у 14 днів і симптоми можуть з'явитись під час нього або ж здати ПЛР тестування і пересвідчитись, що не заражені і не є переносниками хвороби.

З огляду на перелік симптомів можна сказати, що хвороба є досить підступною і небезпечною, адже інфекція може передаватись швидкими темпами і нести великі шкоди здоров'ю всьому населенню планети.

За допомогою моделей дослідження розповсюдження епідемічних захворювань можна прогнозувати подальший розвиток подій і тим самим приймати відповідні міри, щоб швидкість розповсюдження захворювання зменшувалась і тим самим менше людей постраждали від жахливих наслідків хвороби або навіть смертей.

2 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ДЛЯ МОДЕЛЕЙ ДОСЛІДЖЕННЯ РОЗПОВСЮДЖЕННЯ ЕПІДЕМІЧНИХ ЗАХВОРЮВАНЬ

Зміни довкілля, демографії людей, міжнародні подорожі, мікробна еволюція та руйнування закладів охорони здоров'я – все це сприяло зміні спектру інфекційних захворювань, з якими стикається світова спільнота. Існуючі методи нагляду за інфекційними захворюваннями та результати їх виконання недостатні для задоволення зростаючих потреб у профілактиці, виявленні та реагуванні виконавчих органів на ситуацію. Здатність прогнозувати перебіг епідемії забезпечить механізм для урядів і служб охорони здоров'я для своєчасного реагування на спалахи, що дозволить мінімізувати вплив хвороби і заощадити ресурси. Для багатьох інфекційних захворювань, передові технології спостереження та моделювання, створюють потенціал для прогнозування тимчасового та просторового ризику епідемій. У поєднанні з комунікаційними технологіями ці методи можуть надати важливі інструменти, які одночасно є економічно ефективними і своєчасними.

Розвиток систем нагляду за захворюваннями, епідеміологічного моделювання та інформаційних технологій показує, що системи раннього попередження є необхідними інструментами для боротьби з повторним виникненням і поширенням інфекційних захворювань. Незважаючи на те, що є багато даних, які використовуються в цих системах, доступні безкоштовно або за невисоку вартість, якість та доступність епідеміологічних даних надзвичайно різняться. Довжина та просторовий масштаб серій епідеміологічних даних є особливо важливими для дослідження моделей захворювання на довгостроковий термін, наприклад, рік або ж декілька років. Створено велика кількість підходів до прогнозування, і навряд чи можна охопити їх всі. Розглянемо деякі з методів прогнозування епідемічних ситуацій та спалахів.

2.1 Регресійні моделі

Типовим підходом до прогнозування розповсюдження захворюваності є регресійний аналіз, він є і найбільш популярним методом. Основною задачею регресії є знаходження оцінок невідомих параметрів та формуванні функціональної залежності між захворюванням та факторами, які її спричинили. Середи регресійних моделей виділяють декілька груп [10], а саме:

- неадаптивні моделі, які використовують всі існуючі дані, щоб для оцінок параметрів;
- адаптивні моделі, значення параметрів яких розраховується на основі змінного вікна спостережень.

Вид регресійної залежності для неадаптивної моделі, зазвичай, обирається вихід із властивості аналізованого тимчасового ряду. В окремих випадках можна застосовувати поліноміальні або ступеневі функції [11], але частіше модель повинна враховувати сезонний характер захворюваності. Для цих цілей використовують запропоновану Серфлінгом циклічну функцію виду:

$$\check{y}_t = \sum_{j=0}^v \alpha_j t^j + \sum_{j=1}^k (\beta_{2j-1} \sin \theta_j + \beta_{2j} \cos \theta_j), \quad (2.1)$$

де $\check{y}_t$ – оцінка захворюваності в момент часу t ;

α_j, β_j – параметри регресії, ступінь полінома v зазвичай рівні одиниці;

θ_j – лінійна функція часу t .

Також існує багато доповнень до даної моделі, якщо враховують ще додаткові параметри.

Неадаптивні моделі такого виду припускають, що сезонність щороку проявляється схожим чином. Якщо це не так, можливе застосування складніших ієрархічних моделей. Дані моделі враховують всю історію захворюваності, тому якщо деякі властивості захворювання змінились, то застарілі дані будуть не

актуальні для нової моделі, яка буде побудована, що є досить вагомою проблемою, якщо обирати саме цю модель.

Натомість адаптивні регресійні моделі відображають обмежений відрізок часу і мають бути досить чутливими аби показувати зміни на захворюваність відносно нових тенденцій. Адаптивні моделі найчастіше використовують для прогнозів на короткий термін не більше декількох місяців, що також може стати вагомою проблемою, якщо потрібний прогноз на довготривалий строк [12].

2.2 Байєсова мережа

Байєсова мережа або ймовірнісна орієнтована ациклічна графова модель – це графова модель, яка представляє набір випадкових змінних та їхніх умовних залежностей за допомогою орієнтовного ациклічного графу [13]. Мережу представляють у формі спрямованого графу, де вершини – це змінні моделі, а ребра є ймовірнісним залежностям між ними, які задаються за певними законами розподілу.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.2)$$

Перл у 1988 році стверджував, що байєсівські мережі є графічними моделями, які містять інформацію про причинно-наслідкові ймовірності між змінними і часто використовуються для допомоги при прийнятті рішень. Причинно-ймовірнісні зв'язки в байєсівській мережі можуть бути запропоновані експертами або оновлені за допомогою теореми Байєса та нових даних, які збираються.

Байєсова модель набуває великої популярності зараз і використовується у багатьох сферах. В розрізі епідеміології модель може представляти зв'язки між захворюваннями і симптомами (рисунок 2.1), що дає змогу використовувати для обчислення ймовірностей наявності різних захворювань за наявних симптомів.

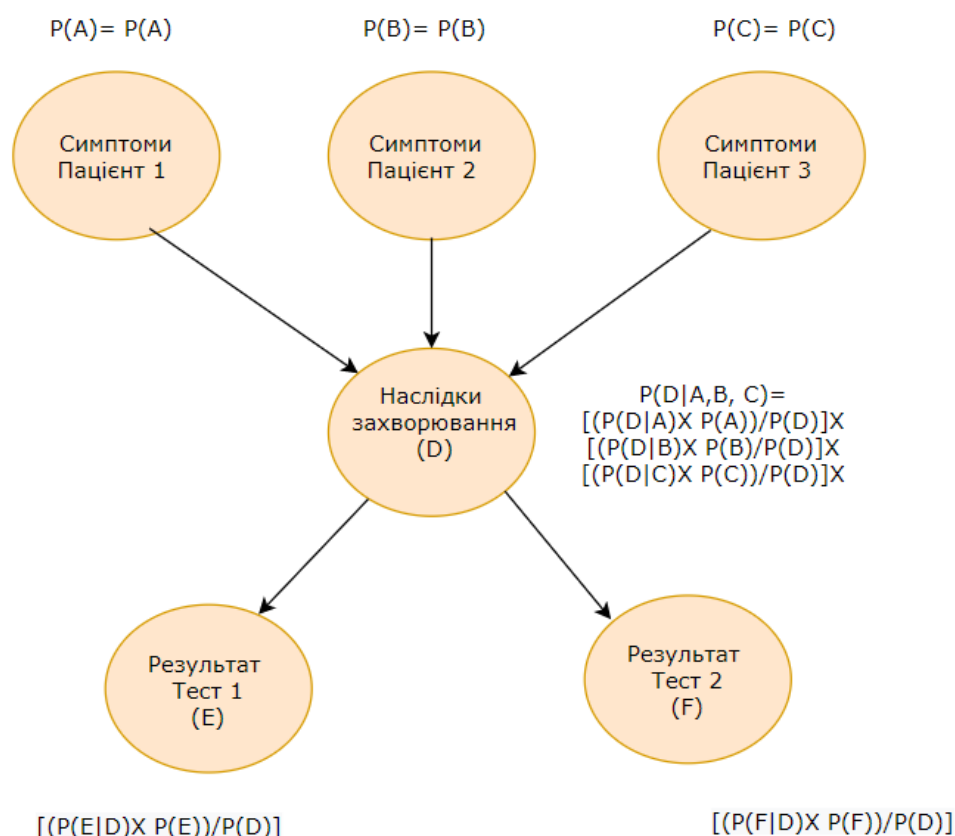


Рисунок 2.1 – Приклад Баєсової мережі та супутні умовні ймовірності

Модель можна будувати за наявності як при великій кількості вихідних даних, так і при малій. Дана мережа дозволяє робити тільки короткотривале прогнозування

2.3 Міркування на основі прецедентів

Міркування на основі прецедентів (case-based reasoning) вирішує нові проблеми шляхом отримання раніше вирішених проблем і повторного використання відповідних рішень. Специфіка підходу, заснованого на прецедентах, полягає в його зосередженні на невіддільності міркування від пам'яті та від навчання [14]. У термінології міркуванні на основі прецедентів, прецедент зазвичай позначає проблемну ситуацію. Ситуація, яка вже відбулась і була зафіксована, і засвоєна таким чином, щоб її можна було повторно використовувати

для вирішення майбутніх проблем, називають попереднім випадком або ж збереженим випадком. Відповідно, новий випадок або невирішений випадок є описом нової проблеми, яку потрібно вирішити. По суті, міркування на основі прецедентів — це циклічний та інтегрований процес вирішення проблеми, навчання на минулому досвіді, вирішення нової проблеми тощо[15].

Загальний життєвий цикл для вирішення проблеми за допомогою міркування на основі прецедентів в основному здійснюється в чотири етапи [16]:

- отримати: один або кілька прецедентів, подібних до нової задачі, витягуються з бази раніше вирішених прецедентів, і намагаються змінити їх, щоб вони відповідали новим параметрам проблеми;
 - повторне використання: рішення вибраних попередніх задач повторно використовуються для вирішення нової;
 - перегляд: вирішена нова проблема потім переглядається з фактичним рішенням;
 - зберегти: після успішної перевірки він додається до бази попередніх прецедентів для повторного використання для вирішення майбутніх проблем.
- Зокрема, продемонстрував, що є відповідною методологією для:
- робота з неструктурованими даними доменів або складна ситуація з отриманням знань, наприклад, формальні моделі або універсально застосовні рекомендації, погано розуміють багато захворювань;
 - системи служби підтримки, які використовуються в сфері обслуговування клієнтів для вирішення проблем з продуктами або послугами;
 - автоматичне здобуття суб'єктивних знань;
 - медична діагностика на основі подібності симптомів бази випадків з поточним випадком виберіть той, який найкраще підходить для постановки діагнозу.

Модель міркувань (рисунки 2.2) на основі прецедентів розрахована на короткотривалий прогноз і передусім передбачає, що мають враховуватись всі зовнішні фактори, які впливають на перебіг захворювання. Використання даного методу для прогнозування на середній термін або великий затребувало б

величезних масивів статистичних даних та більш складних метрик, які б враховували також і ефект старіння інформації і навіть, за наявності всіх метрик пошук близьких випадків буде досить складним і наявність хоч найменших відмінностей один від одного, буде суттєво впливати на точність прогнозування.

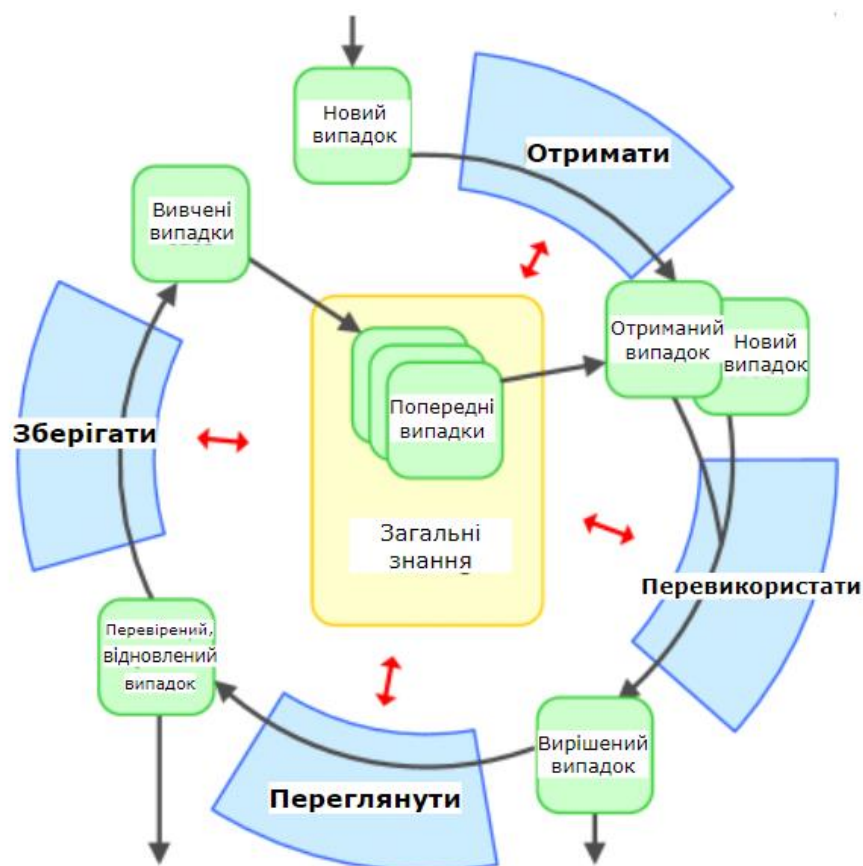


Рисунок 2.2 - Міркування на основі прецедентів

2.4 Штучні нейронні мережі

Штучна нейронна мережа (також називають ще нейронною мережею) — це адаптивна система, яка навчається за допомогою взаємопов'язаних вузлів або нейронів у багат шаровій структурі, що нагадує мозок людини. Нейронна мережа (НМ) може вчитися на даних, тому її можна навчити розпізнавати закономірності, класифікувати дані та прогнозувати майбутні події [17].

Нейронна мережа розбиває вхідні дані на шари абстракції. Її можна навчити на багатьох прикладах розпізнавати закономірності в мові чи зображеннях, наприклад, так само, як це робить людський мозок. Поведінка визначається способом з'єднання окремих елементів і силою або вагою цих зв'язків. Ця вага автоматично коригується під час навчання відповідно до заданого правила навчання, поки штучна нейронна мережа не виконає потрібне завдання правильно.

Методологія НМ широко відома і добре підходить для вирішення задач, де аналітичне дослідження є досить складним. НМ це спрямований зважений граф, вершини якого моделюють функціонування біологічних нейронів. Вершини приймають вхідні сигнали і за досить великому значенні їх вираженої суми перетворюють в вихідний сигнал. Навчання НМ полягає у обчисленні коефіцієнтів зв'язків між вершинами, що визначають силу вхідних сигналів, і виконується на основі емпіричних даних: статистики захворюваності та за наявності, значень факторів, що її визначають[18].

Допустимо безліч варіантів застосування НМ для прогнозування захворюваності: НМ можуть відрізнятися архітектурою, кількістю нейронів та видом функції їх активації, методом навчання. Формального підходу до вибору характеристик НМ не існує, у переважній більшості випадків використовуються мережі прямого розповсюдження сигналу та алгоритм зворотного поширення помилки (рисунок 2.3).

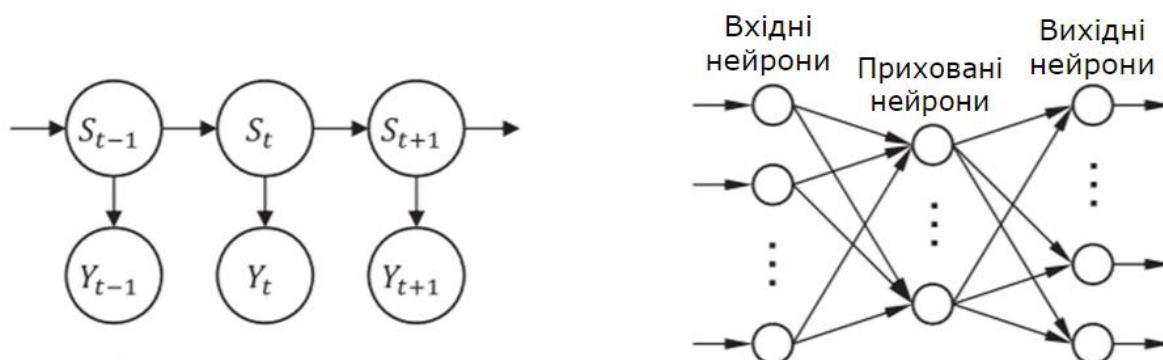


Рисунок 2.3 – Схема залежностей в прихованій марківській моделі ; схема НМ з одним прихованим шаром

Мета навчання НМ — визначення явних та неявних залежностей між вхідними та вихідними даними. Показники захворюваності, як правило, зашумлені, тому для навчання НМ потрібна відносно велика кількість даних. Достатню для довгострокового прогнозування потрібно накопичити велику кількість попередньої історії захворюваності, тому поки що НМ підходять швидше за прогнозування на малі терміни.

2.4 Аналітичні моделі розповсюдження

Одним з базових підходів для аналізу поширення інфекційних захворювань є класична модель SIR (рисунок. 2.4) , спочатку вивчена Хеймером у 1906 році , Россом в 1911 році та Кермаком і Маккендриком у 1927 році [19].

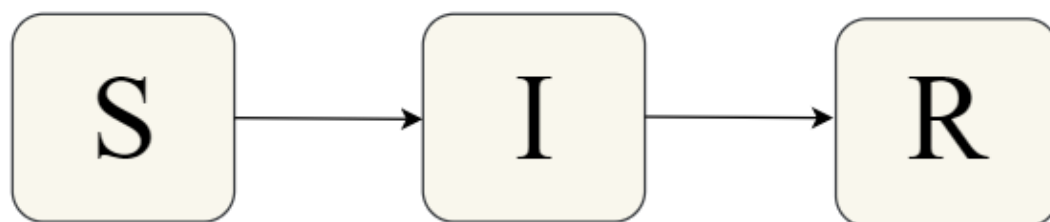


Рисунок 2.4 – Загальна модель SIR

Основне припущення цієї моделі полягає в тому, що популяція, в якій активний патогенний агент, є фіксованою і може бути поділена на три групи, що відповідають різному епідеміологічному статусу: сприйнятливі (від англ. Susceptible) індивідууми, здатні заразитися через контакт з інфікованими (від англ. Infected), а також одужали, які перестали поширювати хворобу (від англ. Recovered) і набули імунітету

Дана модель спирається на аналогію між контактами людей в великій популяції і законами діючих мас в хімічній кінетиці і кожна з груп представляється у вигляді диференційного рівняння [20].

В залежності від вирішуваної задачі модель може бути модифікованою і до неї додаються інші параметри, а також сама популяція може бути розділеною на більше частин, ніж три (рисунок 2.5). Наприклад, може враховуватись інкубаційний період. Кількість рівнянь може бути досить великою, так, наприклад в деяких роботах для моделювання демографічної структури регіону може використовуватись більше сотні або навіть тисячі рівнянь. [21]

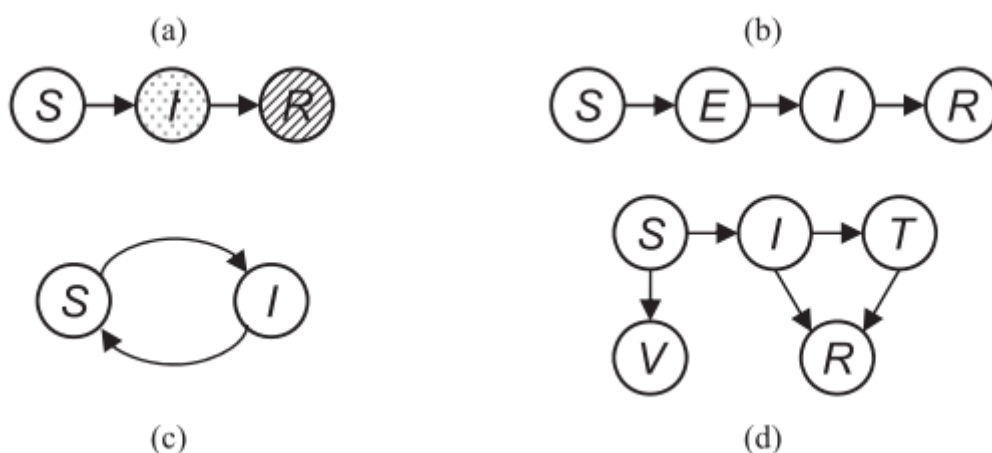


Рисунок 2.5 - . Схема моделі SIR (a); схема моделі SEIR (b); схема моделі SIS (c); схема моделі типу SIR з врахуванням вакцинованих індивідумів, та тих, хто знаходиться на лікуванні (T — treatment) (d)

Важливим питанням при моделюванні епідемій є просторові характеристики розподілу інфекції та вплив географічних факторів. До їхнього обчислення підходять по-різному, наприклад, в рамках методології SIR можна представити різні географічні точки, ввівши додатковий поділ на підгрупи. Тоді систему рівнянь включають механізми переміщення індивідів між територіями із заданою інтенсивністю. Також для цих цілей можуть застосовуватись гравітаційні моделі або їх елементи. Значною перевагою моделей, що базуються на апараті диференціальних рівнянь, є можливість їх аналітичного дослідження.

Також значна перевага в тому, що модель можна легко модифікувати, додаючи відповідні параметри, які деталізують перебіг захворювання або

розділяють популяцію на більше частин, що також може впливати на результати моделювання.

2.5 Висновки до розділу

Розглянуті моделі є досить популярними для дослідження епідемічної ситуації різних масштабів і різних хвороб, у Таблиці 1.1 перераховані основні методи прогнозування інфекційної захворюваності, розглянуті у даному розділі, та вказано кращі терміни побудови.

Таблиця 1.1 – Моделі дослідження розповсюдження захворювання

Назва моделі	Короткотривале прогнозування	Довгострокове прогнозування
Регресійні моделі		✓
Байєсова мережа	✓	
Міркування на основі прецедентів (випадків)	✓	
Нейронні мережі	✓	
Аналітичні моделі розповсюдження захворювання	✓	✓

З алгоритмічної точки зору між середньостроковими та довгостроковими прогнозами немає відмінностей, тому їх обчислення можуть застосовуватися одні й ті самі підходи. Слід зазначити, що поняття терміну прогнозування певною мірою умовне і залежить від кроку дискретизації часового ряду.

Як видно з таблиці, кількість методів, орієнтованих на малі терміни прогнозування, є трохи більшою, ніж кількість методів, які придатні до розрахунку проєкцій на тривалі терміни.

Існує безліч способів обчислення та застосування короткострокових прогнозів, також на цю тему є багато вже готових моделей з прикладами різних захворювань для різних географічних площ та різних груп населення, але довгостроковому прогнозуванню присвячені лише невелика кількість статей, наприклад [22].

Можливо, це пов'язано з тим, що якість короткострокових проєкцій є набагато вищою за довгострокові, а можливості їх уточнення ширші. Тим не менш, практична цінність середньострокових та довгострокових прогнозів не менша, а для багатьох завдань і більше, ніж короткострокових. Адже при прогнозуванні на тривалий термін можна робити більш далекоглядні оцінки ситуації по захворюваності, що дає змогу планувати всі необхідні запобіжні заходи для уникнення критичних ситуацій.

3 РОЗРОБЛЕННЯ МАТЕМАТИЧНОЇ МОДЕЛІ SIR та SEIR

3.1 Математична модель SIR

Модель, яку буде використано, називається моделлю SIR, де SIR означає «Susceptible-Infected-Recovered», тобто «Сприйнятливі – Інфіковані – Видужавши/Померлі» (рисунок 3.1). Це базова модель, яка досить часто зустрічається при моделюванні захворюваності різних захворювань. Дану модель прийнято обирати як базовою, так звану «точкою відліку», адже багато досліджень створено саме на її основі. То ж для початку і за основу було обрано саме цю модель, на базі якої буде створено декілька модифікацій.

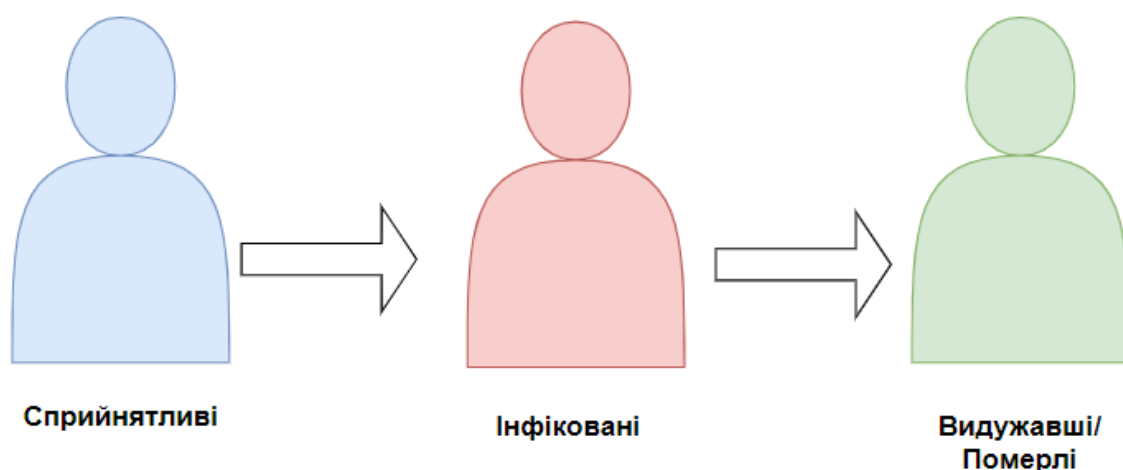


Рисунок 3.1 - Схема для «Сприйнятливі – Інфіковані – Видужавші/Померлі» моделі

Прийнято вважати, що всі люди поділяються на три групи, якщо розглядати в рамках однієї країни, то все населення це загальна сума для трьох груп, тобто будь-яка людина належить до однієї з цих груп. Нижче наведено короткий опис цих груп.

Перший набір залежних змінних підраховує людей у кожній із груп, кожна як функція часу:

$$S = S(t), \quad (3.1)$$

$$I = I(t), \quad (3.2)$$

$$R = R(t) \quad (3.3)$$

Сприйнятливі особи $S(t)$ - це люди, у яких немає імунітету захворювання, але вони не є заразними. Оскільки на початку захворювання зазвичай немає відповідної вакцини, яка б захистила населення від нової хвороби, можна сказати, що всі люди схильні до інфікування цією хворобою, а отже, «Сприйнятливий» відсік може бути представлений цілим населенням.

Людина з групи «Сприйнятливий» може рухатися на наступний рівень моделі «Інфікований» через контакт з інфікованою людиною. За допомогою цієї єдиної передачі, кількість сприйнятливих/інфікованих людей зменшується/збільшується на один, відповідно. «Інфікований» відповідає значенню $I(t)$.

Наступна група — для інфікованих осіб, які мають хворобу і можуть поширити її на сприйнятливих. Інфіковані люди можуть перейти до «Видужавших/Померлих» $R(t)$, видужаючи від хвороби або ж померти.

Група «Видужавших/Померлих» включає тих, хто більше не є заразним і ті, хто має, помирає від хвороби (закриті випадки). І сума цих трьох відсіків у моделі SIR залишається незмінною і дорівнює початковій чисельності населення [23].

Другий набір залежних змінних представляє частку загальної сукупності в кожній із трьох категорій. Отже, якщо N — загальна чисельність населення :

$$s(t) = \frac{S(t)}{N} \quad (3.4)$$

$$i(t) = \frac{I(t)}{N} \quad (3.5)$$

$$r(t) = \frac{R(t)}{N} \quad (3.6)$$

Також, зауважимо, що :

$$s(t) + i(t) + r(t) = 1 \quad (3.7)$$

Далі введемо деякі припущення щодо швидкості зміни залежних змінних. Ніхто не додається до сприйнятливої групи, оскільки ігнорується народження та імміграція. Єдиний спосіб покинути сприйнятливу групу – це заразитися. Припускається, що швидкість зміни $s(t)$, залежить від кількості вже сприйнятливих, кількості вже інфікованих осіб та кількості контактів між сприйнятливими та інфікованими [24].

Зокрема, припустимо, що кожна інфікована особа має фіксовану кількість β контактів на день, достатню для поширення хвороби. Не всі ці контакти відбуваються з уразливими особами. Якщо припустити однорідне перемішування популяції, то частка цих контактів, які є з сприйнятливими, дорівнює $s(t)$. Таким чином, в середньому кожна інфікована особина генерує $\beta s(t)$ нових інфікованих осіб на добу.

Також припускаємо, що фіксована частка γ інфікованої групи відновиться протягом будь-якого дня. Наприклад, якщо середня тривалість зараження становить три дні, то в середньому щодня одужує одна третина нині інфікованого населення. Власне кажучи, те, що маємо на увазі під «Інфікованим», насправді є «інфекційним», тобто здатним поширити хворобу на сприйнятливу людину. «Видужавша» людина все ще може почувати себе недокінця одужавшою, а може навіть померти пізніше від ускладнень після хвороби, тож в називаємо останню групу в моделі «Видужавші/Померлі».

Рівняння для «Сприйнятливої» групи має вигляд :

$$\frac{ds}{dt} = -\beta * i(t) * s(t) \quad (3.8)$$

Рівняння для «Видужавших/ Померлих» :

$$\frac{dr}{dt} = \gamma * i(t) \quad (3.9)$$

Рівняння для «Інфікованої» групи :

$$\frac{di}{dt} = \beta * i(t) * s(t) - \gamma * i(t) \quad (3.10)$$

Отже, за цими рівняннями можна зробити висновок, що

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0 \quad (3.11)$$

Базова модель SIR (рисунок 3.2), де β позначає швидкість зараження або швидкість передачі, а γ позначає швидкість відновлення або видалення. Взагалі кажучи, ці параметри (β , γ) не є постійними; вони є функціями розміру інфекційної та видужавшої/померлої групи [25].

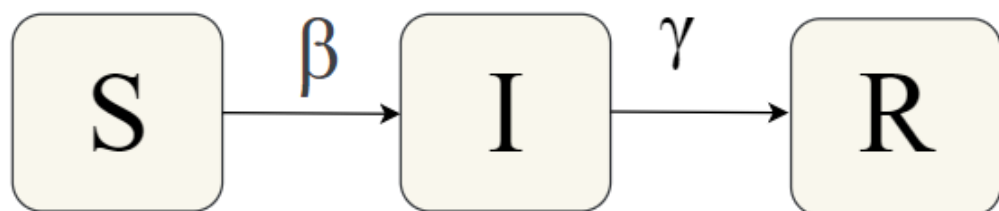


Рисунок 3.2 – Базова модель SIR з відповідними коефіцієнтами

Щоб розв'язати цей набір диференціальних рівнянь, нам потрібні початкові значення для змінних трьох станів S , I і R а саме $S(0)$, $I(0)$, і $R(0)$.

Ось, $R(0)$ можна встановити до нуля, так як на початку захворюваності кількість одужавших і померлих ще є невідомою. Модель SIR — це епідеміологічна модель, яка обчислює теоретичну кількість людей, інфікованих інфекційною хворобою, у закритій популяції за певний час.

Для простоти робимо декілька припущень:

1. Постійна (закрита) чисельність популяції, N ;
2. Постійні норми (наприклад, передача);
3. Відсутність демографії (тобто народження та смерті);
4. Добре змішана популяція(населення).

Добре змішана популяція — це та, де будь-яка інфікована особа має ймовірність контакту будь-яка сприйнятлива особа, яка досить добре апроксимується середнім показником. Це часто буває найбільш проблематичне припущення, але в більш складних моделях його вирішують.

У закритій популяції з N особин, що S є сприйнятливими, I інфікований, а R є видужавші/ померлі. Зауважимо, що N – кількість населення вірний такій рівності [26]:

$$N = S(t) + I(t) + R(t) \quad (3.12)$$

Важливим параметром моделі є R_0 . R_0 . – індекс репродукції – це середня кількість людей, які заразяться від однієї зараженої людини R_0 визначає можливість поширення епідемії в повністю сприйнятливій популяції за відсутності спеціальних заходів контролю. У повністю сприйнятливій популяції кількість нових інфекцій, викликаних індексним випадком, дорівнює ефективному коефіцієнту контакту, помноженому на середній період інфекції ($\beta \times 1Y$), який, за визначенням, дорівнює R_0 (рисунок 3.3).

$$R_0 \propto \left(\frac{\text{інфекція}}{\text{контакт}} \right) * \left(\frac{\text{контакт}}{\text{час}} \right) * \left(\frac{\text{час}}{\text{інфекція}} \right)$$

Рисунок 3.3 – Індекс зараження (репродукції)

В залежності від моделі, а також безпосередньо від захворювання коефіцієнт має різні значення, що суттєво може вплинути на результати моделювання. Також показник може відрізнятися для кожної країни, а також регіону.

Принцип передачі вірусу є досить небезпечним, адже кожна людина, наприклад при коефіцієнті $R_0 = 2$ може заразити дві людини і кожен вже заражений дві людини можуть заразити ще двох і так далі (рисунок 3.4). Аби кількість інфікованих людей зменшувалась потрібно зменшувати і сам показник, який може зменшуватись при меншому контакті людей.

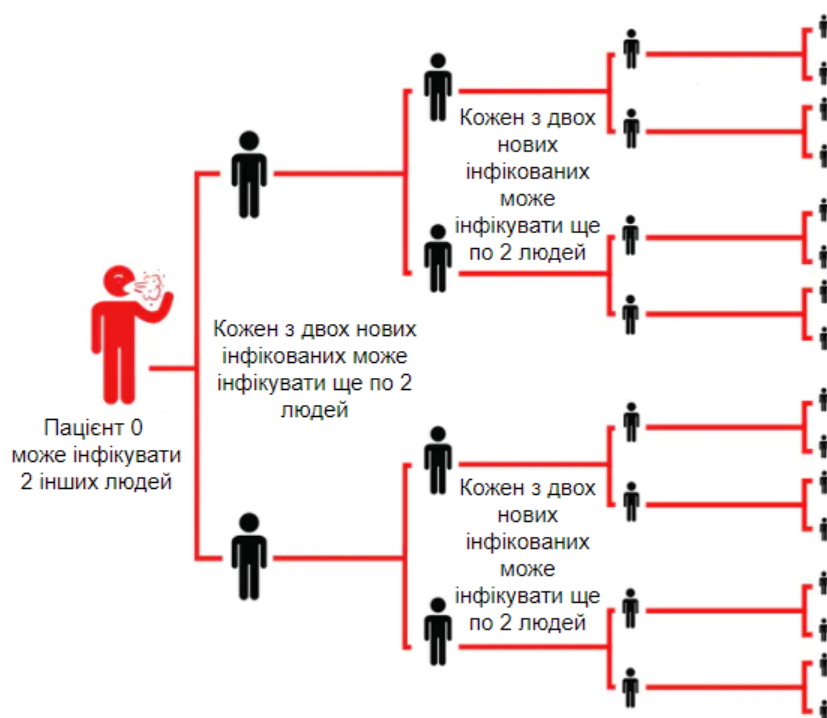


Рисунок 3.4 - Принцип передачі вірусу при $R_0 = 2$

Умови за яких хвороба стає епідемією. Епідемія виникає, якщо кількість заражених особин збільшується, тобто $\frac{dI}{dt} > 0$

$$\beta * s * i - \gamma * i > 0 \quad (3.13)$$

$$\frac{\beta * s * i}{\gamma} > 1 \quad (3.14)$$

На початку епідемії майже всі вразливі. Отже, ми можна сказати, що $S \approx 1$. Підставляючи $S = 1$, приходимо до наступної нерівності

$$\frac{\beta}{\gamma} = R_0 > 1 \quad (3.15)$$

Отже, якщо $R_0 > 1$, то це спалах епідемії

Якщо ж

$$\frac{\beta}{\gamma} = R_0 < 1 \quad (3.16)$$

То спалаху немає.

До прикладу, для різних хвороб значення R_0 можуть сильно різнитись. Для відомих всім епідемії маємо такі значення, які наведені у таблиці 3.1

Таблиця 3.1 - Значення R_0 для різних захворювань [27]

Хвороба	R_0
COVID-19	1.4 - 2.8
Кір	12 - 18
Дифтерія	6 - 7
Натуральна віспа	5 - 7

Продовження таблиці 3.1

Хвороба	R_0
Поліомієліт	5 - 7
Краснуха	5 - 7
Епідемічний паротит	4 - 7
ВІЛ/СНІД	2 - 5
Коклюш	5,5
Тяжкий гострий респіраторний синдром	2 - 5
Грип (штам 1918 року)	2 - 3
Ебола (2014 року)	1,5 – 2,5

Значення дійсні саме для базової моделі SIR, адже при подальших модифікаціях, даний параметр змінюється і може зростати або ж навпаки зменшуватись

3.2 Реалізація SIR моделі

Існує багато дослідження з моделюванням SIR, вони є для різних захворювань та на різну тривалість часу. Так як існує багато досліджень на короткотривалий період, спробуємо реалізувати прогноз на рік, тобто 365 днів. В даному випадку розглянемо модель на досить актуальну тему, а саме COVID-19, адже хвороба має нетиповий перебіг захворювання, тому потрібно також зауважити, що дана модель показує приблизні спалахи епідемії.

Оберемо період, який був найбільш цікавим через те, що це був самий початок епідемії, коли вже багато країн охопила дана хвороба, то ж відповідно і велика кількість даних була зібрана за той час, адже в цей період кількість захворювань почала стрімко збільшуватись по всьому світу і всі почали приймати термінові міри аби попередити масові захворювання і смерті на той час зовсім

невідомої хвороби, а саме початок пандемії – весна 2020 року. Дані були взяті з 13 березня 2020 року з CDC [28] .

Для того аби наглядно зрозуміти чи модель показує коректні значення було обрано країни в яких коронавірус розвивався досить швидко, а саме США та Іспанія, а також для того аби зрозуміти різницю з нашою країною до прикладу була взята ще й Україна.

Зауважимо, що кількість видужавших осіб за даними є невідомим на період з якого починається відлік для прогнозування, тому кількість одужавших і померлих осіб для трьох обраних країн буде дорівнювати 0, в представленій моделі.

Перший параметр, який необхідний для побудови – це кількість населення, адже популяція є незмінною і в той час кількість захворювання безпосередньо залежить від кількості населення. Популяція є змішаною, тому кількість населення буде відповідати кількості жителів відповідної країни і ніяк більше не буде поділятися для такого типу моделей. Так, як кількість населення в моделі розглядається як незмінний параметр, то маємо (Таблиця 3.2) :

Таблиця 3.2 – Кількість населення для обраних країн станом на 2020 рік

Країна	Кількість населення станом на 2020 рік
США	333 466 000
Україна	41 902 400
Іспанія	47 350 000

Кількість інфікованих станом на 13 березня 2020 року за даними епідемії наведено у Таблиці 3.3

Таблиця 3.3 – Кількість інфікованих станом на 13 березня 2020 року

Країна	Кількість захворівши
США	633

Продовження таблиці 3.3

Країна	Кількість захворівши
Україна	2
Іспанія	2955

Для моделі обов'язково треба вказати тривалість хвороби, адже в залежності від цього будуть змінювати результати прогнозування моделлю. Для COVID-19 тривалість хвороби може тривати зовсім різну кількість днів, особливо після того, як почали з'являтися все нові і нові штами. На початку було прийнято, що COVID-19 триває близько 10-14 днів, то ж і для моделювання буде обрана тривалість хвороби 14 днів, адже навіть після того як симптоми могли вже бути не такими явними, але тести на визначення хвороби все ще показували позитивні результати.

$$S(0) = N, \quad (3.17)$$

$$R(0) = 0. \quad (3.18)$$

$I(0)$ — кількість захворівших станом на 13 березня 2020 року в кожній обраній країні відповідно.

3.3 Моделювання MATLAB

3.3.1. Вибір програмного пакету для моделювання

Для побудови моделей обрано програмний пакет MATLAB. MATLAB — це багатоцільове програмне забезпечення з власною мовою програмування.

MATLAB створений для легкої обробки математичних операцій, наприклад як множення та додавання матриці, що можна легко це зробити за допомогою коду, без додаткових обчислень самотужки [29]. Цикли, які можна записати на C, C++ та інших мовах програмування, можна зручно записати в MATLAB, використовуючи навіть код, який є набагато коротшим і лаконічним. Написати визначені

користувачем функції в MATLAB надзвичайно легко, які можна зберігати в окремих файлах і запускати незалежно один від одного, що зручним, коли є багато операцій або функцій, які регулярно використовуються, і їх можна перевикористовувати, щоб не писати їх кожного разу.

Використовуючи MATLAB, можна імпортувати дані з файлу txt або Excel за допомогою однієї команди. В випадку даної роботи, це дуже зручно, адже можна імпортувати дані, які відповідають кількості інфікованих. MATLAB також дозволяє зберігати дані у файлі Excel або .mat. Зберігати дані у змінних дуже просто, оскільки є готові до використання команди, які зберігають числа у векторі або матриці без необхідності записувати цикл або будь-який інший додатковий код.

На додаток до програмування, MATLAB є сумісним з численними інструментами, створеними для конкретних застосувань, таких як підгонка кривих, моделювання та моделювання динамічних систем, цифрова обробка та багато інших застосувань в інженерії та інших сферах, до прикладу, як в роботі, буде застосовано до епідеміології. Цими інструментами можна керувати за допомогою вбудованих кодів, тому не доведеться керувати моделлю вручну.

Підсумовуючи, на відміну від звичайних програм, які можуть бути написані іншими мовами, MATLAB пропонує швидкий спосіб виконання обчислювального аналізу та математичних операцій разом із багатьма корисними інструментами.

В роботі маємо три звичайних диференціальне рівняння (ODE), які містить одну похідну від залежної змінної y відносно однієї незалежної змінної t , яку називають часом.

Методи для розв'язання диференціальних рівнянь у MATLAB мають широкий спектр використання в інженерії та науці. Існують методи розв'язання для звичайних диференціальних рівнянь, поставлених або як задачі на початкове значення, або як крайові задачі, диференціальні рівняння із затримкою та диференціальні рівняння часткових похідних. Крім того, існують функції для інтеграції функціональних виразів за допомогою квадратури або для чисельної інтеграції дискретних наборів даних.

Для вирішення рівнянь, які маємо для моделі SIR MATLAB пропонує декілька рішень `ode45`, `ode23`, `ode113`, `ode78`, `ode89`. Для більшості рівнянь, `ode45` працює найкраще. Однак `ode23` рекомендується для випадків, які допускають дещо більш грубу толерантність до помилок або за наявності помірної жорсткості. Аналогічно, `ode113` може бути ефективнішим, ніж `ode45`, для проблем з більш жорсткими допусками помилок або коли функція ODE є проміжним для обчислення та для оцінки. `ode78` і `ode89` — це вирішувачі високого порядку, які відрізняються тривалими інтеграціями, де точність має вирішальне значення для стабільності. Для обчислень обрано `ode45` [30].

3.4 Моделювання та аналіз

Для моделювання з використанням MATLAB був обраний такий алгоритм:

- створення змінних для коефіцієнтів швидкості розповсюдження захворювання та одужання;
- створення окремого файлу для вирішення трьох диференціальних рівнянь для кожної групи моделі відповідно;
- виклик функції та застосування `ode45` для обрахунку диференціальних рівнянь у результуючому файлі;
- запис методу для побудови графіків.

Рівняння задаються в окремому файлі у вигляді функцій аби можна було використати у подальших розрахунках, що є дуже зручним. Після створення цього файлу його можна просто «викликати» вписуючи тільки назву функції.

Називаємо функцію `ode_fun_simple_new` (рисунок 3.5) і передаємо туди вхідні параметри, а також ряд додаткових потрібних для обчислень даних, в випадку з COVID-19 для побудови нашої моделі обрали період захворюваності 14 днів, в коді за це значення відповідає змінна *Duration*. Також було передано на вхід t , y , які будуть використані для вирішувача диференціальних рівнянь (рисунок 3.6). На вхід передається змінна β , що відповідає коефіцієнту β . Значення Y в

кодi позначено як r , назву змiнною було обрано iншою через те, що синтаксис програмного пакету не передбачає деяких позначень, таких як ця змiнна.

Всi початковi значення задаються згiдно реальних даних .

```
function dydt = ode_fun_simple_new(t, y, beta)
    Duration = 14;
    r = 1/Duration;

    S = y(1);
    I = y(2);

    dS = -beta*I.*S;
    dI = beta*I.*S - r*I;
    dR = r*I;

    dydt = [dS;dI; dR ];

end
```

Рисунок 3.5 – Функція для обчислення диференціалів

```
%% Differential equations
tspan = 0:1:365; % Observe the next year
y0 = [N-633, 633, 0]; % Coronavirus count of USA mar13, 2020 for SIR
[t,y]=ode45(@(t,y) ode_fun_simple_new(t,y,beta), tspan, y0);
```

Рисунок 3.6 – ode45 для SIR моделі США та початкові дані

Країни для моделювання обрані в попередніх підрозділах, а також за наданою інформацією точно параметри та межі в яких мають коливатись всі параметри. Для R_0 відомі тільки проміжки і точно для кожної країни на період моделювання є невідомим.

Для початку оберемо для моделювання параметр по середині наведеного проміжку, адже країни, які розглядаються мали досить велику швидкість розповсюдження і найменший параметр скоріше за все не підійде, то ж параметр $R_0 = 2$

Результати моделювання для США, України та Іспанії показані на рисунках (рисунок 3.7- 3.9)

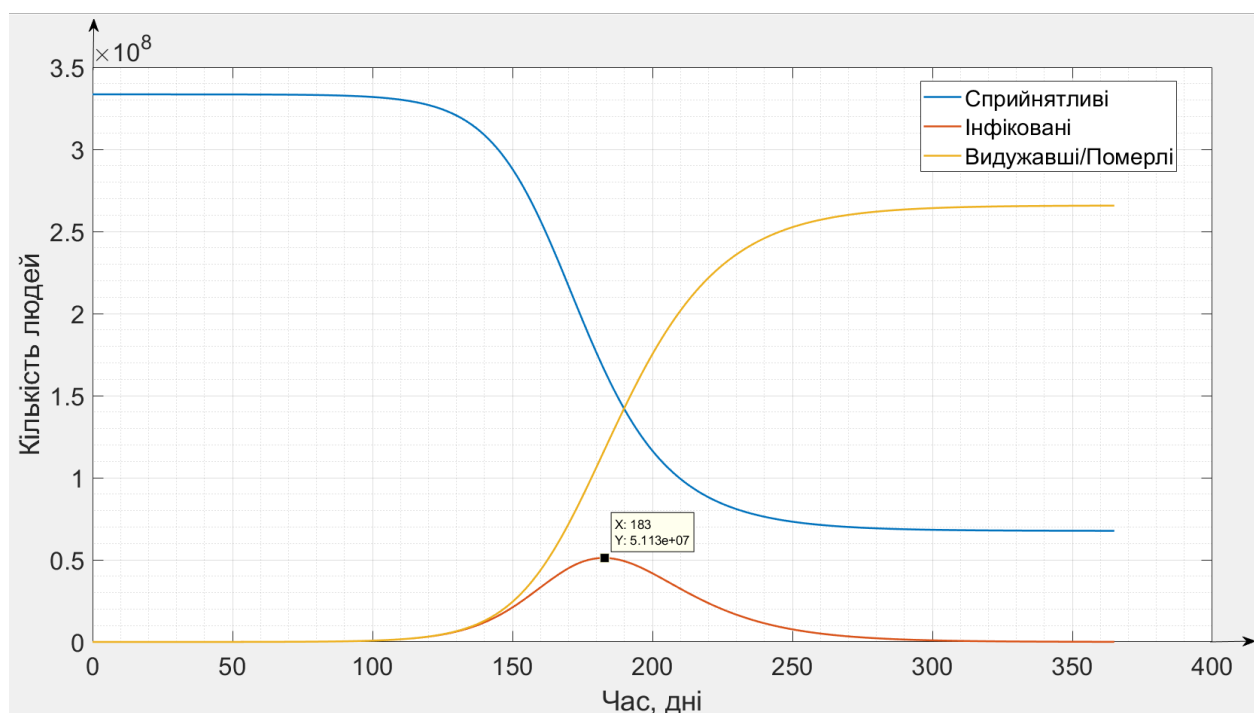


Рисунок 3.7 - SIR модель для ковіду в США

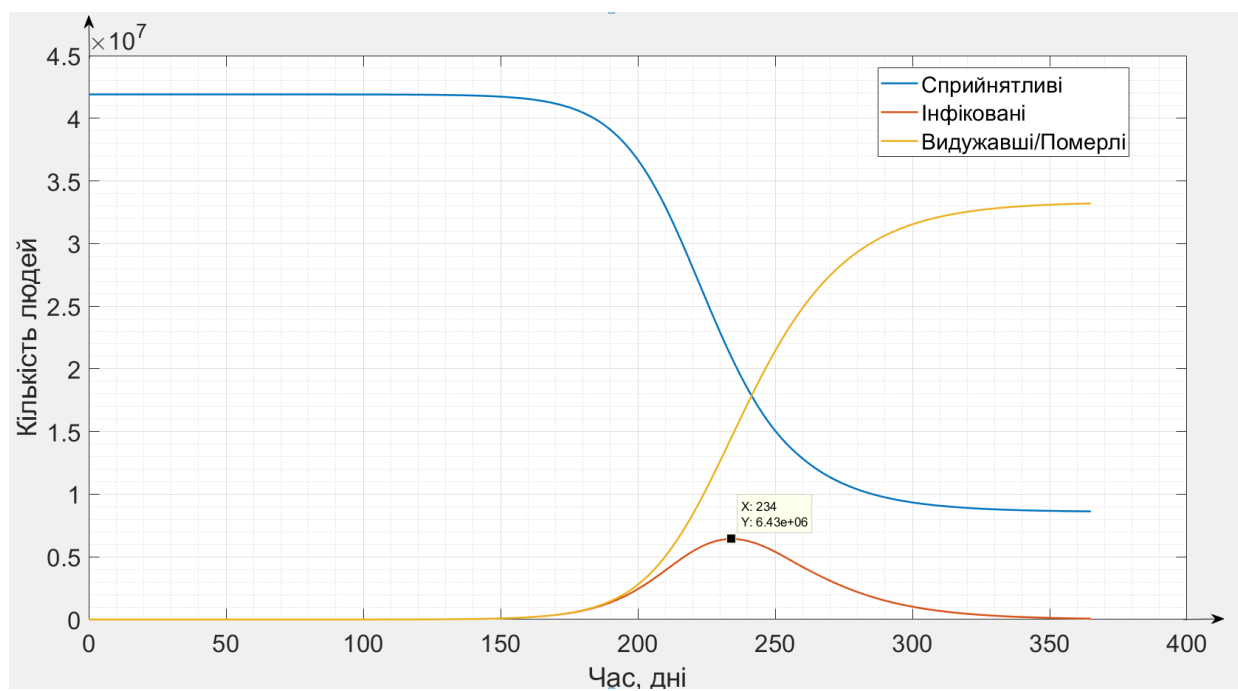


Рисунок 3.8 - SIR модель для ковіду в Україні

По осі абсцис – дні року від 1 до 365 починаючи від 13 березня 2020 року . По осі ордината – кількість населення

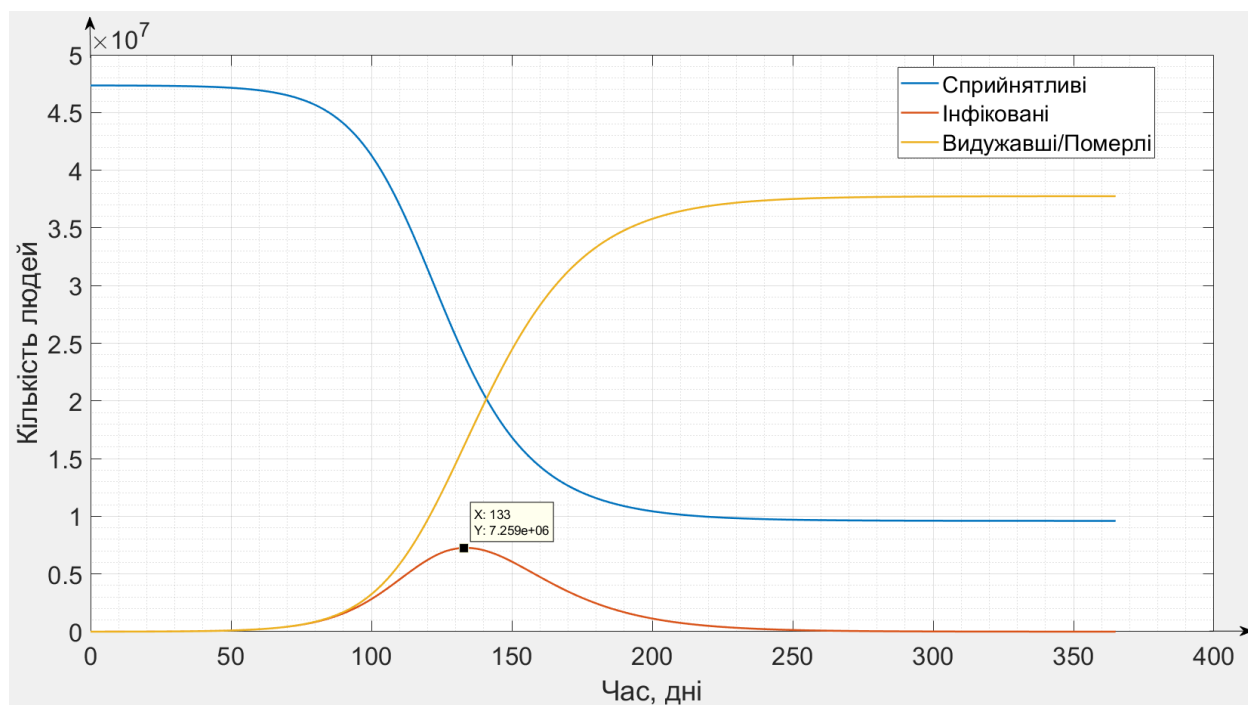


Рисунок 3.9 - SIR модель для ковіду в Іспанії

Виходячи з моделювання за допомогою MATLAB , можна зробити висновок, що модель SIR передбачає пік захворюваності на 183 день у США, на 234 день в Україні та на 133 день в Іспанії.

Результати з кількістю захворівших на спалаху будуть занесені у таблицю 3.4

Таблиця 3.4 – Результати моделювання SIR

Країна	День піку захворюваності	Кількість захворівших
США	183	51 130 000
Україна	234	64 30 000
Іспанія	134	72 61 000

Як видно рисунків, модель коректно відображається кількість уразливих осіб до хвороби, а також в залежності від введених початкових даних про інфікованих осіб відповідно до заданої моделі.

Піки в представлених країнах є різними і кількість інфікованих осіб теж, але це тільки теоретично змодельовані дані. Так, як вже відомо в моделі прийнято, що кількість населення не змінюється і немає впливу коефіцієнт народжуваності та ще багато інших показників можна припустити, що модель відображає не зовсім повну картину по розвитку розповсюдження хвороби.

Для того аби пересвідчитись в вірності моделювання порівняємо з реальними даними. Для того аби відобразити в моделі реальні дані, було використано команду для зчитування файлу та відповідно збережені до масиву. В MATLAB, щоб відобразити графіки, які мають різні межі на осі ординат на одному графіку, потрібно відобразити їх як дві окремих вісі. Розділення на дві окремі вісі ординат можна виконати за допомогою ууахіс команди, щоб обидва графіки були зображені коректно. Спрогнозовані дані зображені звичайною кривою, а реальні за стовпчикою діаграмою за допомогою команди *bar* (рисунок 3.10 -3.12).

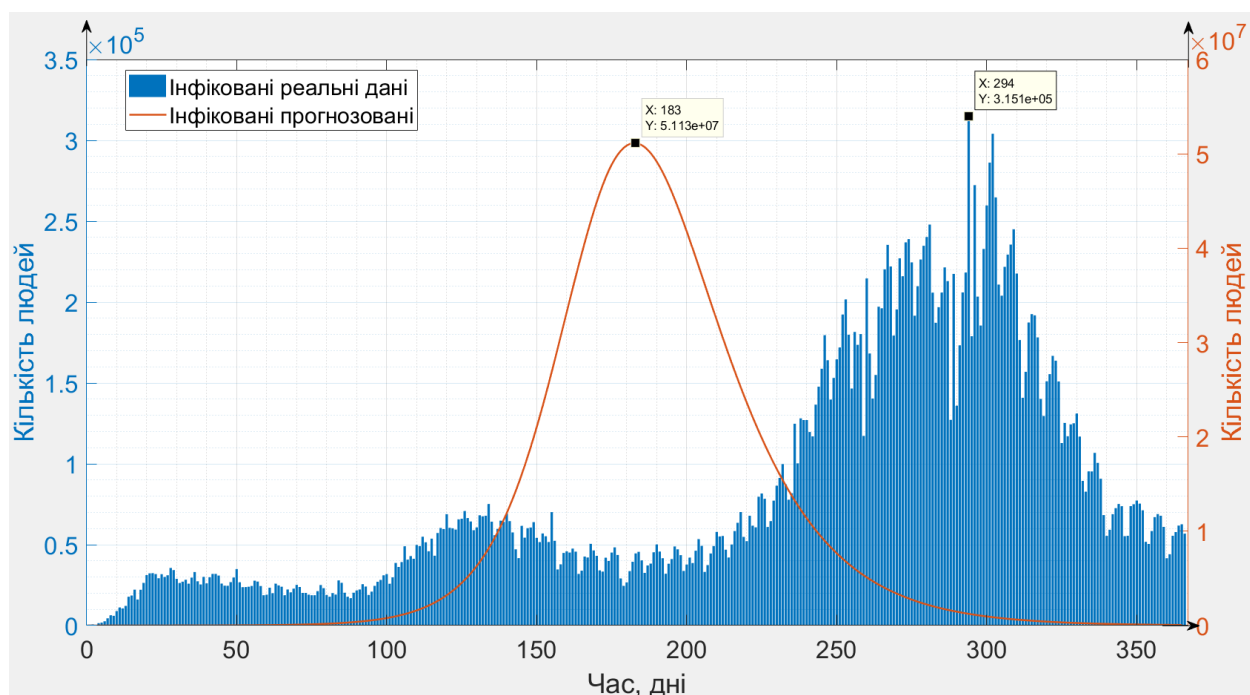


Рисунок 3.10 – Інфіковані особи прогнозовані та реальні для США

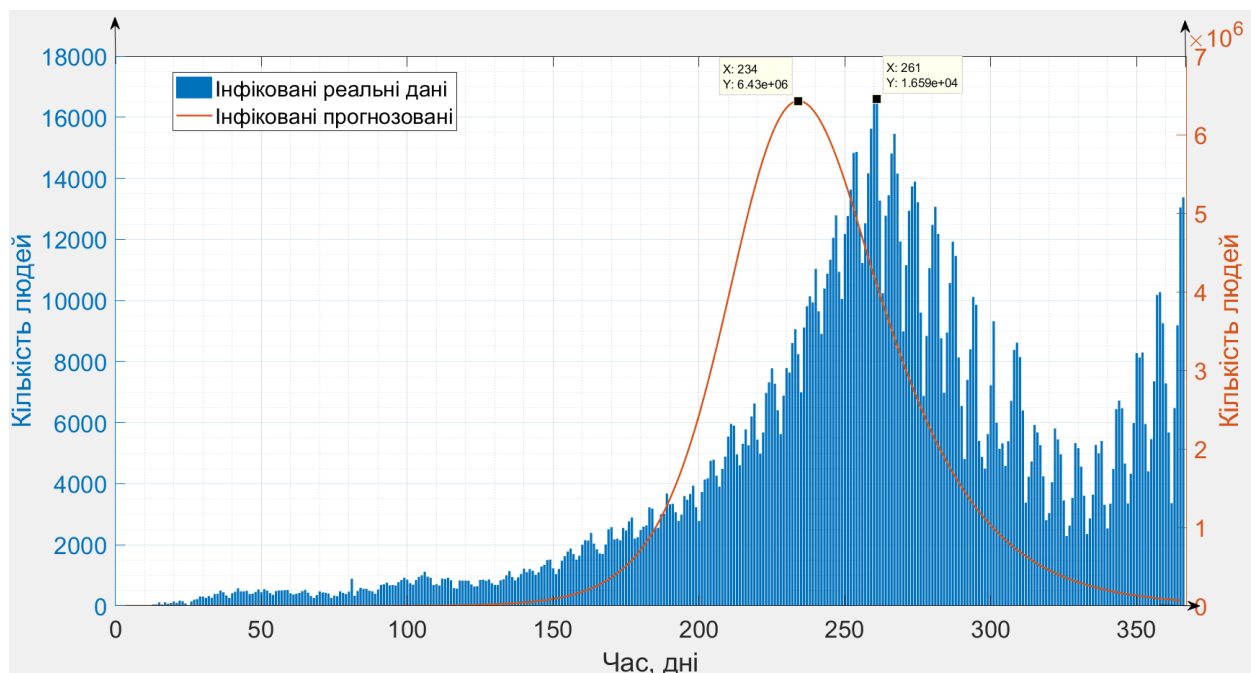


Рисунок 3.11 – Інфіковані особи прогнозовані та реальні для України

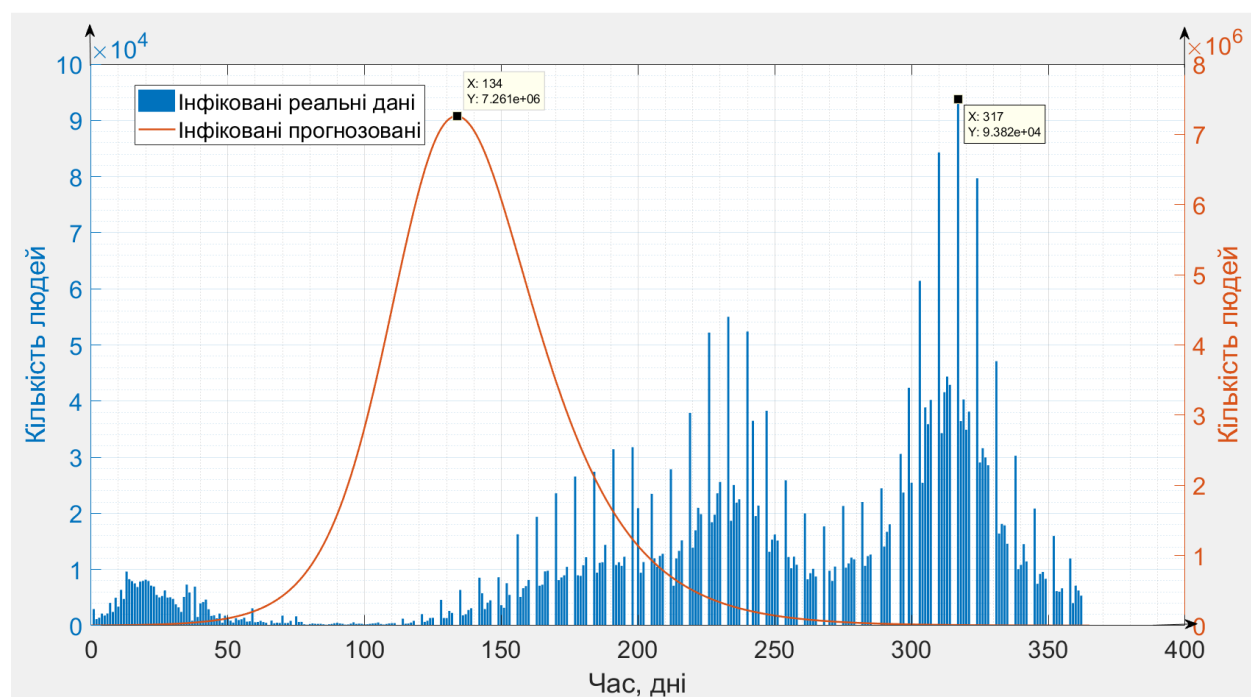


Рисунок 3.12 – Інфіковані особи прогнозовані та реальні для Іспанії

Результати порівняння прогнозування і реальні дані відносно кількості захворівших наведено у таблиці 3.5

Таблиця 3.5 - Результати моделювання базової моделі

Країна	Кількість захворівших на піку (прогнозований)	Кількість захворівших на піку (реальні)	Похибка між реальними і прогнозованими даними
США	51 130 000	315 100	50 814 900
Україна	6 430 000	16 590	6 413 410
Іспанія	72 61 000	93 820	7 167 180

Виходячи з результатів моделювання, можна сказати, що похибка в моделі досить велика відносно реальних даних, адже різниця, до прикладу в США становить більше 50 мільйонів людей чи можна зробити висновок, що модель є недостовірною і має велику похибку

Суттєвим недоліком моделі SIR є те, що безліч параметрів не враховується у цій моделі. Дана модель є базовою, тут не розглядається ймовірність вакцинації населення, тобто колективний імунітет був придбаний природним шляхом, до того ж, за моделлю SIR, до одужавших відносяться люди, які померли від даного захворювання. Це є одним із мінусів даного моделювання.

Базова модель SIR є відправною точкою для розробки більш складних моделей, що включають такі характеристики, як демографічні групи з різними ризиками для здоров'я, вплив заходів від громадських служб охорони здоров'я, природні показники народжуваності та смертності та вплив випадковості.

Також можна побачити, що не для всіх країн пік збігається навіть приблизно, що може свідчити про те, що параметр R_0 був обраний некоретно для цих країн, щоб визначити, який параметр буде коретний, спробуємо змодельовати для США модель SIR з параметрами наведеними у таблиці, а також одразу буде занесено результат моделювання на момент дня піку.

В разі того, якщо значення не співпадають потрібно збільшувати чи зменшувати коефіцієнт, прийнято обирати з кроком 0.1. В даній моделі зрозуміло, що якщо графік потрібно рухати праворуч коефіцієнт зменшують на один або

декілька кроків, тобто хвороба розповсюджується не так швидко, якщо ж ліворуч, тобто швидкість розповсюдження більша, то коефіцієнт збільшується.

Аби зафіксувати всі результати і оцінити чи вийшло підібрати потрібний коефіцієнт, потрібно записати дані. Після моделювання з різними коефіцієнтами отримані результати були занесені до таблиці 3.6

Таблиця 3.6 – Результати моделювання для США при різних R_0

R_0	День піку реальний	Кількість реально інфікованих	День піку змодельований	Кількість інфікованих змодельованих	Помилка в к-сть людей
2	294	315 100	171	51 130 000	50 814 900
1.4	294	315 100	-	14 740 000	-
1.6	294	315 100	286	27 080 000	26 764 900

Після моделювання з різними коефіцієнтами день піку вийшло спрогнозувати як можна ближче на 286 день при коефіцієнті (рисунок 3.13). Для коефіцієнту 1.4 пік епідемії вийшов за межі часу моделювання, тому пік є невизначеним

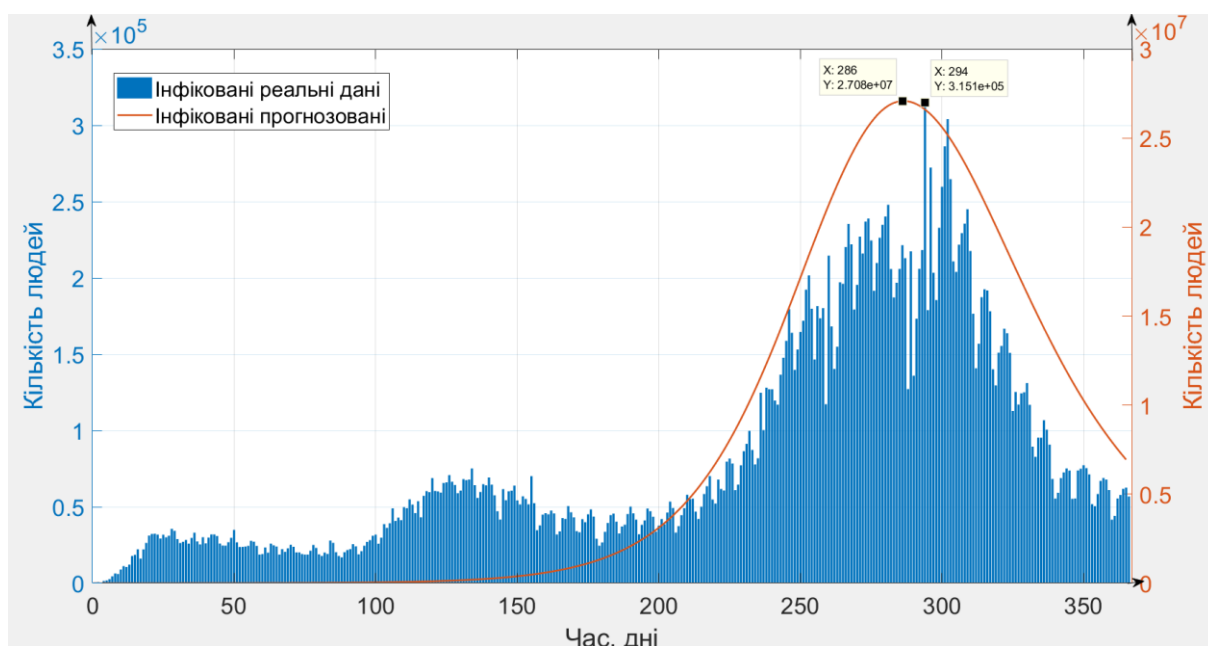


Рисунок 3.13- Моделювання для США при $R_0=1.6$

Для України при підборі коефіцієнту R_0 отримано результати, які наведені в таблиці 3.7.

Таблиця 3.7 – Результати моделювання для України при різних R_0

R_0	День піку реальний	Кількість реально інфікованих	День піку змодельований	Кількість інфікованих змодельовані	Помилка в к-сть людей
2	261	16 590	234	6 430 000	6 413 410
1.4	261	16 590	-	-	-
1.9	261	16 590	257	569 300	552 710

При моделюванні зі значенням $R_0 = 1.4$ графік вийшов за межі обраного періоду моделювання і так само як і для США пік не був зафіксований пік захворюваності протягом року. Найближче значення при моделюванні вийшло на 257 день (рисунок 3.14)

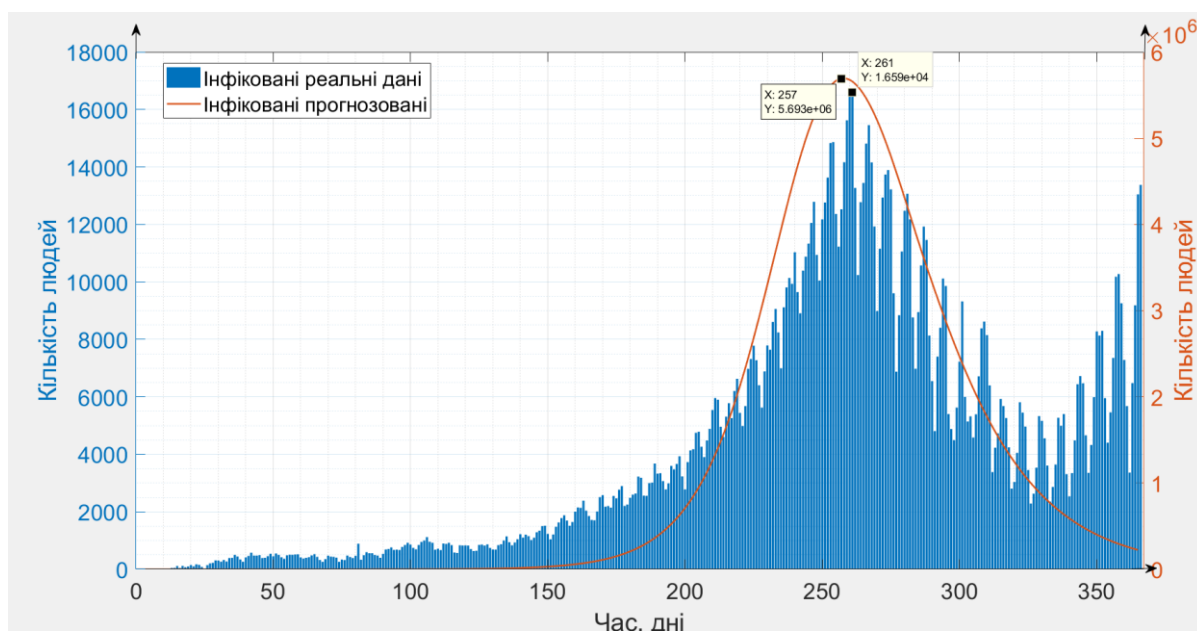


Рисунок 3.14 – Моделювання SIR для України при $R_0 = 1.9$

Також виконаємо моделювання підбираючи потрібний коефіцієнт i для третьої обраної країни – Іспанії, та запишемо результати моделювання до результуючої таблиці 3.8.

Таблиця 3.8 - Результати моделювання для Іспанії при різних R_0

R_0	День піку реальний	Кількість реально інфікованих	День піку змодельований	Кількість інфікованих змодельовані	Помилка в к-сть людей
2	317	93 820	234	72 61 000	7 167 180
1.4	317	93 820	284	2 151 000	2 056 180
1.3	317	93 820	354	1 372 000	1 278 180

Після моделювання найбільш близький пік до реального отримали на 284 день (рисунок 3.15)

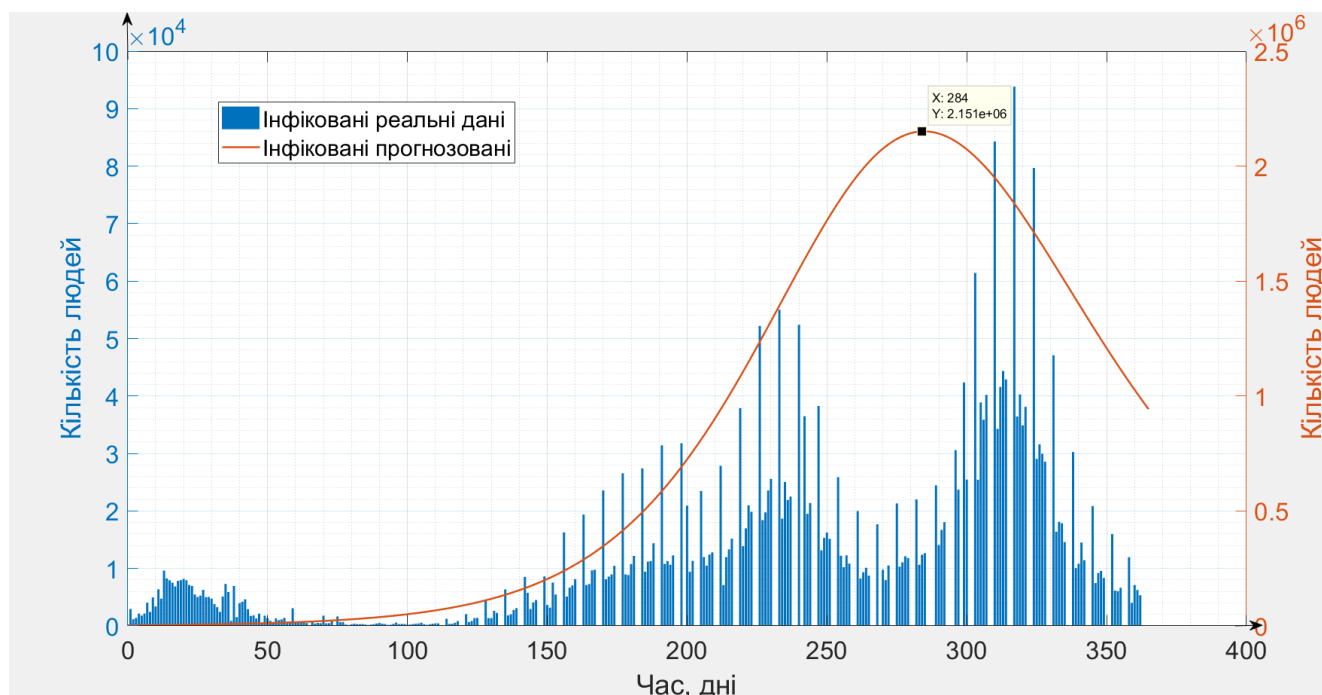


Рисунок 3.15 - Моделювання SIR для Іспанії при $R_0 = 1.4$

Виходячи з результатів моделювання можна зробити висновки, що за допомогою регулювання R_0 можна наблизити результати моделювання до

реальних, але в кількості захворівших людей, прогнозування буде досить не точним. Саме підбирання коефіцієнтів є досить трудомістким в плані затрат часу, а також дає велику помилку.

Так як модель SIR не враховує особливості захворюваності, такі як наприклад, інкубаційний період, під час якого особа ще не має явних симптомів захворювання, але вже є заразною.

3.5 Модифікація моделі SIR - SEIR модель

Хвороби мають різний перебіг та свої особливості, щоб наблизити результати моделювання до реальних було створено безліч моделей з урахуванням деяких, однією з таких прикладів є SEIR модель.

Якщо певний тип інфекційного захворювання має інкубаційний період проміжок часу між моментом зараження і появою перших симптомів інфекційного захворювання. Інша назва — прихований період хвороби: людина вже заражена, але хвороба себе ще не проявила, то модель SIR не відповідає інфекційним характеристикам захворювання [31].

Тривалість інкубаційного періоду може варіюватись від кількох годин (харчова токсикоінфекція, грип та інші) аж до десятків років (пріонові хвороби). Тривалість інкубаційного періоду також залежить від стану імунітету зараженого організму, від заражаючої дози бактерій, вірусів або інших інфекційних агентів, що потрапили в організм, і від їх патогенності й вірулентності. Протягом цього періоду ушкодження, що виникають у зараженому організмі, непомітне через наявність резервів макроорганізму, невелику кількість або недостатню активність інфекційних агентів.

З'являється модель SEIR, яка на основі моделі SIR представляє заражену людину, групу Е-типу, тобто SEIR поділяється на такі групи «Susceptible-Exposed-Infected-Recovered» (рисунок 3.16) , «Сприйнятливі –Інкубаційні - Інфіковані – Видужавші/Померлі» . Основна відмінність між моделлю SEIR і моделлю SIR

полягає в додаванні групи особин, які знаходяться у інкубаційному періоді до моделі SIR. SEIR також базується на трьох припущеннях моделі SIR. Сприйнятливі люди, які контактували з хворими, захворіють не відразу, а стануть носіями збудника захворювання і будуть відноситись до «Інкубаційної» групи.

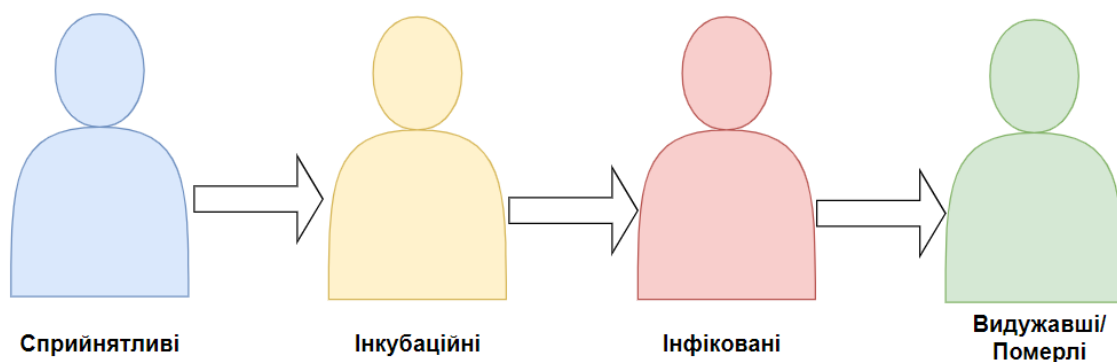


Рисунок 3.16 – Схема для «Сприйнятливі –Інкубаційні - Інфіковані – Видужавші/Померлі» моделі

Якщо розглядати хворобу COVID-19, то дана модель як раз передбачає одну з особливостей захворювання, а саме інкубаційний період, який є досить небезпечним, адже людина, ще може не відчувати симптомів, а вже заразити інших.

У порівнянні з характеристикою моделі SIR, що людина є заразною при контакті, модель SEIR вважає, що заразними є лише деякі люди, сприйнятливі до контакту з інфікованою людиною. Ця особливість подовжує період поширення хвороби.

Однак, модель SEIR все одно не враховує ще більш детальні фактори в процесі прогнозування епідемії, такі як ступінь державного контролю, швидкість і масштаби потоку людей, психологічний стан населення, фізичні характеристики різних людей і складність симптомів хворих. Усі ці фактори впливають на рівень зараження захворюванням, тривалість періоду зараження та кількість інфікованих.

Модель не враховує різні відхилення між різними групами людей у реальній ситуації. Наприклад, чи можуть заражені люди також заразити сприйнятливих

людей, чи однаковий рівень смертності в різних групах та чи одужавши особи знову можуть інфікуватись та через який період.

«Інкубаційна» група є кроком між «Сприйнятливою» та «Інфекційною» групами. Сюди входять особи, які були піддані інфекції, але самі ще не заразні або ж вже можуть заражати, даний період для COVID-19 для кожної особини протікає по різному, адже в деяких випадках люди можуть не бути заразними в цей період, але ряд досліджень доводять, що в інкубаційний період підчас COVID-19 деякі особини можуть заражати інших.

Зауважимо, що сума чотирьох змінних всіх станів в будь-який момент залишається постійною і дорівнює початковому значенню, а γ в рівняннях є зворотним значенням інкубаційному періоду (який може варіюватися від 2 до 14 днів, якщо розглядається хвороба COVID-19). Оскільки в цій моделі маємо чотири групи замість трьох, потрібні чотири диференціальні рівняння для опису поширення інфекції (рисунок 3.17).

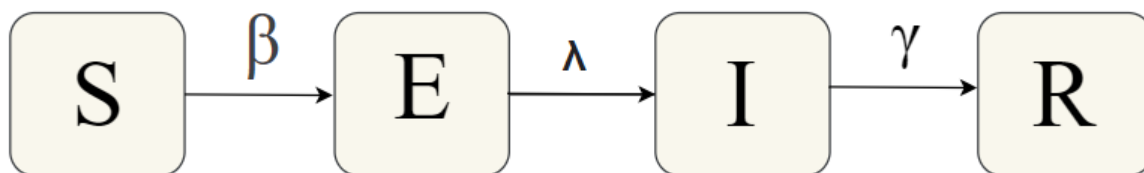


Рисунок 3.17 – Модель SEIR

Як і для попередньої моделі приймаємо, що, перший набір залежних змінних підраховує людей у кожній із груп, кожна як функція часу:

$$S = S(t), \quad (3.19)$$

$$E = E(t), \quad (3.20)$$

$$I = I(t), \quad (3.21)$$

$$R = R(t) \quad (3.22)$$

Другий набір залежних змінних представляє частку загальної сукупності в кожній із чотирьох категорій. Отже, N — загальна чисельність населення:

$$s(t) = \frac{S(t)}{N}, \quad (3.23)$$

$$e(t) = \frac{E(t)}{N}, \quad (3.24)$$

$$i(t) = \frac{I(t)}{N}, \quad (3.25)$$

$$r(t) = \frac{R(t)}{N}. \quad (3.26)$$

Рівняння для «Сприйнятливої» групи має вигляд у моделі з інкубаційним періодом:

$$\frac{ds}{dt} = -\beta * i(t) * s(t) \quad (3.27)$$

Рівняння для «Інкубаційної» групи :

$$\frac{de}{dt} = \beta * i(t) * s(t) - \lambda * e(t) \quad (3.28)$$

Рівняння для «Інфікованої» групи :

$$\frac{di}{dt} = \lambda * e(t) - \gamma * i(t) \quad (3.29)$$

Рівняння для «Видужавших/ Померлих»

$$\frac{dr}{dt} = \gamma * i(t) \quad (3.30)$$

Також, як і в попередній моделі, тільки відтепер для чотирьох груп зауважимо, що

$$N = S(t) + E(t) + I(t) + R(t) \quad (3.31)$$

Та

$$\frac{ds}{dt} + \frac{de}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0 \quad (3.32)$$

Для побудови також обрано три країни, а саме США, Іспанія та Україна. Початкові значення для населення залишаються незмінними та беруться з таблиці 3.3. Кількість інфікованих для цієї моделі також обрано з відкритого джерела даних з CDC. Кількість людей, які знаходяться в інкубаційному періоді є невідомою, а також для початку хвороби її прирівнюємо до 0, так само як і для людей, які одужали, тобто:

$$S(0) = N, \quad (3.33)$$

$$E(0) = 0, \quad (3.34)$$

$$R(0) = 0. \quad (3.35)$$

$I(0)$ – кількість захворівших станом на 13 березня 2020 року в кожній обраній країні відповідно.

Побудова буде виконана за допомогою програмного пакету MATLAB, як і для моделі SIR для трьох країн: США, Україна та Іспанія. При цьому параметр R_0 буде змінюватись, адже модель має тепер іншу специфіку та враховує інкубаційний період хвороби. Для побудови було отримані такі значення R_0 наведені у таблиці 3.9

Таблиця 3.9 - R_0 для моделей SEIR

Країна	R_0
США	2.7
Україна	3.8
Іспанія	2.1

Інкубаційний період в моделі буде дорівнювати 14 днів та буде враховуватись для моделі в рівнянні для «Інкубаційної» групи, але має бути використаний в зворотній формі, тобто

$$r = \frac{1}{Y} \quad (3.36)$$

Де Y – тривалість інкубаційного періоду

На рисунках 3.18-3.20 зображено вже чотири криві, які відповідно позначають кожну з груп на які популяція поділена умовно в рамках обраної моделі, включаючи «Інкубаційних»

Для того аби підсумувати результати моделювання занесемо значення для групи «Інфіковані» до таблиці 3.10.

Таблиця 3.10– Результати моделювання SEIR модель

Країна	День піку реальний	К-сть захворівших реально	День піку змодельований	К-сть захворівших змодельовані	Помилка
США	294	315 100	282	44 870 000	44 554 900
Україна	261	16 590	259	7 795 000	7 778 410
Іспанія	317	93 820	294	3 997 000	3 903 180

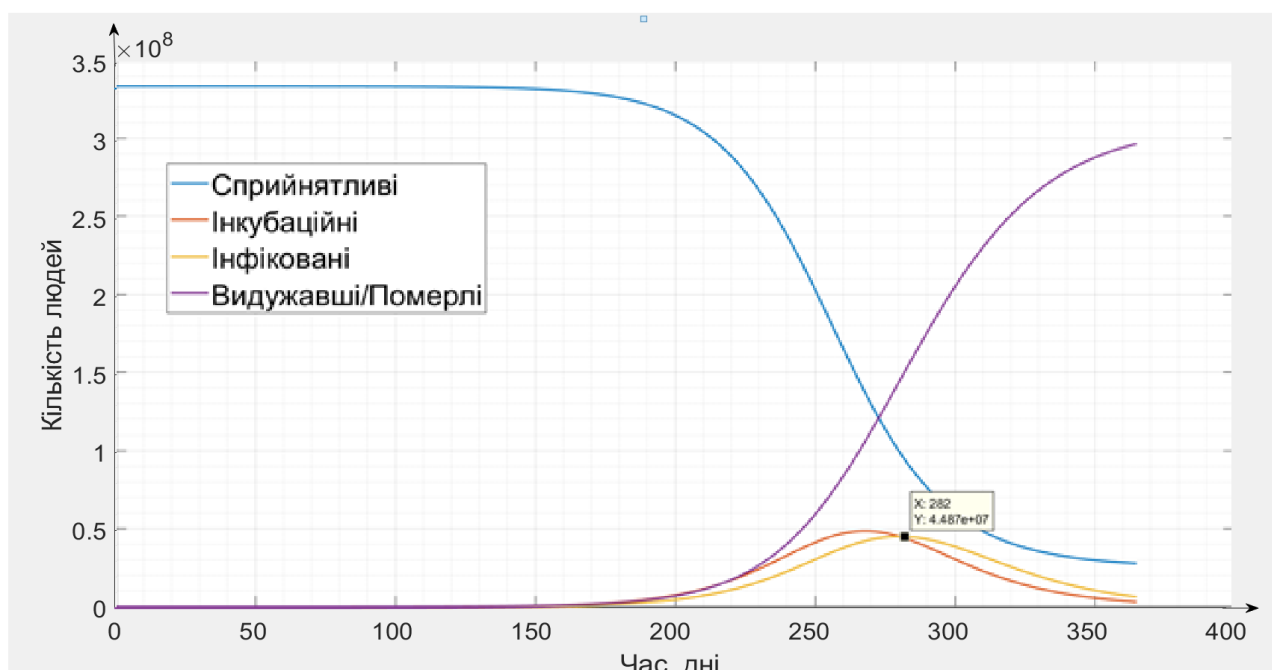


Рисунок 3.18 – SEIR модель для США при $R_0 = 2.7$

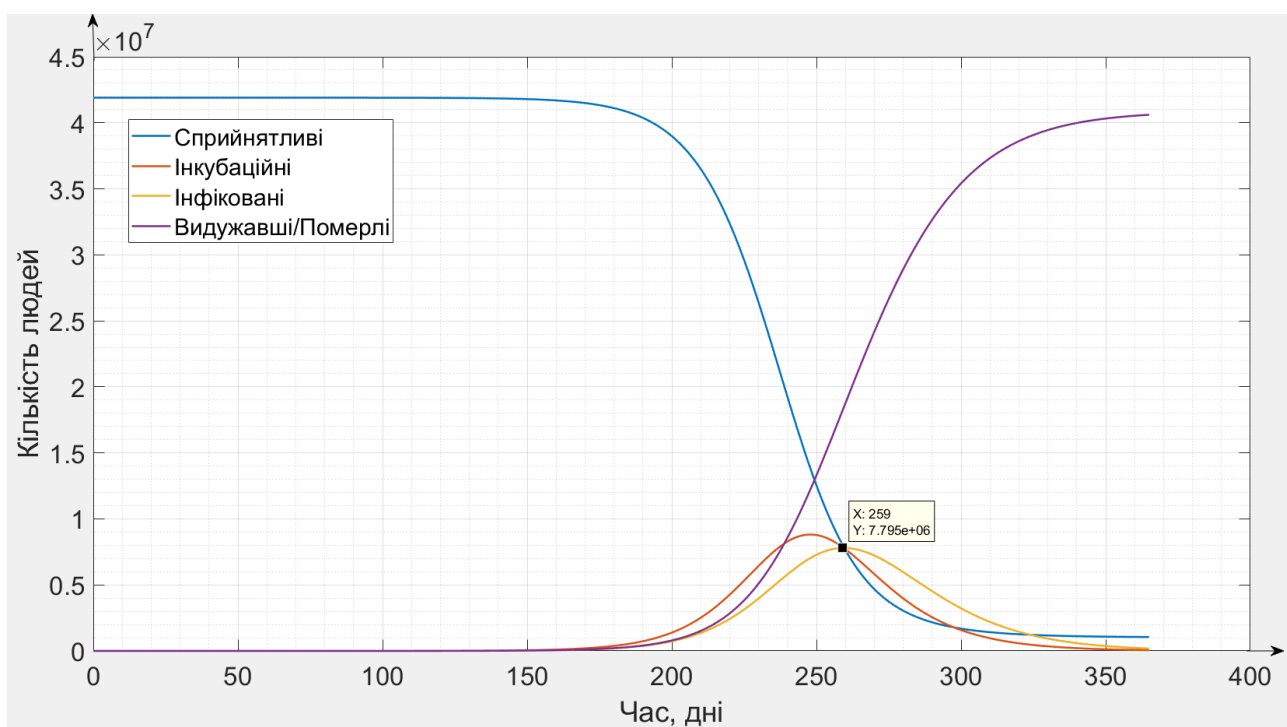


Рисунок 3.19 – SEIR модель для України при $R_0 = 3.8$

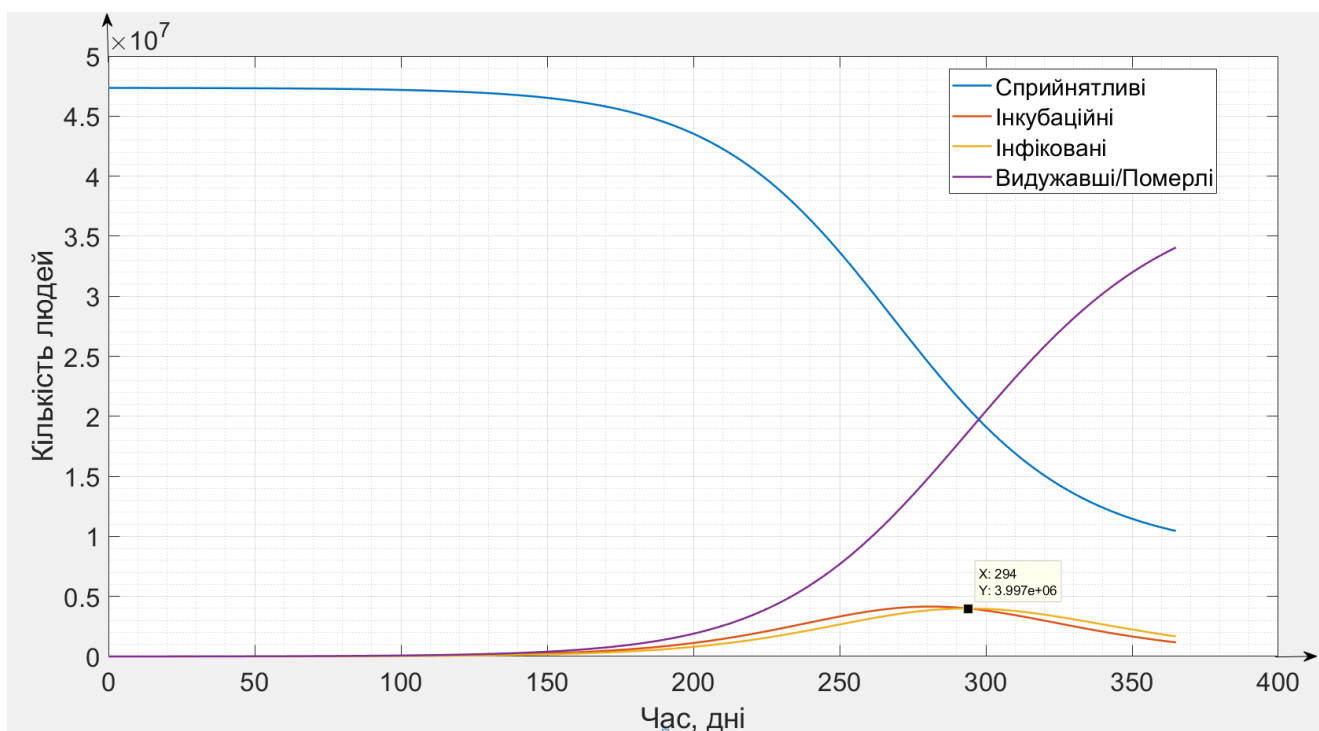


Рисунок 3.20 - SEIR модель для Іспанії при $R_0 = 2.1$

Аби оцінити чи справді модель показує набагато кращі результати потрібно порівняти числа, які були занесені до результуючих таблиць.

Для США отримали помилку в SIR моделі в 26 764 900 людей, а для SEIR моделі в 44 554 900, а також значну різницю в днях і спалах є прогнозованим тільки приблизно. То ж у випадку з даною країною, навіть кращі результати були отримані не враховуючи інкубаційний період.

Для України різниця між реальними і прогнозованими даними за допомогою SIR моделі склала 552 710 людей, що є в рази більше, ніж реальні дані, а після моделювання з інкубаційним період помилка становить 7 778 410, що є ще більшим значенням відносно реальних даних і також показує гірший результат за простішу модель.

Для Іспанії прогнозовані дані використовуючи модель SIR показали відмінність від реальних даних 2 056 180 в день, який був найближчим до реального піку, а ось за допомогою моделі SEIR - 3 903 180.

Також робимо висновки, що був обрано не достатньо коректно тривалість інкубаційного періоду, але знати його точне число не можливо, якщо розглядати саме COVID-19, адже коли почали з'являтися ще й мутації захворювання, то інкубаційний період зовсім змінився. Щоб додати такий параметр до моделі, потрібно будувати модель, яка б враховувала ще один параметр і він був змінний, що зробить модель більш складною для розрахунків, тим самим збільшиться час на збір даних та час на побудову такої моделі.

Дана модель також має велику різницю між реальними і прогнозованими даними. Проблема залишається в тому, що за допомогою цієї моделі неможливо отримати значення наближені, навіть приблизно до реальних, хоч і враховується особливість захворювання, яка суттєво впливає.

То ж за основу для подальших модифікацій немає сенсу обирати ускладнену модель, а можна скоригувати базову моделі потрібним чином, як і більшість вже існуючих модифікацій.

3.6 Висновки до розділу

Базова модель для дослідження розповсюдження захворювань є легкою для моделювання та досить зрозумілою, що є великою перевагою. Також можливість побудувати на будь-який термін, а головне на довгостроковий термін як було розглянуто – на рік, що дає можливість планувати подальші дії для запобігання серйозних наслідків. Модель легко модифікувати розділяючи популяцію на більше груп, що було представлено і в цьому дослідженні, а саме SEIR модель.

Модель була побудована за допомогою програмного пакету MATLAB, який одразу дає можливість побачити результати як аналітично, так і графічно, а також дозволяє легко перевикористати вже існуючий код, що було зручно при створенні модифікованої моделі SEIR.

4 РОЗРОБЛЕННЯ МОДЕЛІ ПРОГНОЗУВАННЯ З ОПТИМІЗОВАНИМИ ПАРАМЕТРАМИ

З попередньо отриманих результатів, видно, що результати моделювання для дослідження розповсюдження захворювань можна спрогнозувати досить приблизно і відмінність між реальними даними і змодельованими є суттєвою, що кардинально може вплинути на ситуацію із захворюваністю.

Адже, якщо спиратись на таке моделювання можна не встигнути попередити швидке зараження і вчасно не впровадити запобіжні заходи або ж навпаки введені карантинні норми будуть діяти занадто довго, що може вплинути вже на психологічний стан населення, а також на економічне становище людей. Також, не останнім пунктом, потрібно відмітити, що обране для моделювання захворювання є новим і дійсно дієвого лікування немає, тому прогнозування надає змогу дослідити розвиток хвороби і дозволяє провести дослідження як буде далі розвиватись вірус і які міри потрібно запровадити, щоб зменшити ризики для населення. То ж модель дослідження розповсюдження епідемічного захворювання, яка б мала невелику помилку є вкрай необхідною.

Щоб покращити результати моделювання і наблизити їх до реальних пропонується оптимізувати параметри, що допоможе зменшити помилку в моделюванні. Існує багато методів для оптимізації параметрів, які б дозволили виконати оптимізацію потрібних для моделювання параметрів, але найчастіше вони використовувались на короткий термін і таких досліджень є велика кількість, але для того аби бачити ситуацію, наприклад, на рік вони не застосовуються, після огляду і аналізу вже застосованих методів, було прийнято рішення розглянути метод, а саме алгоритм Левенберга-Марквардта для визначення моделей епідемії. Цей алгоритм поєднує в собі метод Гаусса–Ньютона та методу градієнтного спуску, що допоможе покращити стабільність параметрів. Розглянемо основні формули і поняття для задачі найменших квадратів та методі Левенберга-Марквардта.

4.1 Метод найменших квадратів

Метод найменших квадратів — це форма математичного регресійного аналізу, що використовується для визначення лінії, яка найкраще підходить для набору даних, що забезпечує візуальну демонстрацію зв'язку між точками даних. Кожна точка даних представляє зв'язок між відомою незалежною змінною та невідомою залежною змінною[32].

Цей метод регресійного аналізу починається з набору точок даних, які наносяться на графік осей x і y . При використанні даного методу, генерується лінія найкращого підходу, яка пояснює потенційний зв'язок між незалежними та залежними змінними.

Метод найменших квадратів дає загальне обґрунтування розміщення лінії, яка найкраще підходить серед досліджуваних точок даних. Найбільш поширене застосування цього методу, яке іноді називають «лінійним» або «звичайним», спрямоване на створення прямої лінії, яка мінімізує суму квадратів помилок, які генеруються результатами пов'язаних рівнянь, наприклад як квадрат залишків, що є результатом відмінностей у спостережуваному значенні, і очікуваного значення на основі цієї моделі [33].

Лінія найкращого підходу, визначена за методом найменших квадратів, має рівняння, яке показує зв'язок між точками даних. Лінія найбільш відповідних рівнянь може бути визначена моделями за допомогою відповідного програмного забезпечення, які включають підсумок результатів для аналізу, де коефіцієнти та підсумкові результати пояснюють залежність змінних, що перевіряються.

Якщо дані показують менший зв'язок між двома змінними, лінія, яка найкраще відповідає цьому лінійному зв'язку, називається лінією регресії за методом найменших квадратів, яка мінімізує відстань по вертикалі від точок даних до лінії регресії. Термін «найменші квадрати» використовується, оскільки це найменша сума квадратів помилок, яку також називають «дисперсією».

У регресійному аналізі залежні змінні зображені на вертикальній осі y , а незалежні змінні – на горизонтальній осі x . Ці позначення утворюють рівняння для лінії найкращого прилягання, яка визначається за методом найменших квадратів.

На відміну від лінійної задачі, нелінійна задача найменших квадратів не має закритого розв'язку і, як правило, розв'язується шляхом ітерацій.

До прикладу, потрібно оптимізувати параметр a . Дано вектор функції $f(x, a)$ для задачі найменших квадратів :

$$F(a) = \sum_{i=1}^m [y_i - f(x_i, a)]^2 \quad (4.1)$$

Для вирішення цієї задачі для початку потрібно задати початкові припущення для параметру вектора a , наприклад :

$$a = (1, 1, 1, \dots, 1) \quad (4.2)$$

Так як вирішення має ітеративний підхід, то потім кожен крок ітеративного циклу буде замінювати вектор a на $a + c$, де c - інкремент. Функції матимуть вигляд:

$$f(x_i, a + c) \quad (4.3)$$

Та ці функції лінійно апроксимуються матрицею Якобі:

$$f(x_i, a + c) \approx f(x_i, a) + J_i * c \quad (4.4)$$

Де $J_i = \frac{\partial f(x_i, a)}{\partial a}$ являє собою оптимальну лінійну апроксимацію функції щодо a .

Матриця Якобі — це матриця часткових похідних. Якобіан є визначником матриці Якобі. Матриця міститиме всі часткові похідні векторної функції. Основне використання якобіана знаходиться в перетворенні координат. Тут йдеться про поняття диференціювання з перетворенням координат.

Для відображення $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$:

$$u = (u_1, u_2, \dots, u_m), \quad (4.5)$$

$$u_i = u_i(x_1, x_2, \dots, x_n), \quad (4.6)$$

$$i=1, \dots, m \quad (4.7)$$

В деякій точці x всі часткові похідні першого порядку. Матриця J , складається з часткових похідних функцій в точці x і має назву матриця Якобі цієї системи та має вигляд

$$J(x) = \begin{pmatrix} \frac{\partial u_1}{\partial x_1}(x) & \frac{\partial u_1}{\partial x_2}(x) & \dots & \frac{\partial u_1}{\partial x_n}(x) \\ \frac{\partial u_2}{\partial x_1}(x) & \frac{\partial u_2}{\partial x_2}(x) & \dots & \frac{\partial u_2}{\partial x_n}(x) \\ \dots & \dots & \dots & \dots \\ \frac{\partial u_m}{\partial x_1}(x) & \frac{\partial u_m}{\partial x_2}(x) & \dots & \frac{\partial u_m}{\partial x_n}(x) \end{pmatrix} \quad (4.8)$$

Щоб сума квадратів $F(a)$ була мінімальною, градієнт F щодо c повинен дорівнювати нулю. Підставляючи номер формули в номер формули отримаємо :

$$F(a + c) \approx \sum_{i=1}^m [y_i - f(x_i, a) - J_i * c]^2 \quad (4.9)$$

Або у векторній формі

$$F(a + c) \|y - f(a) - J * c\|^2 \quad (4.10)$$

Взявши похідну рівняння () щодо c та створення дає результат операції, що дорівнює 0

$$(J^T J)c = J^T [y - f(a)] \quad (4.11)$$

де J це матриця Якобі і J_i представляє i -тий рядок Якобі матриці і i -тий компонент f , $y \in f_i$ та y_i відповідно.

Матриця Якобі, як визначено вище, є не квадратною матрицею як в загальному випадку, а прямокутною матрицею розміром $m \times n$, де n - кількість параметрів (розмір вектору a). Множення матриці $(J^T J)$ дає необхідне $n \times n$ квадратна матриця і виготовлене матриця-вектор у правій частині дають вектор розміру n . В результаті виходить набір лінійних рівнянь, які можна вирішити для s .

4.2 Метод Левенберга-Марквардта

Метод Левенберга-Марквардта є стандартним методом, який використовується для розв'язування нелінійних задач найменших квадратів. Проблеми з найменшими квадратами виникають при підборі параметризованого функції до набору виміряних точок даних шляхом мінімізації суми квадратів помилки між точками даних і функцією. Нелінійні задачі найменших квадратів виникають, коли функція не є лінійною за параметрами [34].

Нелінійні методи найменших квадратів передбачають ітераційне покращення значень параметрів з метою зменшення суми квадратів похибок між функцією та виміряними даними. Метод підгонки кривої Левенберга-Марквардта насправді є комбінацією двох методів мінімізації: методу градієнтного спуску та методу Гаусса-Ньютона. Метод градієнтного спуску, сума квадратів помилок зменшується шляхом оновлення параметрів у напрямку найбільш крутого спуску.

У методі Гаусса-Ньютона сума квадратів помилок зменшуються, якщо припустити, що функція найменших квадратів локально квадратична, і знаходження мінімуму саме квадратичного. Метод Левенберга-Марквардта діє за схожим принципом, але більше схожий на метод градієнтного спуску, коли параметри далекі від оптимальних значень, і працює як метод Гаусса-Ньютона, коли параметри близькі до їх оптимальних значень.

В підрозділі 4.1 наведено рішення традиційного методу найменших квадратів, а от за допомогою методу Левенберга-Марквардта можна покращити вирішення, додавши коефіцієнт демпфіювання λ . Коефіцієнт демпфування, також може називатись як коефіцієнт затухання.

Нехай маємо таку задачу про найменші квадрати:

$$F(a) = \sum_{i=1}^m [y_i - f(x_i, a)]^2 \rightarrow \min \quad (4.12)$$

То при методі Левенберг-Марквардта, додається коефіцієнт λ і отримаємо :

$$(J^T J + \lambda M) * c = J^T [y - f(a)] \quad (4.13)$$

де M – одинична матриця.

У 1971 році, Флетчер дослідив, що коефіцієнт λ є завеликим і що вираз $J^T J + \lambda M$ не зможе бути використаним без вдосконалення цього виразу. Тому в своєму дослідженні Флетчер припустив, що можливо масштабувати кожен компонент градієнту в відповідності до кривизни і тоді буде більший рух вздовж напрямків, де градієнт є менше [35]

Це дозволить уникнути повільного сходження напрямку невеликого градієнту, тому у своїй роботі він змінив одиничну матрицю на діагональну матрицю $(J^T J)$ і отримав такий остаточний результат :

$$(J^T J + \lambda * diag * (J^T J)) * c = J^T [y - f(a)] \quad (4.14)$$

Для кращого вибору параметра демпфування λ було висунуто різні більш-євристичні аргументи. Існують теоретичні аргументи, які показують, чому деякі з цих варіантів гарантують локальну збіжність алгоритму, однак, ці варіанти можуть призвести до того, що глобальна конвергенція алгоритму страждає від небажаних властивостей найкрутішого спуску, зокрема, дуже повільної збіжності, близької до оптимальної.

Абсолютні значення будь-якого вибору залежать від того, наскільки добре масштабована початкова проблема. Марквардт рекомендував починати зі значення λ_0 і коефіцієнта $\vartheta > 1$. Спочатку встановлюється $\lambda = \lambda_0$ і обчислення залишкової суми квадратів після одного кроку від початкової точки з коефіцієнтом демпфування і по-друге, з λ_0/ϑ . Якщо обидва вони гірші за початкову точку, то демпфування збільшується шляхом послідовного множення на ϑ , доки не буде знайдено кращу точку з новим коефіцієнтом демпфування $\lambda_0\vartheta^k$ для деяких k .

Якщо використання коефіцієнта демпфування λ/ϑ призводить до зменшення квадрата залишку, тоді це приймається як нове значення λ та нове оптимальне розташування приймається таким, як отримане з цим коефіцієнтом демпфування і процес продовжується. Якщо використання λ/ϑ призвело до гіршого залишку, але використання λ призвело до кращого залишку, λ залишається незмінною, а новий оптимум приймається як значення, отримане з λ як коефіцієнтом демпфування.

Ефективна стратегія контролю параметра демпфування, що називається «відтермінованим задоволенням», полягає у збільшенні параметра на невелику величину для кожного кроку на підйом і зменшення на велику величину для кожного кроку вниз. Ідея такого підходу полягає в тому, щоб уникнути занадто швидкого спуску на початку оптимізації, отже обмежуючи кроки, доступні в майбутніх ітераціях, і, відповідно, сповільнюючи зближення. Збільшення в 2 рази і зменшення в 3 рази виявилося ефективним у більшості випадків, тоді як для великих проблем більш екстремальні значення можуть працювати краще, зі збільшенням у 1,5 раза та зменшенням у раз. з 5 [36].

4.3 Застосування методу Левенберга-Марквардта для SIR моделі

Задля кращих результатів моделювання у даній роботі, пропонується розглянути метод Левенберг-Марквардта саме для SIR моделі для оптимізації параметрів. Для того аби застосувати цей метод можна було б взяти і SEIR модель, але враховуючи те, що моделі при побудові результати моделі з інкубаційним

періодом мали ще більшу помилку, ніж без його врахування, буде доцільним обрати, так звану, базову модель.

Також зважаючи на те, що інкубаційний період може коливатись і не обов'язково його тривалість може складати таку кількість днів, як було обрано при моделюванні це стає скоріше не допоміжним параметром, а ще однією невідомою змінною. Ще одним немало важливим приводом взяти за основу саме SIR модель є те, що саме її обирають основою для всіх інших моделей і сама SEIR модель скоріше вже є її модифікацією.

Обрано одну з модифікацій SIR моделей. Приймаємо за основу, що диференціали для SIR моделі незмінні як в підрозділі 3.1 .

Припускаємо, що населення кожної країни залишається незмінним під час епідемії і $N = I + S + R$, для отримання модифікованої моделі підставляємо це рівняння в рівняння інфікованих особин di/dt та отримаємо таку систему рівнянь:

$$\frac{di}{dt} = r * \tau * i - \varepsilon * i^2 - \varepsilon * i * r \quad (4.15)$$

$$\frac{dr}{dt} = \Upsilon * i \quad (4.16)$$

$$\text{де } r * \tau = R_0 * \beta - \Upsilon \text{ та } \varepsilon = R_0 * \beta$$

Отримавши рівняння вже для двох груп, після підстановки рівняння населення отримані нові коефіцієнти, ці змінні є ті, які будуть використанні задля оптимізації параметрів. Для остаточного варіанту потрібно підставити змінні в диференційні рівняння і після підстановки r та ε , отримуємо два диференційні рівняння, які будуть використанні для моделі:

$$\frac{di}{dt} = (R_0 * \beta - \Upsilon) * i - (R_0 * \beta) * i^2 - (R_0 * \beta) * i * (R_0 * \beta - \Upsilon) \quad (4.17)$$

$$\frac{dr}{dt} = \Upsilon * i \quad (4.18)$$

З отриманих рівнянь тепер можна застосовувати метод Левенберга-Марквардта, а саме позначаємо вектор як :

$$a = [Y \beta R_0] \quad (4.19)$$

Функція має вигляд як вказано у (ф-лі ЛМ) і після підстановки всіх значень відповідно до алгоритму вдосконаленого Флетчером отримаємо функцію для оптимізації :

$$F(a) = [T_{\text{кзд}} - ((R_0 * \beta - Y) * i - (R_0 * \beta) * i^2 - (R_0 * \beta) * i * (R_0 * \beta - Y))]^2$$

Алгоритм Левенберга–Марквардта вводить затухаючий коефіцієнт у традиційний метод найменших квадратів при обчисленні матриці . Для всіх затухаючих коефіцієнтів, більших за 0, матриця параметрів буде позитивно визначеною, що гарантує, що матриця знаходиться в напрямку спуску. Це покращує надійність і стабільність параметрів традиційного алгоритму.

Визначаємо модифіковану SIR модель з використанням цього алгоритму як статичну модель. Ця статистична модель має добре поєднувати фактичні дані спостережень з прогнозованими даними та згладжувати помилку через динамічність поширення епідемії COVID-19 і може забезпечити тим самим кращі результати для більш точного прогнозу.

4.4 Вибір програмного забезпечення для розробки моделі

Для моделювання буде використано пакет MATLAB. Як вже було описано у роботі, даний програмний пакет дає безліч можливостей для різних математичних операцій, а також для методів прогнозування та використання алгоритмів без додаткового написання коду. Можна просто вибрати відповідну функцію, яка вже буде включати в себе потрібний для задачі алгоритм, що зменшить час на

написання додаткового коду як от в іншому програмному забезпеченні, а також буде швидше, ніж обраховувати дані методи аналітично.

В випадку з SIR моделлю вже перевірили, що це досить зручно обраховувати диференційні рівняння з багатьма змінними, використовуючи MATLAB та вирішувач ode45. Для методу Левенберг-Марквардта також потрібно вирішувати диференційні рівняння, то ж ode45 буде застосовано і тут.

Для того аби реалізувати метод Левенберг-Марквардта можна написати код самому для нашого програмного забезпечення або ж спробувати застосувати вже вбудовані методи. Після детального вивчення документації було знайдено вже готове рішення для задачі методу найменших квадратів з вирішенням за допомогою методу Левенберг-Марквардта.

Даний пакет має назву lsqcurvefit – вирішувач задач підгонки нелінійної кривої (підгонки даних) в задачі найменших квадратів.

В документації наведено приклади, які додаткові параметри потрібні для того, щоб застосувати метод і кількість параметрів може бути різною. То ж для розгляду маємо перелік показаний на рисунку 4.1

```
x = lsqcurvefit(fun,x0,xdata,ydata)
x = lsqcurvefit(fun,x0,xdata,ydata,lb,ub)
x = lsqcurvefit(fun,x0,xdata,ydata,lb,ub,options)
x = lsqcurvefit(problem)
[x,resnorm] = lsqcurvefit( __ )
[x,resnorm,residual,exitflag,output] = lsqcurvefit( __ )
[x,resnorm,residual,exitflag,output,lambda,jacobian] = lsqcurvefit( __ )
```

Рисунок 4.1 – Синтаксис lsqcurvefit [37]

Для поставленої задачі в дисертації перші три використання lsqcurvefit може підійти у вирішення проблеми. Потрібно розглянути їх опис більш детально.

$x = \text{lsqcurvefit}(\text{fun}, x_0, \text{xdata}, \text{ydata})$ починається з x_0 і знаходить коефіцієнти x , які найкраще відповідають нелінійній функції $\text{fun}(x, \text{xdata})$ до даних ydata (у сенсі

найменших квадратів). `ydata` повинні мати такий самий розмір, як вектор (або матриця) `F`, що повертається `fun`.

`x = lsqcurvefit(fun,x0,xdata,ydata,lb,ub)` визначає набір нижніх і верхніх меж проектних змінних у `x`, так що рішення завжди знаходиться в діапазоні $lb \leq x \leq ub$. Ви можете виправити компонент рішення `x(i)`, вказавши `lb(i) = ub(i)`.

`x = lsqcurvefit(fun,x0,xdata,ydata,lb,ub)` визначає набір нижніх і верхніх меж проектних змінних у `x`, так що рішення завжди знаходиться в діапазоні $lb \leq x \leq ub$. Ви можете виправити компонент рішення `x(i)`, вказавши `lb(i) = ub(i)`.

Так як є вхідна функція і початкові значення, а також відомі потрібні межі для параметрів, а саме межі для R_0 , β та γ , які наведені у таблиці 4.1

Таблиця 4.1 - Межі для коефіцієнтів

Значення	Межі
R_0	0 - 5
β	0 - 1
γ	0 - 1

То ж обираємо варіант - `x = lsqcurvefit(fun,x0,xdata,ydata,lb,ub)`.

Також, щоб виключити від'ємні значення та оптимізація була виконана більш точно, використаємо `optimoptions`. `Optimoptions` дозволяє задати опції для оптимізації, що зможе допомогти з проблемою негативних значень, а також використати відповідний алгоритм, який був описаний вище.

`Optimoptions` – це спосіб об'єднання набору пар ім'я-значення. Вони корисні, оскільки дозволяють:

- налаштувати або замінити процес оптимізації.
- вибрати додаткові функції, такі як вихідні функції та функції графіка.
- зберігати і повторно використовувати налаштування.

Вони спрощують синтаксис вирішувача, не потрібно включати багато пар ім'я - значення у виклик до вирішувача (рисунок 4.2).

Syntax

```
options = optimoptions(SolverName)
options = optimoptions(SolverName,Name,Value)

options = optimoptions(olddoptions,Name,Value)

options = optimoptions(SolverName,olddoptions)

options = optimoptions(prob)
options = optimoptions(prob,Name,Value)
```

Рисунок 4.2 – Синтаксис optimoptions [38]

За допомогою optimoptions можна створити опції оптимізації. Для випадку з обраною моделлю, задаємо такі параметри оптимізації : 'lsqcurvefit', 'levenberg-marquardt', 'iter', 'FunctionTolerance', 'MaxFunctionEvaluations', 'MaxIterations', 'OptimalityTolerance'. Що дозволяють задати саме метод Левенберг-Маркварда, кількість ітерації та поставити деякі обмеження на кількість ітерацій.

4.5 Програмна реалізація моделі з використанням методу Левенберг-Марквардта

Для того аби використовувати SIR модель раз з запропонованим методом, спочатку, потрібно:

- взяти попередньо описаний код для SIR моделі
- визначити з рівнянь базової моделі значення для R для подальшого застосування;
- додати рівняння для модифікованої моделі SIR(рисунок);
- використати вирішувач ode45 також і для модифікованої моделі;
- задати початкові параметри для нової моделі;

- записати всі необхідні optimoptions;
- записати функцію для lsqcurvefit використовуючи попередньо визначене значення для R (рисунок 4.2);
- записати метод для побудови графіку.

```
function dydt = ode_fun_params(param, t)

    N = 41902400;

    global ito_pred;
    global rto_pred;

    I = ito_pred ;
    R = rto_pred;

    beta = param(1);
    R_0 = param(2);
    r = param(3);

    dI = I.*(beta*R_0-r) - ((beta*R_0)/N)*I.^2 - R.*I*((R_0*beta)/N);

    dydt = dI;
end
```

Рисунок 4.2 – Модифікована модель SIR для методу Левенберг-Марквардта

Опції задаються відповідно до синтаксису MATLAB (рисунок 4.3)

```
options = optimoptions('lsqcurvefit','Algorithm','levenberg-marquardt','Display','iter','FunctionTolerance')
```

Рисунок 4.3 – Задані опції для optimoptions

Межі для кожного з коефіцієнтів будуть задані в вектор згідно з синтаксисом (рисунок 4.4)

```
lb =[0,0,0];
ub=[1,5,1];
```

Рисунок 4.4- Задані межі для кожного з параметрів, які будуть оптимізовані

Вхідні параметри для lsqcurvefit представлені на рисунку 4.5

```
lsqcurvefit(@ode_fun_params, x0, tspan, data',lb,ub,options);
```

Рисунок 4.5 - Вирішувач для оптимізації параметрів lsqcurvefit

Дані кроки виконуємо використовуючи дані для трьох країн по черзі, а саме, США, України та Іспанії.

Також будемо підставляти початкові дані і для параметрів, які потрібно оптимізувати (R_0, β, γ). Початкові значення в методі є вектором, записуємо їх в векторній формі в відповідному порядку до формули. Початкові значення для кожної країни окремо обрані відповідно до моделі SIR. Після запуску описаного вище алгоритму для США отримуємо графіки (рисунок 4.6)

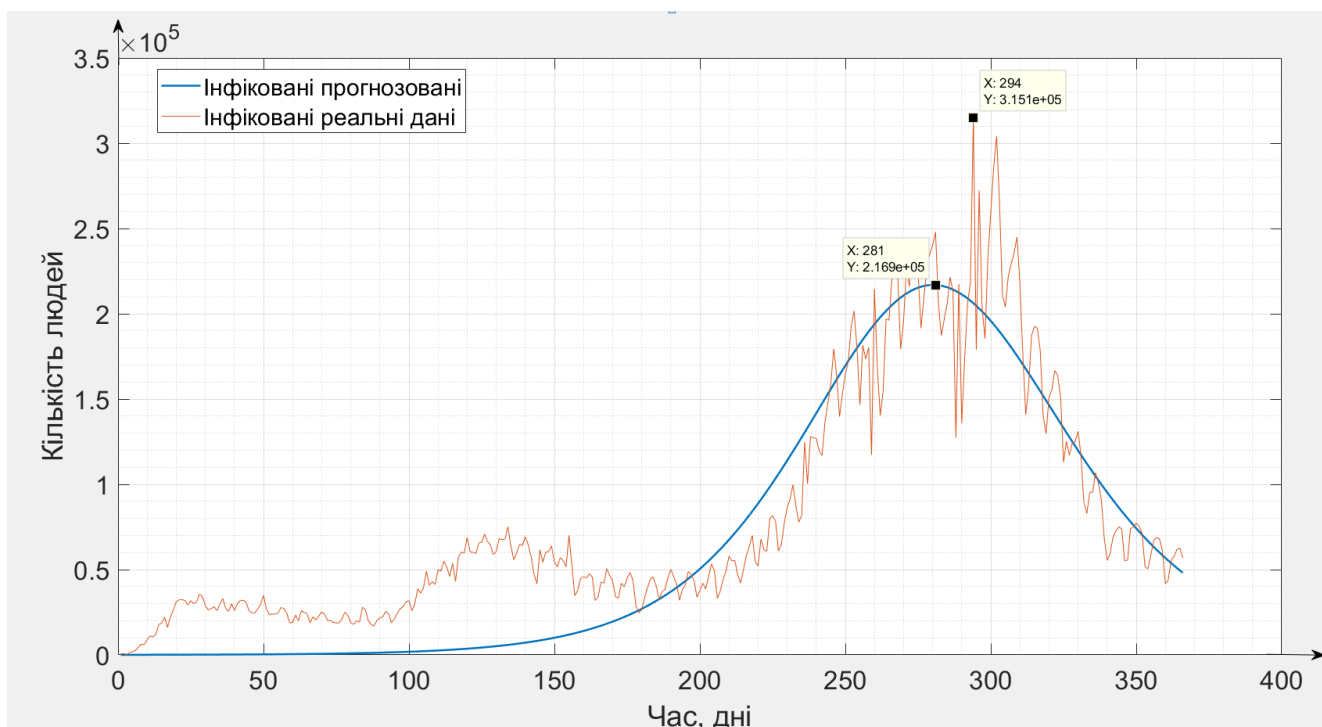


Рисунок 4.6 - Оптимізація параметрів для США

Як видно з побудованих графіків спалах, який відбувся на 294 день не співпадає з графіком, який отримано після оптимізації параметрів, але має співпадіння з графіком з попереднім відносно меншим спалахом на 281 день.

Можна припустити декілька причин такого неспівпадіння, а саме те, що COVID-19 за такий тривалий період мав декілька спалахів, як видно і на графіку в час між 100 і 150 днем, в свою ж чергу SIR модель немає такої властивості, щоб створювати прогноз для декількох спалахів.

Наступне припущення заключається в тому, що спалах на 281 день є досить великим і наступний умовний пік на 294 день є, можна сказати одиничним випадком і поступовий спад від 280 дня до кінця, окрім спалахів є досить поступовим, то ж на такий несподіваний спалах на 294 могли вплинути, наприклад, якесь непердабачуване скупчення людей на свята або навіть мітинги чи звичайний вихідний день, що точно не враховується у обраній моделі.

Модель розповсюдження COVID-19 з оптимізованими параметрами в порівнянні з реальним даними представлена на рисунку 4.7

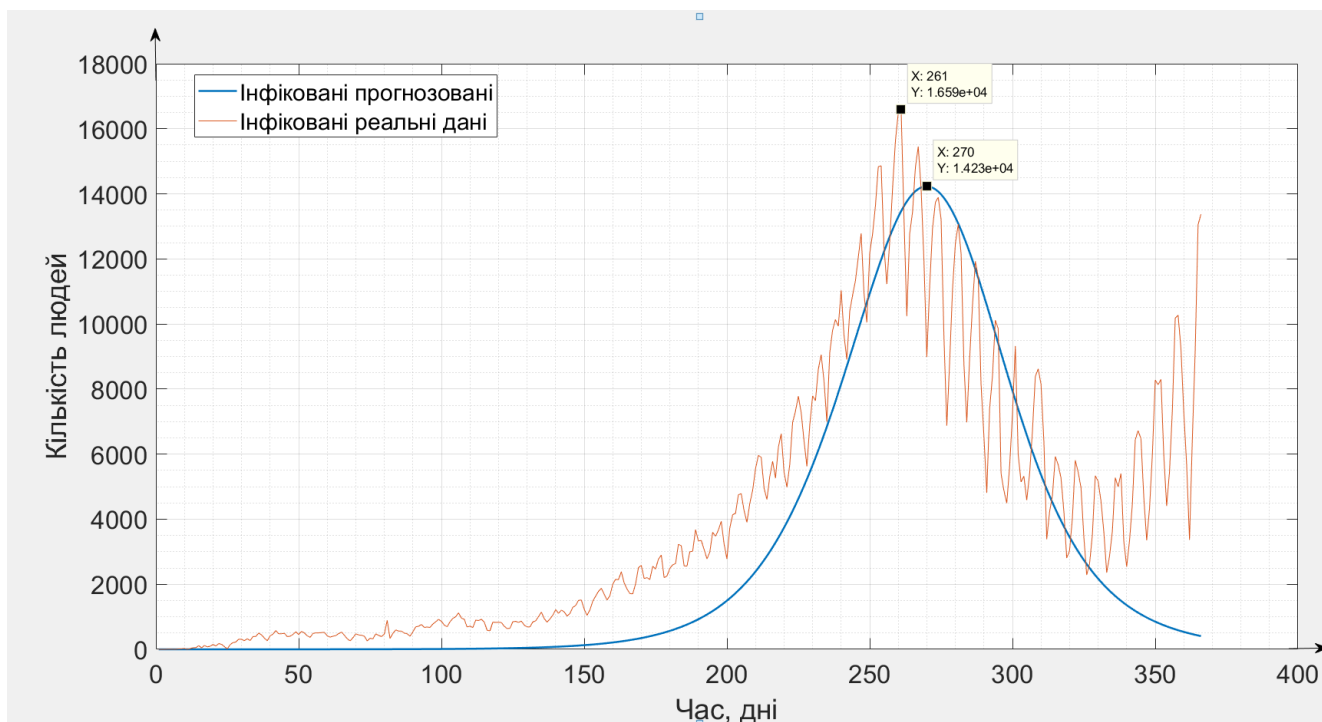


Рисунок 4.7 – Оптимізація параметрів для України

Отримані результати, використовуючи оптимізовані параметри мають суттєво меншу помилку, порівнюючи з реальними даними, також бачимо, що сам спалах захворювання майже співпав, з різницею у 10 днів. Великою перевагою є те, що кількість прогнозованих інфікованих людей є близькою до реальних даних, також відхилення від дня піку з реальними даними є не суттєвим, що дозволяє оцінити ситуацію прозоро і похибка в 10 днів буде не значною, якщо використовувати дану модель у реальному житті.

Моделювання з прогнозу розповсюдження вірусної інфекції COVID-19 для Іспанії з використанням оптимізованих параметрів, показано на рисунку 4.8

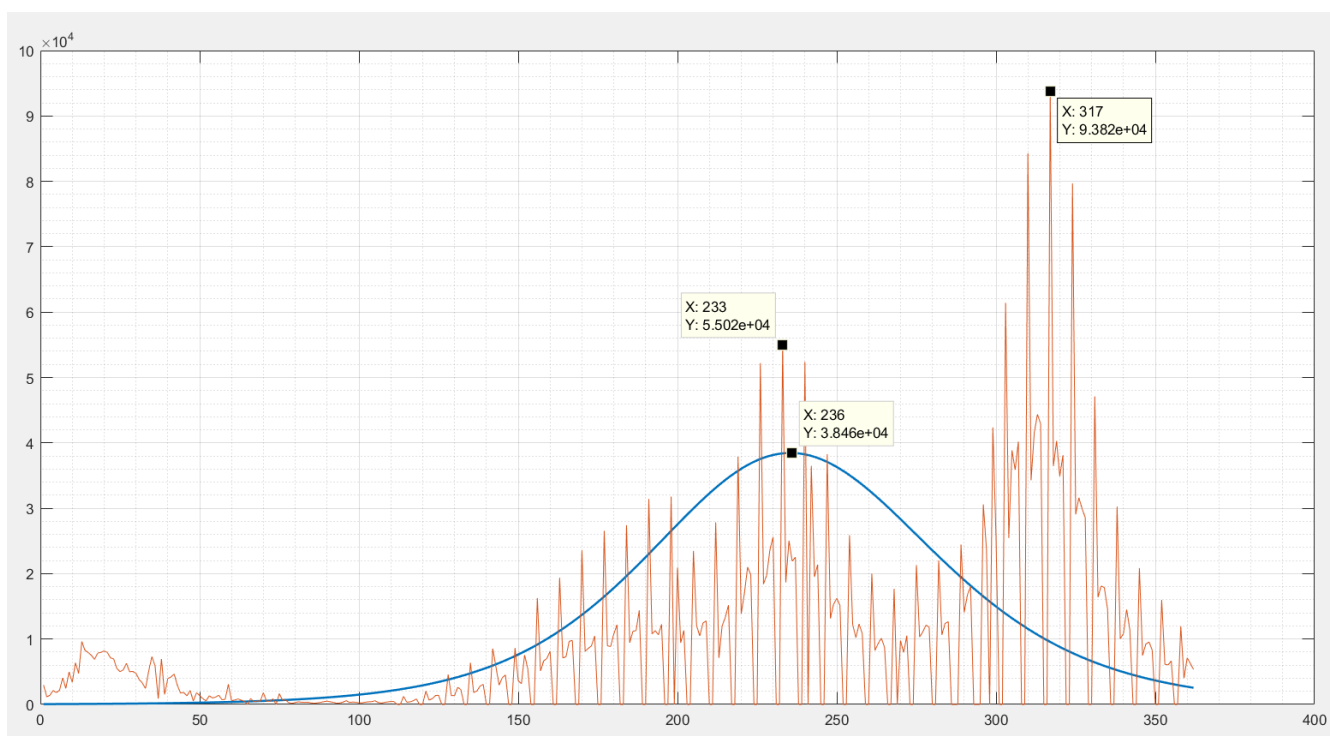


Рисунок 4.8 - Оптимізація параметрів для Іспанії

На графіку з оптимізованими параметрами видно, що пік прогнозований показано на 236 день та має деяку помилку людей відносно спалаху на 317 день за реальними даними.

Дана похибка не стала меншою використовуючи метод Левенберга-Марквардта через ряд причин. Як відомо, однією з країн, яка найскладніше переживала пандемію COVID-19, це була саме Іспанія, велика кількість обмежень

і карантинних норм не приводила до покращення статистики захворювання, також не останню роль грав менталітет та темперамент даної країни, де прийняті часті обійми і поцілунки, що досить легко сприяє розповсюдженню вірусу.

Попередньо названа проблема є звісно не математичним підтвердження похибки моделі, але має значний вплив на статистику захворюваності, щодо проблем через, які за допомогою оптимізації все одно залишилась досить велика похибка. Також, як видно по графіку з даними є досить багато значень, які дорівнюють нулю, скоріше за все це свідчить про некоректно зібрані дані, адже з такою кількістю захворівших навряд можна припустити, що одного дня було захворювань більше 2000 людей, а вже наступних два днів було 0. В контексті обраного методу, вхідні дані є дуже важливими, адже саме за допомогою інформації про них відбувається оптимізація.

Отримані результати після моделювання занесемо до результуючої Таблиці

Таблиця 4.2- Результати для моделей з оптимізованими параметрами

Країна	Кількість інфікованих (реальні дані)	Прогнозовані дані
США	315 100	216 900
Україна	16 590	14 230
Іспанія	93 820	38 460

Порівнюючи всі отримані результати після оптимізації параметрів, можна підсумувати, що вирішення задачі найменших квадратів за допомогою методу Левенберг-Марквардта зменшує в десяткі разів помилку відносно звичайної моделі SIR, також показує близький результат відносно спалахів, за умови чітко заданих вхідних параметрів та статистичних даних. Також показує вірно і термін спадання кількості захворівших, що дозволяє оцінити приблизно закінчення жорстких карантинних мір або зовсім їх відміни.

Недоліком звісно є те, що можна спрогнозувати тільки один спалах в такому довготривалому прогнозі, що не дає змогу побачити повну картину розвитку захворювання, пропускаються менші спалахи, які теж мали велику кількість захворівших.

4.6 Вплив температури на перебіг захворюваності

На рахунок захворюваності на COVID-19 вже є безліч досліджень та припущень. Наприклад, на тему того, що COVID-19 є досить схожим захворюванням до грипу.

Грип і COVID-19 є заразними респіраторними захворюваннями, але вони є викликаними різними вірусами. COVID-19 викликаний коронавірусною інфекцією, вперше був виявленим у 2019 році. В свою ж чергу грип є викликаним зараженням вірусом грипу.

І COVID-19, і грип можуть передаватися від людини до людини, які знаходяться в тісному контакті один з одним (приблизно на відстані 1.5 метрів). Обидва поширюються в основному за допомогою великих і дрібних частинок, що містять вірус, які виділяються, коли хворі (COVID-19 або грип) кашляють, чхають або розмовляють. Ці частинки можуть потрапити в рот або ніс людей, які знаходяться поблизу, і тим самим чином потрапити в легені. У деяких випадках, наприклад, у приміщенні з поганою вентиляцією, дрібні частинки можуть поширюватися на довжину більше ніж 6 футів і викликати інфекцію з більш високою ймовірністю.

Хоча найбільше поширюється через вдихання, можливо, що людина може заразитися, доторкнувшись до іншої людини (наприклад, потиснувши руку людині, яка має вірус на руках), або доторкнувшись до поверхні чи предмета, на якому є вірус, а потім торкатися власного рота, носа чи очей.

І віруси грипу, і вірус, що викликає COVID-19, можуть передаватися іншим людям до того, як у них почнуть проявлятися симптоми; людьми з дуже легкими симптомами; і людьми, які ніколи не відчують симптомів (безсимптомні люди).

Наскільки відомо на даному етапі досліджень, COVID-19 поширюється легше, ніж грип.

Порівняно з грипом, COVID-19 може викликати більш серйозні захворювання у деяких людей. COVID-19 також може тривати довше, перш ніж у людей з'являться симптоми, і люди можуть залишатися заразними протягом більш тривалого періоду.

Не можливо відрізнити грип від COVID-19, просто опираючись на симптоми, оскільки вони мають деяку схожість між собою. Для цього необхідно пройти обстеження, щоб визначити, що це за хвороба, і підтвердити діагноз, а саме пройти спеціальне тестування на виявлення, до прикладу, COVID-19, тим самим підтвердивши діагноз або навпаки схилитись до діагнозу грипу.

Хоча тема COVID-19 і вірус, який його викликає, дізнаються все більше, є ще речі, такі як стани після COVID (також відомі як тривалий COVID), які невідомі.

Зважаючи на припущення, що COVID-19 все таки має багато схожого з грипом, існують також припущення на рахунок того, що COVID-19 може мати деяку сезонність. Як відомо грип має деяку сезонність і з'являється тільки в холодну пору року на цю тему є велика кількість досліджень.

Схожі дослідження були проведені і з COVID-19, а саме провели дослідження і пов'язали температуру повітря і швидкість розповсюдження вірусу. Дослідження показали, що на передачу коронавірусу можуть вплинути кілька метеорологічних факторів, включаючи температуру і вологість, що впливає на виживання вірусу в шляхи передачі. Численні епідеміологічні та лабораторні дослідження досліджували зв'язок між передачею COVID-19 і метеорологічних факторів, хоча ці висновки є суперечливими виявили, що 60,0% підтверджених випадків COVID-19 відбувалося в місцях, де температура повітря коливається від 5 °C до 11 °C [39]. Дослідження 2020 року виявили, що значні спалахи COVID-19 в населення

поширені в регіонах із середньою температура від 5 до 11 °C у поєднанні з низькою вологістю (3–6г/кг) [39].

Спробуємо пересвідчитись чи насправді погода в проміжку впливає на COVID-19. Аби виконати порівняння потрібно взяти середню температуру по трьом обраним країнам та порівняти з даними про інфікованих людей.

Відповідно до дослідження, яке було взято за основу вказано, що під час того як температура має значення від 5 до 11 градусів, найчастіше зустрічаються спалахи.

Розглянемо температуру повітря, яка є середньою для кожного місяця для обраної з країн. Для кожної окремо представлені графіки на рисунку

Середня температура для США протягом 2020 року представлена на рис. В моделі ж буде використано набір даних завантажений безпосередньо з ресурсу. Графіки температур представлені для наглядності коливання температур країн.

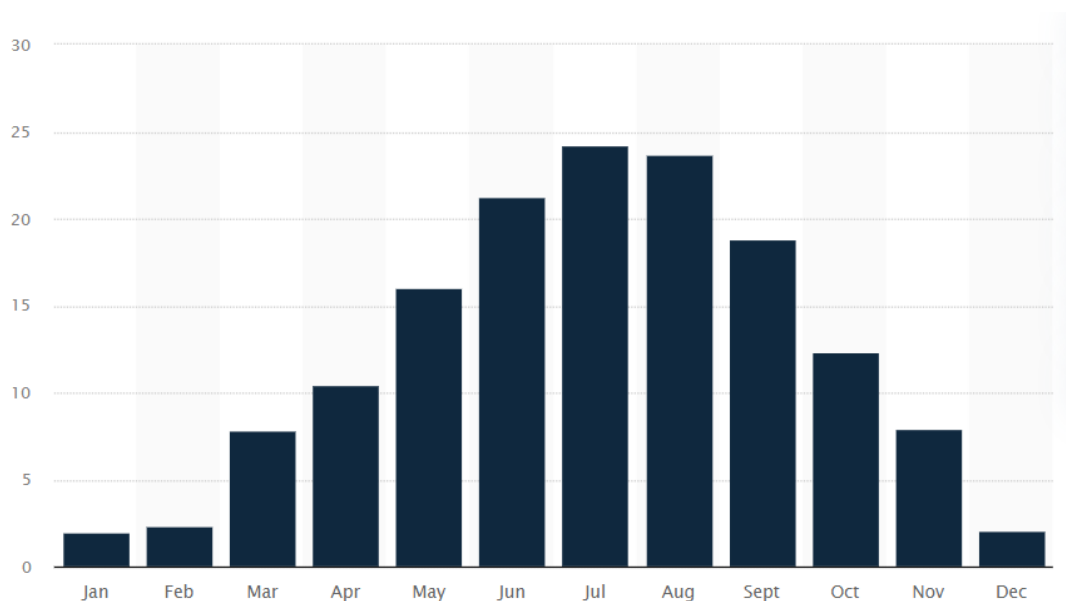


Рисунок 4.10 – Средньомісячна температура на 2020 рік для США[40]

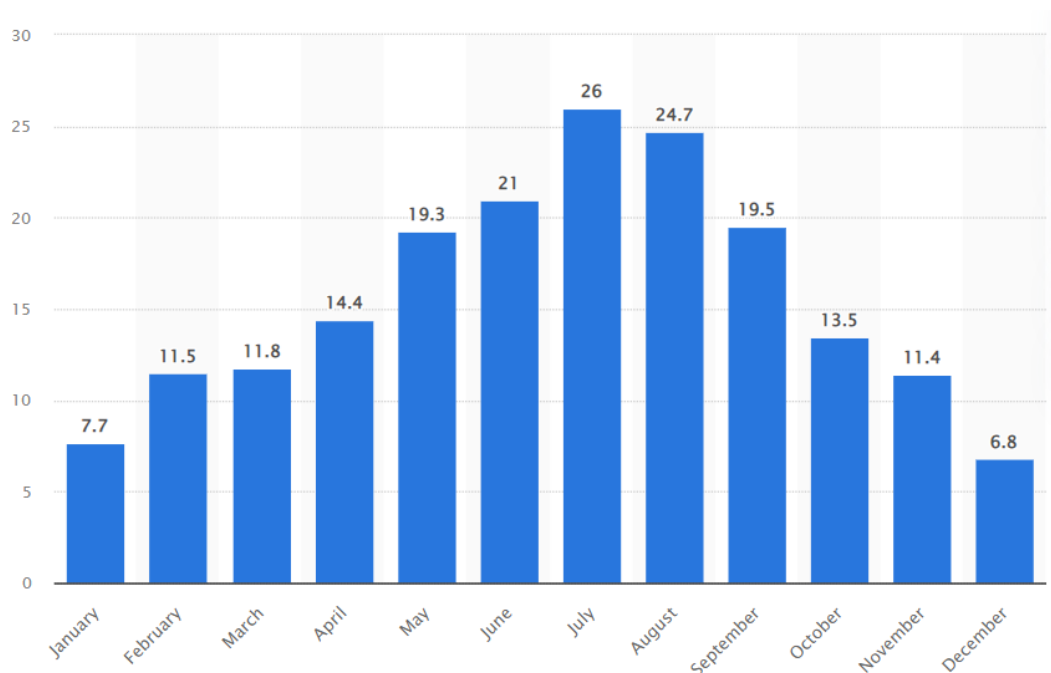


Рисунок 4.11 – Средньомісячна температура на 2020 рік для Іспанії [41]

Дані для України не були представлені у вигляді графіку, але дані можливо було зберегти в текстовому форматі, які будуть використані для моделювання.

Зауважимо, що було обрано час моделювання від 13 березня 2020 року, тобто прогнозування було виконано до 13 березня 2020 року. Обрано було три різні країни, де клімат є зовсім різним по всій території, а саме, в Іспанії набагато тепліший, порівняно з Україною, також в Іспанії відсутня холодна зима, тобто температура рідко опускається нижче 0 градусів.

Потрібно проаналізувати в який же з періодов відносно обраних даних відповідно до дослідження Хуанга очікується найбільша кількість захворювань. Для цього відповідно зібраних даних по температури середній за місяць для кожної країни обираємо місяці, які включають в себе відповідні температурні межі. Зауважимо, що обрані відрізки будуть різними для кожної країни, зважаючи на різний клімат.

Середня температура для США в 5-11 градусів припадає на такі місяці : березень, квітень, жовтень і листопад. Так як для моделювання було обрано не початок року, а саме середина березня, то потрібно звернути увагу на перші 47 дні в моделі – це буде проміжок березнь-квітень, жовтень-листопад - це з 201 дня по

261. Обираємо дані інфікованих людей на цих проміжках, зобразимо відповідні відрізки на графіку за весь рік (рисунок 4.12)

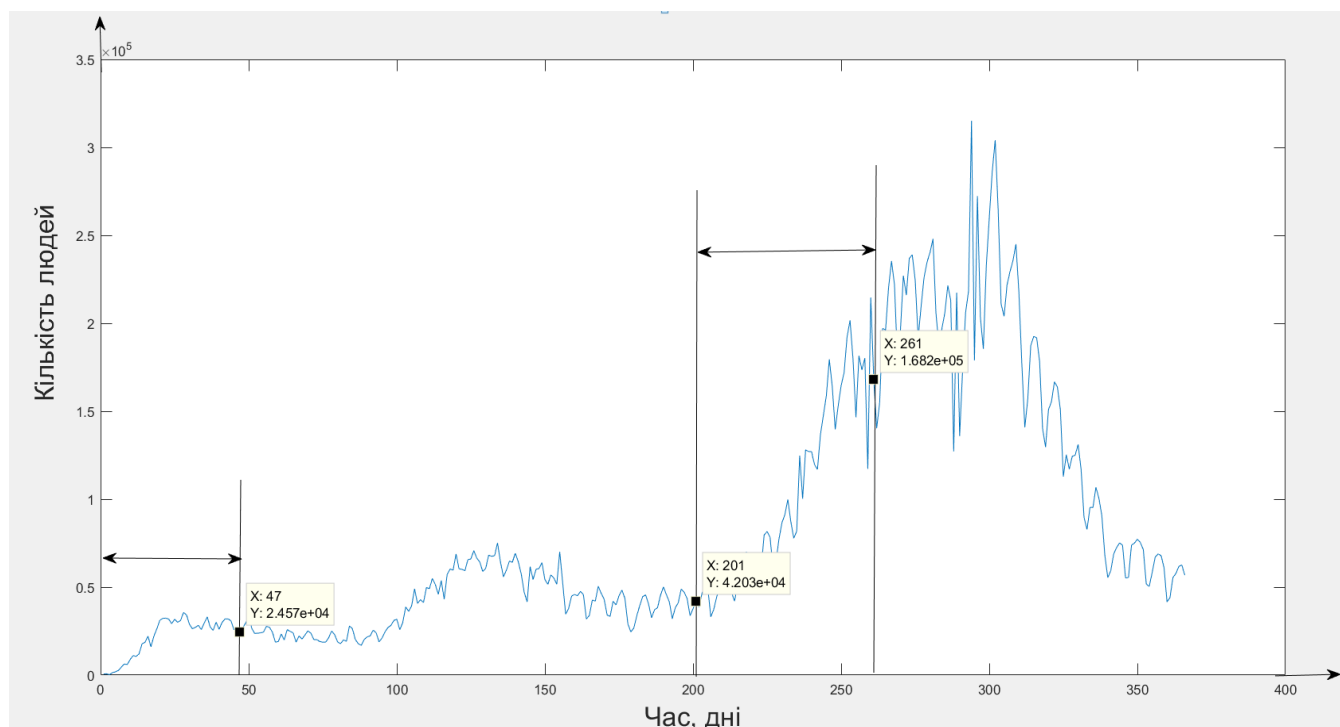


Рисунок 4.12 - Графік інфікованих в США та температурні зміни

Для України період в який температура з 5 до 11, це весна, а саме березень, квітень, жовтень, листопад. Розділимо на такі зони і графік інфікованих для України. Для обраних даних це дні з першого по 47 та 201 і так само як і для США, це буде по 261 день (рисунок 4.13)

Для Іспанії проміжки дещо відрізняються, адже клімат відрізняється від попередніх країн. В Іспанії досить багато місяців в яких температура повітря є набагато вищою за 11 градусів. Місяці в які температура коливається приблизно від 5 до 11 градусів це : січень, лютий, листопад та грудень, тобто всі місяці зими та листопад. В рамках нашого дослідження дні це такі дні з 232 дня по 352 день. Після нанесення на графік відповідних ліній, які вказують на ці дні, де коливалась температура в припустимі параметри отримаємо результати, які зображені на рисунку 4.14.

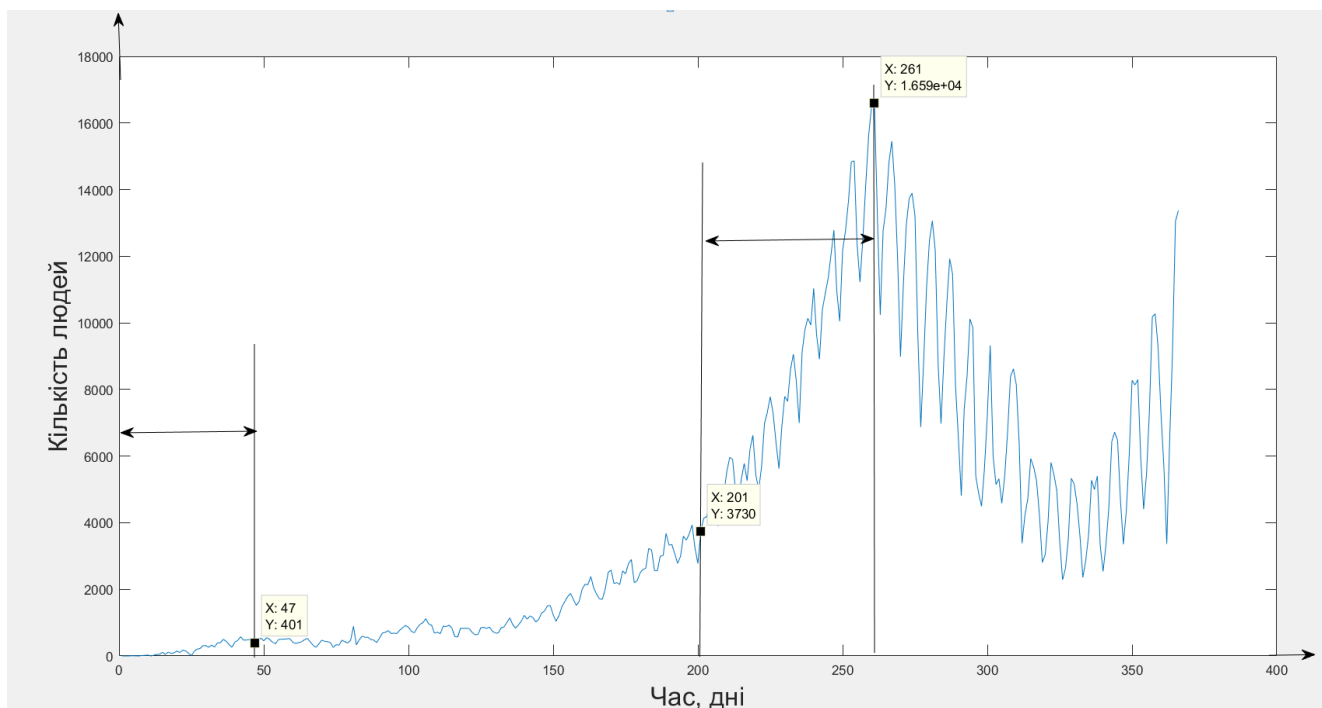


Рисунок 4.13 – Графік інфікованих в Україні та температурні зміни

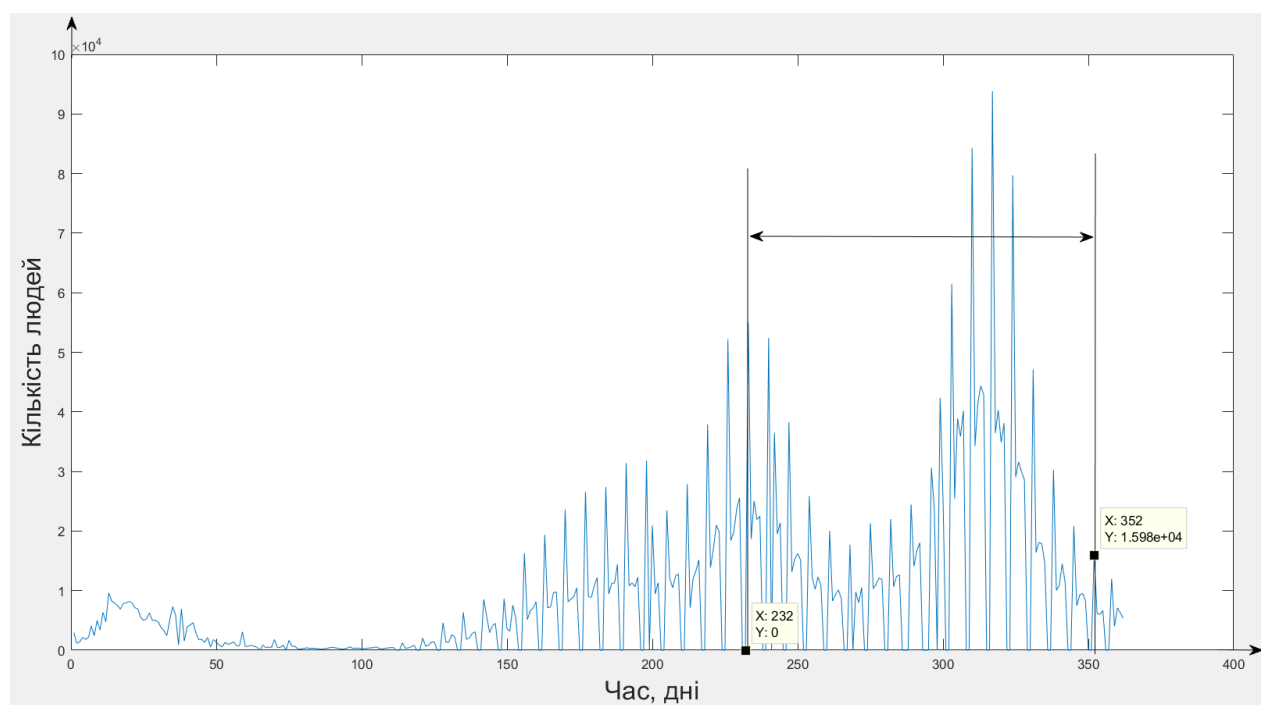


Рисунок 4.14 – Графік інфікованих в Іспанії та температурні зміни

Після побудови графіків можна оцінити чи припадають спалахи хвороби саме в проміжках де температура була в межах від 5 до 11 відповідно до дослідження [39].

Для США на початку захворюваності на COVID-19 було зовсім небагато захворювань і перший, так званий, температурний коридор не відповідає тому, що припадає багато інфікувань на цей період. В наступний же коридор з графіку можна зробити висновок, що інфікування збільшується швидкими темпами, але в обрані межі не потрапляє найбільший спалах.

Для України в обраних межах відносно температури, також видно, що в першому проміжку тільки починається захворюваність і на тому етапі захворювань ще не прогресувало так швидко, а от вже на наступному відрізку часу потрапляє велика кількість захворювань і навіть сам момент спалаху на 2020 рік. Для Іспанії на відрізку і справді потрапляє і спалах і велика кількість захворювань.

З отриманих графіків та аналізу даних відносно температури повітря можна сказати, що насправді деякий вплив на розповсюдження хвороби. Звичайно температурні межі були взяті досить скориговані і проміжки не були такими точними, адже середньомісячна температура дещо виходить за межі 5 та 11 градусів, але тільки в декілька десятих. На основі дослідження з температурою було сказано, що найбільша кількість захворівших припадає на період з зазначеною температурою приблизно для 60% хвороб .

Для обраних країн і справді більша кількість хвороб припадає на ці проміжки, але температура там досягає і більше градусів, до прикладу для США в жовтні середня температура повітря досягає 11.7 градусів, станом на 2020 рік, тому можна сказати, що обрані проміжки для температури краще обирати від 5 градусів і до 12, адже також і для України межі дещо більші.

Незважаючи на те, що обрано більший температурний коридор, ніж було зазначено у дослідженні Хуанга, зв'язок значення температури повітря має зв'язок з швидкістю зараження COVID-19 і так само має деяку сезонність як і грип, має досить великий відсоток.

4.7 Математична модель для моделі розповсюдження з врахуванням температури

Цікавим припущенням є те, що температура повітря може впливати на розвиток COVID-19, для того аби пересвідчитись в цьому потрібно побудувати модель.

Для моделі також буде використано задачу для найменших квадратів, яку буде вирішено за допомогою методу Левенберг-Марквардта лише з деякими відмінностями.

За основу також модифікована модель SIR, яка представлена у розділі, розроблена, яка включає в себе потенційний вплив температури на захворюваність COVID-19.

Також приймається, що все населення це N , все населення поділяється на декілька груп, які вже були розглянуті у попередніх розділах, а саме, йде мова про базову модель, де був прийнятий поділ на три групи – сприйнятливі, інфіковані, видужавші/померлі, при підстановці в рівняння формули ($N = S + I + R$) буде отримана модифікована модель.

Розглядаємо модель, в якій будуть модифіковані саме параметри, які мають безпосередній вплив на розповсюдження захворювання і впливають на результати моделювання. Для того аби отримати авторську модель, де буде враховано температурний режим для країни, а саме середньомісячна температура повітря протягом року. цього буде додано вплив температури в коефіцієнтах передачі вірусу. Після таких модифікацій буде отримано два диференціальні рівняння, які будуть вирішені за схожим принципом до попереднього підрозділу.

Тобто тепер рівняння для визначення інфікованої групи матиме вигляд:

$$\frac{di}{dt} = (R_0 * \beta - \gamma) * i - (R_0 * \beta) * i^2 - (R_0 * \beta) * i * (R_0 * \beta - \gamma)$$

$$\frac{dr}{dt} = \gamma * i$$

де $\beta = \beta_0 + \beta_1 * F(T)$.

Для того аби отримати β , приймаємо, що β_0 - це початкове значення, без використання значення температури, тобто значення, яке було взяте в попередніх моделюваннях для кожної країни відповідно. В свою чергу $F(T)$ - це попердньо визначені функції місцевої температури.

Значення температури будуть взяти для кожної країни окремо з відкритих даних. Температура повітря буде обрана саме середня для кожної країни та відповідно взята на довготривалий термін, а саме – рік, як і було в попередніх моделях дослідження. Середня температуру обирається окремо для кожного місяця. Дану модель також можливо модифікувати, якщо брати середню температуру на день, що може дати більш реальну картину, але в даній обрано на місяць. Звичайно можна було б обрати і середньорічну, але тоді б не було можливим прослідкувати чи справді на проміжку від 5 і приблизно до 11 градусів за Цельсієм отримуємо більшість випадків захворюваності.

4.8 Моделювання моделі в програмному пакеті MATLAB

Алгоритм моделювання моделі у програмному середовищі MATLAB :

- задаються диференціальні рівняння для моделі SIR;
- обраховуються значення для диференційних рівнянь за допомогою ode45;
- вибирається вже обрахований параметр R з базової моделі для подальшого використання;
- описується вже модифікована модель SIR з використанням температури;
- задаються вхідні параметри для температури;
- задається опис параметра з використанням функції відносно температури повітря;
- обраховуються диференціали для модифікованої моделі використовуючи ode45;

- задаються початкові дані для методу Левенберга-Марквардта;
- виконується оптимізація параметрів за допомогою `lsqcurvefit`;
- за результатами виконаних попередніх кроків описується метод для побудови графіків.

Всі початкові значення були взяті з моделей, які вже були зазначені у дисертації, а саме: початкові дані для інфікованих осіб, параметри R_0 , β та γ . Середня температура кожної країни задана в форматі текстового документу зі значеннями, який за допомогою MATLAB можна відкрити, використовуючи команди для зчитування файлу та обрати потрібний стовпчик даних.

Після виконання всіх кроків використовуючи задані параметри для кожної з трьох країн були побудовані графіки (рисунки 4.15-4.17).

Для США модель в якій використано додатковий параметр як температура показала результат, де видно не тільки одну хвилю, а й менші піки хвороби, які теж є важливими для того аби можна було робити подальші прогнози на рахунок розвитку ситуації з захворюваністю.

Також бачимо, що результати стали максимально наближеними до реальних, що свідчить про те, що все таки модель з використанням впливу температури покращила результати.

Після моделювання для України, отримали графік, який як і для США показує досить хороші результати. З графіку бачимо, що вийшло відокремити невеликі спалахи та основний пік, що є значним покращенням, адже тепер можна спостерігати розвиток хвороби на всіх фазах, що допоможе в подальшому робити більш вірні прогнози використовуючи дану модель.

Модель з оптимізованими параметрами для Іспанії на рисунку , також має декілька піків, але все одно містить помилку і не досягає такого сильного наближення як для попередні змодельовані для інших країн.

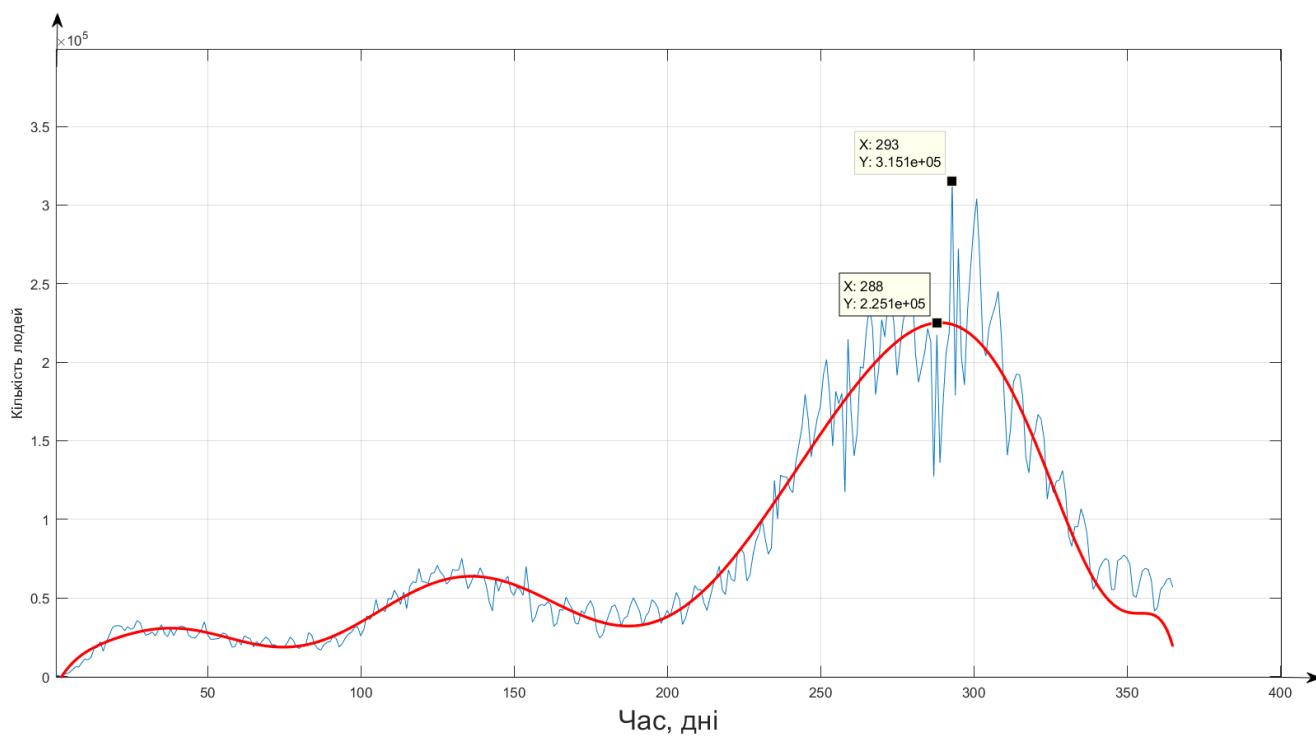


Рисунок 4.15 - Модель дослідження враховуючи T для США

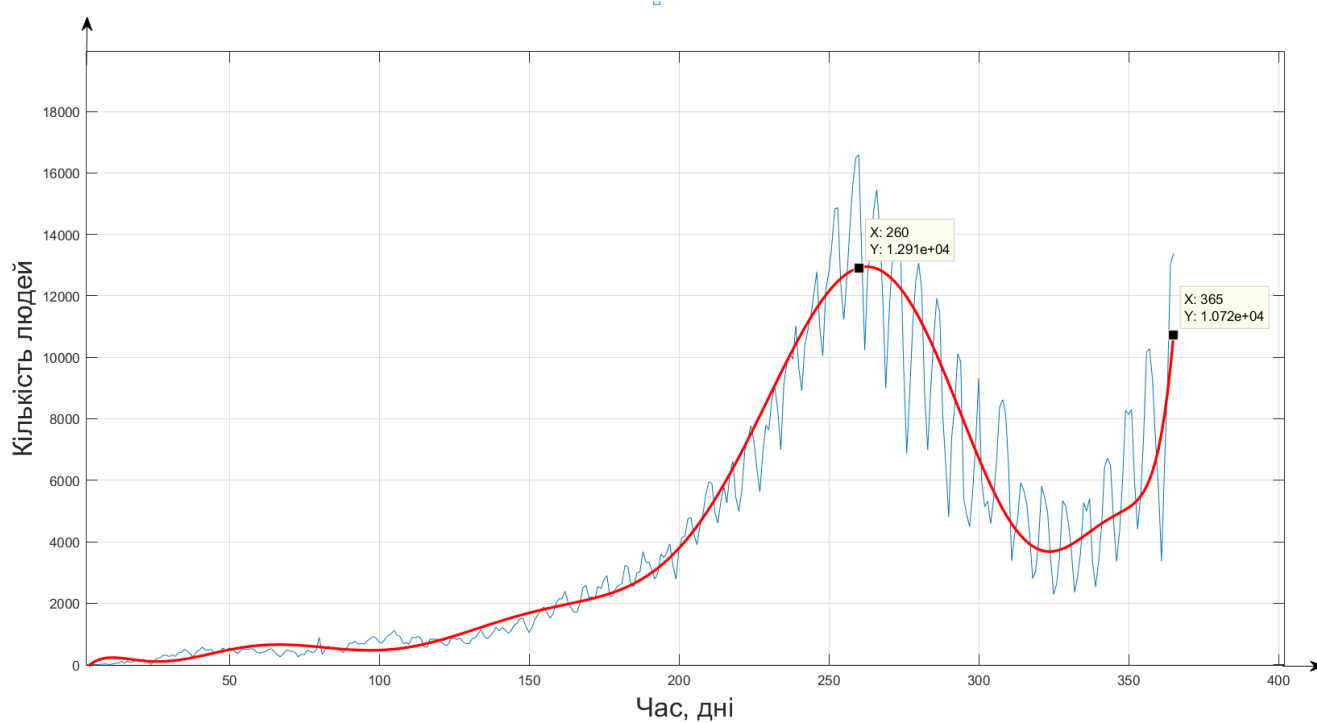


Рисунок 4.16- Модель дослідження враховуючи T для України

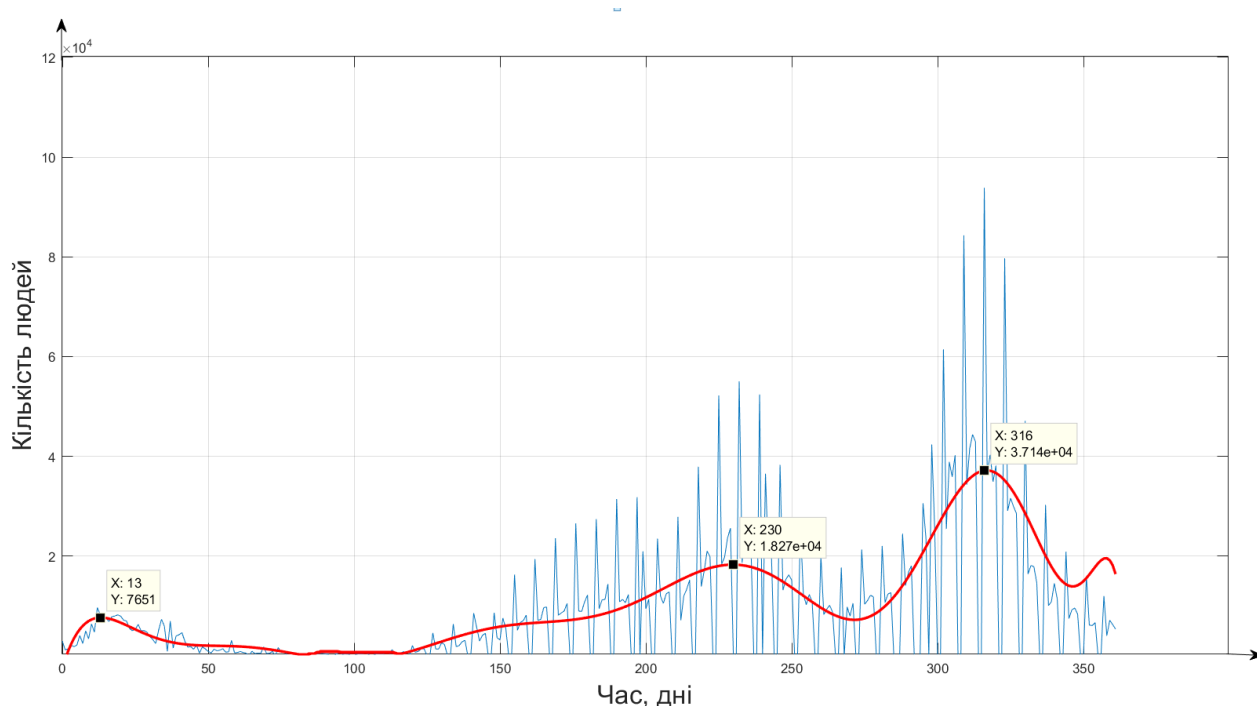


Рисунок 4.17 - Модель дослідження враховуючи T для Іспанії

У випадку з моделюванням для Іспанії припускається, що є недосконалість в зібраних даних, вони є не досить точним, адже бачимо часті перепади до нуля та різкі спалахи, насправді, така ситуація можлива була б, але зважаючи на те, що дані беруться на цілу країну та з таким

Також потрібно враховувати, що пандемія для Іспанії мала зовсім не стандартний розвиток як в усіх інших країнах та була непередбачуваною, зважаючи на те, що велика кількість карантинних норм була запроваджена у країні, а розвиток хвороби не вщухав.

Для більш точного порівняння результатів занесемо деякі пікові точки спалахів для кожної країни у таблиці 4.3.

Для США обираємо три пікові точки, які були зафіксовані в прогнозованих даних та обрахуємо відхилення від реальних даних у цей день, адже як прийняли в минулих розділах, що деякі раптові спалахи є непередбачуваними і суттєво не впливають на результати всієї моделі загалом.

Таблиця 4.3 – Результати моделювання враховуючи температуру на спалахах для США

День	Дані реальні	Прогнозовані
38	30 080	30 810
135	55 840	63 780
288	217 400	225 100

Для України обрано дві точки в 262 день та вже на 365, адже друга хвиля на графіку видно, що починається тільки з 365 дня, а до 262 маємо поступове зростання без інших зростань. Результати в ці два пікові моменти на графіку занесені до Таблиці 4.4

Таблиця 4.4 - Результати моделювання враховуючи температуру на спалахах для України

День	Дані реальні	Прогнозовані
260	16 590	12 910
365	13 380	12 620

Для Іспанії обрано три різні точки (дні), враховуючи те, що в цій країні було зафіксована найбільше випадків і не одно разові спалахи, які були в декілька хвиль., то три піки є цілком закономірними і відповідають пікам відносно даних. Зростання за 365 днів відбувається декілька разів, вони і занесені до таблиці 4.5

Таблиця 4.5 – Результати моделювання враховуючи температуру на спалахах для Іспанії

День	Дані реальні	Прогнозовані
13	8 271	7 651
230	22 352	18 270
316	93 820	37 140

Виходячи з результатів моделювання можна виділити декілька основних пунктів, які вказують на переваги методу з використанням температурного параметру. За допомогою вдосконаленого алгоритму, який враховує температуру вийшло змодельовати не тільки один спалах, а й відтворити деякі з коливань захворювань, що також є важливим, так як у випадку з COVID-19, прослідковується деяка сезонність і спалахи можуть відбуватись не тільки один раз на рік, то ж за допомогою такої моделі можна виконувати прогнозування і бачити декілька піків, що є дуже цінним.

Для кожної країни отримали різну кількість коливань, які припадають на той період, коли були коливання температури приблизно в межах від 5 до 11 градусів. Побудовані за допомогою вдосконаленого алгоритму графіку відображають більш точну картину перебігу подій під час захворювання на COVID-19. Для кожної з країн після моделювання бачимо, що графіки мають відповідні коливання як і в реальних даних.

За допомогою вдосконаленого алгоритму, графіки наблизились до реальних даних, але не для кожної країни. Існує багато чинників, які б могли вплинути на результати моделювання за допомогою цієї моделі дослідження. Один з них, це те, що брали температуру середню по всій країні, адже температура у кожному регіону є різною, що також може впливати, важливим є те, що точність моделювання можна було б підвищити використовуючи дані, які враховують температуру

сердьодобову, але відкритих даних для обраних трьох країн не було у відкритому доступі задля використання в даній дисертації.

4.9 Висновки до розділу

В даному досягнуто такі цілі як :

- розглянуто задачу найменших квадратів та метод вирішення її;
- за допомогою методу Левенберг-Марквардта була виконана оптимізація параметрів, що призвело до покращення результатів прогнозування;
- представлений метод для удосконалення вирішення задачі методом Левенберг-Марквардта, а саме додаткове значення температури для коефіцієнта розповсюдження вірусу
- виконане моделювання для авторської моделі дослідження розповсюдження епідемічного захворювання COVID-19 на базі SIR моделі з використанням додаткового параметру, який враховує температурні зміни.

5 ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛІ З ВРАХУВАННЯМ ТЕМПЕРАТУРИ ВІДНОСНО ВЖЕ ІСНУЮЧИХ МОДЕЛЕЙ

Після виконаних модифікацій моделей доцільним було б порівняти чи насправді авторська модель показала кращі результати за базові вже існуючі моделі.

Для того аби порівняння було коректним потрібно обрати на основі чого буде виконане порівняння. Не раціональним буде порівнювати в даній роботі по групі сприйнятливих осіб, адже не було враховано ніякий коефіцієнт, який би вплив на цю групу, а саме враховувалась смертність не тільки по обраній хворобі, а й за інших обставин або ж враховувалась народжуваність, що теж не було представлено в роботі. Тому сприйнятливую групу порівнювати не будемо.

Для інкубаційної групи була побудована лише один вид моделі і всі наступні модифікації не містили її, тому до порівняння вона не буде розглянута.

Для група видужавших і померлих була представлена у кожній моделі, але для того аби оцінити точність потрібні також і реальні дані. В роботі не було представлено джерел з точною статистикою по смертності або одужанні пацієнтів, тому дану групу теж доцільно не розглядати.

Група інфікованих була представлена у кожній моделі, також є багато джерел, де доступні відкриті дані і представлена кількість захворівших на кожного дня протягом року для трьох обраних країн, які розглядались у роботі.

Для того аби оцінити, яка модель показала найкращий результат відносно реальних даних потрібно обрати спосіб оцінки. Так як формул саме для такого порівняння запишемо для кожної країни таблицю, де занесемо всі результати для кожного методу та подивимось, де відхилення від реальних даних було найменшим. Тобто виконаємо віднімання від реальних даних на піку від прогнозованих.

Спершу розглянемо всі методи та порівняємо, який показав найбільш наближений результат до реальних даних для США.

Запишемо всі дані в одну таблицю 5.1 порівняти, який метод показав найкращий результат, тобто найбільш близький до реальних даних, тим самим, який буде кращим для більш точних прогнозів.

Таблиця 5.1 – Результати всіх досліджень для США

Назва методу	Кількість інфікованих (реальні дані)	Кількість інфікованих прогнозована	Помилка
Базова модель	315 100	27 080 000	26 764 900
Авторська модель з коефіцієнтом демпфування	315 100	216 900	98 200
Авторська модель з температурним параметром	315 100	225 100	90 000

За таким самим принципом запишемо всі отримані дані для України до однієї таблиці 5.2.

Таблиця 5.2 - Результати всіх досліджень для України

Назва методу	Кількість інфікованих (реальні дані)	Кількість інфікованих прогнозована	Помилка
Базова модель	16 590	569 300	552 710
Авторська модель з коефіцієнтом демпфування	16 590	14 430	2 160
Авторська модель з температурним параметром	16 590	12 910	3 680

Та для порівняння результатів, також запишемо всі дослідження для Іспанії та обрахуємо помилку (Таблиця 5.3)

Таблиця 5.3 - Результати всіх досліджень для Іспанії

Назва методу	Кількість інфікованих (реальні дані)	Кількість інфікованих прогнозована	Помилка
Базова модель	93 820	2 150 000	2 056 180
Авторська модель з коефіцієнтом демпфування	93 820	38 460	55 360
Авторська модель з температурним параметром	93 820	37 140	56 680

Обрахуємо помилку у відсотках та порівняємо, де ж була найменша помилка при моделюванні відносно реальних даних, обчислення виконаємо за такою формулою :

$$e = \frac{|(\text{реальні дані} - \text{прогнозовані})|}{\text{реальні дані}} * 100\%$$

Для США обрахуємо помилку для базової моделі

$$e_{\text{США}_{\text{баз}}} = \frac{|315100 - 27080000|}{315100} * 100\% = \frac{26764900}{315100} * 100\% = 8494,09\%$$

Також обрахуємо для США авторської моделі з коефіцієнтом демпфування :

$$e_{\text{США}_{\text{автдем}}} = \frac{|315100 - 216900|}{315100} * 100\% = \frac{98200}{315100} * 100\% = 31,16\%$$

Та знаходимо значення помилки для авторської моделі з температурним параметром

$$e_{\text{США}_{\text{авттем}}} = \frac{|315100 - 225100|}{315100} = \frac{90000}{315100} * 100\% = 28,56\%$$

Наступною країною в дослідженні була Україна, то ж обраховуємо помилку для кожної з трьох моделей

Перше значення помилки обраховуємо для базової моделі:

$$e_{\text{укр}_{\text{баз}}} = \frac{|16590 - 569300|}{16590} * 100\% = \frac{552710}{16590} * 100\% = 3331,59\%$$

Для моделі з коефіцієнтом демпфування для України отримаємо :

$$e_{\text{укр}_{\text{автдем}}} = \frac{|16590 - 14430|}{16590} * 100\% = \frac{2160}{16590} * 100\% = 13,02\%$$

Та обраховуємо для моделі з температурним коефіцієнтом для України:

$$e_{\text{укр}_{\text{авттем}}} = \frac{|16590 - 12910|}{16590} * 100\% = 22,18\%$$

Такі ж кроки для обрахування помилки зробимо і для моделей, які були побудовані для Іспанії

Базова модель для Іспанії має таку помилку:

$$e_{\text{Ісп}_{\text{баз}}} = \frac{|93820 - 2150000|}{93820} * 100\% = \frac{2056180}{93820} * 100\% = 2191,62\%$$

Помилка для Іспанії в авторській моделі з коефіцієнтом демпфування склала :

$$e_{\text{Іспавтдем}} = \frac{|93820 - 38460|}{93820} * 100\% = \frac{55360}{93820} * 100\% = 59\%$$

Для моделі з параметром, який враховує середньомісячну температуру для Іспанії: отримаємо:

$$e_{\text{Іспавтдем}} = \frac{|93820 - 37140|}{93820} * 100\% = \frac{56680}{93820} * 100\% = 60,41\%$$

З отриманих результатів можна зробити висновки, що моделювання за допомогою SIR моделі дає велику помилку і точно погано підходить для відображення реальної ситуації по розповсюдженню COVID-19. Адже, при використанні такої моделі не враховується особливість перебігу захворювання, а також ще безліч параметрів, які мають бути враховані для відображення повної картини і змоги передбачити подальші наслідки розповсюдження вірусів. Помилка склала для кожної з обраних країн від 2191,62% аж до 8494,09%, що є неприпустимим в моделі для дослідження розповсюдження епідемічного захворювання і несе небезпеку для всієї планети.

Набагато меншу помилку, а саме від 31,16% до 59% показують результати моделювання при використанні методу з коефіцієнтом демпфування, де виконується оптимізація параметрів. Виходячи з результатів, помилка стає меншою в декілька тисяч відсотків і є це дуже добрий показник, адже при такому моделюванні змоделювати приблизно як буде розвиватись ситуація і передбачити всі потрібні запобіжні міри. Помилка, наприклад, для США є меншою, ніж базова модель аж на декілька тисяч відсотків, тому більш доцільним для моделювання розповсюдження епідемічних захворювань було б виконати саме за допомогою оптимізації параметрів, а не базової моделі, якщо порівнювати саме ці дві моделі.

Якщо ж брати до уваги наступний запропонований метод, а саме авторська модель, де була використана не тільки оптимізація параметрів, а ще й був врахований параметр температури, то бачимо, що для різних країн модель показала різні результати, до прикладу, для США помилка стала меншою відносно обох попередніх моделей і результат з авторською моделлю з параметром температури показав найменшу помилку в 13.02%, що є хорошим показником і є найближчим до реальних даних. Натомість для України та Іспанії помилка стала дещо більшою для авторської моделі з температурою, ніж для моделі з коефіцієнтом демпфування. Звичайно, це є деяким недоліком цієї моделі, але також потрібно враховувати, що температуру повітря було обрано досить приблизно і тільки середньомісячну, тобто не виключено, що якщо обрати наприклад середньодобову температуру, то прогнозування мало б меншу різницю відносно реальних даних.

Після підрахунку значення помилки можна зробити висновок, що найкращі результати показав метод з коефіцієнтом демпфування, хоч метод з врахуванням температури показав найкращий результат для США. Для Іспанії бачимо, що всі методи показують значну помилку і причиною цього є дані на основі яких були побудовані моделі, як бачимо, часті спади до 0 не дають точної інформації про весь перебіг хвороби і не відображають реальну ситуацію для цієї країни. Для України метод з врахуванням температури має незначне відхилення, але досить точно відображає повну ситуацію по всіх спалахах захворюваності.

Також потрібно враховувати, що загалом всі моделі для дослідження розповсюдження епідемічних захворювань, які виконуються на такий довготривалий термін мають велику помилку через це і існує так мало досліджень тому, що звести до мінімуму цю помилку досить складно.

На перебіг захворювань впливає безліч чинників і для того аби як можна краще зробити прогнозування за допомогою математичних моделей, потрібно розділяти населення на більшу кількість підгруп за віком, статтю, можливо, за вродженим імунітетом, також потрібно б було врахувати всі способи набуття імунітету від саме COVID-19, адже з часом почала з'являтися вакцина, але яка також має різний рівень захисту. Дана вакцина з'явилась не на початку обраного в

даний дисертації проміжку часу, також вона була не така розповсюджена по світу і була тільки на етапі розробки і деяких з них на етапі тестувань. Враховуючи всі перелічені фактори, а також ще ті, які були тут не згадані, зменшення помилки до декількох десятків тисяч відносно реально захворівших людей є досить хорошим результатом, який надасть можливість спрогнозувати ситуацію, яка буде більше походити саме на приблизно реальний перебіг подій і надасть змогу вводити карантинні норми у відповідних межах.

Між методом з оптимізацією параметрів і тим самим методом ще й додаванням температури, помилка і стає дещо більшою відповідно реальних даних, але основною перевагою є те, що метод Левенбарг-Марквардта і з додавання функції залежної від температури дає змогу оцінити не тільки результати на піку і побачити, яка ж буде там ситуація, а й показує ще одні важливі результати, а саме які коливання буде мати хвороба. Коливання показують в якій мірі потрібно запроваджувати міри безпеки, а саме, карантинні норми.

Також показують на які дні потрібно звернути увагу, до прикладу, якщо після прогнозування отримуємо, що спалахи будуть відбуватись на зимній період, до прикладу, на період початку січня, то можна зробити висновки, що цьому сприяли свята і люди збирались на святкування разом, також причиною спалахів прослідковувались всі масові збори, такі як мітинги тощо.

Якщо ж коливання прослідковуються в період з певною температурою повітря і вони мають повторювальний характер, то звичайно для дослідження це буде свідчити про те, що це буде повторюватись не тільки протягом обраного для моделювання року, а й надалі.

Тобто сама особливість захворювання є такою, що спалахи можуть бути не один раз за такий довгий проміжок і як було виявлено в дослідженні Хуанга і також перевірено відносно обраних країн в дослідженні в рамках дисертаціях, що захворювання все таки має залежність від температури повітря. Зважаючи на те, що ситуація є змінною і в моделі з використанням методу з коефіцієнтом демпфування та функцією відносно температури є більш корисною аби можна було дослідити

ситуацію з розповсюдженням COVID-19 більш детально і бачити на кожному етапі який буде розвиток захворювання, окрім піку.

Проводити дослідження використовуючи моделі прогнозування є досить зручним способом і незамінним в реаліях частих епідемій, враховуючи ситуацію з COVID-19, використання обраних методів в даній дисертації можуть бути вдосконалені та використовуватись у подальших дослідженнях для епідемічних ситуацій.

ВИСНОВКИ

В результаті виконання магістерської дисертації було вдосконалено базову модель дослідження розповсюдження епідемічних захворювань, створеної за допомогою програмного пакету Matlab, а саме :

- розроблено модель дослідження розповсюдження епідемічного захворювання – COVID-19 на базі SIR моделі Кермака-Маккендріка ;
- розроблену розширену модель для застосування методу Левенберга - Марквардта;
- розширено модель з використанням методу Левенберга-Марквардта враховуючи температурні зміни для кожної з країн;
- виконано моделювання відносно реальних даних.

Проведено аналіз моделі дослідження розповсюдження епідемічних захворювань та порівняно з модифікованою моделлю, яка враховує інкубаційний період захворювання. Запропоновано вдосконалення базової моделі, а саме, виконання оптимізації параметрів розповсюдження захворювання.

За результатами моделювання авторської моделі зменшено помилку та отримано такі значення похибки:

- для США 28,56%;
- для України 13,02%;
- для Іспанії 59%.

Окрім змін у функціональній частині моделі, виконано наступне:

1. Написано програмний код, котрий дозволяє виконати прогнозування за допомогою базової моделі.
2. Написано програмний код, котрий дозволяє виконати прогнозування використовуючи авторську модель з виконанням оптимізації параметрів
3. Написано програмний код, котрий дозволяє виконати прогнозування на базі модифікованої базової моделі та застосуванням температурного параметру
4. Досліджено і розраховано помилку відносно реальних даних.

Головна особливість моделі з температурним показником – це можливість виконувати прогнозування на довготривалий термін та відобразити при моделюванні всі спалахи захворювання.

За результатами проведених досліджень було опубліковано статтю:

Репнікова Н. Б Вдосконалення функціональних можливостей моделі прогнозування епідемічних захворювань / Н. Б. Репнікова, П. В. Капченко - Міжнародний науковий журнал "Інтернаука". — 2022. — №4

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. WHO. Health Disease outbreaks. [Електронний ресурс] – Режим доступу: https://www.who.int/topics/disease_outbreaks/en/
2. Global alert to avian influenza virus infection [Електронний ресурс] - Режим доступу: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001451/>
3. Кодекс цивільного захисту України від 02.10.2012 № 5403-VI (Редакція від 12.05.2017). [Електронний ресурс] – Режим доступу: <https://zakon.rada.gov.ua/laws/show/5403-17>
4. P. Spreeuwenberg. Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic/ P. Spreeuwenberg //American Journal of Epidemiology – 2018. - 187 (12) – p. 2561–2567.
5. What Is a Pandemic? [Електронний ресурс] – Режим доступу: <https://www.webmd.com/cold-and-flu/features/what-is-pandemic>
6. Pandemic Influenza Preparedness and Response: A WHO Guidance Document. [Електронний ресурс]- Режим доступу : <https://www.ncbi.nlm.nih.gov/books/NBK143061/>
7. WHO Director-General's opening remarks at the media briefing on COVID-19— 11 March 2020. World Health Organization [Електронний ресурс] – Режим доступу: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
8. ПРО НОВИЙ КОРОНАВІРУС (nCoV), ВІЯВЛЕНИЙ У ПРОВІНЦІЇ ХУБЕЙ, КИТАЙ [Електронний ресурс] – Режим доступу: <http://www.vin.gov.ua/departament-okhorony-zdorovia/23836-pro-novij-koronavirus-ncov-viyavlenij-u-provintsiji-khubej-kitaj>
9. Тестування на COVID-19: різновиди [Електронний ресурс] – Режим доступу: <https://moz.gov.ua/article/health/testuvannja-na-covid-19-riznovidi>

10. Боев Б. В. Прогнозно-аналитические модели эпидемий / Б. В. Боев. – М., 2005. [Электронный ресурс] – Режим доступа : <http://www.armscontrol.ru/course/lectures05a/bvb050324.pdf>
11. Forecasting methods and models of disease spread [Электронный ресурс] – Режим доступа : https://www.researchgate.net/publication/335165992_Forecasting_methods_and_models_of_disease_spread
12. Online detection and quantification of epidemics spread [Электронный ресурс] – Режим доступа : https://www.researchgate.net/publication/5907243_Online_detection_and_quantification_of_epidemics
13. Кондратьев М. А Совершенствование методологии моделирования распространения инфекционных заболеваний // Вычислительные, измерительные и управляющие системы: сборник научных трудов.— СПб. : Политехн. ун-та, 2011. – С. 37–44.
14. Bichindartiz Isabelle, A case-based reasoning system using control case-base, 1994 – p. 38-42
15. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System, Approaches [Электронный ресурс] – Режим доступа : https://www.researchgate.net/publication/225070522_Case-Based_Reasoning_Foundational_Issues_Methodological_Variations_and_System_Approaches
16. Classification of Effort Estimation Methods [Электронный ресурс] – Режим доступа : https://www.researchgate.net/publication/300577568_Classification_of_Effort_Estimation_Methods
17. Штучна нейронна мережа [Электронный ресурс] – Режим доступа : https://uk.wikipedia.org/wiki/%D0%A8%D1%82%D1%83%D1%87%D0%BD%D0%B0_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0_%D0%BC%D0%B5%D1%80%D0%B5%D0%B6%D0%B0
18. Штучні нейронні мережі: обчислення [Электронный ресурс] – Режим доступа : http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/ShtuchnNejronMeregNester2004.pdf

19. Kermack W. O., McKendrick A. G. A Contribution to the Mathematical Theory of Epidemics // Proceedings of the Royal Society. 1927. P. 700–721
20. Brauer, F. Mathematical Models in Population Biology and Epidemiology / F. Brauer, C. Castillo-Chavez // Texts in Applied Mathematics. – Vol. 40. – N.Y. : Springer-Verlag New York, 2012. – 508 p.
21. The impact of contact structure on infectious disease control: influenza and antiviral agents [Электронный ресурс] – Режим доступа : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870680/>
22. Гальченко В. Я., Попов К. Р., Приземина И. Н, Качур Н. В. Прогнозирование временных рядов в задаче оценки эпидемической ситуации заболеваемости ОРВИ и гриппом по данным Луганской области // Український медичний альманах. — 2010. — Т. 13, № 2. — С. 20–22.
23. Coburn B. J., Wagner B. G., Blower S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1) // BMC Medicine. — 2009. — Vol. 7, № 30.
24. Kermack, W; McKendrick, A (1991). Contributions to the mathematical theory of epidemics – I. [Электронный ресурс] – Режим доступа : <https://pubmed.ncbi.nlm.nih.gov/2059741/>
25. Kermack, W; McKendrick, A (1991). Contributions to the mathematical theory of epidemics – II. The problem of endemicity. [Электронный ресурс] – Режим доступа : <https://pubmed.ncbi.nlm.nih.gov/2059742/>
26. Kermack, W; McKendrick, A (1991). Contributions to the mathematical theory of epidemics – III. Further studies of the problem of endemicity [Электронный ресурс] – Режим доступа : <https://pubmed.ncbi.nlm.nih.gov/2059743/>
27. History and Epidemiology of Global Smallpox Eradication [Электронный ресурс] – Режим доступа : <http://emergency.cdc.gov/agent/smallpox/training/overview/pdf/eradicationhistory.pdf>
28. COVID Data Tracker [Электронный ресурс] – Режим доступа : <https://covid.dc.gov/covid-data-tracker/#datatracker-home>

29. MATLAB Documentation [Электронный ресурс] – Режим доступа : <https://www.mathworks.com/help/matlab/>
30. MATLAB . Ordinary Differential Equations [Электронный ресурс] – Режим доступа : <https://www.mathworks.com/help/matlab/ordinary-differential-equations.html>
31. . Virus infections in infants as protection against subsequent infections / F. R. Velazquez, D. O. Matson, J. J. Calva [et al.] // N. Engl. J. Med. - Vol. 335. (1996), - P. 1022-1028.
32. Лоусон Ч., Хенсон Р. Численное решение задач методом наименьших квадратов. — М.: Наука, 1986
33. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. — Т. 2: Айвазян С А. Основы эконометрики. — М.: ЮНИТИ-ДАНА, 2001. — 432 с
34. Метод Левенбрг-Марквардта Equations [Электронный ресурс]–Режим доступа : <http://alglib.sources.ru/optimization/levenbergmarquardt.php>
35. Солдатова, О.П., Применение нейронных сетей для решения задач прогнозирования / О.П. Солдатова, В.В. Семенов – Саратов: СГАУ им. ак. С.П. Королева, 2010. – С. 56-58
36. MODIFIED MARQUARDT SUBROUTINE FOR NON-LINEAR LEAST SQUARES [Электронный ресурс] – Режим доступа : <https://www.osti.gov/biblio/4667484>
37. MATLAB Documentation . Lsqcurvefit [Электронный ресурс] – Режим доступа : <https://www.mathworks.com/help/optim/ug/lsqcurvefit.html>
38. MATLAB Documentation . Optimoptions [Электронный ресурс] – Режим доступа : <https://www.mathworks.com/help/optim/ug/optim.problemdef.optimizationproblem.optimoptions.html>
39. Optimal temperature zone for the dispersal of COVID-19 [Электронный ресурс] – Режим доступа : <https://www.sciencedirect.com/science/article/pii/S0048969720330047?via%3Dihub>

40. Average Monthly Temperature for USA[Электронный ресурс] – Режим доступа : <https://www.statista.com/statistics/513644/monthly-average-temperature-in-the-us-celsius/>

41. Average Monthly Temperature for Spain[Электронный ресурс] – Режим доступа : <https://www.statista.com/statistics/802681/average-monthly-temperature-spain/>

ДОДАТОК А

Перелік опублікованих матеріалів

International Scientific Journal "Internauka" <https://doi.org/10.25313/2520-2057-2022-4>

Технічні науки

УДК 32.973.3

Репнікова Наталія Борисівна

кандидат технічних наук, доцент,

викладач факультету інформатики та обчислювальної техніки

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Репникова Наталья Борисовна

кандидат технических наук, доцент,

преподаватель факультета информатики и вычислительной техники

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского»

Repnikova Nataliya

PhD, Associated Professor of the

Faculty of Informatics and Computer Science

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Капченко Поліна Володимирівна

студент факультету інформатики та обчислювальної техніки

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Капченко Полина Владимировна

студент факультета информатики и вычислительной техники

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Kapchenko Polina

Student of the Faculty of Informatics and Computer Science of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

International Scientific Journal "Internauka" <https://doi.org/10.25313/2520-2057-2022-4>

**ВДОСКОНАЛЕННЯ ФУНКЦІОНАЛЬНИХ МОЖЛИВОСТЕЙ
МОДЕЛІ ПРОГНОЗУВАННЯ ЕПІДЕМІЧНИХ ЗАХВОРЮВАНЬ
УСОВЕРШЕНСТВОВАНИЕ ФУНКЦИОНАЛЬНЫХ
ВОЗМОЖНОСТЕЙ МОДЕЛИ ПРОГНОЗИРОВАНИЯ
ЭПИДЕМИЧЕСКИХ ЗАБОЛЕВАНИЙ
IMPROVING THE FUNCTIONAL CAPABILITIES OF THE EPIDEMIC
DISEASE PREDICTION MODEL**

***Анотація.** Розглядається проблема прогнозування розвитку епідемічних захворювань. Пропонується використання базової моделі прогнозування епідемічних захворювань та методу оптимізації з метою наближення прогнозу розвитку епідемії до її реальних параметрів.*

В ході вдосконалення використовуються методи найменших квадратів для покращення точності результатів прогнозування. Приведено результати моделювання моделі з використанням прикладного пакету Matlab на базі статистичних даних захворювання на COVID-19 Сполучених Штатів Америки.

***Ключові слова:** віруси, моделі дослідження, епідемія, метод найменших квадратів, прогнозування.*

***Аннотация.** Рассматривается проблема прогнозирования развития эпидемических заболеваний. Предлагается использование базовой модели прогнозирования эпидемических заболеваний и метода оптимизации с целью приближения прогноза развития эпидемии к реальным параметрам.*

В ходе усовершенствования используются методы наименьших квадратов для улучшения точности результатов прогнозирования. Приведены результаты моделирования модели с использованием прикладного пакета Matlab на базе статистических данных заболевания COVID-19 Соединенных Штатов Америки.

Ключевые слова: вирусы, модели исследования, эпидемия, метод наименьших квадратов, прогнозирование.

Summary. Considered the problem of predicting the epidemic development of diseases. It is proposed to use a basic model for predicting epidemic diseases and an optimization method in order to bring the epidemic development forecast closer to real parameters.

The improvement uses least squares methods to improve the accuracy of the prediction results. The results of modeling are presented in the Matlab application package on the basis of statistical data on the COVID-19 disease in the United States of America are presented.

Key words: viruses, research models, epidemic, least squares method, forecasting.

Вступ. Методи прогнозування інфекційної захворюваності активно розвиваються в наші дні. В останні роки кількість робіт на цю тему стрімко зростає завдяки розгортанню інформаційних систем нагляду та появи великих обсягів статистики, доступної для аналізу. Епідеміологічні прогнози виконуються для різних термінів і в залежності від них служать різним цілям.

Дана тема є досить актуальною в наші дні, враховуючи спалахи епідемій нових штамів грипу, а також зовсім нових хвороб, таких як COVID-19.

Для того, аби як можна швидше попереджувати зараження великої кількості людей використовуються спеціальні заходи зі сторони виконавчих органів країн. Для передбачення того, як швидко ситуація може стати критичною використовують статистичні дані та різні види моделей прогнозування на їхній основі. Існують полігамні моделі в епідеміології, регресійні моделі та багато інших. Недоліком моделей є прогноз на великі

проміжки часу, адже найчастіше, вони показують велике відхилення від реальних.

Основна частина. Найчастіше використовують полігамні моделі (Рис.1) для прогнозування розповсюдження захворювання. Для дослідження цих моделей використовують звичайні диференціальні рівняння (які є детермінованими). В моделі населення поділяється на три групи: люди, які ще не захворіли, інфіковані та видужавші або померлі.

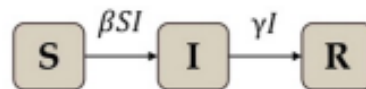


Рис. 1. Базова модель прогнозування розповсюдження захворювання

Модель описується такими рівняннями:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta IS}{N} \\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

де S - люди, які ще не захворіли, I – інфіковані, R – видужавші або померлі

β - інтенсивність зараження, середня кількість контактів на людину протягом періоду часу, γ - інтенсивність одужання .

В даній моделі є припущення, що всі значення є постійними. Це означає, що очікувана тривалість інфекції прямо пропорційна швидкості одужання/смертності.

N – це кількість населення, що є сталим і незмінним числом, і є сумою всіх груп моделі

$$N = S(t) + I(t) + R(t) = Constant$$

Для побудови моделі обрана хвороба COVID-19 та країна Сполучені Штати Америки. З відомих нам параметрів це буде кількість населення станом на 2020 рік і це 333 466 000 осіб.

R_0 – коефіцієнт поширення інфекції. R_0 є безрозмірним числом. Даний параметр для різних країн і різних інфекцій є різним. В залежності від етапу хвороби, також є різним. У період епідемій R_0 завжди більше 1. У випадку з ковідом параметр коливається від 1.4 та 2.8 [1].

Також потрібно враховувати, кількість днів, скільки триває хвороба. Ковід в середньому триває 14 днів, то ж $D=14$. Обираємо $R_0=2$. Побудуємо модель (Рис. 2) за допомогою програмного пакету Matlab на період 365 днів.

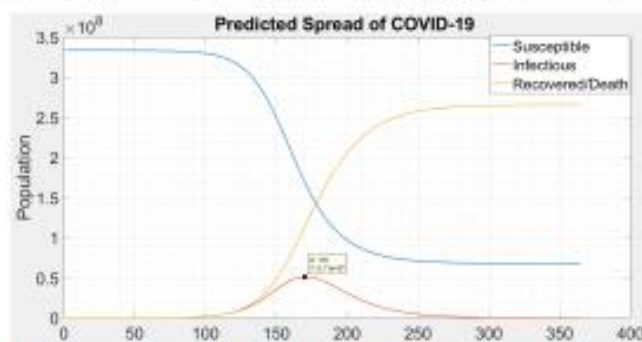


Рис. 2. Модель прогнозування розповсюдження ковіду в США

Джерело: авторська розробка

Для оцінки результатів порівняємо з реальними даними інфікованих від березня 2020 року [2] та побудуємо модель тільки для інфікованих випадків

Після моделювання (Рис. 3) можна зробити висновки, що прогноз має велику похибку, а також не співпадає пік з реальними даними. Проблема полягає в тому, що в базовій моделі не враховується безліч факторів, таких як перебіг хвороби, інкубаційний період, коли особа вже заразна, але ще не відчуває симптомів, а також не враховуються карантинні міри.

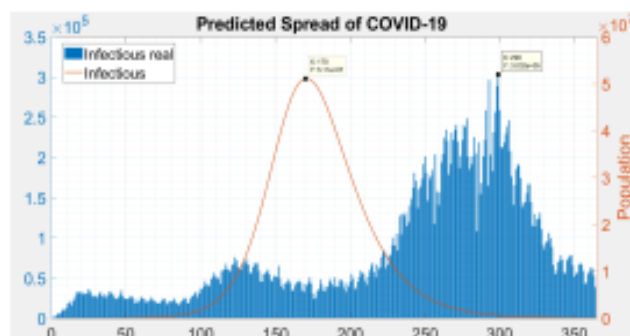


Рис. 3. Модель прогнозування розповсюдження ковіду в США з реальними даними

Джерело: авторська розробка

Існує багато модифікацій полігамної моделі для епідемії, де враховуються деякі з перелічених факторів. Будемо використовувати метод найменших квадратів для покращення результатів прогнозування.

Так як протягом епідемії в моделі кількість населення є сталою, переписемо рівняння для знаходження кількості інфікованих та одужавших або померлих осіб. Підставляючи N в диференційні рівняння отримаємо :

$$\frac{dI}{dt} = (R_0\beta - \gamma)I - \frac{R_0\beta}{N}I^2 - \frac{R_0\beta}{N}IR$$

$$\frac{dR}{dt} = \gamma I$$

Для оптимізації параметрів R_0 , β та γ використаємо метод найменших квадратів. Для вектор функції $f(x, \alpha)$, оптимізуємо параметр α

$$\alpha = [\gamma \beta R_0]$$

$$F(\alpha) = \sum_{i=1}^m [y_i - f(x_i, \alpha)]^2$$

Задаємо початкові параметри в α та виконуємо оптимізацію, використовуючи програмний пакет Matlab, а також ode45 для вирішення диференціальних рівнянь та метод найменших квадратів [3], який реалізовано за допомогою lsqcurvefit.

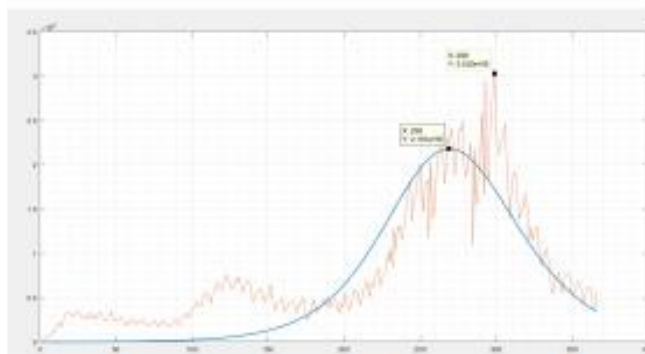


Рис. 4. Моделювання з оптимізованими параметрами

Джерело: авторська розробка

Таблиця 1

Результати моделювання

Метод	Кількість інфікованих	День піку	Відхилення від реальних даних
Реальні дані	303 200	299	-
Базова модель	51 100 000	170	50 796 800
Вдосконалена модель	218 100	269	85 100

Джерело: авторська розробка

На Рис.4 видно, що метод найменших квадратів досить сильно наблизив прогнозування розповсюдження хвороби до реальних даних в порівнянні з вихідною моделлю. Результати порівняння наведені в Таблиці 1.

Хоча вдосконалена модель дає набагато кращі результати, які наблизені до реальних, вона не враховує велику кількість параметрів і не передбачуваних ситуацій, наприклад: скупчення людей на мітингах, концерти і таке інше, в один з днів, там де ймовірність зараження є більшою. Пік прогнозований на 269 день є близьким до реальних прогнозів, адже

одразу йде плавний спад інфікованих осіб і є тільки два непередбачуваних спалахи.

Висновки. В даній науковій роботі запропоновано використання методу найменших квадратів для вдосконалення існуючої моделі прогнозування розповсюдження вірусів. Виконано прогнозування розповсюдження епідемічних захворювань та порівняння з статистичними даними реальної епідемії коронавіруса у Сполучених Штатах Америки, що дало високий відсоток співпадання.

Результати отримані в ході дослідження можуть використовуватися надалі у якості бази для подальших модифікацій та покращень полігамних моделей прогнозування розповсюдження епідемічних захворювань.

Література

1. Batista M. Estimation of the final size of the coronavirus epidemic by the SIR model / M.Batista. RegeGate, 2020.
2. WHO Emergency Committee. Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov). 2020. URL: [https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))
3. Линник Ю.В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений / Ю.В. Линник. Москва: Государственное издательство физико-математической литературы, 1962. 326 с.