

В. Б. Мокін<sup>1</sup>  
В. В. Родінкова<sup>2</sup>  
М. В. Драгований<sup>1</sup>  
Т. С. Вуж<sup>2</sup>  
М. О. Бортник<sup>2</sup>

## СИСТЕМНИЙ АНАЛІЗ ТА ПРОГНОЗУВАННЯ КОНЦЕНТРАЦІЙ СПОР МІКРОМІЦЕТІВ *ALTERNARIA* ЗА ДАНИМИ ПОСТА ЄВРОПЕЙСЬКОЇ АЕРОБІОЛОГІЧНОЇ МЕРЕЖІ НА ОСНОВІ PROPHET

<sup>1</sup>Вінницький національний технічний університет

<sup>2</sup>Вінницький національний медичний університет ім. М. І. Пирогова

### Анотація

Розглянуто питання прогнозування кількості спор мікроміцетів *Alternaria* за допомогою системного аналізу та моделі Prophet на основі даних поста ЄАМ. Проведено аналіз даних, необхідних для даного прогнозування, та здійснено моделювання за допомогою програми на Python.

**Ключові слова:** прогнозування, системний аналіз, моделювання, адекватність моделі, алергія, спори грибів.

### Abstract

The issue of forecasting the concentrations of fungal spores of *Alternaria* using system analysis and the Prophet model based on data from EAN posts is considered. The analysis of the data necessary for the given simulation is carried out and modeling by means of programming language Python is performed.

**Keywords:** forecasting, system analysis, modeling, model adequacy, allergy, fungal spores.

### Вступ

В сучасному світі все більш шкідливий вплив на людський організм спричиняють алергени довкілля. Одним із лідерів серед важливих чинників сезонних алергічних захворювань є спори грибів. А одним із найвідоміших представників фунгальних алергенів є спори грибів *Alternaria*. Такі гриби – одні з найбільш поширених у природі. За даними молекулярної діагностики алергії, до спор альтернативі чутливі близько 23% хворих на сезонну алергію українців. Ця хвороба виникає у будь-якому віці. А причина алергічної реакції може бути як генетично обумовленою, так і зумовленою підвищеною індивідуальною чутливістю до алергенів гриба.

Тому контроль наявності та спостереження за змінами концентрації спор *Alternaria* у повітрі є важливим для прогнозування можливих загострень сезонних алергічних захворювань у населення. Таким спостереженням займаються пости моніторингу Європейської аеробіологічної мережі (ЄАМ). Один з таких постів знаходиться у Вінниці на території Вінницького національного медичного університету ім. М. І. Пирогова.

Важливою задачею сучасної науки є прогнозування поширення спор *Alternaria* за допомогою засобів системного аналізу [1-3]. Вирішення цієї задачі полегшить прогноз захворюваності населення на алергію, спричинену даним алергеном, та надасть змогу такого прогнозу навіть за недостатньої кількості або відсутності даних спостережень за певний період. Адже, у зв'язку із зміною клімату, концентрації спор грибів можуть залишатись високими у повітрі і в листопаді-лютому, коли аеробіологічний моніторинг, зазвичай, не ведеться, внаслідок припущення про відсутність в атмосфері високих концентрацій пилоквих алергенів.

До відомих засобів системного аналізу належить моделювання часових рядів за допомогою моделей бібліотеки Prophet компанії Facebook [4-6], яка є вільною для використання у програмах на R і Python.

Метою даного дослідження є визначення оптимальної структури моделі Prophet для передбачення концентрацій спор гриба *Alternaria* у повітрі за даними ЄАМ.

## Результати дослідження

Моделі Prophet, як відомо, дозволяють змінювати такі параметри:

- сезонність (річна, квартальна (чи пори року), місячна, тижнева тощо);
- адитивність чи мультиплікативність ряду;
- лінійна чи логістична регресія;
- порядок рядів Фур'є, якими моделюється кожна сезонність.

Для дослідження було використано дані ЄАМ за 2019 р. (рис. 1).

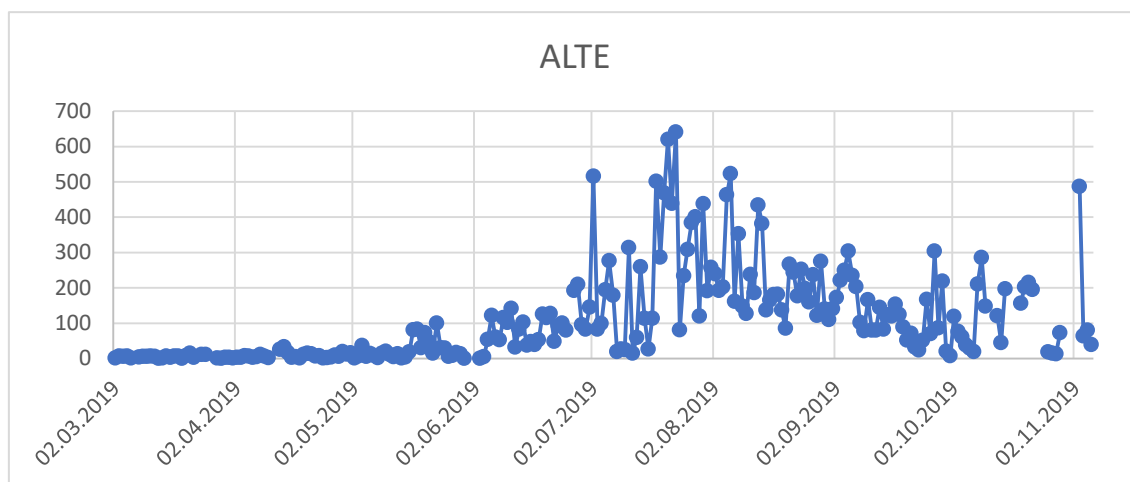


Рис. 1. Дані ЄАМ по спорам *Alternaria* за 2019 рік

Було проведено розвідувальний аналіз даних, який показав, що:

1) доцільно враховувати місячну, квартальну, декадну сезонності, але немає сенсу враховувати річну (дані є тільки за 1 рік) та денну сезонність (дані – сумарні за день);

2) є сенс застосовувати тільки мультиплікативну сезонність, а порядок рядів Фур'є варто варіювати в діапазоні від 2 до 7;

3) варто дослідити і лінійну (найбільшу підходить до моделювання рядів, на які впливає велика кількість досить мінливих факторів), і логістичну (часто застосовується для моделювання саме біологічних процесів) регресії;

4) в результаті прогнозування регресійними моделями часто можуть генеруватись від'ємні значення, тому варто слідкувати, щоб від'ємні значення були замінені нулями.

Особливо перспективним є застосування логістичної регресії з багатьма різними піками. Наявних значень концентрацій замало для того, щоб надійно спрогнозувати усі коливання – для цього потрібні дані за багато років. Тому більш адекватною буде модель, якщо задати піки на певних інтервалах за наявними даними і вимагати, щоб результати моделювання не перевищували їх пікових значень.

Було проведено моделювання на Python. Використано повний перебір варіантів порядку рядів Фур'є в діапазоні від 2 до 7 для усіх трьох видів сезонності і для двох варіантів моделей (лінійної та логістичної). Модель будувалась за даними за березень-вересень. Початок другого піку був встановлений на 24 вересня 2019 року. Для перевірки точності були використані наявні дані за жовтень-листопад. Точність визначалась за метриками:

- RMSE («Root Mean Square Error» або корінь із середньоквадратичної похибки);
- MAE («Mean Absolute Error» або середня абсолютна помилка).

Обидві ці метрики мають ту саму розмірність, що й вхідні дані. Але метрика MAE враховує усі відхилення прогнозованих значень від даних моніторингу незалежно від їх значень, а метрика RMSE враховує квадрати цих відхилень, тобто «карає» за помилки, передусім на найбільших таких відхиленнях. Перед визначенням метрики усі від'ємні значення замінялись нулями. Цікаво, що визначені параметри і структура виявились оптимальними (мінімальними), незалежно від метрики.

Оптимальними параметрами моделей (числа зеленим кольором означають період у днях і порядок ряду Фур'є, відповідно) виявились такі, як показано на рис. 2, 3.

Отже, оптимальною і за структурою, і за параметрами моделлю є логістична модель з параметрами як на рис. 2.

```
seasonality_opt = [['weekly',7,7,'multiplicative'],  
                  ['monthly',365.25/12,2,'multiplicative'],  
                  ['quarterly',365.25/4,3,'multiplicative']]
```

Рис.2 Параметри логістичної моделі

Аналіз показав, що модель є недостатньо адекватною. Крім того, пошук піків для логістичної кривої носить емпіричний характер (можливо, варто було брати не 2 піки, а більше). Для поліпшення точності прогнозування варто використовувати більше даних (хоча б за 4-6 років) та більш адаптивний вибір параметрів цих піків (рис. 3, 4).

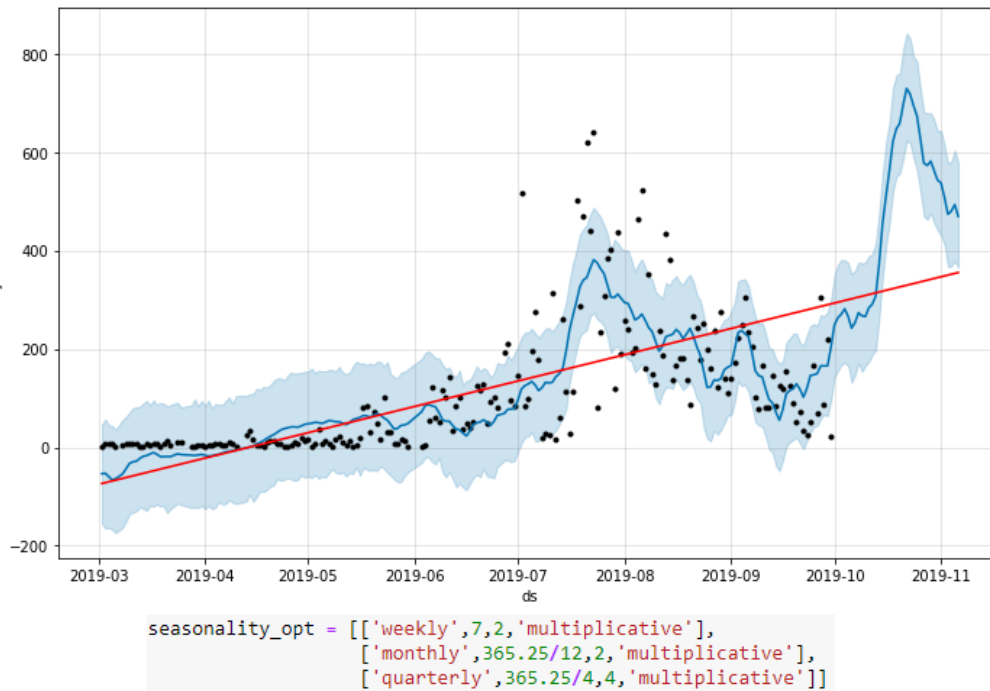


Рис. 3. Лінійна модель (RMSE : оптимальне значення = 72,61, MAE : оптимальне значення = 316,59)

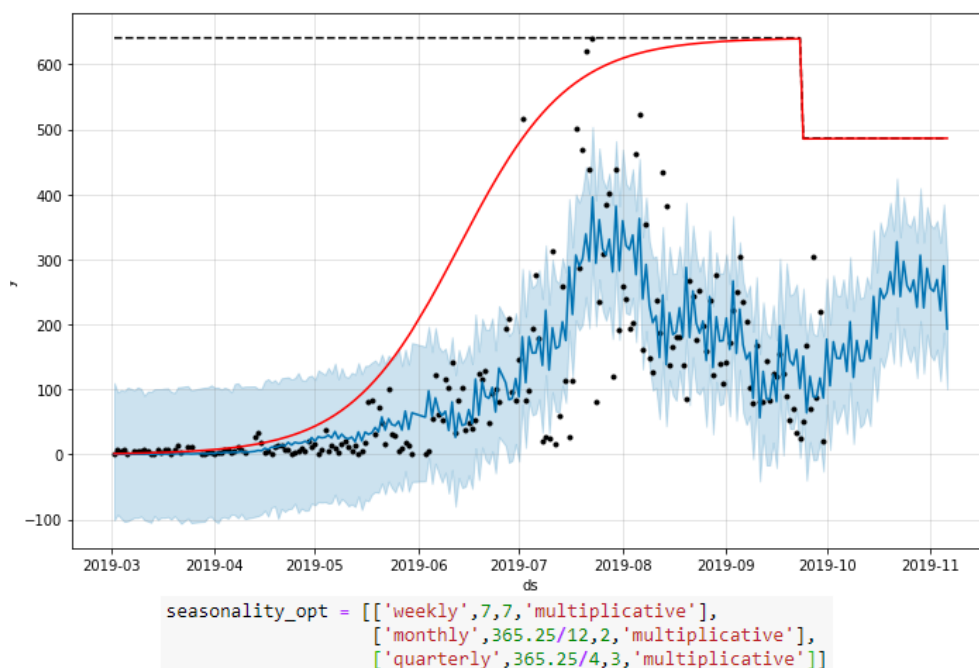


Рис. 4. Логістична модель з двома піками (RMSE : оптимальне значення = 28,95, MAE : оптимальне значення = 125,22)

Проведено аналіз даних поста САМ, які необхідно використовувати для прогнозування концентрацій спор мікроміцетів, зокрема *Alternaria*, у повітрі. Встановлено, що доцільно опрацьовувати місячну, квартальну, декадну сезонності та є сенс застосовування тільки мультиплікативної сезонності. Варто дослідити лінійну і логістичну лінійність. Найбільш перспективним є застосування логістичної регресії з багатьма різними піками. А для підвищення точності прогнозування необхідно використовувати дані за більш тривалий період (не менше 4 років) та більш адаптивний вибір параметрів піків.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. В. Мокін, В. Родінкова, і М. Дратований, «Аналіз експериментальних даних, необхідних для синтезу математичної моделі прогнозування розповсюдження алергенних спор грибів *Alternaria*», *Вісник Вінницького політехнічного інституту*, № 3, с. 50-58, Чер 2019.
2. Viktoriya Rodinkova. Environmental Factors Which Increase *Alternaria* spores in Central Ukraine / Viktoriya Rodinkova, V. Mokin, O. Bilous, Lawrence Dubuske, M. Dratovanyj // *Journal of Allergy and Clinical Immunology* (ISSN 0091-6749, Scopus CiteScore of journal: 6.87). – Elsevier Science Ltd., 2018. – V. 141. – # 2. – p. AB30.
3. Viktoriya Rodinkova. Mapping of Ambrosia Pollen Emission Sources Using Spline Interpolation in Ukraine / V. Rodinkova, V. Mokin, T. Vuzh, M. Dratovanyj // *Aerobiologia* (ISSN 0393-5965, Scopus CiteScore of journal: 2.17), 2019.
4. Документація бібліотеки Prophet / Facebook [Електронний ресурс] – Режим доступу до ресурсу: [https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html).
5. Vitalii Mokin. Data Science for tabular data: Advanced Techniques [Електронний ресурс] / Vitalii Mokin. – 2020. – Режим доступу до ресурсу: <https://www.kaggle.com/vbmokin/data-science-for-tabular-data-advanced-techniques>.
6. Time Series based Air Pollution Forecasting using SARIMA and Prophet Model / K. Krishna Rani Samal, Korra Sathya Babu, Santosh Kumar Das, Abhirup Acharaya. // *ITCC 2019: Proceedings of the 2019 International Conference on Information Technology and Computer Communications*. – 2019. – С. 80–85. URL: <https://dl.acm.org/doi/abs/10.1145/3355402.3355417>.

**Мокін Віталій Борисович** — д-р техн. наук, професор, завідувач кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки, Вінницький національний технічний університет, Вінниця, e-mail: vbmokin@gmail.com.

**Родінкова Вікторія Валеріївна** — д-р біол. наук, професор кафедри фармації, Вінницький національний медичний університет ім. М. І. Пирогова, Вінниця, e-mail: vikarodi@gmail.com

**Дратований Михайло Володимирович** — асистент та аспірант кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки, Вінницький національний технічний університет, Вінниця, mishadratovany@gmail.com.

**Вуж Тетяна Євгенівна** — к.т.н., доцент кафедри біологічної фізики, медичної апаратури та інформатики, Вінницький національний медичний університет ім. М. І. Пирогова, Вінниця, e-mail: tatiana.vuzh@gmail.com

**Бортник Микита Олегович** — аспірант, Вінницький національний медичний університет ім. М. І. Пирогова, Вінниця, e-mail: nikita.bortnik@gmail.com

**Mokin Vitalii B.** — Dr. Sc. (Eng.), Professor, Head of the Department of System Analysis, Computer Monitoring and Engineering Graphics, Vinnytsia National Technical University, Vinnytsia, e-mail: vbmokin@gmail.com.

**Rodinkova Victoria V.** — Dr. Biol. Sciences, Professor of the Department of Pharmacy, National Pirogov Memorial Medical University, Vinnytsia, e-mail: vikarodi@gmail.com

**Dratovany Mikhail V.** – Assistant and postgraduate student of the Department of System Analysis, Computer Monitoring and Engineering Graphics, Vinnytsia National Technical University, Vinnytsia, mishadratovany@gmail.com.

**Vuzh Tetiana Y.** — Ph.D., Associate Professor of the Department of Biological Physics, Medical Equipment and Informatics, National Pirogov Memorial Medical University, Vinnytsia, e-mail: tatiana.vuzh@gmail.com

**Bortnyk Mykyta O.** — graduate student, National Pirogov Memorial Medical University, Vinnytsia, e-mail: nikita.bortnik@gmail.com