

Міністерство освіти і науки України
Департамент освіти і науки виконавчого органу Київської міської ради
(Київської міської державної адміністрації)
Комунальний позашкільний навчальний заклад «Київська Мала академія наук
учнівської молоді»

Відділення математики
Секція: математичне моделювання

ПРОГНОЗУВАННЯ РОЗВИТКУ ЕПІДЕМІЇ КОРОНАВІРУСНОЇ ХВОРОБИ НА ТЕРИТОРІЇ УКРАЇНИ

Роботу виконала:
Степаненко Юлія Вікторівна,
учениця 11 класу
Ліцею №100 «Поділ» м.Києва
КПНЗ «Київська Мала академія наук
Учнівської молоді»

Науковий керівник:
Єфімова Тетяна Леонідівна,
керівник секції математичного
моделювання КПНЗ «Київська Мала
академія наук учнівської молоді»,
вчитель математики гімназії №283 м.Києва,
старший науковий співробітник,
кандидат фізико-математичних наук

Комунальний позашкільний навчальний заклад
«Київська Мала академія наук учнівської молоді»

Анотація



Степаненко Юлія Вікторівна
учениця 11 класу
ліцею №100 «Поділ» м. Києва

Науковий керівник: Єфімова Тетяна Леонідівна, керівник секції математичного моделювання КПНЗ «Київська Мала академія наук учнівської молоді», вчитель математики гімназії №283 м. Києва, старший науковий співробітник, кандидат фізико-математичних наук

**Прогнозування розвитку епідемії коронавірусної хвороби на території
України**

Коронавірусна хвороба — хвороба, що впливає на життя мільйонів людей в усьому світі. Прогнозування її розвитку є дуже актуальним, бо дає можливість планувати навантаження на систему охорони здоров'я, необхідну кількість медперсоналу, медичного та лабораторного обладнання, кількості вільних ліжок у лікарнях та введення обмежувальних заходів, таких як локдаун.

Метою даної роботи є дослідження методів прогнозування розвитку епідемії та застосування методу Хольта – Вінтерса для складання власного прогнозу з найменшим відхиленням від статистичних даних.

У роботі розглянуто класичну SIR модель прогнозування розвитку епідемій та виявлено, що ця модель у класичному вигляді підходить лише для довгострокового прогнозування загального напрямку розвитку епідемії за умови наявності достатньої кількості достовірних статистичних даних. У роботі розглянуто статистичний підхід до прогнозування із застосуванням методу потрійного експоненційного згладжування часових рядів за Хольтом – Вінтерсом. Вдалий вибір сезонності та коефіцієнтів методу дозволив досягти малого відхилення від статистичних даних. Створено вебпортал для динамічного відображення результатів прогнозування.

Ключові слова: коронавірусна хвороба, SIR модель прогнозування розвитку епідемій, аналіз часових рядів, метод Хольта-Вінтерса.

ЗМІСТ

| | |
|--|----|
| ВСТУП | 4 |
| РОЗДІЛ 1. ТЕОРЕТИЧНІ ВІДОМОСТІ ПРО ПРОГНОЗУВАННЯ ЕПІДЕМІЙ | 6 |
| 1.1. Сутність прогнозування та класична модель | 6 |
| 1.2. Змінні дані у класичній моделі | 8 |
| 1.3. Параметри у класичній моделі | 8 |
| Висновки до розділу 1 | 10 |
| РОЗДІЛ 2. БАЗОВА SIR МОДЕЛЬ ДЛЯ КОРОНАВІРУСНОЇ ХВОРОБИ | 11 |
| 2.1. Застосування базової моделі для коронавірусної хвороби в Україні та аналіз отриманих результатів | 11 |
| Висновки до розділу 2 | 14 |
| РОЗДІЛ 3. ПРОГНОЗУВАННЯ З ВИКОРИСТАННЯМ МЕТОДУ ХОЛЬТА–ВІНТЕРСА | 15 |
| 3.1. Екстраполяція та експоненційне згладжування | 15 |
| 3.2. Метод аналізу часових рядів Хольта – Вінтерса | 17 |
| 3.3. Використання метода Хольта-Вінтерса для прогнозування кількості тих, що захворіли в Україні та аналіз отриманих результатів | 18 |
| Висновки до розділу 3 | 21 |
| РОЗДІЛ 4. ВЕБПОРТАЛ ДЛЯ ДИНАМІЧНОГО ПРОГНОЗУВАННЯ | 22 |
| 4.1. Призначення вебпорталу | 22 |
| 4.2. Можливості вебпорталу | 22 |
| Висновки до розділу 4 | 23 |
| ВИСНОВКИ | 24 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ | 25 |

ВСТУП

Коронавірусна хвороба — хвороба, що впливає на життя мільйонів людей в усьому світі. Прогнозування її розвитку є дуже **актуальним**, бо дає можливість планувати навантаження на систему охорони здоров'я, необхідну кількість медперсоналу, медичного та лабораторного обладнання, кількості вільних ліжок у лікарнях та введення обмежувальних заходів, таких як локдаун. Темою даної наукової роботи є прогнозування розвитку коронавірусної хвороби на території України та за її межами.

Метою даної роботи є дослідження методів прогнозування розвитку епідемії, застосування методу Хольта-Вінтерса для складання власного прогнозу з найменшим відхиленням від статистичних даних.

Досягнення зазначеної мети передбачає виконання таких **завдань**:

- 1) Розглянути класичну модель прогнозування розвитку епідемії;
- 2) Розглянути метод тренд-аналізу часових рядів з використанням експоненційного згладжування за Хольтом-Вінтерсом;
- 3) Побудувати власні прогнози на основі вищевказаних методів;
- 4) Створити вебпортал для динамічного відображення результатів прогнозування.

Об'єкт дослідження: епідемія COVID-19.

Предмет дослідження: прогнозування тенденцій розвитку епідемії коронавірусної хвороби в Україні.

Методи дослідження:

- 1) Застосування класичної SIR моделі.
- 2) Застосування методу аналізу часових рядів з використанням експоненційного згладжування Хольта-Вінтерса.
- 3) Розробка алгоритму динамічного налаштування моделі з урахуванням впливу зовнішніх факторів.

Наукова новизна роботи полягає у наступному:

- 1) автором **самостійно** проведено чисельний експеримент для різних співвідношень параметрів базового репродуктивного числа (R_0) у класичній SIR моделі із застосуванням статистичних даних МОЗ;
- 2) автором **самостійно** підібрано коефіцієнти для методу Хольта – Вінтерса;
- 3) автором **самостійно** проведено чисельний експеримент та визначено величину похибки для прогнозування за допомогою методу Хольта – Вінтерса із застосуванням статистичних даних МОЗ;
- 4) автором **самостійно** досліджено ефективність використання для короткострокового прогнозування класичної SIR моделі для епідемій та методу статистичного аналізу Хольта-Вінтерса;
- 5) **вперше** створено вебпортал з наочним відображенням прогнозу кількості осіб, що захворіли та величиною похибки.

Робота носить **дослідницький характер**.

Достовірність отриманих результатів забезпечується порівнянням результатів прогнозування із наявними статистичними даними МОЗ.

Теоретичне значення отриманих результатів полягає у дослідженні методів прогнозування епідемій та обрання оптимального методу.

Практичне значення отриманих результатів полягає в тому, що отриманий прогноз дозволяє прогнозувати результати обмежувальних заходів як локдаун, планувати майбутні заходи в умовах карантину, обираючи оптимальні дати, тощо.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ВІДОМОСТІ ПРО ПРОГНОЗУВАННЯ ЕПІДЕМІЙ

1.1. Сутність прогнозування та класична модель

На сьогодні сучасний стан моделювання епідемій можна охарактеризувати як такий, що активно розвивається через поширення COVID-19. Проте, оскільки такий активний розвиток почався нещодавно, більшість вчених беруть за основу вже відомі моделі та методології, що тестувалися за допомогою теоретичних та статистичних відомостей про попередні епідемії, такі як Гонконгський грип чи пандемії чуми.

Розглянемо основні підходи до прогнозування розвитку епідемій на прикладі класичної моделі SIR. Вони дають змогу зрозуміти складну динаміку епідемій та їхні особливості. Вперше така модель була запропонована О. Кермаком та Андерсоном Греєм МакКендріком [5].

У найпростішому випадку населення поділяють на дві групи: особи, що потенційно схильні до хвороби (позначають, як S — від англ. susceptible), та вже інфіковані особи (позначають, як I — від англ. infected). Для дослідження цих моделей використовують звичайні диференціальні рівняння (які є детермінованими). Пізніше дані моделі почали враховувати й осіб, що одужали (позначають, як R — від англ. recovered).

У даній роботі використовувалася стандартна модель, без урахування додаткових факторів, таких як: латентний (інкубаційний) період хвороби, смертність від хвороби, імунітет, надбаний за допомогою вакцинації, життєвий цикл населення (демографія: народження/смертності), безсимптомні випадки та інші. Існують адаптації моделі, які тим чи іншим шляхом враховують ці фактори впливу, наприклад, SEIR, E — від англ. exposed, що враховує інкубаційний період, SIRD, D — від англ. Deceased і т.д. Ці моделі, хоч і не суттєво, але дають трохи інші результати.

Модель, що ми розглядаємо описується наступною системою диференціальних рівнянь:

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

де N —загальна кількість населення території, що досліджується;

S — кількість “вразливих” на день дослідження;

I — кількість інфікованих на день дослідження;

R — кількість таких, що одужали на день дослідження;

β — кількість контактів, що призвели до зараження;

γ — період “заразності” у днях;

$\frac{\beta}{\gamma}$ —базове репродуктивне число (R_0).

Можна зробити доволі очевидний висновок: сума інфікованих, “підозрюваних”, та одужавших є константою та дорівнює загальній кількості населення досліджуємої групи, адже, будь-яку людину, незалежно від того, чи хворіє вона зараз, можна віднести до однієї з трьох вищеописаних категорій.

Також динаміка поширення хвороби дуже серйозно залежить від коефіцієнта поширення інфекції R_0 . Це співвідношення кількості людей, яку заражена людина може “заразити” за період своєї хвороби до “періоду заразності” тієї самої людини, тобто, скільки днів людина є переносником інфекції.

1.2. Змінні дані у класичній моделі

Як вже було зазначено раніше, класичну модель можна використовувати як у розширеному, так і у більш “вузькому” розумінні. Проте є певні змінні, що використовуватимуться незалежно від того, який підвид моделі буде обраний.

S — Susceptible;

I — Infected;

R — Recovered.

Саме з цих трьох складових і складається назва моделі.

S —Susceptible (з англійської — сприйнятливий, вразливий). Це особи, що “мають шанс заразитися”. Тобто ще не перехворіли, не одужали, не є інфікованими, а лише є людьми, які ще не контактували з хворобою взагалі.

I —Infected (з англійської — заражений, інфікований). Це особи, що наразі переносять хворобу, є зараженими та заразними.

R —Recovered (з англійської — одужав). Це група осіб, що вже перенесли хворобу і більше не її переносниками. Попри те, що деякі люди можуть продовжувати боротися з ускладненнями чи побічними наслідками перенесеної хвороби, вони все одно потрапляють у цю категорію щойно вони вже не є інфікованими досліджуваною хворобою.

У цій моделі, сума усіх трьох складових є константою (N) та дорівнює населенню території, до якої ця модель застосовується

$$N=S+I+R$$

1.3. Параметри у класичній моделі

Параметри — це сталі величини, значення яких можуть змінюватись за певних умов та впливати на загальний результат.

Одним із параметрів класичної моделі SIR є базове репродукційне число (R_0) — середня кількість осіб, що безпосередньо інфікуються хворим упродовж усього заразного періоду хвороби за умови потрапляння хворого до повністю незараженої популяції. Важливість цього числа полягає в тому, що у залежності від його значення, можна не лише робити висновки про “активність” розповсюдження хвороби, але й “коригувати” швидкість її розповсюдження.

Рональд Росс, Альфред Лотки та інші були одними з перших вчених, що вивчали та ввели поняття репродуктивного числа [8]. Проте, вперше його застосуванням зайнявся Джордж Макдональд, який будував моделі поширення малярії у 1952 році [6].

Репродуктивне число складається ще з двох складових та фактично є їхнім відношенням:

β — кількість контактів, або кількість людей, яких переносник заразить за одиницю часу, та

γ — період “заразності” цієї самої людини, період, коли особа може заразити інших

Якщо індивід у заразному періоді хвороби здійснює β контактів за одиницю часу, продукуючи нові інфікування з середнім заразним періодом $1/\gamma$, тоді базове репродукційне число розраховується так:

$$R_0 = \beta/\gamma$$

Базове репродуктивне число не змінюється за допомогою вакцинацій чи інших факторів, що змінюють сприйнятливість населення до хвороби. Проте воно досить стрімко змінюється у разі введення таких запобіжних засобів як соціальна дистанція, “локдаун” та інші.

Для прикладу візьмемо тяжкий гострий респіраторний синдром (ТГРС) (англ. SevereAcuteRespiratorySyndrome / SARS) та кір. Для першої хвороби репродуктивне число — 0.19–1.08, а для другого — 12-18. Це говорить про те, що

ТГРС поширюється повільніше, є не таким швидким і активним у порівнянні з кором, що є надзвичайно швидким у поширенні.

З вищевказаної інформації можемо зробити наступні висновки:

- якщо репродуктивне число менше ніж 1, то інфекція вже знаходиться на стадії згасання та через деякий прогнозований час майже цілком завершить своє існування на території дослідження.
- якщо репродуктивне число більше ніж 1, то дана інфекція активно поширюватиметься серед населення та продовжуватиме своє існування. Що більше це число, то складніше контролювати поширеність хвороби.

Висновки до розділу 1

Проаналізувавши теоретичні відомості, можна зробити висновок, що моделювання епідемій досить добре вивчено на прикладі попередніх епідемій. Проте прогнозування COVID-19 досліджено не настільки добре. Тому ця тема є новою і цікавою для дослідження.

РОЗДІЛ 2

БАЗОВА SIR МОДЕЛЬ ДЛЯ КОРОНАВІРУСНОЇ ХВОРОБИ

2.1. Застосування базової моделі для коронавірусної хвороби в Україні та аналіз отриманих результатів

Сформуємо задачу Коші для системи диференціальних рівнянь (1.1). Для цього візьмемо статистичні дані Міністерства Охорони Здоров'я України та РНБО на конкретну дату, наприклад, 7 листопада:

Таблиця 2.1

| Дата | 07/11 | 08/11 | 09/11 | 10/11 | 11/11 | 12/11 |
|-----------|--------|--------|--------|--------|--------|--------|
| Хворіє | 238393 | 245015 | 251310 | 255784 | 259402 | 264026 |
| Одужало | 204229 | 206866 | 209143 | 214657 | 221459 | 227694 |
| Захворіло | 10746 | 9397 | 8687 | 10179 | 10611 | 11057 |

Використовуючи додаткову інформацію, а саме: кількість населення України 41983564 осіб, можна порахувати такі початкові значення для прогнозу на 8 листопада:

S 41540943,

I 238393,

R 204229

Додатково нам потрібні такі параметри як базове репродуктивне число, “період заразності” та кількість “заражених” людей інфікованим.

За даними Вікіпедії на осінь 2020 року, вважається, що репродуктивне число може змінюватись у діапазоні від 1.5 до 5.7, інкубаційний період може складати від 2 до 15 днів, в середньому 5 днів.

Спираючись на систему рівнянь (1.1), проведемо розрахунки та побудуємо можливі графіки розвитку епідемії. Розрахунки для всіх початкових даних, наведених вище, проведемо із різними параметрами. У таблиці 2.2 деякі дробі не скорочувалися для кращого розуміння змісту параметрів. Розрахунки

виконувались за допомогою метода Ейлера та метода Рунге – Кута четвертого порядку. Розрахунки проводилися як вручну, так і з використанням Microsoft Excel та пакету Maple.

Метод Рунге-Кута полягає у наступному.

Нехай дано задачу Коші для системи звичайних диференціальних рівнянь

$$\bar{y}' = \bar{f}(t, \bar{y}), \quad \bar{y}(t_0) = \bar{y}_0,$$

де $\bar{y}, \bar{f}, \bar{k} \in R^n$, $t, h \in R^1$, h – крок по змінній t .

Тоді наближене значення в наступних точках обраховується за ітераційною схемою

$$\bar{y}_{n+1} = \bar{y}_n + \frac{h}{6}(\bar{k}_1 + 2\bar{k}_2 + 2\bar{k}_3 + \bar{k}_4),$$

$$\bar{k}_1 = \bar{f}(t_n, \bar{y}_n),$$

$$\bar{k}_2 = \bar{f}\left(t_n + \frac{h}{2}, \bar{y}_n + \frac{h}{2}\bar{k}_1\right),$$

$$\bar{k}_3 = \bar{f}\left(t_n + \frac{h}{2}, \bar{y}_n + \frac{h}{2}\bar{k}_2\right),$$

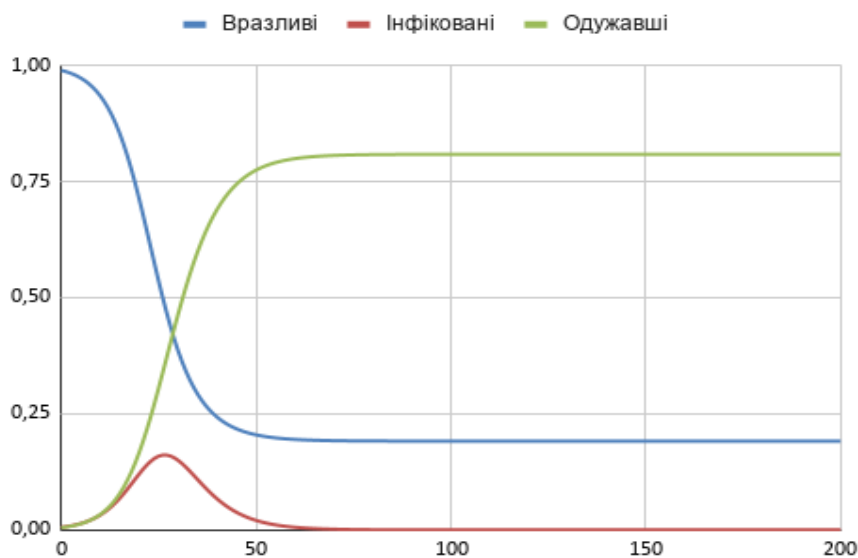
$$\bar{k}_4 = \bar{f}(t_n + h, \bar{y}_n + h\bar{k}_3).$$

Цей метод має четвертий порядок точності, сумарна похибка є $O(h^4)$.

Таблиця 2.2

| Налаштування для моделювання | | | Пояснення налаштувань: захворіє | Результат: кількість захворівши/день | | | | |
|------------------------------|-------|----------|---------------------------------|--------------------------------------|-------|-------|-------|-------|
| β | R_0 | γ | | 08/11 | 09/11 | 10/11 | 11/11 | 12/11 |
| 2/15 | 2 | 1/15 | 2 людини за 15 днів | 15558 | 16548 | 17597 | 18709 | 19887 |
| 2/5 | 2 | 1/5 | 2 людини за 5 днів | 46673 | 55555 | 66016 | 78291 | 92628 |
| 1/10 | 1,5 | 1/15 | 1.5 людини за 15 днів | 7695 | 7930 | 8170 | 8418 | 8671 |
| 3/10 | 1,5 | 1/5 | 1.5 людини за 5 днів | 23085 | 25189 | 27456 | 29896 | 32513 |
| Дані від РНБО | | | | 9397 | 8687 | 10179 | 10611 | 11057 |

На рис.2.1, 2.2 представлено графіки розвитку захворюваності. По осі Y, для кращої візуалізації відображено відсоткову частку населення, що відповідає вразливим (синя лінія), інфікованим (червона лінія) та тих, що одужали (зелена лінія).

Рис 2.1 - Графік розвитку захворюваності при $\beta=2/5$, $R_0=2$.

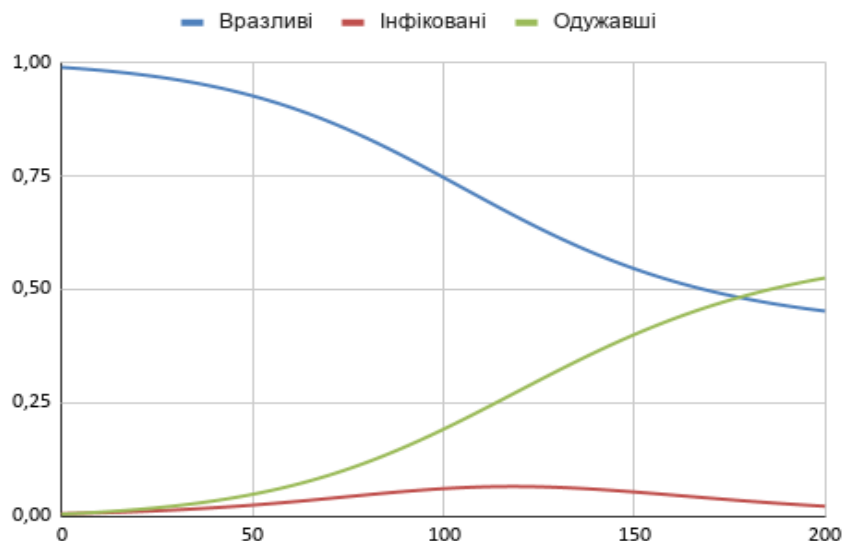


Рис. 2.2 - Графік розвитку захворюваності при $\beta=1/10$, $R_0=1.5$.

Аналіз результатів показує, що отримані дані дуже відрізняються від статистичних в 1,5 раза та більше у більшу сторону. Для $R_0=1.5$, різниця буде ще більше. Виключення складає тільки розрахунок, де репродуктивне число взято мінімальне (вважається, що один хворий за час 15 днів спроможний заразити 1.5 людини).

Застосування SIR моделі призводить до графіків, що плавно змінюються, а реальні дані змінюються циклічно та хвилеподібно.

Висновки до розділу 2

Згідно з отриманими даними був зроблений наступний висновок:

Базова SIR модель у прямому вигляді не може дати бажаних результатів. Вона може використовуватись для довгострокового прогнозування та аналізу загального тренду розвитку хвороби, проте, без використання додаткових пристосувань не видаватиме достатньо точний та достовірний прогноз на короткий проміжок часу — на декілька днів або тижнів.

Можливо, ця модель дає прогноз наближений до реального, але оскільки невідома реальна кількість тих, що захворіли, з урахуванням безсимптомних, незареєстрованих та інших випадків, ми не маємо можливості верифікувати отримані результати.

РОЗДІЛ 3

ПРОГНОЗУВАННЯ З ВИКОРИСТАННЯМ МЕТОДУ ХОЛЬТА–ВІНТЕРСА

3.1. Екстраполяція та експоненційне згладжування

Екстраполяція — це метод наукового дослідження, який заснований на поширенні минулих і справжніх тенденцій, закономірностей, зв'язків на майбутній розвиток об'єкта прогнозування. До методів екстраполяції відносяться метод ковзної середньої, метод експоненційного згладжування, метод найменших квадратів.

Експоненційне згладжування — метод математичних перетворень при прогнозуванні часових рядів. Його сутність полягає у тому, що за кожного повторного застосування математичної операції з оновленими даними, з метою наближення до найбільш точного/правильного результату, враховуються й усі попередні значення даного ряду, проте ступінь врахування зменшується за експоненційним законом.

Вважається, що експоненційну функцію у 17 столітті запропонував використовувати Пуассон як продовження методу чисельного аналізу, а згодом її почали застосовувати для обробки сигналів у 1940-х. Експоненційне згладжування вперше було запропоновано в статистичній літературі Робертом Брауном в 1956 р. [1], який застосував його для прогнозування кількості запасів, які треба зберігати на складах.

Метод простого експоненційного згладжування є найефективнішим для розробки короткострокових прогнозів. Він може бути застосований для прогнозування лише на один період. Його основні переваги: простота обчислень і можливість виявлення помилок та відхилень у статистичних даних, які використовуються у прогнозуванні.

Формула методу експоненційного згладжування:

$$U_{t+1} = \alpha \times y_t + (1 - \alpha) \times U_t,$$

де t — період, що передує прогнозованому;

$t + 1$ — прогнозний період;

U_{t+1} — прогнозований показник;

α — параметр згладжування $\alpha \in (0, 1)$;

Y_t — фактичне значення досліджуваного показника за період, що передує прогнозованому;

U_t — експоненційно зважена середня для періоду, що передує прогнозованому.

Під час прогнозування за допомогою даного метода виникає дві проблеми:

- вибір значення параметра згладжування α ;
- визначення початкового значення U_0 .

Від величини α залежить, як швидко знижується вага впливу попередніх спостережень. Що більше α , то менше позначається вплив попередніх часів. Якщо значення α близьке до одиниці, це призводить до врахування під час прогнозування в основному впливу останніх спостережень. Якщо значення α близьке до нуля, під час прогнозування враховуються всі (або майже всі) минулі спостереження.

Таким чином, якщо є впевненість, що початкові умови, на підставі яких розробляється прогноз, достовірні, слід використовувати невелику величину параметра згладжування ($\alpha \rightarrow 0$). Коли параметр згладжування малий, то досліджувана функція поводить як середня з великого числа минулих рівнів. Якщо немає достатньої впевненості в початкових умовах прогнозування, то слід використовувати велику величину α , що призведе до обліку при прогнозі в основному впливу останніх спостережень.

Точного методу для вибору оптимальної величини параметра згладжування α немає. В окремих випадках Браун пропонував визначати величину α , виходячи з довжини інтервалу згладжування. При цьому α обчислюється за формулою:

$$\alpha = \frac{2}{n+1},$$

де n — число спостережень, що входять в інтервал згладжування.

Завдання вибору U_0 (експоненційно зваженого середнього початкового) розв'язується наступними способами:

- якщо є дані про розвиток явища в минулому, то можна скористатися середньої арифметичної й прирівняти до неї U_0 ;
- якщо таких відомостей немає, то в якості U_0 використовують вихідне перше значення бази прогнозу. Також можна скористатися експертними оцінками.

3.2. Метод аналізу часових рядів Хольта – Вінтерса

Удосконалення методу Брауна було запропоновано американським вченим Хольтом (Holt, 1957) [3] та Вінтерсом (Winters, 1960) [10]. Хольт запропонував модернізовану подвійну модель експоненційного згладжування, яка враховувала тренд у тимчасовому ряду. Вінтерс запропонував ще більш вдосконалену модель потрійного експоненційного згладжування, яка крім тренду враховує ще й сезонність. Вона складається з наступних частин:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1},$$

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L},$$

де Y_t — значення часового ряду в періоді t ,

S_t — згладжені значення в період t , компонент рівня,

β — фактор/компонент тренду,

I — сезонний індекс/компонент сезонності,

F — прогноз на m періодів вперед,

t — індекс, що позначає період розрахунку,

L — довжина сезону в періодах,

m — бажаний проміжок прогнозу,

α, β і $\gamma \in (0, 1)$ — коефіцієнти згладжування, константи, які необхідно визначити.

Сезонність визначається як тенденція даних часового ряду проявляти поведінку, яка повторюється кожні періоди. Термін «сезон» використовується для позначення проміжку часу до того, як поведінка починає повторюватися.

3.3. Використання метода Хольта-Вінтерса для прогнозування кількості тих, що захворіли в Україні та аналіз отриманих результатів

Для прогнозування кількості хворих використовуються офіційні дані про осіб, що захворіли щоденно, які надає Міністерства охорони здоров'я України та Ради національної безпеки та оборони.

Проаналізувавши дані, було зроблено висновок про наявність в них семиденного періоду коливань, пов'язаного з робочими та вихідними днями тижня, що відповідає налаштуванням довжини сезону у 7 днів.

Для прогнозування враховуються дані максимум за останні 6 тижнів, що відповідає довжині інкубаційного періоду (до 2-х тижнів), та середньому строку хвороби (до 4-х тижнів).

Для розрахунків обирались значення коефіцієнтів $\alpha=0,67$, $\beta=0,3$ і $\gamma=0,3$.

Оскільки на час написання роботи є у наявності достатня кількість даних щодо захворюваності, які накопичились з початку епідемії у лютому 2020 року, то

можна створити прогноз на будь-який проміжок часу та перевірити дієвість методу (див. розділ 4). Для ілюстрації, нижче, на рис. 3.1, 3.2, 3.3, 3.4 наведено деякі результати прогнозування на різні дати 2020 та 2021 років.

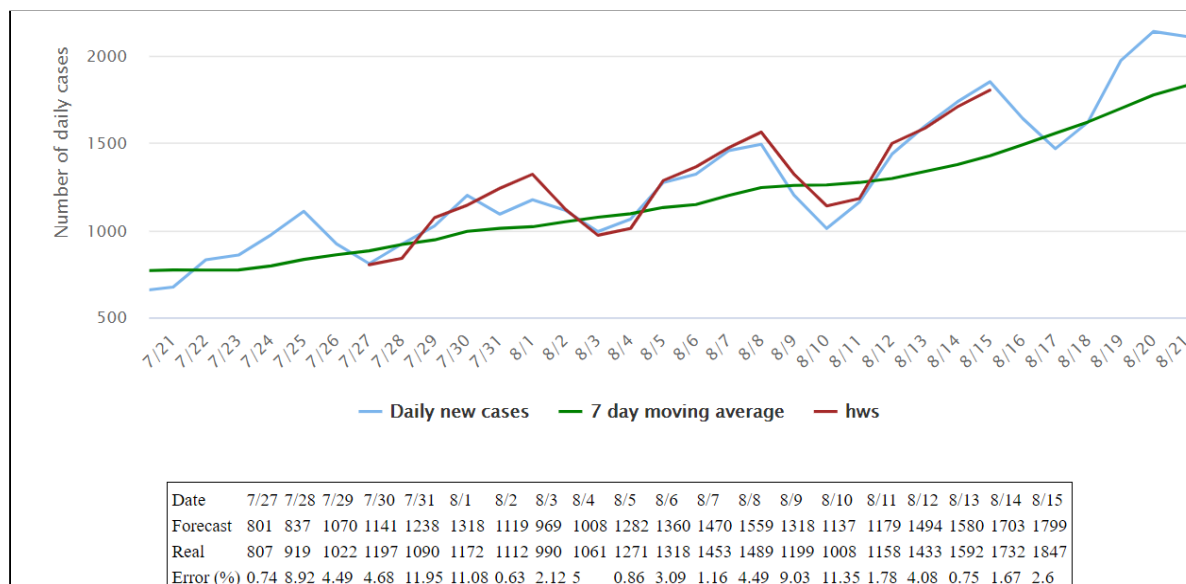


Рис. 3.1 - Прогноз на 27 липня 2020 р. методом Хольта–Вінтерса.

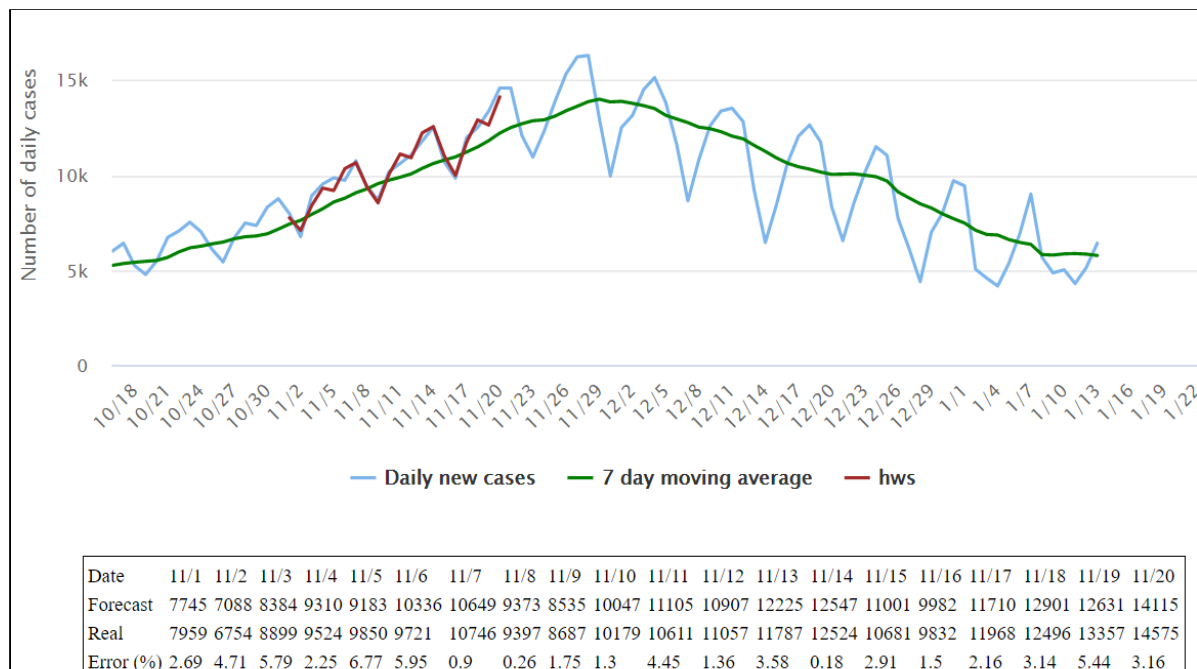


Рис.3.2 - Прогноз на 1 листопада 2020 р. методом Хольта–Вінтерса.

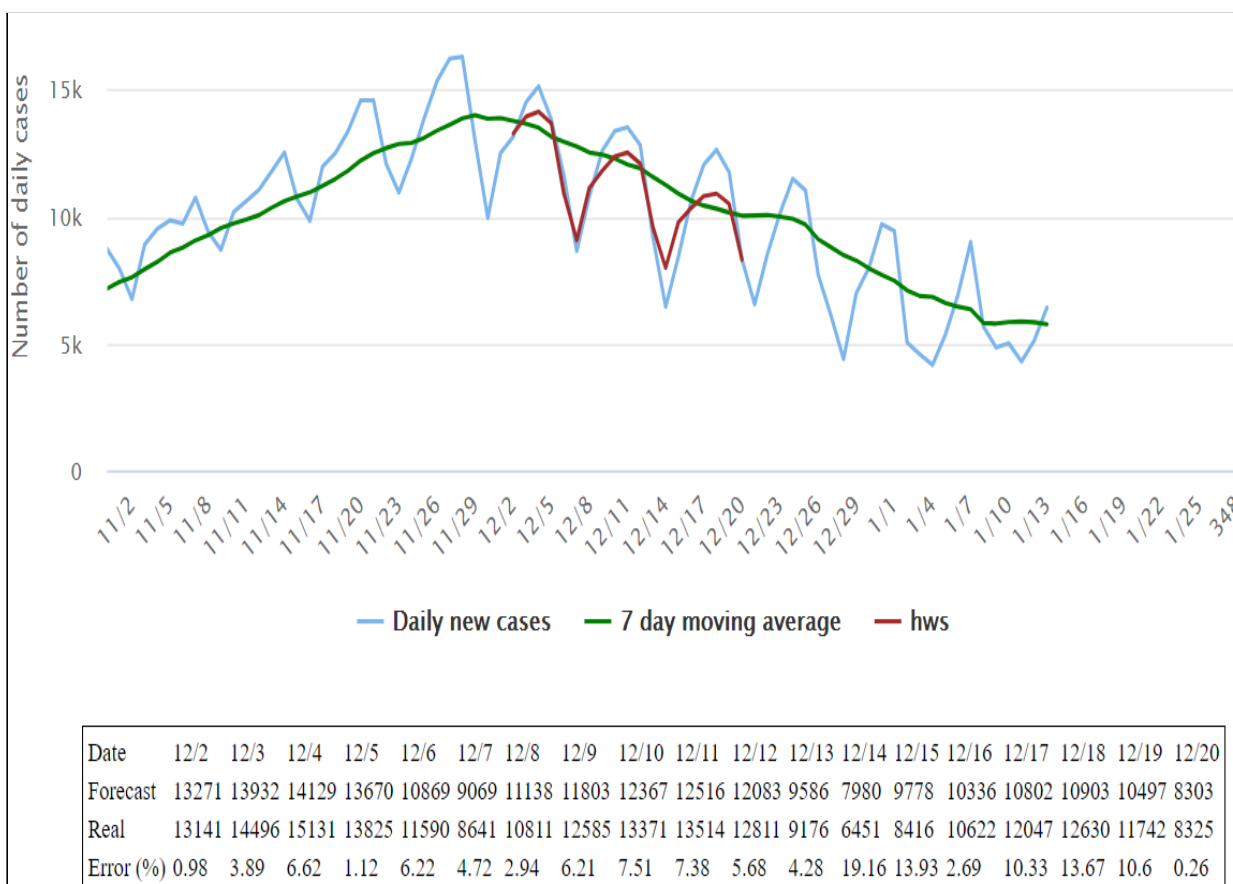


Рис.3.3 - Прогноз на 02 грудня 2020 методом Хольта-Вінтерса.

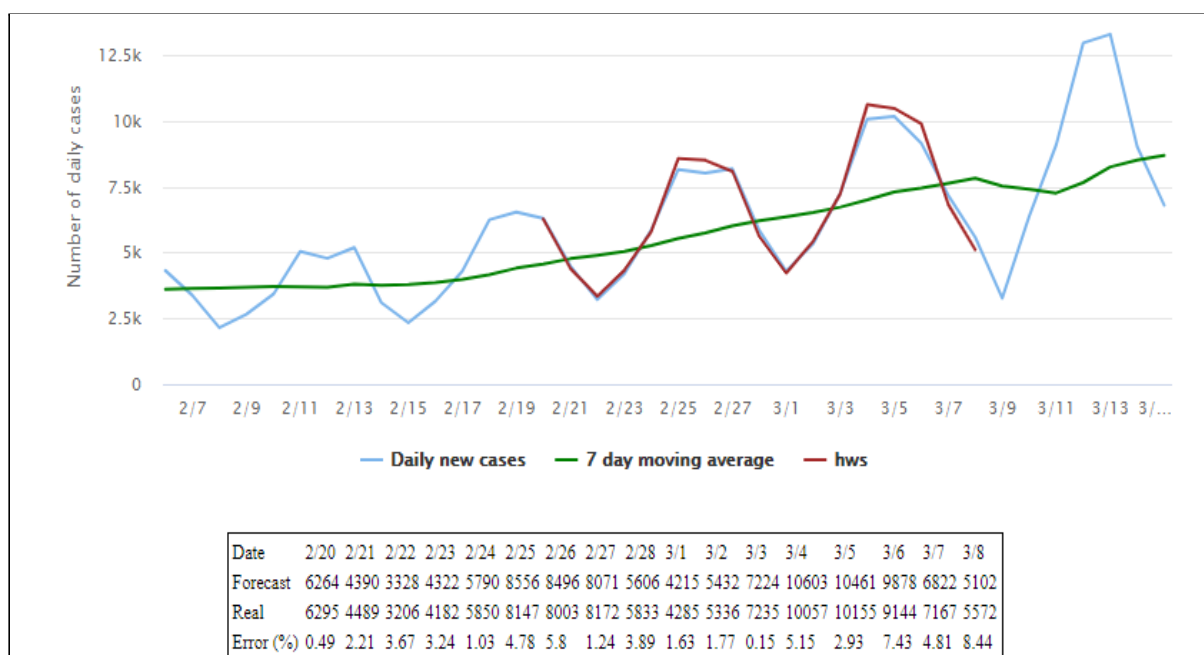


Рис.3.4 - Прогноз на 20 лютого 2021 методом Хольта-Вінтерса.

Висновки до розділу 3

З аналізу отриманих результатів бачимо, що помилка прогнозу за період (нормалізована середньоквадратична помилка/відхилення), в деяких випадках, не перевищує 0.15, а помилка на конкретний день, може не перевищувати 1%.

Метод потрійного експоненційного згладжування за Хольтом-Вінтером з урахуванням сезонної компоненти показує досить точні результати прогнозу статистичних даних за умови короткострокового прогнозу та сталого тренду. Отже, цей метод прогнозування можемо назвати дієвим. На період зміни напрямку тренду можуть знадобитися додаткові налаштування.

РОЗДІЛ 4

ВЕБПОРТАЛ ДЛЯ ДИНАМІЧНОГО ПРОГНОЗУВАННЯ

4.1. Призначення вебпорталу

У роботі створено вебпортал, який слугує для:

- наочної демонстрації можливостей прогнозування кількості інфікованих осіб за певний проміжок часу широкому колу користувачів мережі Інтернет;
- аналізу якості прогнозування, спираючись на попередні дані за будь-який час.

Вебпортал реалізовано на платформі GitHub. та знаходиться за адресою: <https://covid19-info.github.io/covid/>

4.2. Можливості вебпорталу

На порталі знаходиться інформація щодо кількості осіб, що захворіли на коронавірусну хворобу в Україні та ще у кількох країнах світу з початку епідемії. Обираючи країни, можливо зробити:

- короткостроковий прогноз кількості осіб, що захворіли на майбутнє, починаючи з поточної дати;
- прогноз з будь-якої попередньої дати, для порівняння точності прогнозу з реальними даними. Похибка у відсотках відображається для кожної дати прогнозу. Відображається також нормалізована середньоквадратична помилка/відхилення (англ. root-mean-squareerror /deviation) для обраного періоду аналізу даних;
- налаштування моделі для прогнозу, які дозволяють самостійно обрати часовий проміжок, на основі даних якого створюватиметься прогноз, також можна обрати кількість днів, на які варто зробити прогноз;
- увімкнути режим автоматичного налаштування параметрів, що відобразить найточніший прогноз серед тих, що будуть створені методом Хольта-Вінтерса,

спираючись на різні проміжки минулих даних. Виходячи з аналізу відхилення від реальних даних у минулому, буде обрано прогноз з найменшими відхиленнями, що забезпечує точність.

Висновки до розділу 4

З створенням вебпорталу для демонстрації результатів дослідження з'явилася можливість ефективно і швидко дізнатися прогноз поширення хвороби. Портал працює за посиланням <https://covid19-info.github.io/covid/> у режимі реального часу з актуальними статистичними даними МОЗ та РНБО України.

ВИСНОВКИ

Автором **самостійно**:

1) проведено чисельний експеримент для різних співвідношень параметрів базового репродуктивного числа (R_0) у класичній SIR моделі із застосуванням статистичних даних МОЗ;

2) підібрано коефіцієнти для методу Хольта – Вінтерса;

3) проведено чисельний експеримент та визначено величину похибки для прогнозування за допомогою методу Хольта – Вінтерса із застосуванням статистичних даних МОЗ;

4) досліджено ефективність використання для короткострокового прогнозування класичної SIR моделі для епідемій та методу статистичного аналізу Хольта-Вінтерса;

Вперше було створено вебпортал з наочним відображенням прогнозу кількості осіб, що захворіли та величиною похибки. Портал працює за посиланням <https://covid19-info.github.io/covid/> у режимі реального часу з актуальними статистичними даними МОЗ та РНБО України.

Достовірність отриманих результатів забезпечується порівнянням результатів прогнозування із наявними статистичними даними МОЗ.

Теоретичне значення отриманих результатів полягає у дослідженні методів прогнозування епідемій та обрання оптимального методу.

Практичне значення отриманих результатів полягає в тому, що отриманий прогноз дозволяє прогнозувати результати обмежувальних заходів як локдаун, планувати майбутні заходи в умовах карантину, обираючи оптимальні дати тощо.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Brown, Robert G. (1956). Exponential Smoothing for Predicting Demand. Cambridge, Massachusetts: Arthur D. LittleInc. 15 с.
2. Compartmental Models In Epidemiology.
URL:https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology
(дата звернення: 10.11.2020)
3. Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). CarnegieInstituteofTechnology, Pittsburgh USA.
4. Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia.
5. Kermack, W. O. and McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics // Proc. Roy. Soc. Lond. — 1927.
6. Macdonald, G. (September 1952). "The analysis of equilibrium in malaria". TropicalDiseasesBulletin.
7. NIST/SEMATECH e-Handbook of Statistical Methods, April, 2012 (DOI) URL:<https://doi.org/10.18434/M32189> (дата звернення: 20.12.2020).
8. Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE (April 5, 2012). "Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens". PLOS Pathogens. 8 (4): e1002588.
9. The SIR Model for Spread of Disease. David Smith and Lang Moore, "The SIR Model for Spread of Disease," Convergence (December 2004)
10. Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, C. 324–342.
11. Прогнозирование на основе метода экспоненциального сглаживания.
URL:<http://www.ekonomika-st.ru/drugie/metodi/metodi-prognoz-1-4.html> (дата звернення: 20.12.2020).
12. Степаненко Ю.В. Прогнозування випадків захворюваності на COVID-19 в

Україні // Матеріали ХІХ Міжнародної науково-практичної конференції
«Шевченківська весна – 2021» квітень 2021 р., м. Київ, Україна. – С.28-29
(англ).