

Міністерство освіти і науки України
Харківський національний університет імені В.Н. Каразіна
Факультет математики і інформатики
Кафедра прикладної математики

До захисту допущено

кафедрою прикладної математики, протокол № 5 від 12 червня 2026 р.

завідувач кафедри

прикладної математики

доктор фіз.-мат. наук, професор

Валерій КОРОБОВ

КВАЛІФІКАЦІЙНА РОБОТА

здобувача першого (бакалаврського) рівня вищої освіти

«Математичні моделі

поширення інфекційних захворювань»

Спеціальність 113 Прикладна математика

Освітня програма Прикладна математика

Здобувач

Олексій СТЕПАНЕНКО

Науковий керівник

кандидат фіз.-мат. наук,

доцент кафедри прикладної математики

Тетяна СМОРЦОВА

Харків – 2026

Анотації

Степаненко Олексій Олександрович. Математичні моделі поширення інфекційних захворювань. Кваліфікаційна робота присвячена моделюванню детермінованих (MSEIR) та стохастичних (TSIR) моделей розповсюдження кору в Японії. Після наочної ілюстрації нездатності методу найменших квадратів апроксимувати криву інфікованих, робиться перехід до методу максимальної правдоподібності. Оптимізується зареєстрована кількість випадків кору в Японії і пояснюється, чому для подальшого аналізу некоректно використовувати зареєстровані дані. Наочно ілюструється, що модель MSEIR не здатна відобразити описані ВООЗ тренди поведінки кору за відсутності вакцинації і робиться перехід до моделі TSIR. Розглядаються різні розподіли кількості інфікованих і пояснюється, чому для практичного застосування моделі найкращим розподілом є від'ємний біноміальний, тоді як для теоретичного моделювання розподіл Пуассона є найпрактичнішим. На гіпотетичному прикладі двох міст демонструється, як метапопуляційна модель може уточнити стандартну модель. Також математично демонструється, як завчасне запровадження вакцинації може допомогти запобігти епідемії.

Ключові слова: епідеміологія, кір, моделювання, MSEIR, TSIR, розподіли, оптимізація

Stepanenko Oleksii. Mathematical models of infectious disease spread. This qualification paper focuses on the modelling of deterministic (MSEIR) and stochastic (TSIR) models of measles transmission in Japan. After clearly demonstrating the inability of the least squares method to ap-

proximate the infection curve, the study transitions to maximum likelihood estimation. The reported number of measles cases in Japan is optimized, and a rationale is provided as to why utilizing strictly reported data for subsequent analysis is methodologically incorrect. It is visually illustrated that the MSEIR model fails to capture the measles behavioural trends described by the WHO in the absence of vaccination, prompting a shift to the TSIR model. Various distributions for the number of infected individuals are examined, explaining why the negative binomial distribution is optimal for the model's practical application, whereas the Poisson distribution is the most practical choice for theoretical modelling. Using two cities as a hypothetical case study, the work demonstrates how a metapopulation model can refine the standard approach. Finally, the study shows mathematically how the early implementation of vaccination can help prevent an epidemic.

Keywords: epidemiology, measles, modelling, MSEIR, TSIR, distributions, optimization

Зміст

Вступ	6
1. Аналіз та підготовка даних	8
1.1. Аналіз проблеми	8
1.1.1. Модель SIR	8
1.1.2. Недоліки моделі	9
1.2. Підготовка даних	10
1.2.1. Демографічні дані	10
1.2.2. Соціологічні дані	11
1.2.3. Епідеміологічні дані	13
1.3. Інтерполяція даних	13
1.3.1. Інтерполяція демографічних даних	13
1.3.2. Інтерполяція охоплення вакцинацією	14
1.4. Підготовка моделі	15
2. Детермінована модель MSEIR	17
2.1. Опис оригінальної моделі	17
2.2. Модифікація моделі	18
2.3. Застосування моделі	21
2.3.1. Початкова ініціалізація	21
2.3.2. Оптимізація методом найменших квадратів	24
2.3.3. Оптимізація методом максимальної правдоподібності	26
3. Стохастична модель TSIR	30
3.1. Опис оригінальної моделі	30

	5
3.2. Модифікація моделі	31
3.3. Застосування моделі	34
3.3.1. Початкова ініціалізація	34
3.3.2. Оптимізація методом максимальної правдоподібності	34
4. Порівняльний та сценарний аналіз	37
4.1. Проблема звітування (Underreporting)	37
4.2. Сценарій А: Відсутність вакцинації	38
4.2.1. Узгодження параметрів	38
4.2.2. Моделювання моделей MSEIR та TSIR	39
4.2.3. Виведення формули ендемічного центру рівноваги моделі TSIR	40
4.2.4. Моделювання моделі TSIR без врахування демографічних змін	42
4.2.5. Моделювання моделі TSIR без врахування демографічних змін та сезонності	44
4.3. Сценарій Б: Прогнозування майбутнього	45
4.4. Метопопуляційна модель	46
Висновки	51
Список використаних джерел	53

Вступ

Моделювання розповсюдження епідемій є важливим аспектом економік країн світу, оскільки це допомагає їм завчасно спрогнозувати потенційні спалахи хвороб та запобігти подальшим епідеміям.

У цій кваліфікаційній роботі досліджено різні моделі поширення кору в Японії. Зокрема, показано, що для отримання якісних прогнозів базова модель SIR має бути удосконалена шляхом урахування додаткових даних. Так, для побудови складніших моделей, які точніше відображають історичні дані та дають якісніші прогнози захворюваності, зібрано демографічні та соціологічні дані Японії. У таких моделях враховано материнський імунітет та введено латентний період.

У роботі розглянуто ускладнену детерміновану модель $MSEIR$ та стохастичну модель $TSIR$. Зокрема, продемонстровано недоліки використання детермінованих моделей для моделювання кору і необхідність використання стохастичних моделей. Для цього проведено сценарний аналіз цих моделей з метою прогнозування захворюваності на кір у майбутньому за імпорту нових інфікованих осіб та за відсутності вакцинації. Наведено приклад, який доводить необхідність використання метапопуляційної моделі для більш точного аналізу локальної поведінки кору. Крім того, продемонстровано, як вакцинація дозволяє уникнути епідемії.

Результати цієї роботи були частково опубліковані у [24], де було розглянуто модель SIR для Японії і зроблено висновок про необхідність використання складніших моделей для точнішого аналізу розвитку епідемії.

Вибір Японії зумовлений низкою факторів. Інтерес до епідеміологічної динаміки країн Азії був сформований під час проходження стажування

в Сінгапурі у 2025 році. Аналіз регіональних трендів виявив епідеміологічний парадокс Японії: будучи високорозвиненою країною, за останні 50 років вона стикалася з масштабними спалахами кору через історичні провали у національних програмах імунізації. Наявність таких якісних історичних даних у поєднанні з етапами падіння та відновлення колективного імунітету робить Японію оптимальною країною для тестування складних стохастичних епідеміологічних моделей.

Розділ 1

Аналіз та підготовка даних

В цьому розділі буде розглянуто модель SIR та пояснено, чому її не варто використовувати для аналізу розповсюдження кору. Також будуть розглянуті всі дані, використані для тренування моделей з метою надання якомога точнішої інформації про епідемію кору в Японії. Будуть розглянуті методи обробки демографічних, соціологічних та епідеміологічних даних.

1.1. Аналіз проблеми

1.1.1. Модель SIR

Нагадаємо стандартну модель SIR [20]:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta SI}{N} \\ \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1.1)$$

Тут β – коефіцієнт захворюваності, γ – коефіцієнт одужання, S – абсолютна кількість населення, вразливого до інфекції, I – кількість інфікованих, R – кількість тих, хто одужав або помер від захворювання, $N = S + I + R$ – загальна чисельність населення.

1.1.2. Недоліки моделі

У [20] було зазначено, що $N = \text{const}$. У випадку Японії чисельність населення стрімко зростала на кінці ХХ століття, але вийшла на плато і потроху почала спадати у ХХІ столітті, як зазначено в [3].

З метою точнішого аналізу розвитку епідемії кору в Японії потрібно враховувати цю інформацію, а не брати кількість населення константою.

З іншого боку, модель (1.1) не враховує, що кір має так званий латентний період, який для кору становить близько 10 днів, як зазначено в [5]. Латентний період — це період, в який людина вже є інфікованою, але все ще не є заразливою.

До того ж, у [6] зазначено, що діти віком до 6 місяців мають материнський імунітет і, відповідно, не є вразливими. Для вірусу кору, 90% випадків якого відбувались у віці до 10 років до винайдення вакцини [17], це є важливим фактором, оскільки діти до 6 місяців становлять велику частку заразливих, якщо не враховувати материнський імунітет.

Окрім цього, модель (1.1) припускає, що кожна особа з однаковою ймовірністю контактує з будь-якою іншою особою. В реальності це не завжди так: в 2014 році було проведено дослідження [4], яке показало, скільки контактів в середньому на день (у будні та вихідні дні) кожна особа має з представниками кожної вікової групи. Через нерівномірність цих даних та високий ризик захворіти на кір в ранньому віці виникає потреба враховувати ці соціологічні дані.

Також, у 1978 році в Японії було запроваджено обов'язкову вакцинацію дітей [8], що стало поштовхом для поступової елімінації кору. Проте, оскільки у 1994 році вакцинація стала добровільною і кількість вакцинованих різко впала через недовіру населення до нової вакцини, на початку ХХІ століття епідемія спалахнула знову [8].

Оскільки проста *SIR*-модель не здатна проілюструвати описані вище тенденції, пропонується удосконалити модель так, щоб врахувати якомога більше деталей.

1.2. Підготовка даних

1.2.1. Демографічні дані

У [3] надано базу даних про перепис населення Японії за наступні роки:

$years = \{1970, 1975, 1980, 1986-1989, 1991-1994, 1996-2004, 2006-2024\}$.

У кожному переписі населення міститься точна інформація про чисельність населення кожного віку (з округленням до тисяч).

Для подальшого аналізу розіб'ємо все населення на 4 вікові групи: 0-1 рік, 1-4 роки, 5-14 років, 15+ років. Таке розбиття обумовлене тим, що відокремлюються діти до 6 місяців з материнським імунітетом (будемо вважати, що їх приблизно половина від загальної кількості людей 0-1 року), діти 1-4 років ще не ходять до школи, але за відсутності вакцинації вони є найбільш вразливими, бо починають контактувати з навколишнім середовищем, діти 5-14 років збираються в школах і таким чином передають вірус один одному, а у віці 15 років більшість осіб вже перехворіла і має позитивний імунітет.

Для того, щоб визначити чисельність населення кожної з вищезазначених вікових груп, потрібно взяти дані з кожної таблиці так:

$$N_{0-1,t} = N_{0,t} \tag{1.2}$$

$$N_{1-4,t} = \sum_{i=1}^4 N_{i,t} \tag{1.3}$$

$$N_{5-14,t} = \sum_{i=5}^{14} N_{i,t} \quad (1.4)$$

$$N_{15+,t} = N_{total,t} - N_{0-1,t} - N_{1-4,t} - N_{5-14,t} \quad (1.5)$$

Тут $N_{i,t}$ — кількість осіб віком i років за рік t , $N_{total,t}$ — загальна чисельність населення Японії за рік t .

1.2.2. Соціологічні дані

У [4] наведено середню кількість фізичних контактів між представниками кожної вікової групи в будні та вихідні дні. З 15 наявних вікових груп (0-4, 5-9, 10-14, 15-19, ..., 70+) потрібно змінити таблицю так, щоб її можна було застосувати для розбиття, обраного вище.

Читати матрицю, описану в [4], потрібно так: C_{ij} — це кількість контактів на день (будній чи вихідний) представником вікової групи i з представниками вікової групи j .

Оскільки перший рядок таблиці відноситься до вікової групи 0-4 роки, то для того, щоб відокремити вікову групу 0-1 рік, припускаємо, що кількість їхніх контактів з довільною віковою групою дорівнює кількості контактів вікової групи 1-4 з тою самою віковою групою. Відповідно, результувна матриця контактів буде складатися з двох однакових перших рядків.

Спочатку зведемо ці матриці (за будні та за вихідні дні) до однієї, беручи до уваги, що кожен тиждень в середньому складається з п'яти будніх та двох вихідних днів:

$$C_{avg} = \frac{5}{7}C_{weekday} + \frac{2}{7}C_{weekend} \quad (1.6)$$

Далі, застосуємо наступне перетворення матриці для стовпців $C = C_{avg}$: якщо кількість контактів середньостатистичного представника групи i з

групою j дорівнює C_{ij} , то для $j \neq k$: $C_{i,(j \cup k)} = C_{ij} + C_{ik}$. Відповідно, для будь-якої вікової групи i маємо:

$$C_{i,5-14} = \sum_{k \in K_1} C_{i,k} \quad (1.7)$$

$$C_{i,15+} = \sum_{k \in K_2} C_{i,k}, \quad (1.8)$$

де $K_1 = \{5-9, 10-14\}$, $K_2 = \{15-19, 20-24, \dots, 70+\}$ — множини відповідних 5-річних вікових інтервалів.

Перетворення рядків відбувається так: якщо N_i — чисельність вікової групи i , то для $i \neq k$:

$$C_{i \cup k, j} = \frac{C_{ij}N_i + C_{kj}N_k}{N_i + N_k} \quad (1.9)$$

Рівняння (1.9) показує, скільки контактів середньостатистична особа в віковій групі $i \cup k$ робить з представниками будь-якої вікової групи j . Таким чином, для вікової групи 5-14 маємо наступне рівняння:

$$C_{5-14, j} = \frac{C_{5-9, j}N_{5-9} + C_{10-14, j}N_{10-14}}{N_{5-9} + N_{10-14}} \quad (1.10)$$

Відповідно, останній рядок матриці контактів для вікової групи 15+ має вигляд:

$$C_{15+, j} = \frac{\sum_{m \in K_2} C_{m, j}N_m}{\sum_{m \in K_2} N_m}, \quad (1.11)$$

де $K_2 = \{15-19, 20-24, \dots, 70+\}$ — множина відповідних 5-річних вікових інтервалів.

Таким чином, отримуємо матрицю $C = \{C_{ij}\}$, що буде використовуватися в подальшому аналізі для обчислення сили інфекції (див. розділ 2).

1.2.3. Епідеміологічні дані

Використовуємо оцінку ВООЗ [2] для даних про *охоплення щепленнями* в Японії за 1978-2024 роки:

$$v_t = \frac{N_{vaccinated,t}}{N_{total,t}} \quad (1.12)$$

Тут $N_{vaccinated,t}$ — кількість вакцинованих за рік t , $N_{total,t}$ — загальна кількість осіб за рік t , що потребують вакцинації (для кору це діти віком 1 рік).

Зауваження. $N_{vaccinated,t}$ — загальна кількість вакцинованих, яка включає в себе дітей, що вакцинувалися після рекомендованого строку у 12 місяців. Так, після запровадження добровільної вакцинації у 1994 році [8] та через велику кількість побічних дій нової вакцини ММР велика кількість дітей або йшла вакцинуватися пізніше, або не йшла вакцинуватися взагалі [8]. Тому дані ВООЗ потрібно використовувати не як абсолютний ідеал, а як певну оцінку.

Вважається, що саме провал вакцинації на кінці ХХ–початку ХХІ століття спричинив нову хвилю хворих [8]. Оскільки дані ВООЗ [2] розходяться з реальними даними [9], для якіснішого аналізу розвитку епідемії кору будемо використовувати саме реальні дані.

1.3. Інтерполяція даних

1.3.1. Інтерполяція демографічних даних

Оскільки тривалість хвороби на кір та його латентний період не перевищують двох тижнів [16], даних про щорічну кількість нових випадків кору недостатньо для аналізу моделі. У базі даних [3] вже зібрані демографічні дані майже за всі роки, починаючи з 1970-го. Відповідно, є можливість оці-

нити чисельність населення кожної вікової групи за кожен день. Для цього використовується лінійна інтерполяція — якщо на початку року t_{start} чисельність населення певної вікової групи i становить $N_i(t_{start})$, а наприкінці року t_{end} — $N_i(t_{end})$, тоді для будь-якого моменту часу t всередині цього інтервалу ($t_{start} \leq t \leq t_{end}$) чисельність $N_i(t)$ обчислюється за наступною формулою:

$$N_i(t) = N_i(t_{start}) + (t - t_{start}) \cdot \frac{N_i(t_{end}) - N_i(t_{start})}{t_{end} - t_{start}} \quad (1.13)$$

Лінійність формули 1.13 забезпечує демографічний баланс в будь-який момент часу t :

$$N_{total}(t) = \sum_{i=1}^4 N_i(t) \quad (1.14)$$

1.3.2. Інтерполяція охоплення вакцинацією

Інтерполяція частки вакцинованих буде проводитись наступним чином:

1. До введення вакцини в 1978 році [8] кількість вакцинованих фактично дорівнювала нулю.
2. В період 1994–2002 років через недовіру населення до вакцинації процентна частка вакцинованих була доволі низькою. Оскільки з літератури [9] маємо дані за 2002 рік (61.7%), ці дані будуть взяті для всього періоду 1994–2002 років.
3. Решта даних беруться з оцінки ВООЗ (WUENIC), але для врахування відмов від вакцинації верхній поріг охоплення вакцинацією був обраний у 98%.

Формально, алгоритм формування щорічного рівня вакцинації можна записати як кусково-задану функцію:

$$\hat{p}_y = \min(p'_y, 0.98), \quad (1.15)$$

де базове значення p'_y визначається історичним контекстом:

$$p'_y = \begin{cases} 0, & y < 1978 \\ 0.617, & 1994 \leq y \leq 2002 \\ P_{\text{WUENIC}}(y), & \text{якщо дані } P_{\text{WUENIC}}(y) \text{ існують} \\ \hat{p}_{y-1}, & \text{в іншому випадку} \end{cases} \quad (1.16)$$

Для отримання щоденного значення частки вакцинованих $p(t)$ для будь-якого дня $t \in [t_{start}, t_{end}]$ всередині року застосовувалась лінійна інтерполяція:

$$p(t) = \hat{p}_{y_{start}} + (t - t_{start}) \cdot \frac{\hat{p}_{y_{end}} - \hat{p}_{y_{start}}}{t_{end} - t_{start}} \quad (1.17)$$

де t_{start} та t_{end} — дні, що відповідають початку поточного та наступного років відповідно.

Зауваження. Кількість осіб, що набувають імунітет після щеплення, насправді дорівнює $p(t) \cdot VE$, де VE — так звана ефективність вакцини (*Vaccine Efficacy*). Для вакцини MCV1 для дітей віком < 9 місяців вона становить близько 94% [13]. Для запобігання перенавчанню моделі через різні показники ефективності вакцини, що збільшуються з часом, було обрано обмежити $p(t)$ числом 0.98 та брати в якості $p(t)$ історичні оцінки ВООЗ.

1.4. Підготовка моделі

Оскільки епідемія кору виникла раніше, ніж у 1970-х роках, а демографічні та епідеміологічні дані починаються у 1970 році, для більш точного та реалістичного аналізу пропонується ввести так звану фазу розігріву.

В початковий момент часу (наприклад, 1 січня 1950 року) обираємо певну кількість інфікованих та осіб в латентному періоді і симулюємо моделі, описані в подальших розділах. Це дає більш правдоподібну картину, оскільки, як буде видно далі, епідемії виникають не одразу після виникнення перших випадків.

Розділ 2

Детермінована модель MSEIR

В розділі 1 було описано недоліки моделі SIR і необхідність використовувати складніші моделі. Далі в цьому розділі буде розглядатися модель $MSEIR$, що являє собою детерміновану модель, що враховує материнський імунітет (M) та латентний період кору (E).

2.1. Опис оригінальної моделі

Означення 2.1. *Моделлю MSEIR називається наступна система диференціальних рівнянь [20]:*

$$\left\{ \begin{array}{l} \frac{dM}{dt} = B - \delta_m M - \mu M \\ \frac{dS}{dt} = \delta_m(1-p)M - \frac{\beta SI}{N} - \mu S \\ \frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E - \mu E \\ \frac{dI}{dt} = \sigma E - \gamma I - \mu I \\ \frac{dR}{dt} = \delta_m p M + \gamma I - \mu R \end{array} \right. \quad (2.1)$$

Тут M, S, E, I, R — чисельність осіб у відповідних епідемічних станах (з материнським імунітетом, вразливі, інфіковані в латентному періоді, заразні та вилучені/вакциновані відповідно), N — загальна чисельність населення. Параметри системи:

- B — абсолютна кількість народжень за одиницю часу;

- μ — коефіцієнт демографічного відтоку (природна смертність/старіння);
- δ_m — швидкість втрати материнського імунітету (обернена до середньої тривалості захисту немовлят);
- p — рівень охоплення вакцинацією (частка дітей, які отримують штучний імунітет одразу після втрати материнського);
- β — коефіцієнт розповсюдження вірусу;
- σ — швидкість переходу з латентного стану в інфекційний (обернена до латентного періоду);
- γ — швидкість одужання (обернена до періоду інфекційності).

Оскільки нас цікавить денна інцидентність, то значення параметрів δ_m , γ та σ можемо порахувати одразу. Так, оскільки середня тривалість захисту немовлят становить ≈ 180 діб, період інфекційності кору становить ≈ 7 діб та латентний період становить ≈ 10 діб, маємо наступні значення параметрів моделі:

$$\delta_m = \frac{1}{180}, \quad \gamma = \frac{1}{7}, \quad \sigma = \frac{1}{10} \quad (2.2)$$

2.2. Модифікація моделі

Оскільки система (2.1) припускає, що населення є однорідним, то для того, щоб урахувати соціологічні показники, потрібно удосконалити систему, переписавши її для кожної вікової групи, обговореної в розділі 1 окремо, таким чином, отримуючи 20 рівнянь.

Для компактного запису повної вікової моделі введемо вектори станів $\mathbf{M}, \mathbf{S}, \mathbf{E}, \mathbf{I}, \mathbf{R} \in \mathbb{R}^n$, де $n = 4$ — кількість вікових груп. Динаміка системи описується наступним диференціальним рівнянням:

$$\left\{ \begin{array}{l} \frac{d\mathbf{M}}{dt} = \mathbf{B} - \delta_m \mathbf{M} + \mathbf{A}\mathbf{M} \\ \frac{d\mathbf{S}}{dt} = \delta_m(1-p)\mathbf{M} - \mathbf{\Lambda}\mathbf{S} + \mathbf{A}\mathbf{S} \\ \frac{d\mathbf{E}}{dt} = \mathbf{\Lambda}\mathbf{S} - \sigma\mathbf{E} + \mathbf{A}\mathbf{E} \\ \frac{d\mathbf{I}}{dt} = \sigma\mathbf{E} - \gamma\mathbf{I} + \mathbf{A}\mathbf{I} \\ \frac{d\mathbf{R}}{dt} = \delta_m p \mathbf{M} + \gamma\mathbf{I} + \mathbf{A}\mathbf{R} \end{array} \right. \quad (2.3)$$

Тут $\mathbf{B} = [B, 0, 0, 0]^T$ — вектор народжуваності (поповнює лише першу комірку), $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ — діагональна матриця сили інфекції (див. рівняння 2.10), а $\mathbf{A} \in \mathbb{R}^{n \times n}$ — матриця демографічних переходів (тобто, старіння та смертності):

$$\mathbf{A} = \begin{pmatrix} -(a_1 + \mu_1) & 0 & 0 & 0 \\ a_1 & -(a_2 + \mu_2) & 0 & 0 \\ 0 & a_2 & -(a_3 + \mu_3) & 0 \\ 0 & 0 & a_3 & -\mu_4 \end{pmatrix} \quad (2.4)$$

Тут a_i — швидкість переходу до наступної вікової групи (старіння), а μ_i — коефіцієнт смертності у відповідній віковій групі.

Надалі, припускаємо, що кожної доби однакова частка дітей переходить з однієї вікової групи в іншу. Ця частка обернено пропорційна до загальної тривалості перебування кожної особи у відповідній віковій групі (тут і надалі припускаємо, що кожний рік містить в собі 365.25 днів):

$$a_1 = \frac{1}{1 \cdot 365.25} \quad (2.5)$$

$$a_2 = \frac{1}{4 \cdot 365.25} \quad (2.6)$$

$$a_3 = \frac{1}{10 \cdot 365.25} \quad (2.7)$$

$$a_4 = 0 \quad (2.8)$$

Якщо знехтувати дитячою смертністю, коефіцієнт смертності $\mu_4 = \mu$ можна обчислити з демографічних даних за наступною формулою:

$$\mu(t) = B_t + N_{total,t-1} - N_{total,t} \quad (2.9)$$

Тут B_t — народжуваність в момент часу t , $N_{total,t}$ — загальна чисельність населення в момент часу t .

Фізичний зміст цієї формули: приріст населення дорівнює різниці народжуваності та смертності.

Відповідно, матриця \mathbf{A} перетворюється на таку:

$$\mathbf{A} = \begin{pmatrix} -a_1 & 0 & 0 & 0 \\ a_1 & -a_2 & 0 & 0 \\ 0 & a_2 & -a_3 & 0 \\ 0 & 0 & a_3 & -\mu \end{pmatrix} \quad (2.10)$$

Сила інфекції для кожної групи обчислюється через матрицю соціальних контактів $\mathbf{C} = \{c_{ij}\}$, обчислену в розділі 1.2.2:

$$\lambda_i(t) = \beta \sum_{j=1}^n c_{ij} \frac{I_j(t)}{N_j(t)} \quad (2.11)$$

Ця формула є узагальненням формули для сили інфекції в моделі SIR , яка тепер подається як лінійна комбінація сил інфекції для кожної вікової групи.

Це перетворення можна записати через обернену діагональну матрицю населення:

$$\vec{\lambda}(t) = \beta \mathbf{C} \text{diag}(\mathbf{N})^{-1} \mathbf{I}(t) \quad (2.12)$$

де $\text{diag}(\mathbf{N})$ — матриця, на головній діагоналі якої розташовані значення загальної чисельності кожної вікової групи N_i .

Надалі, для використання в системі (2.3), з вектора $\vec{\lambda}(t)$ формується діагональна матриця $\mathbf{\Lambda} = \text{diag}(\vec{\lambda})$. Такий матричний запис є узагальненням більш простого випадку, описаного в другому доданку другого рівняння системи (2.1).

Будемо оновлювати матрицю контактів на кожному кроці:

$$C_{i,j}^{adj}(t) = \frac{C_{i,j}N_{i,t} + C_{j,i}N_{j,t}}{2N_{i,t}} \quad (2.13)$$

Це рівняння береться з того, що в ідеальній середовищі сума кількостей контактів представників групи i з представниками групи j дорівнює сумі кількостей контактів представників групи j з представниками групи i , але оскільки наша матриця контактів не ідеальна, беремо середнє арифметичне загальних контактів і ділимо його на чисельність населення групи i , щоб отримати кількість контактів для однієї людини.

2.3. Застосування моделі

Беремо модель, отриману в розділі 2.1, і застосуємо її до відзвітованих даних [1]. Оскільки неможливо сказати точно, скільки реальних випадків кору було виявлено (див. розділ 4.1), будемо намагатися описати тенденції поведінки кривої кількості інфікованих. При цьому обираємо сталий коефіцієнт реєстрації 0.1 (див. розділ 4.1).

2.3.1. Початкова ініціалізація

Застосовуємо фазу розігріву тривалістю 20 років (див. розділ 1.4) до моделі, описаної в розділі 2.1. Матрицю контактів було обчислено в розділі 1.2.2, інтерполяцію чисельності населення було описано в розділі 1.3.1, а інтерполяцію охоплення вакцинацією — в розділі 1.3.2. Значення параметрів

δ_m, γ та σ обираються з рівнянь (2.2).

На початку роботи програми задаємо $seed = 500$ (це кількість інфікованих та осіб, що перебувають в латентному періоді станом на 1 січня 1950 року). Оскільки в той час майже всі хворіли на кір і отримували пожиттєвий імунітет після одужання, задаємо наступну початкову частку вразливих у кожній віковій групі з такими припущеннями:

$$S_{0,0-1} = 0.5N_{0,0-1} \quad (\approx 50\% \text{ дітей мають материнський імунітет}) \quad (2.14)$$

$$S_{0,1-4} = 0.6N_{0,1-4} \quad (\approx 40\% \text{ дітей дошкільного віку перехворіли}) \quad (2.15)$$

$$S_{0,5-14} = 0.15N_{0,5-14} \quad (\approx 85\% \text{ дітей шкільного віку перехворіли}) \quad (2.16)$$

$$S_{0,15+} = 0.05N_{0,15+} \quad (\text{майже всі дорослі перехворіли}) \quad (2.17)$$

Обираємо початкове значення β наступним чином: у [19] наведено формулу для обчислення базового репродуктивного числа R_0 :

$$R_0 = \rho(K(S)), \quad (2.18)$$

де ρ — спектральний радіус оператора наступного покоління $K(S)$. Для дискретних випадків $K(S)$ являє собою матрицю, що за відсутністю смертності для хворих обчислюється за наступною формулою [19]:

$$K(S) = T(S)D^{-1}, \quad (2.19)$$

де D^{-1} — середній час перебування особи в інфікованому стані, що в нашому випадку дорівнює γ^{-1} , $T(S)$ — матриця передачі хвороби, що в нашому випадку обчислюється за формулою:

$$T(S) = \beta \mathbf{C}, \quad (2.20)$$

де β — коефіцієнт розповсюдження інфекції моделі (2.3), \mathbf{C} — матриця

контактів, задана в розділі 1.2.2.

Підсумовуючи, маємо загальну формулу для обчислення базового репродукційного числа кору з моделі (2.3):

$$R_0 = \rho(\beta \mathbf{C} \gamma^{-1}) \quad (2.21)$$

Оскільки β та γ в моделі 2.3 є сталими, їх можна винести за дужки, таким чином отримуючи фінальну формулу:

$$R_0 = \beta \gamma^{-1} \rho(\mathbf{C}) \quad (2.22)$$

Базове репродукційне число для кору зазвичай становить $\approx 12-18$ [7]. Для розрахунків обираємо значення, що лежить посередині цього інтервалу: $R_0 = 15$.

Матриця контактів \mathbf{C} має наступний вигляд:

$$\mathbf{C} = \begin{pmatrix} 0.2904 & 1.1939 & 0.3686 & 1.4585 \\ 0.2904 & 1.1939 & 0.3686 & 1.4585 \\ 0.0384 & 0.1580 & 4.2467 & 1.7996 \\ 0.0417 & 0.1714 & 0.4823 & 4.4631 \end{pmatrix} \quad (2.23)$$

Її максимальне власне значення було обчислено в середовищі Python за допомогою модуля `np.linalg` і становить приблизно 5.38. Відповідно, якщо підставити це значення в рівняння (2.22), отримаємо значення для β :

$$\beta \approx 0.398 \quad (2.24)$$

Це значення буде обрано як стартове значення.

Програмна реалізація цього та всіх наступних алгоритмів проводилась за допомогою використання інструментів генеративного штучного інтеле-

кту і наведено за посиланням [22]. Чисельне інтегрування з метою отримання річної інциденції проводилось в середовищі Python за допомогою модуля `scipy.integrate.odeint`. На 2.1 показано графік кількості інфікованих в залежності від часу для початкового значення β . Час виконання програми становить 0.97 секунд.

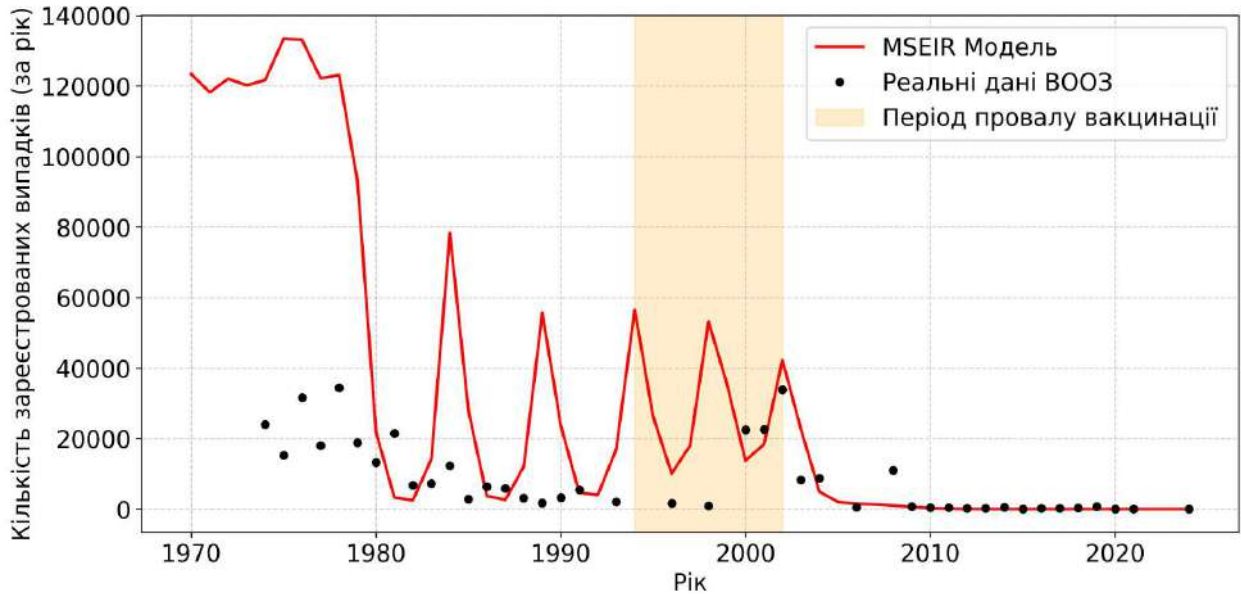


Рис. 2.1: Модель MSEIR для початкового значення $\beta = 0.398$.

2.3.2. Оптимізація методом найменших квадратів

У розділі 2.3.1 було застосовано модель *MSEIR* до певного початкового значення параметра β . У цьому розділі буде оптимізовано цей параметр за допомогою метода найменших квадратів.

Задача метода найменших квадратів полягає в тому, щоб мінімізувати середньоквадратичну похибку:

$$MSE = \frac{1}{k} \sum_{y=1974}^{2024} (I_{sim}(y) \cdot 0.1 - I_{real}(y))^2, \quad (2.25)$$

де I_{sim} — кількість випадків, спрогнозованих моделлю, а I_{real} — кількість зареєстрованих випадків. Число 0.1 — це так званий коефіцієнт реєстрації

(див. розділ 4.1). Число $k = 51$ в знаменнику дроби береться як кількість років симуляції:

$$k = t_{end} - t_{start} + 1, \quad (2.26)$$

де $t_{start} = 1974$, $t_{end} = 2024$ — початок і кінець симуляції.

Відповідно, шукаємо таке значення β , для якого $MSE \rightarrow \min$.

Оптимізація реалізована в середовищі Python за допомогою модуля `scipy.optimize`. В якості основного інструменту пошуку глобального мінімуму тут і надалі було обрано алгоритм L-BFGS-B (Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bounds). Такий вибір обумовлений тим, що значення β лежать в інтервалі $(0,1)$ через фізичний зміст параметра. Окрім цього, на відміну від класичного методу Ньютонa, який вимагає аналітичного обчислення або скінченно-різницевого наближення повної матриці Гессе $\nabla^2 f(x)$ на кожній ітерації (що є обчислювально складним для жорстких систем диференціальних рівнянь), алгоритм L-BFGS-B буде апроксимацію оберненої матриці Гессе. Використання модифікації з обмеженою пам'яттю (Limited-memory) дозволяє зберігати не всю матрицю розмірності $n \times n$, а лише історію оновлень для m останніх ітерацій, що забезпечує швидку збіжність при мінімальних витратах обчислювальних ресурсів.

Зауваження. Оскільки стан без захворювань є глобальною асимптотично стійкою точкою спокою [20], оптимізатор часто намагався якомога швидше наблизитися до нуля, замість того, щоб охопити пік початку XXI століття і зробити кращу форму кривої. Тому було прийнято рішення модифікувати функцію сили інфекції, додавши до неї маленьке число ($\alpha \approx 10^{-8}$), яке було підібране таким чином, щоб оптимізатор не згладжував криву.

На рис. 2.2 показано результат виконання оптимізації методом наймен-

ших квадратів. Час виконання програми становить 112.89 секунд.

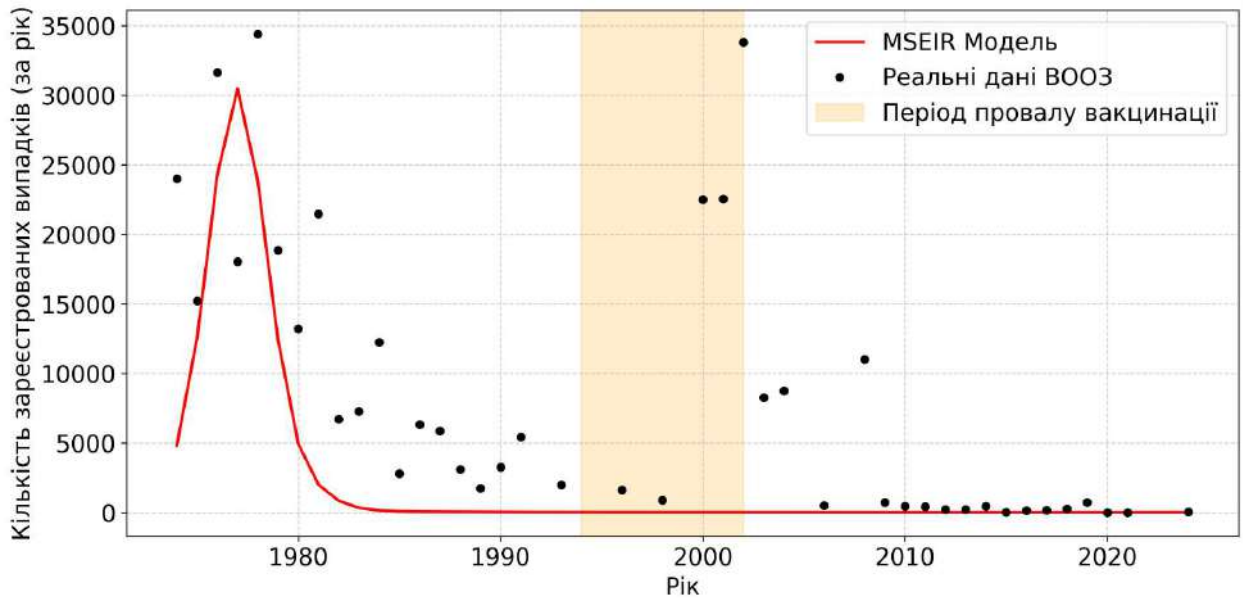


Рис. 2.2: Модель MSEIR, оптимізована за допомогою метода найменших квадратів ($\beta \approx 0.05$).

2.3.3. Оптимізація методом максимальної правдоподібності

Недоліки методу найменших квадратів Проблема методу найменших квадратів полягає в тому, що крива кількості інфікованих не охоплює другий пік початку XXI століття, а також згладжується. З одного боку, грає роль сталий коефіцієнт реєстрації, але з іншого боку, оптимізація методом найменших квадратів намагається зберегти якомога простішу форму кривої, жорстко штрафуючи модель за відхилення.

Щоб покращити аналіз, розглянемо більш загальний метод оптимізації — метод максимальної правдоподібності (Maximum Likelihood Estimate, MLE). Метод максимальної правдоподібності базується на знаходженні такого вектора параметрів θ , який максимізує спільну ймовірність отримання наявного набору емпіричних спостережень $Y = \{y_1, y_2, \dots, y_n\}$. У загальному вигляді функція правдоподібності $L(\theta)$ для незалежних спостережень

записується як добуток функцій маси ймовірності P :

$$L(\theta | Y) = \prod_{i=1}^n P(y_i | \theta) \quad (2.27)$$

Зауваження. Зазначимо, що метод найменших квадратів являє собою метод максимальної правдоподібності за умови, що похибки мають нормальний розподіл [15]. Насправді, кількість інфікованих є дискретною величиною, тому для її оцінки краще використовувати саме дискретні розподіли (див. розділ 2.3.3).

З метою уникнення обчислювальної нестабільності при множенні малих ймовірностей, а також для адаптації задачі під стандартні алгоритми градієнтної мінімізації, замість максимізації $L(\theta)$ здійснюється мінімізація негативної логарифмічної правдоподібності (Negative Log-Likelihood, NLL). Загальний вигляд функції NLL, незалежний від розподілу, має вигляд [15]:

$$NLL(\theta | Y) = -\ln L(\theta | Y) = -\sum_{i=1}^n \ln P(y_i | \theta) \quad (2.28)$$

Вибір розподілу. Проблема біноміального розподілу полягає в тому, що він припускає незалежність подій (тобто, в нашому випадку незалежність захворювання двох різних осіб). Звичайно, ці події в реальному житті є залежними, оскільки контакти між особами впливають на передачу інфекції.

Проблема розподілу Пуассона полягає в тому, що він припускає, що дисперсія є сталою: $Var = \mu$ [21]. У випадку кору, який поширюється групами (наприклад, в школах, дитсадках тощо), дисперсія не є сталою, оскільки якщо в якийсь момент один учень стає інфікованим, майже всі невакциновані мають великий ризик заразитися, через що дисперсія перевищить середнє значення ($Var > \mu$).

З огляду на це, розглянемо від'ємний біноміальний розподіл. Вектор параметрів моделі $\theta = [\beta, k]$, де k — додатний параметр дисперсії, який буде оптимізуватися разом з коефіцієнтом розповсюдження інфекції β . Для від'ємного біноміального розподілу дисперсія описується наступною формулою:

$$Var = \mu + \frac{\mu^2}{k}, \quad (2.29)$$

При $k \rightarrow \infty$ $Var \rightarrow \mu$, і отримуємо розподіл Пуассона. З іншого боку, при $k \approx 0$ дисперсія швидко зростає, що дозволяє охопити кластерну структуру розповсюдження інфекції. Відповідно, параметр k відповідає саме за те, наскільки великою є дисперсія і допомагає визначити стохастичний шум.

Функція маси ймовірності негативного біноміального розподілу через гамма-функцію $\Gamma(\cdot)$ записується як [21]:

$$P(y_i | \mu_i, k) = \frac{\Gamma(y_i + k)}{\Gamma(y_i + 1)\Gamma(k)} \left(\frac{k}{k + \mu_i}\right)^k \left(\frac{\mu_i}{k + \mu_i}\right)^{y_i}, \quad (2.30)$$

де $\mu_i(\beta)$ — динамічне математичне сподівання для i -го року, обчислене за допомогою інтегрування системи диференціальних рівнянь MSEIR.

Відповідно, фінальна цільова функція NLL, яка була мінімізована алгоритмом L-BFGS-B, має наступний вигляд:

$$NLL(\beta, k) = - \sum_{i=1}^n \left[\ln \Gamma(y_i + k) - \ln \Gamma(y_i + 1) - \ln \Gamma(k) + k \ln \left(\frac{k}{k + \mu_i}\right) + y_i \ln \left(\frac{\mu_i}{k + \mu_i}\right) \right] \quad (2.31)$$

Програмна реалізація. Програмна реалізація оптимізації параметрів β і k була описана в середовищі Python за допомогою модуля `scipy.optimize`. Результат роботи програми та оптимізовані значення параметрів наведено на рис. 2.3. Час виконання програми становить 247.53 секунд.

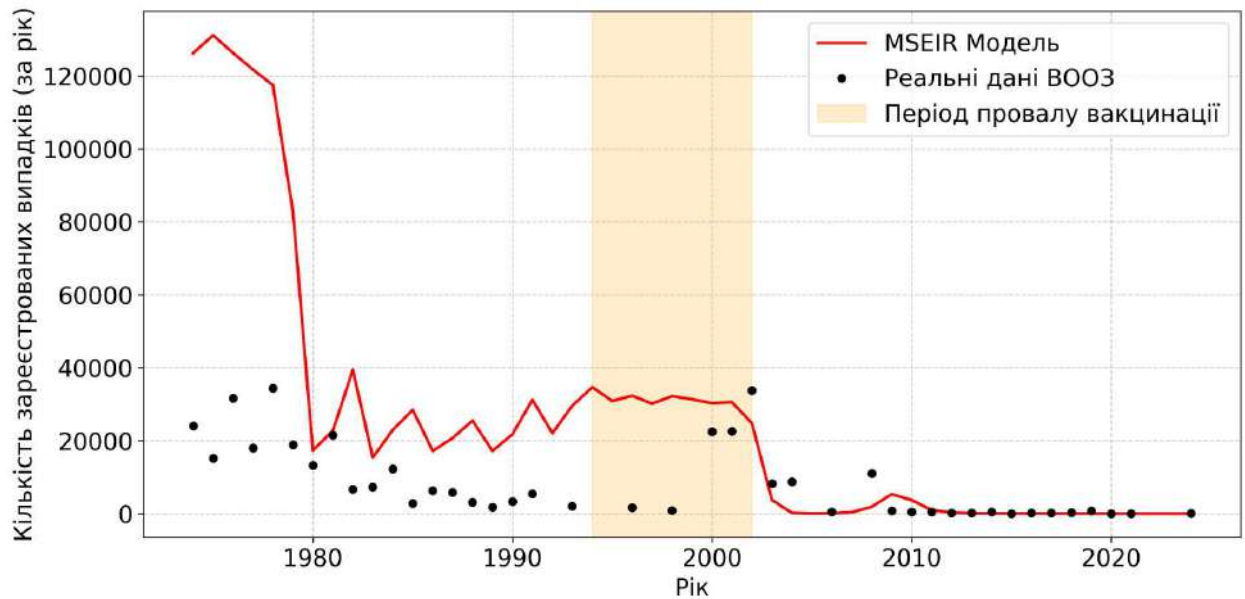


Рис. 2.3: Модель MSEIR, оптимізована за допомогою метода максимальної правдоподібності ($\beta \approx 0.98$, $k = 0.26$).

Зазначимо, що метод максимальної правдоподібності не згладжує криву так, як це робить оптимізація MSE. Так, одразу після провалу вакцинації у 1994 році крива зростає і залишається приблизно на одному і тому самому рівні доти, доки не вводиться тотальна вакцинація. До введення вакцинації крива поводить себе так само, як це описують тренди ВООЗ [16], що показує, що від'ємний біноміальний розподіл дає кращі результати, ніж нормальний розподіл похибок.

Зауваження. Значення, отримані оптимізацією за MSE та MLE, не відповідають фізичній реальності кору. Дійсно, ці значення сильно відрізняються від історичних даних – підставивши їх в рівняння (2.22), бачимо, що отримане базове репродукційне число буде або завеликим, або замалим. Отже, модель MSEIR перенавчена, що показує, що для кращого аналізу потрібно мати або якісніші дані про коефіцієнт реєстрації (див. розділ 4.1), або застосувати інші моделі. В наступних розділах розглянемо стохастичну модель TSIR і зробимо порівняльний аналіз наведених моделей.

Розділ 3

Стохастична модель TSIR

3.1. Опис оригінальної моделі

На відміну від детермінованих моделей, таких як MSEIR, можна розглядати стохастичні моделі. Так, у 2000 році [11] була створена модель TSIR (Time-Susceptible-Infected-Recovered). Її основна ідея полягає в тому, щоб працювати безпосередньо з даними, а не просто їх симулювати.

З кроком $t = 14$ діб (це число обирається саме тому, що 14 діб — тривалість інфікованості на кір, і відповідно, всі особи, хто за попередні 14 діб захворіли, в наступні 14 діб одужають) маємо наступні рівняння:

$$S_{t+1} = S_t + B_t - I_{t+1} \quad (3.1)$$

Фізичний зміст: кількість вразливих змінюється на різницю народжених та інфікованих.

$$E[I_{t+1}] = \beta_t \cdot I_t^\alpha \cdot S_t \quad (3.2)$$

Фізичний зміст: очікувана кількість нових хворих залежить від сезонної бети β_t , кількості вразливих S_t та кількості хворих I_t . Тут $\alpha \approx 0.97$ — параметр, що вводиться для врахування неоднорідності контактів у популяції та ефекту локального вичерпання сприйнятливих осіб (local depletion of susceptibles).

В оригінальній роботі ([11]) береться логарифм від рівняння 3.2 і за допомогою лінійної регресії оптимізується параметр β .

3.2. Модифікація моделі

У нашому випадку, оригінальну модель TSIR потрібно налаштувати під дані, отримані раніше. В оригінальній моделі все населення вважається однорідним, а в нашій моделі існують 4 різні вікові групи, що взаємодіють одна з одною згідно з матрицею контактів.

До того ж, автори [11] рахували S_t у вигляді суми ряду, обираючи замість S_0 певним чином середнє \bar{S} . В нашому випадку з чотирма віковими групами, потрібно врахувати старіння і вакцинацію, а також втрату материнського імунітету. Таким чином, формула 3.2 перетворюється на таку:

$$S_{i,t+1} = S_{i,t} - I_{i,t+1} - V_{i,t} - \text{AgeOut}_{i,t} + \text{AgeIn}_{i,t} + \text{WaningM}_{i,t}, \quad (3.3)$$

де $V_{i,t}$ — кількість вакцинованих, $\text{AgeOut}_{i,t}$ — кількість людей, що внаслідок старіння перейшли в наступну вікову групу, $\text{AgeIn}_{i,t}$ — кількість людей, що внаслідок старіння перейшли в вікову групу i , $\text{WaningM}_{i,t}$ — кількість дітей, що втратили материнський імунітет (для $i > 0$ $\text{WaningM}_{i,t} = 0$).

Проблема обчислення швидкості інфікування за формулою λS , як у детермінованих моделях (див. розділ 2.1), полягає в тому, що TSIR використовує дискретні кроки довжиною 14 діб, і відповідно, може трапитись так, що лінійно апроксимована кількість інфікованих може перевищити кількість вразливих, що фізично не є можливим. Модифікацією процесу обчислення швидкості інфікування є формула Ріда-Фроста: ймовірність зараження на кожному кроці оцінюється як $1 - e^{-\lambda}$. Таким чином, математичне сподівання кількості нових інфікованих у кожній віковій групі визначається нелінійним рівнянням:

$$E[I_{i,t+1}] = S_{i,t} (1 - \exp(-\lambda_{i,t})) \quad (3.4)$$

Таке формулювання гарантує, що кількість нових інфікованих асимптотично наближається до загальної кількості вразливих $S_{i,t}$, але ніколи не перевищує її. Це забезпечує фізичну та біологічну коректність симуляції.

Окрім цього, замість лінійної регресії застосуємо оптимізацію методом максимальної правдоподібності. В порівнянні з методом найменших квадратів, MLE працює швидше і зберігає форму кривої, що ми і хочемо побачити.

Окрім цього, оскільки найбільшу частку інфікованих на кір становлять діти, має сенс розглянути сезонність — в залежності від того, чи перебувають діти на канікулах, чи ні, змінюється кількість контактів. Щоб запобігти перенавчанню моделі для 26 різних значень параметрів β (з кроком 14 днів за рік), було параметризовано два стани системи:

$$\beta_t = \begin{cases} \beta_{school}, & \text{якщо крок } t \text{ припадає на навчальний семестр} \\ \beta_{school} \cdot (1 - c), & \text{якщо крок } t \text{ припадає на канікули} \end{cases} \quad (3.5)$$

Тут параметр $c \in (0, 1)$ показує, на скільки зменшується кількість контактів на канікулах. Для спрощення моделі вважається, що сама матриця контактів не змінюється.

Для кращого аналізу потрібно врахувати, в який час в Японії відбуваються шкільні канікули [18]: кінець липня-кінець серпня (3 кроки: 15-17), кінець грудня-початок січня (2 кроки: перший та останній), кінець березня-початок квітня (1 крок: 6).

Сила інфекції в кожній віковій групі i розраховується з урахуванням контактів зі всіма групами:

$$\lambda_{i,t} = d_{inf} \cdot \beta_t \sum_{j=1}^4 C_{i,j}^{adj}(t) \left(\frac{I_{j,t}^\alpha}{N_{j,t}} \right) \quad (3.6)$$

Тут $d_{inf} = 7$ днів — тривалість періоду заразливості в межах 14-денного кроку, а параметр $\alpha \approx 0.97$ береться аналогічно до формули 3.2. Застосовуючи формулювання Ріда-Фроста, математичне сподівання кількості нових випадків інфікування $\mu_{i,t}$ визначається як:

$$\mu_{i,t} = S_{i,t} (1 - \exp(-\lambda_{i,t})) + \text{Imp}_{i,t} \quad (3.7)$$

Тут $\text{Imp}_{i,t}$ — кількість імпортованих випадків (кожні 2 тижні будемо 'імпортувати' трохи інфікованих, щоб вірус не згасав повністю і оптимізатор не показав пряму лінію $I = 0$).

Реальна кількість нових інфікованих $I_{i,t+1}^{new}$ моделюється як випадкова величина з від'ємного біноміального розподілу:

$$I_{i,t+1}^{new} \sim \text{NegBinomial}(\mu_{i,t}, k) \quad (3.8)$$

При цьому виконується наступна умова:

$$I_{i,t+1}^{new} \leq S_{i,t} \quad (3.9)$$

Маємо чотири комірки (M — материнський імунітет, S — вразливі, I — інфіковані, R — ті, хто одужав), чисельність населення яких описується рекурентними відношеннями. Нехай ω — частка немовлят, що втрачають материнський імунітет за один крок dt , v_t — поточний коефіцієнт охоплення вакцинацією, B_t — кількість народжених, а $\delta_{i,1}$ — дельта Кронекера (лише перша комірка поповнюється новонародженими). Нехай Φ_X позначає старіння для комірки X , тобто різницю між кількістю осіб, що постаріли і прийшли з молодшої вікової групи, та тими, що перейшли до старшої

вікової групи. Тоді фінальна система рівнянь має вигляд:

$$\begin{cases} M_{i,t+1} = M_{i,t} - \omega M_{i,t} + B_t \delta_{i,1} + \Phi_M \\ S_{i,t+1} = S_{i,t} + \omega M_{i,t}(1 - v_t) - I_{i,t+1}^{new} + \Phi_S \\ I_{i,t+1} = I_{i,t+1}^{new} + \Phi_I \\ R_{i,t+1} = R_{i,t} + I_{i,t} + \omega M_{i,t} v_t + \Phi_R \end{cases} \quad (3.10)$$

3.3. Застосування моделі

3.3.1. Початкова ініціалізація

Спочатку вважаємо, що $\beta_{school} = 0.4$, $c = 0.3$, $k = 2$ (тобто, на канікулах інтенсивність передачі вірусу зменшується на 30%). $\beta_{school} = 0.4$ береться як округлене значення параметра, отримане в рівнянні 2.24. Матрицю контактів і всі раніше використані дані обрано таким самим чином, як в розділі 2.3.1.

Програмну реалізацію виконано в середовищі Python. Результат роботи програми при початкових значеннях β , c , k наведено на рис. 3.1. Час роботи програми становить 0.15 секунд, що показує, що TSIR працює набагато швидше, ніж MSEIR.

3.3.2. Оптимізація методом максимальної правдоподібності

Метод максимальної правдоподібності, описаний в розділі 2.3.3, не згладжує криву і краще відображає її форму, навіть якщо розв'язок не є оптимальним з точки зору середньоквадратичної похибки.

Як зазначено в [23], оптимізаційні алгоритми, що покладаються на обчислення градієнта (зокрема, L-BFGS-B, розглянутий у розділі 2.3.3), є непридатними для оптимізації цільових функцій, де похідні математично недоступні. У випадку моделі TSIR, оскільки кількість інфікованих на ко-

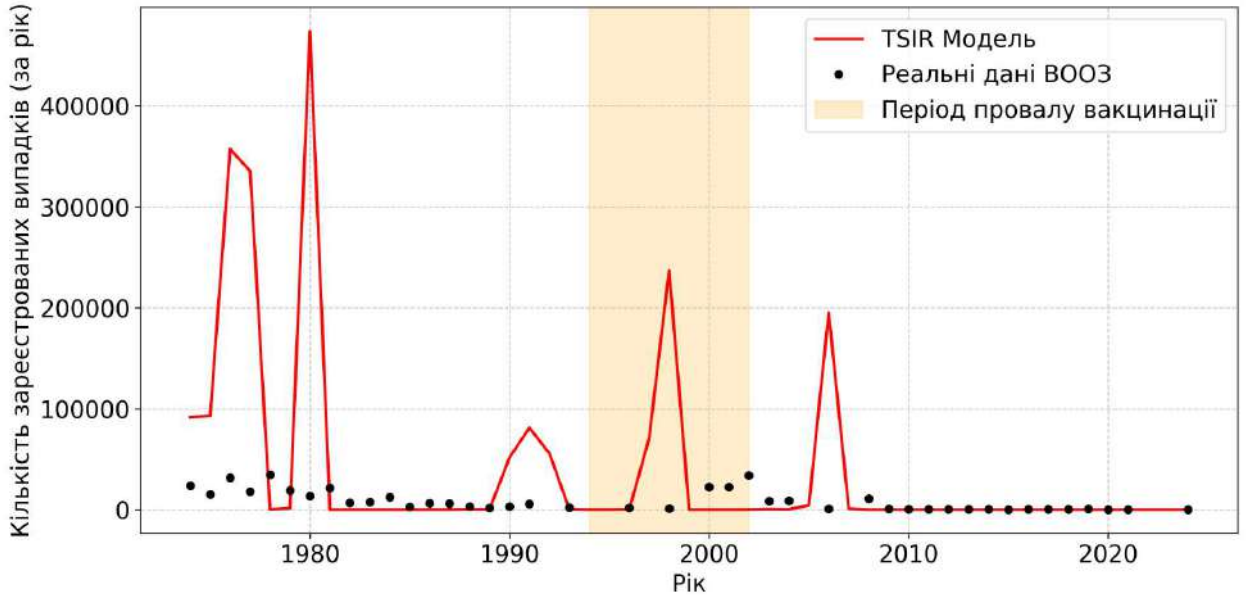


Рис. 3.1: Модель TSIR при початкових значеннях $\beta_{school} = 0.4$, $c = 0.3$, $k = 2$.

жному кроці є цілим числом, цільова функція (NLL) набуває форми недиференційованої ступінчастої функції відносно неперервних параметрів. Через це оптимізатор перестає працювати на етапі чисельного наближення градієнта за допомогою скінченних різниць:

$$\nabla f(x_k) \approx \frac{f(x_k + \epsilon) - f(x_k)}{\epsilon}, \quad (3.11)$$

де ϵ — крок скінченної різниці (зазвичай порядку 10^{-8} , що відповідає кореню з машинного нуля). При невеликому зсуві параметра на величину ϵ , дискретний результат симуляції (кількість інфікованих) залишається незмінним. Відповідно, значення цільової функції також не змінюється ($f(x_k + \epsilon) = f(x_k)$), і чисельний градієнт дорівнює нулю: $\nabla f(x_k) = 0$. Оскільки напрямок пошуку в квазіньютонівських методах (таких як BFGS) обчислюється як $p_k = -B_k^{-1} \nabla f(x_k)$, де B_k — наближення матриці Гессе, отримання нульового градієнта призводить до відсутності напрямку оптимізації ($p_k = 0$). Через це крок алгоритму $s_k = \alpha_k p_k = 0$, і точка наступної ітерації залишається в початковій позиції: $x_{k+1} = x_k$. Оптиміза-

тор хибно ідентифікує нульовий градієнт як досягнення локального мінімуму і достроково завершує роботу, видаючи початкові параметри як оптимальні. З огляду на це, для параметризації таких стохастичних моделей необхідно використовувати прямі методи оптимізації, які не потребують обчислення градієнта. Одним із найбільш надійних підходів для цієї задачі є метод повного перебору простору параметрів — пошук по сітці (grid search). Обираємо декілька значень-кандидатів, що відповідають $R_0 \approx 15$: $\beta \in \{0.38, 0.4; 0.43; 0.45; 0.48\}$, $c \in \{0.2; 0.3; 0.4; 0.5; 0.6; 0.7\}$, $k \in \{1; 2; 3; 4; 5; 7; 10; 15; 20; 30; 50; 100; 1000; 2000; 10000\}$. Перебираючи кожну комбінацію (β, c, k) , знаходимо ту, для якої NLL буде найменшою.

Результат роботи програми в середовищі Python показано на рис. 3.2. Оскільки метод пошуку по сітці перебирає всі значення-кандидати по черзі, він є відносно повільним. Час виконання програми становить 216.38 секунд.

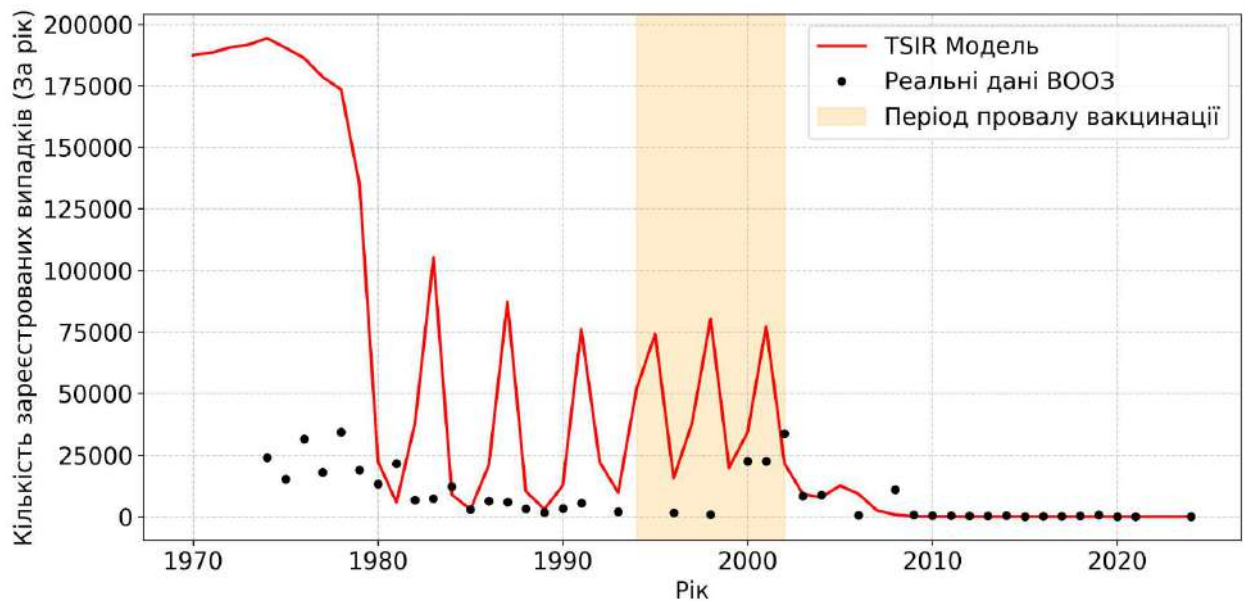


Рис. 3.2: Модель TSIR, оптимізована за допомогою метода максимальної правдоподібності ($\beta_{school} = 0.43$, $c = 0.2$, $k = 1$).

Розділ 4

Порівняльний та сценарний аналіз

4.1. Проблема звітування (Underreporting)

У випадку Японії, як і для багатьох інших країн світу, великим фактором, що ускладнює аналіз та оцінку роботи моделі є звітування. Так, вважається [9], що на початку XXI століття кількість реальних випадків перевищувала кількість відзвітованих випадків у 5-10 разів, а після запровадження активного спостереження у 2006 році майже всі випадки кору були відомі.

Під час локальної епідемії у 2002 році кількість відзвітованих випадків становила близько 34000, тоді як вважається, що реальна кількість випадків оцінюється у 100-300 тисяч [10].

Означення 4.1. *Коефіцієнтом реєстрації* називається частка зареєстрованих випадків і реальних випадків:

$$R = \frac{N_{reported}}{N_{real}}, \quad (4.1)$$

де $N_{reported}$ — кількість зареєстрованих випадків, N_{real} — реальна кількість випадків.

Обрані моделі мали сталий коефіцієнт реєстрації 10%. Це було зроблено для того, щоб подивитися на форму кривої та навести гіпотезу про вплив динамічного коефіцієнта реєстрації на її форму.

Через те, що неможливо встановити реальну кількість випадків, описані

в розділах 2 та 3 моделі не можна перевірити на правильність, використовуючи лише зареєстровані випадки. Для подальшого порівняння моделей пропонується розглянути сценарні аналізи, такі як відсутність вакцинації. Також пропонується подивитися на те, як моделі передбачають кількість випадків кору в Японії в майбутньому, якщо в країну потраплять декілька інфікованих туристів, і зробити висновок про стійкість колективного імунітету країни.

4.2. Сценарій А: Відсутність вакцинації

4.2.1. Узгодження параметрів

В розділах 2 та 3 було зроблено оптимізацію параметрів. Оскільки оптимізація моделі MSEIR показала фізично нереальні результати, для подальшого розрахунку приймаємо результати оптимізації моделі TSIR (див. розділ 3.3.2). Оскільки модель TSIR враховує сезонність і має два різних параметри $\beta_{school} = 0.43$ та $\beta_{holiday} = 0.8 \cdot \beta_{school} = 0.344$, для отримання одного параметра для моделі MSEIR пропонується взяти зважене середнє. Усереднення було виконано шляхом розрахунку середньозваженого значення за річний епідемічний цикл (26 кроків моделі, з яких 20 припадають на навчання, а 6 — на канікули):

$$\bar{\beta} = \frac{20 \cdot \beta_{school} + 6 \cdot \beta_{school}(1 - c)}{26} \approx 0.41 \quad (4.2)$$

Це значення $\bar{\beta} = 0.41$ буде обрано для моделювання сценарного аналізу моделі MSEIR. Воно демонструє узгодженість з теоретичним значенням $\beta = 0.398$, отриманим в рівнянні 2.24.

4.2.2. Моделювання моделей MSEIR та TSIR

В розділах 2 та 3 при програмній реалізації було враховано частку вакцинованих. Якщо замінити її на нуль та запустити програми для моделей MSEIR ($\beta = 0.41$) та TSIR ($\beta_{school} = 0.43$, $c = 0.2$, $k = 1$), маємо результати, показані на рис. 4.1 та рис. 4.2.

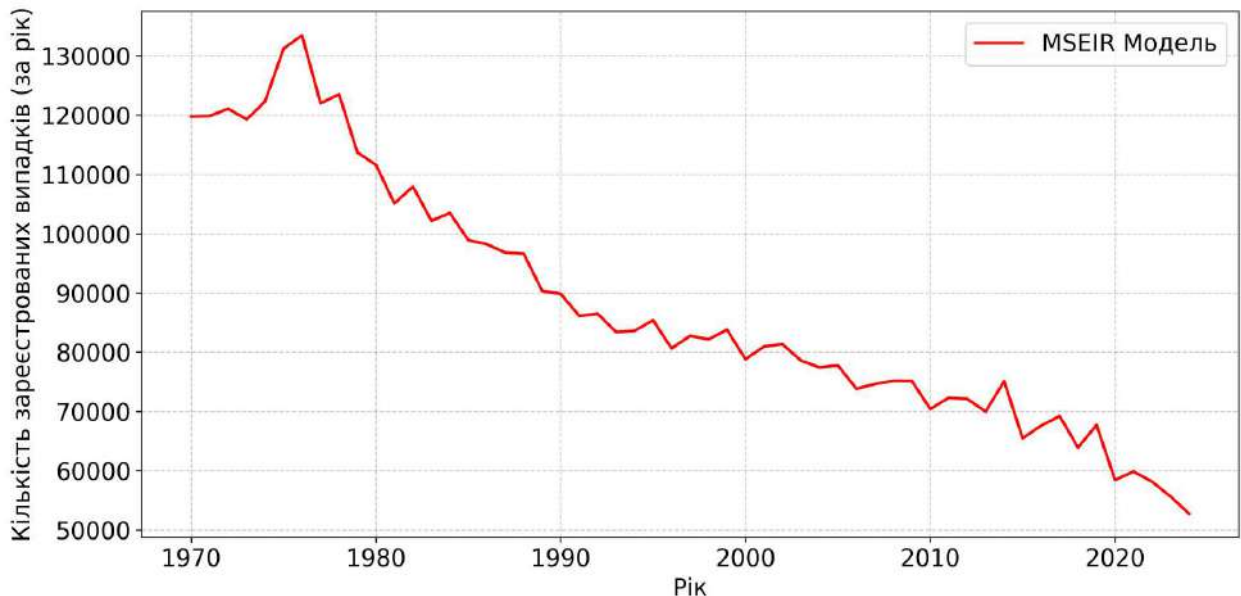


Рис. 4.1: Модель MSEIR за відсутності вакцинації ($\beta = 0.41$).

Як бачимо, детермінована модель не здатна відтворити циклічну природу кору [16], що доводить необхідність переходу до стохастичних моделей. Натомість, графік стохастичної моделі нагадує коливання навколо спадаючої осі. Усі подальші сценарні аналізи буде реалізовано на базі модифікованої моделі TSIR.

Оскільки метою цього дослідження є оцінити реальну кількість випадків кору у випадку відсутності вакцинації, для максимальної схожості з реальністю для моделі TSIR було обрано від'ємний біноміальний розподіл.

Для моделі TSIR також було підраховано, скільки випадків кору в Японії могло б бути, якщо вакцину не винайшли б. Сумарна кількість випадків кору становить 66173247, що становить більше ніж половину поточного на-

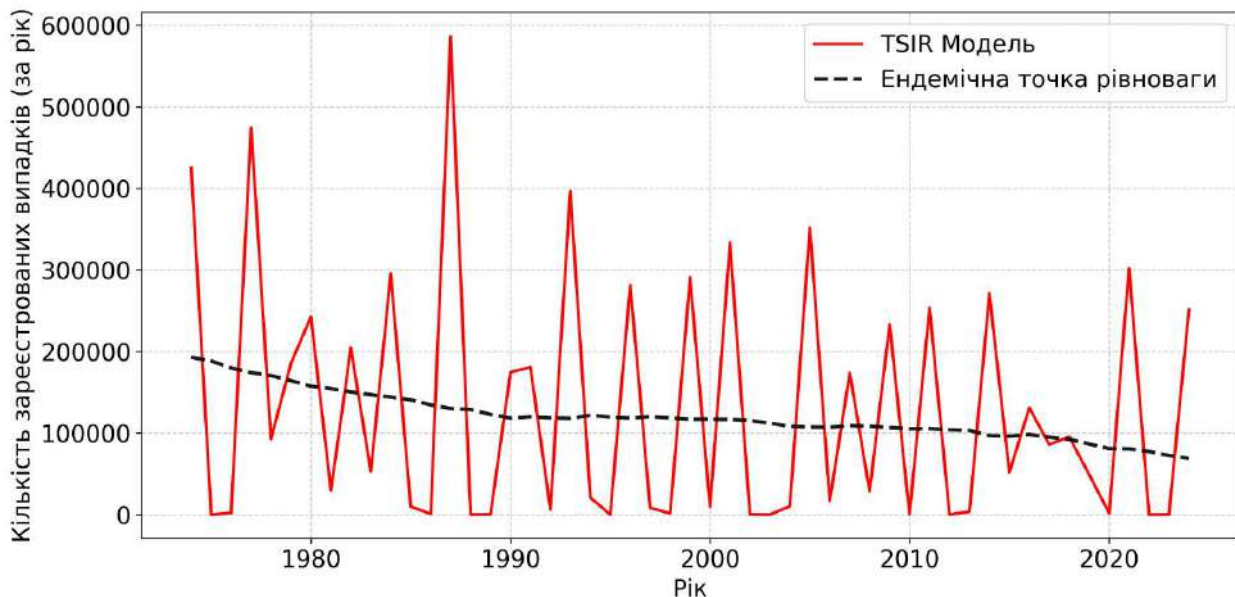


Рис. 4.2: Модель TSIR за відсутності вакцинації ($\beta_{school} = 0.43$, $c = 0.2$, $k = 1$).

селення Японії. Навіть якщо вважати, що смертність від кору становить 1%, вакцинація допомогла зберегти життя сотень тисяч осіб.

4.2.3. Виведення формули ендемічного центру рівноваги моделі TSIR

Аналогічно до [12], виведемо формулу для ендемічного центру рівноваги моделі TSIR.

У класичній моделі TSIR для закритої популяції (без смертності та еміграції) базове рівняння динаміки вразливих осіб має вигляд:

$$S_{t+1} = S_t + B_t - I_{t+1} \quad (4.3)$$

У стані ендемічної рівноваги змінні системи перестають змінюватися з часом, тобто досягають своїх стаціонарних значень:

$$S_{t+1} = S_t = S^* \quad (4.4)$$

$$I_{t+1} = B_t = I^* \quad (4.5)$$

За таких умов рівняння спрощується до наступного:

$$I^* = B \quad (4.6)$$

Це означає, що в стані рівноваги для закритої популяції кількість інфікованих дорівнює рівню народжуваності. Проте, в модифікованій моделі TSIR населення є відкритою системою, і вразливі особи постійно 'витікають' з епідемічного процесу (переходять у старші вікові групи або помирають до моменту інфікування). Позначимо цей сумарний демографічний витік на кроці t як функцію від кількості вразливих $\mu_{leak}S_t$. Тоді модифіковане рекурентне рівняння набуває вигляду:

$$S_{t+1} = S_t + B - I_{t+1} - \mu_{leak}S_t \quad (4.7)$$

Для знаходження справжнього динамічного центру епідемії застосуємо умову стаціонарності ($t \rightarrow \infty$):

$$S^* = S^* + B - I^* - \mu_{leak}S^* \quad (4.8)$$

Віднімаючи S^* з обох частин рівняння, отримуємо:

$$0 = B - I^* - \mu_{leak}S^* \quad (4.9)$$

Переносячи I^* у ліву частину, отримуємо фінальний вираз для справжньої ендемічної рівноваги ($\mu = \mu_{leak}$):

$$I^* = B - \mu S^* \quad (4.10)$$

Зауваження. Формула 4.10 не залежить від розподілу кількості інфікованих.

Розв'язок рівняння 4.10 показано чорною пунктирною лінією на рис. 4.2.

4.2.4. Моделювання моделі TSIR без врахування демографічних змін

Використаємо певні спрощення: візьмемо населення Японії 100 мільйонів осіб, без урахування вікових груп, матриці контактів та демографічних змін (беремо сталу народжуваність 0.01).

Оскільки оптимізоване значення $\beta_{school} = 0.43$, то підставляючи це значення в формулу для базового репродукційного числа (2.22), отримаємо:

$$R_0 \approx 0.43 \cdot 7 \cdot 5.38 \approx 16.2 \quad (4.11)$$

Оскільки розподілу на вікові групи немає, матриця контактів перетворюється на матрицю 1×1 з єдиним елементом (1): в цій матриці не може стояти ніяке інше число, оскільки це суперечить фізичному змісту матриці контактів (якщо замість 1 взяти будь-яке інше число, сума кількості контактів всіх осіб не буде дорівнювати сама собі, чого не може бути). Тоді, обираючи $R_0 = 16.2$, значення β_{school} для цього і наступного прикладу становить:

$$\beta_{school}^* = \frac{16.2}{7} \approx 2.31 \quad (4.12)$$

Це число буде обрано як значення β_{school} . Значення $c = 0.2$ залишиться без змін.

На рис. 4.2 видно, що кількість інфікованих коливається навколо епідемічної точки рівноваги, але не видно, як саме вона наближається до чорної прямої. Метою цього і наступного дослідження було продемонструвати, як саме поводить себе графік моделі TSIR за ідеальних умов. Натомість, при від'ємному біноміальному розподілі за відносно невеликих значень k дисперсія являє собою дуже велике число, оскільки $Var = \mu + \frac{\mu^2}{k}, k \ll \mu$ (μ — це математичне сподівання кількості інфікованих, тому це число ви-

мірюється тисячами). На рис. 4.3 показано, що застосування негативного біноміального розподілу з високою дисперсією у моделі TSIR демонструє явище стохастичного вимирання. Це пояснюється тим, що як тільки стохастичний генератор видає значення $I = 0$, сила інфекції λ стає нульовою. Оскільки модель описує закриту популяцію без імпорту випадків ззовні, подальше інфікування стає неможливим, і вірус зникає, незважаючи на постійне народження нових вразливих дітей. Натомість, якщо обрати достатньо велике значення k , від'ємний біноміальний розподіл майже не буде відрізнятися від розподілу Пуассона.

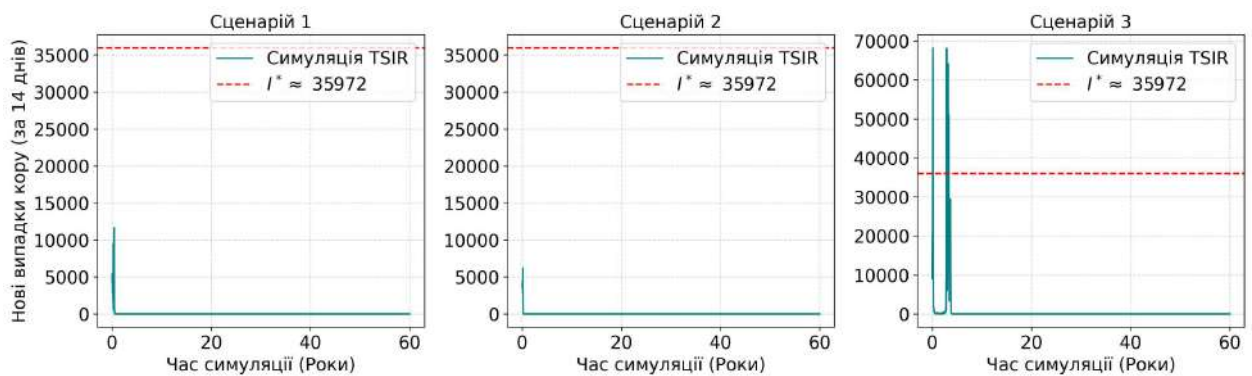


Рис. 4.3: Модель TSIR за відсутності вакцинації, без врахування демографічних змін та вікових груп ($\beta_{school} = 2.31$, $c = 0.2$, $k = 1$, від'ємний біноміальний розподіл, 3 незалежні сценарії). $I^* \approx 35972$ — ендемічна точка рівноваги.

Оскільки мета цього і наступного дослідження — продемонструвати базову механіку моделі TSIR за умови гіпотетичного сценарію відсутності вакцинації та демографічних змін і перевірити, що кількість інфікованих дійсно наближається до ендемічної точки рівноваги, було прийнято рішення обрати розподіл Пуассона замість від'ємного біноміального. Дисперсія при такому розподілі є сталою, і як видно з рис. 4.4, кількість інфікованих коливається, але з часом наближається до епідемічної точки рівноваги, і сила інфекції ніколи не падає до нуля.

На рис. 4.4 показано, як з часом змінюється кількість інфікованих. Ана-

логічно до розділу 4.2.3, ендемічна точка рівноваги має рівняння $I^* = B - \mu S^*$.

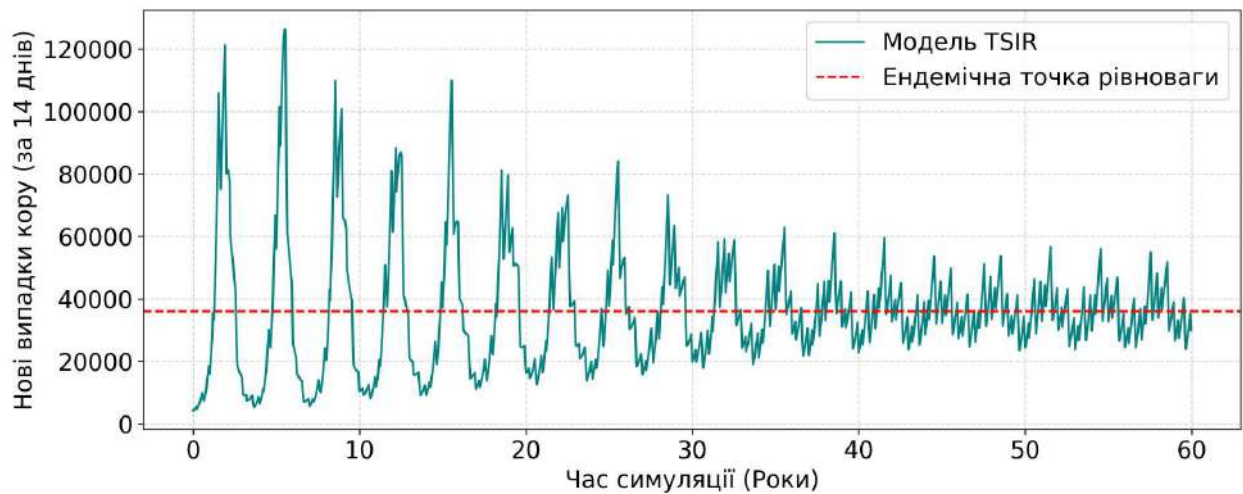


Рис. 4.4: Модель TSIR за відсутності вакцинації, без врахування демографічних змін та вікових груп ($\beta_{school} = 0.43$, $c = 0.2$, розподіл Пуассона).

4.2.5. Моделювання моделі TSIR без врахування демографічних змін та сезонності

Якщо до моделі, описаної в розділі 4.2.4, додати припущення про відсутність сезонності, то замість β_{school}^* та $\beta_{school}^* \cdot (1 - c)$ треба буде взяти зважене середнє за рік. Воно обчислюється за наступною формулою (оскільки 20 кроків моделі TSIR щорічно припадають на навчальний період і 6 — на канікули):

$$\bar{\beta}^* = \frac{20 \cdot \beta_{school}^* + 6 \cdot \beta_{school}^* (1 - c)}{26} \approx 2.2 \quad (4.13)$$

Це значення було обрано для подальшого розрахунку.

Оскільки в розділі 4.2.4 з'ясувалося, що для теоретичного дослідження розподіл Пуассона дає кращі результати, для цього дослідження також було обрано цей розподіл.

Залежність кількості інфікованих від часу показана на рис. 4.5. При цьому ендемічна точка рівноваги задовільняє рівняння $I^* = B - \mu S^*$. За-

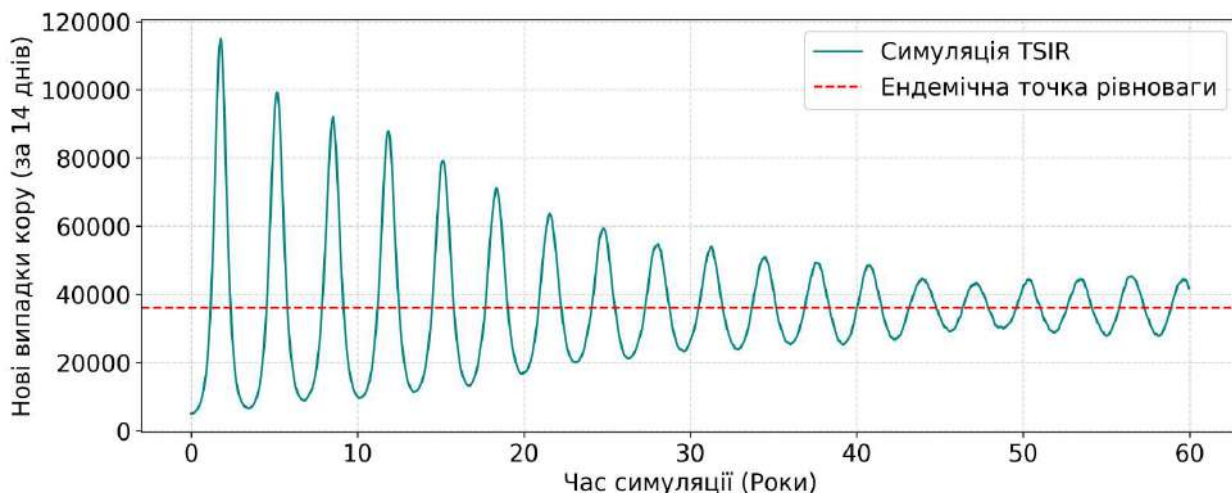


Рис. 4.5: Модель TSIR за відсутності вакцинації, без врахування демографічних змін, вікових груп та сезонності ($\bar{\beta}^* = 2.2$, розподіл Пуассона).

уваження. В [20] та [12] було доведено, що за відсутністю сезонності та стабільною демографією розв'язок системи має наступний вигляд:

$$I(t) \approx I^* + e^{-at} (C_1 \cos(\omega t) + C_2 \sin(\omega t)) \quad (4.14)$$

Це — рівняння затухаючих коливань навколо ендемічного центру I^* . Рис. 4.5 ілюструє ці коливання і їхній центр.

4.3. Сценарій Б: Прогнозування майбутнього

Візьмемо раніше отримані оптимізовані значення параметрів β , c та k , отримані в розділі 3.3.2 для моделі TSIR, і подивимось, як будуть виглядати графіки кількості інфікованих у випадку імпорту в Японію 100 інфікованих і наявності в країні ще 100 інфікованих. При цьому вважаємо, що поріг вакцинації становить 95%. На рис. 4.6 показано прогноз моделі TSIR на 2025-2034 роки. Оскільки метою цього дослідження є перевірка реальних ризиків потенційної епідемії кору в Японії, використовувався від'ємний біноміальний розподіл. Було візуалізовано 30 різних незалежних один від одного сценаріїв, а також виділено найгірший та медіанний сценарії. Ха-

рактальною ознакою прогнозу є той факт, що для будь-якого сценарію за один-два роки всі інфіковані перехворіють і кір зникне. Це означає, що колективний імунітет Японії наразі стабільний і невелика кількість інфікованих не зможе спричинити епідемію.

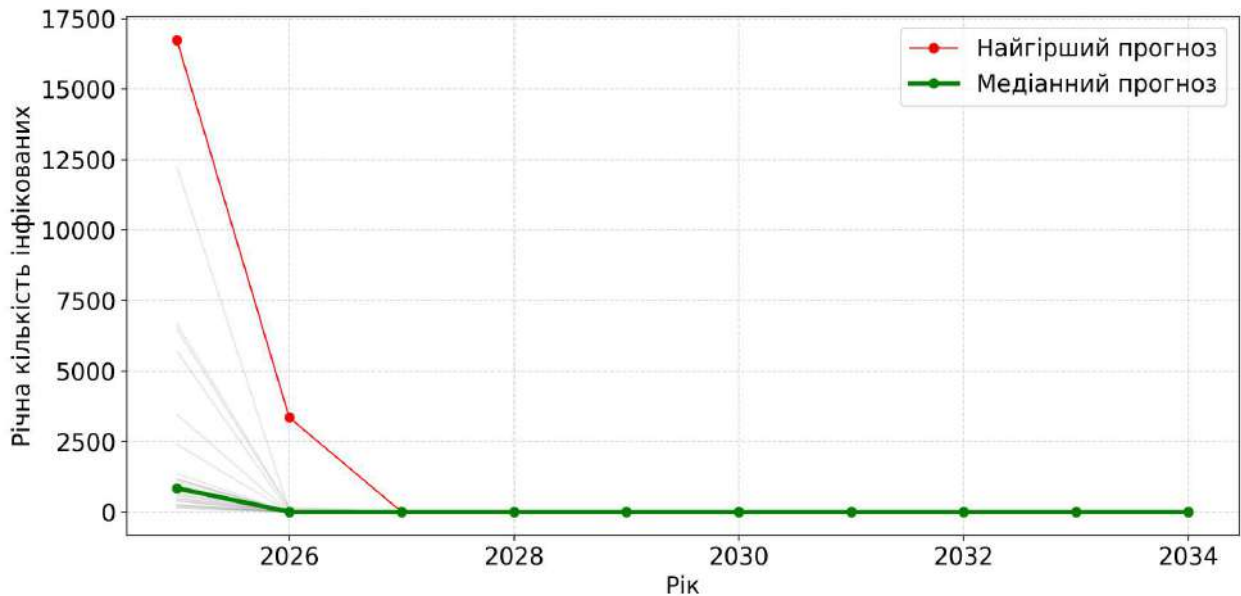


Рис. 4.6: Прогноз моделі TSIR кору в Японії за появи 200 інфікованих (від'ємний біноміальний розподіл).

4.4. Метапопуляційна модель

Всі моделі, описані раніше, розглядали населення кожної комірки як однорідне. Іншими словами, будь-яка особа, що живе в Токіо, мала такий самий шанс контактувати з особою, що живе в Токіо, як і з особою, що живе в провінції. Метапопуляційна модель — це модель, яка розглядає країну не як одне ціле, а як зв'язний граф, вершинами якого є населені пункти (для простоти можна обрати міста). Ідея розглядати метапопуляційні моделі для моделювання епідемій була описана в [14].

Далі буде наведено програмну реалізацію такої метапопуляційної моделі, для простоти обмеженої двома містами (мегаполісом з населенням 1

мільйон осіб і провінцію з населенням 10 тисяч осіб). Оскільки мета цього дослідження — показати необхідність використання метапопуляційних моделей, для спрощення будемо брати розподіл Пуассона кількості інфікованих. Припустимо, що населення мегаполіса має рівень колективного імунітету 95%, а провінція — лише 70% (наприклад, населення провінції не довіряє вакцині). Ця частка охоплює як вакцинованих осіб, так і осіб, що перехворіли і мають пожиттєвий імунітет. Для моделювання буде використано модель TSIR. Кожні 14 діб (тобто, на кожному кроці моделі TSIR) будемо імпортувати 5 інфікованих в мегаполіс і подивимось, чи вистачить цього для того, щоб створити локальну епідемію.

Для того, щоб побачити, як працює транспортний ефект, модель було спрощено і сезонність та вікову структуру не було враховано. Стохастичність процесу моделювалася за допомогою розподілу Пуассона, оскільки використання від'ємного біноміального розподілу потрібно було для опису дисперсії реальних історичних даних, що є надлишковим для теоретичного сценарію.

Важливим параметром метапопуляційної моделі є так званий параметр просторового зв'язку ρ : він показує, яку частку контактів жителя мегаполіса становлять контакти з жителями провінції. Якщо брати населення всієї системи (мегаполіс A + провінція B) як єдине ціле, параметр ρ дорівнював б 0.01 (оскільки чисельність населення провінції становить 1% від чисельності населення міста). Для подальшого аналізу оберемо $\rho = 0.005$ (тобто, житель мегаполіса вдвічі частіше контактує з жителями мегаполіса (якщо брати кількість контактів на душу населення), ніж з жителями провінції).

За умови симетричного зв'язку, сила інфекції $\lambda_{A,t}$ для міста A визначається як лінійна комбінація внутрішнього та зовнішнього інфекційного

ВПЛИВУ:

$$\lambda_{A,t} = \beta_t \left((1 - \rho) \frac{I_{A,t}}{N_{A,t}} + \rho \frac{I_{B,t}}{N_{B,t}} \right) \quad (4.15)$$

Результати симуляції, наведеної на рис. 4.7, наочно демонструють, що за

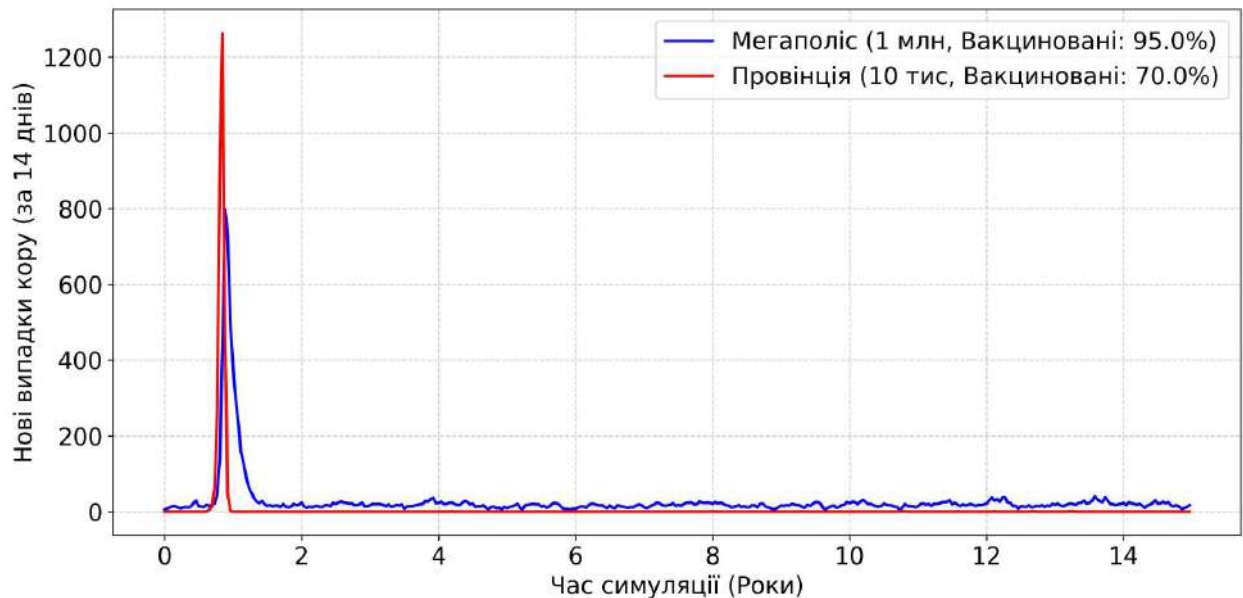


Рис. 4.7: Метапопуляційна модель розповсюдження кору на прикладі двох міст ($R_0 = 15$).

наявності навіть мінімального транспортного зв'язку, спалах у регіоні з низьким рівнем вакцинації неминуче передається в захищений регіон, що спричиняє вторинні епідемічні хвилі. При цьому, бачимо, що спалах вірусу передається майже одразу, тобто, через 2 тижні після того, як вірус спалахнув в провінції, він спалахує у місті, і швидко слабшає.

Оскільки поріг колективного імунітету провінції становить 70%, то на початку симуляції провінція мала 3000 вразливих, серед яких в піковий момент одночасно захворіли більше, ніж 1200 осіб (тобто, $> 40\%$ всіх заразливих). Через 2 тижні після цього епідемія передається в місто, в якому на піку одночасно хворіють близько 800 осіб. Якщо для міста 800 хворих — це невелика частка вразливих (оскільки 5% населення міста є вразливими, це становить 50000 осіб), і скоріш за все, в лікарнях міста вистачить місць для всіх інфікованих. Натомість, якщо в маленькому місті з населенням 10000

осіб одночасно хворіють 1200, місткості місцевих лікарень не вистачить.

З іншого боку, потрібно зазначити, що вірус спалахує не одразу після запуску симуляції, йому потрібен деякий час (в симуляції це приблизно 1 рік), щоб вийти на максимум. Це означає, що моніторинг населених пунктів з низьким порогом колективного імунітету дозволить запобігти розповсюдженню хвороби і перенавантаженню лікарень. На рис. 4.8 показано, як

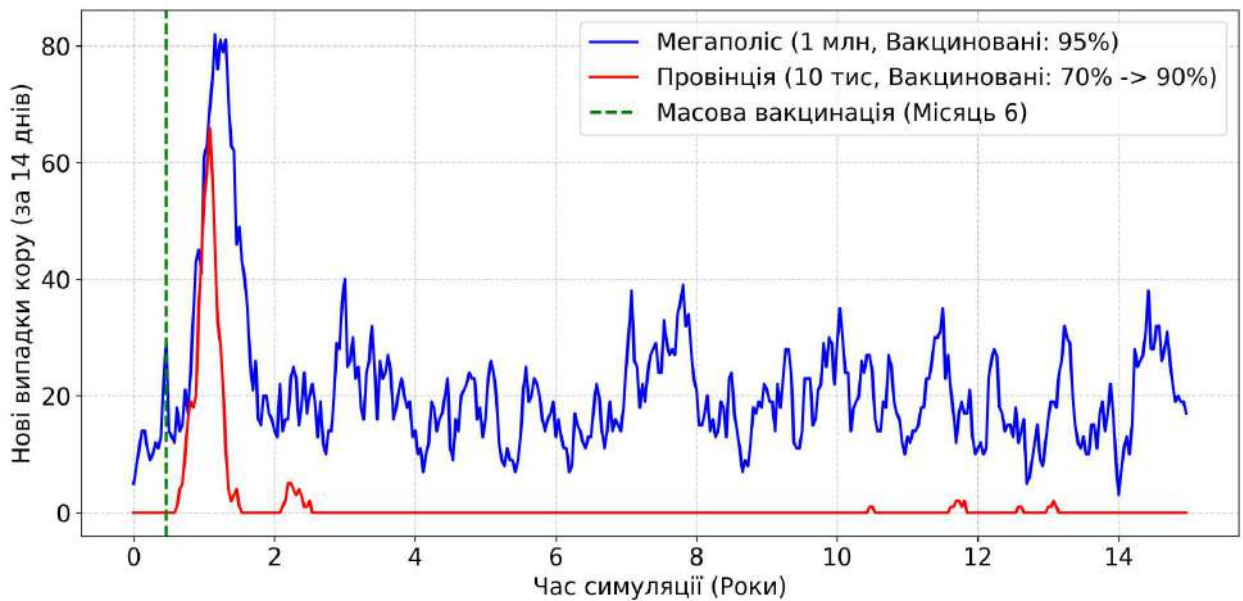


Рис. 4.8: Метопуляційна модель розповсюдження кору на прикладі двох міст ($R_0 = 15$) за умови завчасного запровадження примусової вакцинації.

змінюється кількість інфікованих з часом за умови, що уряд завчасно (через 6 місяців, або 12 кроків TSIR) побачив низький поріг колективного імунітету в провінції і запровадив масову вакцинацію, тим самим підвищивши поріг колективного імунітету до 90%. Бачимо, що хоч вірус і спалахне приблизно через рік після початку симуляції, кількість інфікованих в провінції буде на рівні 65 на піку, а для міста — на рівні 80. В таблиці 4.1 показано, скільки випадків кору виникає в мегаполісі і провінції короткостроково (за 2 роки) і за 15 років, і на скільки відсотків кількість випадків зменшується у випадку запровадження вакцинації. Це показує, що завчасне запровадження вакцинації може запобігти захворюванню не лише жителів провінції,

а й жителів мегаполіса.

Табл. 4.1: Статистика без та після впровадження вакцинації в провінції

Період	Локація	Без вакцинації	Після вакцинації	Зниження (%)
За 2 роки	Мегаполіс	4661	1732	62.8%
	Провінція	2994	563	81.2%
	Разом	7655	2295	70.0%
За 15 років	Мегаполіс	10662	8211	23.0%
	Провінція	3008	612	79.7%
	Разом	13670	8823	35.5%

Це доводить, що розгляд кору як виключно локальної проблеми окремих міст є математично некоректним, і довгострокові стратегії елімінації повинні оцінюватися на рівні національних метапопуляційних мереж.

Висновки

В процесі виконання кваліфікаційної роботи було обґрунтовано недоліки базової моделі SIR. Окрім цього, було ускладнено детерміновану модель шляхом врахування материнського імунітету, латентного періоду, демографічних, соціологічних та епідеміологічних даних. Було виявлено, що оптимізація методом найменших квадратів не є влучним інструментом для моделювання епідемій, оскільки оптимізатор згладжує криву. Також було отримано інструмент, що дозволяє краще проаналізувати динаміку розповсюдження епідемій — метод максимальної правдоподібності.

Було розібрано основні розподіли кількості нових інфікованих на кожному кроці моделі та проаналізовано, які з них влучно використовувати в різних випадках в залежності від мети дослідження. Було розібрано, що при дискретному розподілі квазіньютонівський алгоритм BFGS не спрацює через ступінчасту структуру цільової функції.

Було виявлено недоліки детермінованих моделей шляхом моделювання сценарію відсутності вакцинації. Оскільки форма кривої, отриманої в результаті запуску моделі MSEIR, не відповідає формі кривої, описаної ВООЗ, було прийнято рішення використовувати для подальшого аналізу стохастичну модель TSIR, яка показала правдоподібну форму кривої у випадку відсутності вакцинації.

Було наочно продемонстровано, що за умов відсутності вакцинації, стабільної демографії та відсутності розподілу за віковими групами розв'язок моделі TSIR являє собою комплексну експоненціальну функцію. Також було спрогнозовано поведінку кору за умови імпорту декількох інфікованих і продемонстровано, що вірус згасає менше, ніж за рік.

Окрім цього, було розглянуто метапопуляційну модель на прикладі двох міст і продемонстровано, що низький рівень вакцинації одного міста має безпосередні наслідки для обох міст, і навіть за умов імпорту небагатьох інфікованих в місто з кращим порогом колективного імунітету приблизно через рік відбудеться короткостроковий спалах вірусу в обох містах послідовно, що спричинить перенавантаження лікарень. Окрім цього, було наведено приклад, що за умов завчасного запровадження примусової вакцинації в місті з нижчим рівнем колективного імунітету кількість інфікованих в обох містах короткостроково зменшиться в декілька разів. Відповідно, моніторинг рівня вакцинації міст допоможе запобігти локальним спалахам вірусу.

Таким чином, для подальшого аналізу кору пропонується ускладнити наявну модель TSIR, використавши метапопуляційну модель у вигляді зв'язного графу між найбільшими містами Японії. Незважаючи на те, що оцінити якість моделі на історичних даних важко через невідомий коефіцієнт реєстрації, розібрані в роботі моделі являють собою міцний фундамент для розглядання і прогнозування нових епідемій для довільної країни.

Список використаних джерел

- [1] Measles - number of reported cases [Електронний ресурс] / World Health Organization. — Режим доступу: <https://www.who.int/data> (дата звернення: 12.04.2026).
- [2] Measles vaccination coverage [Електронний ресурс] / World Health Organization. — Режим доступу: <https://immunizationdata.who.int/> (дата звернення: 16.04.2026).
- [3] Population estimates (1970-2024) [Електронний ресурс] / Statistics of Japan. — Режим доступу: <https://www.e-stat.go.jp/en> (дата звернення: 11.04.2026).
- [4] Munasinghe L. Quantifying heterogeneous contact patterns in Japan: a social contact survey / L. Munasinghe, Y. Asai, H. Nishiura // *Theor Biol Med Model.* — 2019. — Vol. 16, No. 6. — DOI: 10.1186/s12976-019-0102-8.
- [5] Griffin D. E. Measles virus, immune control, and persistence / D. E. Griffin, W. H. Lin, C. H. Pan // *FEMS Microbiol Rev.* — 2012. — Vol. 36, No. 3. — P. 649–662. — DOI: 10.1111/j.1574-6976.2012.00330.x.
- [6] The waning of maternal measles antibodies: A multi-country maternal-infant seroprevalence study / K. S. Tiley, H. Ten Hulscher-van Overbeek, S. Basnet [et al.] // *J Infect.* — 2025. — Vol. 91, No. 2. — P. 106531. — DOI: 10.1016/j.jinf.2025.106531.
- [7] The basic reproduction number (R_0) of measles: a systematic review / F. M. Guerra, S. Bolotin, G. Lim [et al.] // *Lancet Infect Dis.* — 2017. — Vol. 17, No. 12. — P. e420–e428. — DOI: 10.1016/S1473-3099(17)30307-9.

- [8] Nakayama T. Vaccine chronicle in Japan / T. Nakayama // *J Infect Chemother.* — 2013. — Vol. 19, No. 5. — P. 787–798. — DOI: 10.1007/s10156-013-0641-6.
- [9] Nakayama T. Current status of measles in Japan / T. Nakayama, J. Zhou, M. Fujino // *J Infect Chemother.* — 2003. — Vol. 9, No. 1. — P. 1–7. — DOI: 10.1007/s10156-002-0206-6.
- [10] Takahashi K. The economic disease burden of measles in Japan and a benefit cost analysis of vaccination, a retrospective study / K. Takahashi, Y. Ohkusa, J. Y. Kim // *BMC Health Serv Res.* — 2011. — Vol. 11. — P. 254. — DOI: 10.1186/1472-6963-11-254.
- [11] Finkenstädt B. F. Time series modelling of childhood diseases: a dynamical systems approach / B. F. Finkenstädt, B. T. Grenfell // *Journal of the Royal Statistical Society: Series C (Applied Statistics).* — 2000. — Vol. 49. — P. 187–205. — DOI: 10.1111/1467-9876.00187.
- [12] Anderson R. M., May R. M. *Infectious Diseases of Humans: Dynamics and Control.* Oxford : Oxford University Press, 1991. 757 p. DOI: 10.1093/oso/9780198545996.001.0001.
- [13] Immunogenicity, effectiveness, and safety of measles vaccination in infants younger than 9 months: a systematic review and meta-analysis / L. M. Nic Lochlainn, B. de Gier, N. van der Maas [et al.] // *Lancet Infect Dis.* — 2019. — Vol. 19, No. 11. — P. 1235–1245. — DOI: 10.1016/S1473-3099(19)30395-0.
- [14] Apolloni A. Metapopulation epidemic models with heterogeneous mixing and travel behaviour / A. Apolloni, C. Poletto, J. J. Ramasco [et al.] // *Theor Biol Med Model.* — 2014. — Vol. 11. — P. 3. — DOI: 10.1186/1742-4682-11-3.

- [15] Bishop C. M. Pattern Recognition and Machine Learning / C. M. Bishop. — Springer, 2006.
- [16] Measles [Электронный ресурс] / World Health Organization. — Режим доступа: <https://www.who.int/news-room/fact-sheets/detail/measles> (дата звернения: 21.04.2026).
- [17] Measles - Diseases and Epidemiology [Электронный ресурс] / World Health Organization Eastern Mediterranean Region. — Режим доступа: <https://www.emro.who.int/health-topics/measles/> (дата звернения: 14.04.2026).
- [18] Overview of Japanese School System [Электронный ресурс]. — Режим доступа: <https://education.jnto.go.jp/en/school-in-japan/japanese-education-system/> (дата звернения: 18.04.2026).
- [19] Li M. Y. An Introduction to Mathematical Modeling of Infectious Diseases / M. Y. Li. — Springer, 2018.
- [20] Keeling M. J. Modeling infectious diseases in humans and animals / M. J. Keeling, P. Rohani. — Princeton University Press, 2008.
- [21] Johnson N. L. Univariate Discrete Distributions / N. L. Johnson, A. W. Kemp, S. Kotz. — John Wiley & Sons, 1992.
- [22] Measles-Japan [Электронный ресурс] / O. Stepanenko. — Режим доступа: <https://github.com/OleksiiStepanenko-main/Measles-Japan> (дата звернения: 23.04.2026).
- [23] Kochenderfer M. J. Algorithms for Optimization / M. J. Kochenderfer, T. A. Wheeler. — MIT Press, 2019.

- [24] Степаненко О. О. Моделювання розповсюдження інфекцій за допомогою диференціальних рівнянь / О. О. Степаненко // Сучасні проблеми математики та її застосування в природничих науках та інформаційних технологіях : тези XX Міжнар. наук.-практ. конф. студ. та молодих вчених (м. Харків, 5–6 трав. 2026 р.). — Харків : ХНУ ім. В. Н. Каразіна, 2026. — С. 118–120.