

# НОВИЙ ПІДХІД ДО АВТОМАТИЗАЦІЇ СТАТИСТИЧНОГО АНАЛІЗУ ДИНАМІКИ ЗАБРУДНЕННЯ ПОВІТРЯ СПОРАМИ ГРИБІВ ЗА ДАНИМИ АЕРОБІОЛОГІЧНОГО МОНІТОРИНГУ З ВИКОРИСТАННЯМ ЗАСОБІВ МОВИ R

Вінницький національний технічний університет

## Анотація

Запропоновано новий підхід до автоматизації статистичного аналізу динаміки забруднення повітря алергенними спорами грибів за даними Європейської системи аеробіологічного моніторингу з використанням засобів мови R.

**Ключові слова:** статистичний аналіз, моделювання алергенних спор грибів у повітрі, часові ряди, мова R, аеробіологічний моніторинг.

## Abstract

A new approach to the automation of statistical analysis of the air pollution dynamics by allergenic spores of fungi is proposed based on the data of the European aerobiological monitoring system with use of means of language R.

**Keywords:** statistical analysis, modeling of allergenic fungal spores in the air, time series, R, aerobiological monitoring.

Алергія є однією з основних причин розвитку хронічних захворювань у розвинених країнах. Найбільший вплив на розвиток алергічних захворювань мають повітряні алергени (пилки рослин, епідерміс тварин, частки рослин та комах, а також спори грибів) [1-10]. З перерахованих повітряних алергенів найменш дослідженими є спори грибів [1, 9, 10].

Використовуючи методи системного аналізу є можливість дослідити маршрути переміщення спор грибів під час їх активного розповсюдження по місцевості. Також методами системного аналізу є можливість зробити прогноз вмісту спор в атмосфері. Як правило, в Європі для цього використовують дані Європейської системи аеробіологічного моніторингу.

Є величезна кількість досліджень та виявлених закономірностей такого роду з використанням методів статистичного аналізу. Але не існує єдиного підходу для визначення цих залежностей для різних видів спор, часу та місцевості.

Дослідники використовували різні пакети та мови програмування для обробки цих даних. Таким чином, новий підхід до автоматизації статистичного аналізу динаміки забруднення повітря алергенними спорами грибів за даними Європейської системи аеробіологічного моніторингу є актуальною областю дослідження.

Одним із найбільш потужних апаратів для дослідження та прогнозування забруднення повітря спорами грибів мають пакети (бібліотеки) мови програмування R, але немає єдиних підходів для їх застосування щодо такого прогнозування.

Задачею даного дослідження є уніфікація та пришвидшення налаштування засобів мови R для нового підходу до автоматизації статистичного аналізу динаміки забруднення повітря алергенними спорами грибів за даними Європейської системи аеробіологічного моніторингу.

Як відомо, статистичний аналіз даних здійснюється за таким алгоритмом [11-13]:

1. Первинний статистичний аналіз — формування і подання масивів даних у вигляді варіаційних рядів з наступним обчисленням статистичних характеристик.
2. Одновимірний статистичний аналіз — ідентифікація та відтворення одновимірних розподілів.
3. Перевірка статистичних гіпотез — порівняння двох і більше вибірок на основі процедур перевірки статистичних гіпотез.
4. Багатовимірний статистичний аналіз — пошук і відображення залежностей між даними, що аналізуються, — кореляційний, регресійний та факторний аналіз.
5. Статистичний аналіз випадкових процесів (часових рядів).

У випадку, коли проведений аналіз не підпорядковується нормальному закону розподілу, його потрібно нормалізувати. Якщо нормалізацію закону розподілу не вдалося досягти, в такому випадку потрібно використовувати кореляційний аналіз даних.

Для аналізу даних спор грибів *Cladosporium* використаємо лінійну декомпозицію часового ряду даних. (рис.1).

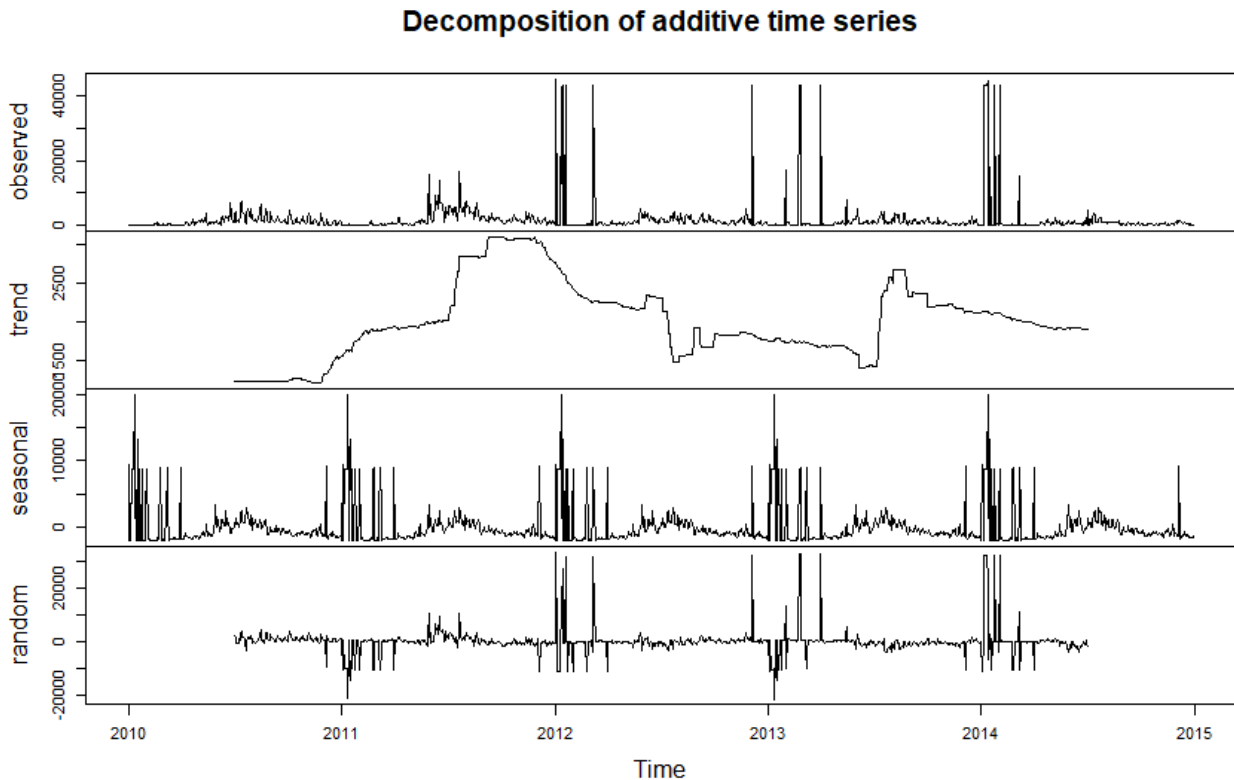


Рис. 1. Результат декомпозиції часового ряду даних за 2010-2014 рр. по спорах *Cladosporium* у повітрі м. Вінниці

З рис. 1 видно, що є чітка періодична складова, є нелінійний тренд і стохастична (random) складова. Але, видно, що стохастична складова (random) охоплює майже усю амплітуду значень ряду. Отже, така декомпозиція не є коректною – виділена сезонність та тренд є малозначущими.

Використовуючи бібліотеки мови R для роботи з часовими рядами, кореляційного та регресійного аналізу можна автоматизувати процес аналізу динаміки спор грибів за даними Європейської системи агробіологічного моніторингу [9, 10, 13].

Висуваємо припущення, що між кількістю спор грибів в атмосфері та метеорологічними факторами є певний зв'язок. Для перевірки припущення використаємо класичний кореляційний аналіз та провести множинний регресійний аналіз та ідентифікацію регресії з використанням операторів `lm` (лінійна регресія) та `glm` (нелінійна регресія) [14, 15].

Засоби мови програмування R дозволяють перебір різних варіантів регресії та підбір найкращих з них для розв'язання заданої задачі. Найбільш складним на даному етапі аналізу забруднення повітря спорами грибів є виділення даних, що вплинули на розповсюдження алергенних спор грибів з урахуванням специфікації та вегетаційних особливостей грибів у різних місцевостях (один тип грибів розповсюджують спори при високій температурі повітря та високій вологості, інші – такі, як *Cladosporium*, розповсюджують свої спори весь теплий сезон). Часто спори, що спостерігаються на одному пості спостережень, є спорами, принесеними з різних ареалів, а не вибіркою одного процесу – перевірити це можна за критерієм Кохрена [12]. Тому цей процес потребує окремого інтелектуального аналізу даних, що є темою окремої статті.

Даний підхід випробуваний на даних моніторингу деяких спор грибів, таких як *Alternaria* та *Cladosporium* [9, 10].

### Висновки

Отже, була розглянута проблема уніфікації та пришвидшення автоматизації статистичного аналізу динаміки забруднення повітря алергенними спорами грибів за даними Європейської системи аеробіологічного моніторингу. Запропоновано новий підхід до такої автоматизації за рахунок використання пакетів (бібліотек, засобів) мови програмування R, який вже був успішно випробуваний для спор грибів *Alternaria* та *Cladosporium*.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Родінкова В. В., Білоус О. С. Розповсюдження аероалергенних спор у повітрі м. Вінниці // Природничий альманах, 2015. – 13 с.

2. Мокін В. Б. Контроль та мінімізація впливу алергенних рослин на стан захворюваності населення на бронхіальну астму на основі просторово-хронологічної моделі / Т. Є. Вуж, В.Б. Мокін // Зб. тез доп. XII Міжнародної конференції “Контроль і управління в складних системах (КУСС-2014)”, м. Вінниця, 14-16 жовтня 2014 р. – Вінниця : ВНТУ, 2014. – С. 205.
3. Мокін В. Б. Аналіз ризику впливу алергенних рослин на здоров’я дітей чи дорослих у населених пунктах на основі просторово-хронологічної моделі даних / Т. Є. Вуж, В. Б. Мокін // Екологічна безпека та природокористування: Зб. наук. праць. — К., 2014. – Вип. 16. – С. 57-67.
4. Мокін В.Б. Оцінювання впливу просторово-розподілених шкідливих факторів на стан здоров’я населення на основі просторово-хронологічної моделі даних / Т. Є. Вуж, В. Б. Мокін // Оптико-електронні інформаційно-енергетичні технології. – Том 28. – 2014. – № 2. – С. 71-80.
5. Мокін В. Б. Виявлення та аналіз закономірностей щодо можливих причин захворюваності дитячого населення Вінницької області на бронхіальну астму за допомогою інформаційних технологій / В. Б. Мокін, Т. Є. Вуж // Сборник докладов международной научно-технической конференции «Геоинформационные системы, компьютерные технологии эколого-экономического мониторинга, ГИСКТЭЭМ – 2014», 9 – 11 апреля 2014. – Днепропетровск: НГУ, 2014. [Електронне видання на DVD]. – Систем. вимоги: 32 Mb RAM; DVD, Windows 98/2000/NT/XP; MS Word 2000/XP/2003.
6. Цимбалюк В. А. Метод побудови оптимальної просторово-хронологічної інформаційної моделі формалізації впливу алергенних рослин на захворюваність людей / В. А. Цимбалюк, В. Б. Мокін, Т. Є. Вуж // XLIV Регіональна науково-технічна конференція професорсько-викладацького складу, співробітників та студентів університету з участю працівників науково-дослідних організацій та інженерно-технічних працівників підприємств м. Вінниці та області, електронне наукове видання матеріалів конференції, 11—13 березня 2015. – Режим доступу: <http://conf.vntu.edu.ua/allvntu/2015/ineek/txt/tsimbalyuk.pdf>
7. Мокін В. Б. Побудова геоінформаційної моделі природних шкідливих факторів, які впливають на стан здоров’я людей, на основі просторово-хронологічної моделі даних у циліндричній системі координат / В. Б. Мокін, Т. Є. Вуж // XIV Міжнародна науково-практична конференція: Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: 5-9 жовтня 2015 р. – К., 2015. – С. 121-125.
8. Мокін В. Б. Виявлення зон і періодів часу підвищеної небезпеки захворюваності дітей м. Вінниці бронхіальною астмою внаслідок впливу осередків амброзії за певних метеоумов / В. Б. Мокін, Т. Є. Вуж, В. А. Цимбалюк // V-й Всеукраїнський з’їзд екологів з міжнародною участю (Екологія/Ecology-2015), 23-26 вересня, 2015. Збірник наукових праць. – Вінниця: ТОВ «Нілан-ЛТД», 2015. – С. 107.
9. Rodinkova, V., Mokin, V., Bilous, O., Dubuske, L., Dratovanyj, M. Environmental Factors Which Increase *Alternaria* spores in Central Ukraine // Journal of Allergy and Clinical Immunology. – 2018. – 141(2). – AB30.
10. Мокін В. Б. Статистичний аналіз динаміки спор грибів *Alternaria* за даними Європейської системи аеробіологічного моніторингу / В. Б. Мокін, В.В. Родінкова, М.В. Дратованій, О.С. Білоус // Вісник Вінницького політехнічного інституту. — 2017. — № 4 — С. 33-42.
11. Бабак, В. П., Білецький, А. Я., Приставка, О. П., Приставка, П. О. (2001). Статистична обробка даних. – К.: МІВВІЦ, 388.
12. Методологія та організація наукових досліджень : навчальний посібник / Б. І. Мокін, О. Б. Мокін. – 2-е вид., змін. та доп. – Вінниця : ВНТУ, 2015. – 317 с.
13. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. – 2015 [Електронний ресурс]. – Режим доступу: [https://github.com/ranalytics/r-tutorials/tree/master/Edition\\_2015/Book](https://github.com/ranalytics/r-tutorials/tree/master/Edition_2015/Book)
14. Package ‘stats’ [Електронний ресурс]. – Режим доступу: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/1m.html>
15. Package ‘gclus’ [Електронний ресурс]. – Режим доступу: <https://cran.r-project.org/web/packages/gclus/index.html>

**Дратованій Михайло Володимирович** — аспірант кафедри системного аналізу, комп’ютерного моніторингу та інженерної графіки, e-mail: mishadratovany@gmail.com;

**Dratovanyi Mykhailo V.**— Post-Graduate Student of the Chair of System Analysis, Computer Monitoring and Engineering Graphic, Vinnytsia National Technical University, Vinnytsia, e-mail: mishadratovany@gmail.com