

В. Б. Мокін<sup>1</sup>  
В. В. Родінкова<sup>2</sup>  
М. В. Дратованій<sup>1</sup>  
О. С. Білоус<sup>2</sup>

## СТАТИСТИЧНИЙ АНАЛІЗ ДИНАМІКИ СПОР ГРИБІВ *ALTERNARIA* ЗА ДАНИМИ ЄВРОПЕЙСЬКОЇ СИСТЕМИ АЕРОБІОЛОГІЧНОГО МОНІТОРИНГУ

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

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

*Розглянуто актуальну задачу статистичного аналізу динаміки вмісту спор грибів Alternaria за даними Європейської системи аеробіологічного моніторингу та залежності цих даних від метеофакторів. Досліджено типові проблеми цього аналізу на прикладі щодобових даних для м. Вінниця за 2009—2014 роки. Запропоновано алгоритм їх вирішення та проведено апробацію цього алгоритму з використанням інструментарію мови програмування R шляхом побудови лінійної та нелінійної регресій залежності цих даних від трьох метеофакторів. Виявлено нові закономірності та зроблено рекомендації щодо ефективнішого застосування цього алгоритму.*

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

### Вступ та вихідні передумови

Алергічний риніт та інші алергічні захворювання стають дедалі частішими серед населення, особливо міського, обумовлені дією різних, у т. ч. природних факторів. Одним з поширених у містах алергенів є спори грибів *Alternaria* [1—4].

В Європі діє потужна мережа аеробіологічного моніторингу, у межах якого у містах багатьох країн, у т. ч. в Україні, регулярно фіксується вміст спор грибів *Alternaria* у повітрі. Як правило, потім ці дані узагальнюються по годинах і добах. У м. Вінниця вимірювальний пост Європейської аеробіологічної мережі діє на базі Вінницького національного медичного університету імені М. І. Пирогова [5]. Дані на ній по спорах грибів *Alternaria* накопичуються з 17.04.2009 р.

У світі вже давно проводяться дослідження вмісту цих та інших алергенних спор у повітрі з метою виявлення закономірностей щодо динаміки їх концентрацій та можливостей прогнозування, а отже, формування першочергових заходів для боротьби з цими грибами та превентивних заходів для потенційних пацієнтів (зміна щоденного маршруту пересування, зміна місць відпочинку тощо). Однак, на динаміку спор грибів *Alternaria* впливає багато факторів, в першу чергу, метеорологічних, динаміку яких досі не вдається надійно прогнозувати. А отже, досі не існує надійних універсальних методів аналізу статистичних даних щодо концентрації спор у містах та для їх прогнозування з високою точністю.

*Метою дослідження є проведення статистичного аналізу концентрації спор грибів Alternaria у повітрі за даними Європейської системи аеробіологічного моніторингу, ідентифікації її регресійної моделі між цією концентрацією та метеопказниками для заданого посту, що дозволить виявити нові закономірності утворення цих спор та вибирати оптимальні заходи щодо зниження негативного впливу цих спор на населення.*

### Етапи статистичного аналізу даних

Для автоматизації аналізу використаємо мову програмування даних R, яка станом на кінець серпня 2017 містить вже біля 11300 пакетів спеціалізованих підпрограм (R-пакетів), тобто надзвичайно потужний інструментарій для роботи з будь-якими даними.

За даними EAN, вміст спор грибів *Alternaria* у повітрі для міста Вінниця вже визначено за пе-

ріод 2009—2014 років. (рис. 1). За 2015—2017 роки дані ще оброблені не повністю (така обробка полягає у підрахунку спор на відповідних ділянках плівок вимірювального пристрою типу «Буркард» з використанням потужних мікроскопів — процес довготривалий, враховуючи, що щороку визначається 60 видів пилку рослин та 20 видів спор грибів, у т.ч. *Alternaria*, і більшість операцій здійснюється вручну). Як правило, щороку вимірювання проводяться з 1 квітня по 31 жовтня (245 днів), коли ймовірність появи спор у повітрі найбільша.

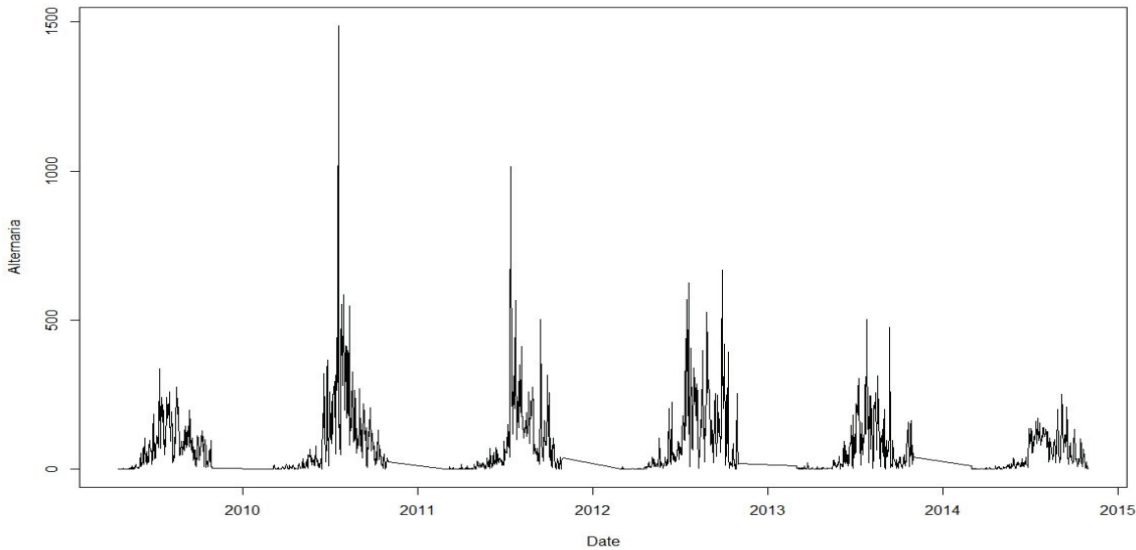


Рис. 1. Дані EAN щодо вмісту спор грибів *Alternaria* у повітрі м. Вінниця за період 2009—2014 рр.

Застосуємо сучасні засоби R та типові підходи до статистичного аналізу даних (див., наприклад, роботу [6]), показаних на рис. 1 (позначимо вміст спор *Alternaria* у повітрі як  $S$ ).

### Етап 1. Тест Кохрена

Для того, щоб з'ясувати чи можна вважати вибірки  $S$  за кожен рік як вибірки з однієї генеральної сукупності і розглядати їх як єдиний часовий ряд, застосуємо тест Кохрена з використанням методології робіт [7, 8]. Аналіз показав, що розрахункове значення показника Кохрена  $K_{Kp} = 0,5132$ . У той час, як табличне значення для  $k = 6$  (вбірок) та  $N - 1 = 244$ , тобто більше 144, для значення довірчої ймовірності 0,95 ( $\alpha = 0,05$ ) дорівнює 0,1667. Отже, ці вибірки порівнювати не можна і не можна розглядати як єдиний часовий ряд — треба шукати інші підходи до аналізу.

Що цікаво, подібна ситуація з чималими відхиленням дисперсії по роках має місце у більшості постів (міст), але у більшості статей, де аналізуються зібрані дані, див. наприклад роботи [2—4], тест Кохрена не проводиться і ці різні вибірки помилково вважаються єдиним часовим рядом. Можливо, саме це і пояснює той факт, що у більшості статей, у т.ч. статтях [2—4], зазначається, що побудовані регресії пояснюють тільки 30,6...41 % даних — ці регресії просто не варто було будувати для таких первинних даних.

### Етап 2. Декомпозиція даних

Суттєві відхилення дисперсії по роках показують, що ряд  $S$  — це не один ряд, а сукупність багатьох, кожен з яких зумовлений різними факторами, які діють, в загальному випадку, з різною періодичністю, динамікою та законами розподілу.

З рис. 1 видно, що часовий ряд  $S$  має сезонну складову і є періодичним, але неоднорідним. Припустимо, що цей ряд є лінійною композицією (сумою) декількох часових рядів — тренду (trend), періодичної (seasonal) та випадкової (random) складової. Здійснимо декомпозицію цього часового ряду засобами R (рис. 2).

З рис. 2 видно, що є чітка періодична складова амплітудою біля 400 (26 %), є нелінійний тренд амплітудою до 100 (7 %) і стохастична (random) складова амплітудою від  $(-400)$  до 1000 (67 %). Звичайно, така декомпозиція не є коректною. На основі неї можна зробити такі висновки:

— цей ряд не є лінійною сумою декількох часових рядів — у ньому присутня явна нелінійність та залежність від багатьох факторів, які динамічно змінюються;

- можна виділити щорічну періодичну складову, яка щороку має місце, її можна інтерполювати та прогнозувати;
- цей часовий ряд має нелінійний тренд, який ще треба досліджувати — наврядчи він буде такий, як показано на рис. 2, оскільки він погано відображає загальну тенденцію зміни значень основного ряду;
- стохастична складова є зовеликою — треба з'ясувати її причини і постаратись зменшити, а вже потім пробувати процедуру лінійної декомпозиції знов.

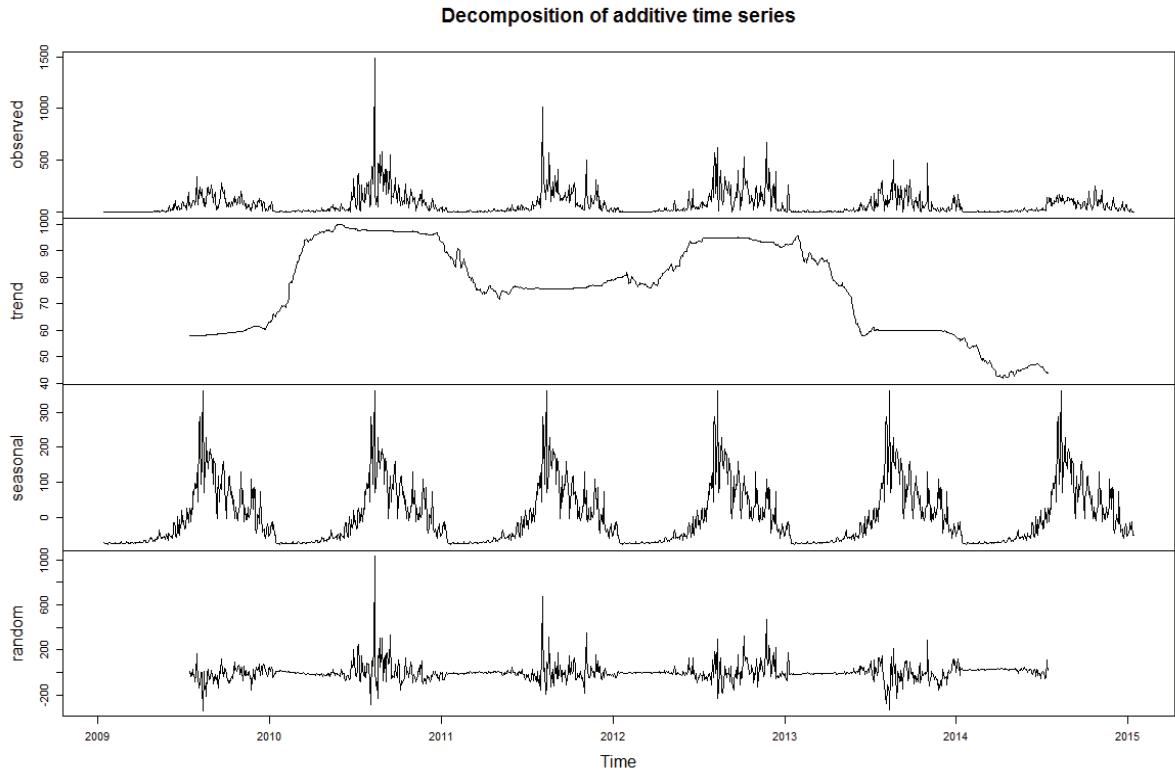


Рис. 2. Результат декомпозиції даних EAN про вміст спор грибів *Alternaria* у повітрі м. Вінниці за період 2009—2014 рр

### Етап 3. Пошук варіантів фільтрації даних

Аналіз ряду  $S$  показує, що він розподілений за пуасонівським або експоненціальним законами розподілу. Складнощі в аналізі вмісту спор *Alternaria* у повітрі зумовлені, серед іншого, великою кількістю малих значень. Водночас, практичне значення мають дні, у які концентрація спор *Alternaria* перевищує 100 спор/м<sup>3</sup> повітря, адже, саме така концентрація вважається граничною, тобто, здатною провокувати симптоми у пацієнтів. Здійснимо фільтр ряду  $S$  і утворимо з нього ряд  $S_{100}$ , значення якого дорівнюють або є більшими за 100 (рис. 3).

З рис. 3 видно, що ряд  $S_{100}$  зберіг свою періодичність, став дещо більш згладженим, але не набагато.

Аналогічно, як і на етапі 2, застосуємо до нового ряду  $S_{100}$  тест Кохрена. Аналіз показав, що розрахункове значення показника Кохрена  $K_{Kp} = 0,4833$ , тобто дисперсія стала менше відрізнятись, але не набагато менше. Кількість ступенів вільності визначається кількістю даних у кожен рік. У ряді  $S_{100}$  ця кількість коливається від 42 до 79, тобто вони усі потрапляють в інтервал (36, 144], а для цього інтервалу табличне значення для  $k = 6$  (вибірок), у разі довірчої ймовірності 0,95 ( $\alpha = 0,05$ ) дорівнює 0,2119. Отже, ці вибірки знову ж таки порівнювати не можна, а  $S_{100}$  не можна розглядати як єдиний часовий ряд — треба шукати інші підходи до аналізу.

Як один із варіантів вирішення цієї задачі — це пошукати таку границю (не 100 одиниць), яка дасть такі вибірки, для яких тест по Кохрену буде позитивним, але ця задача може і не мати розв'язку, тому варто шукати інші підходи.

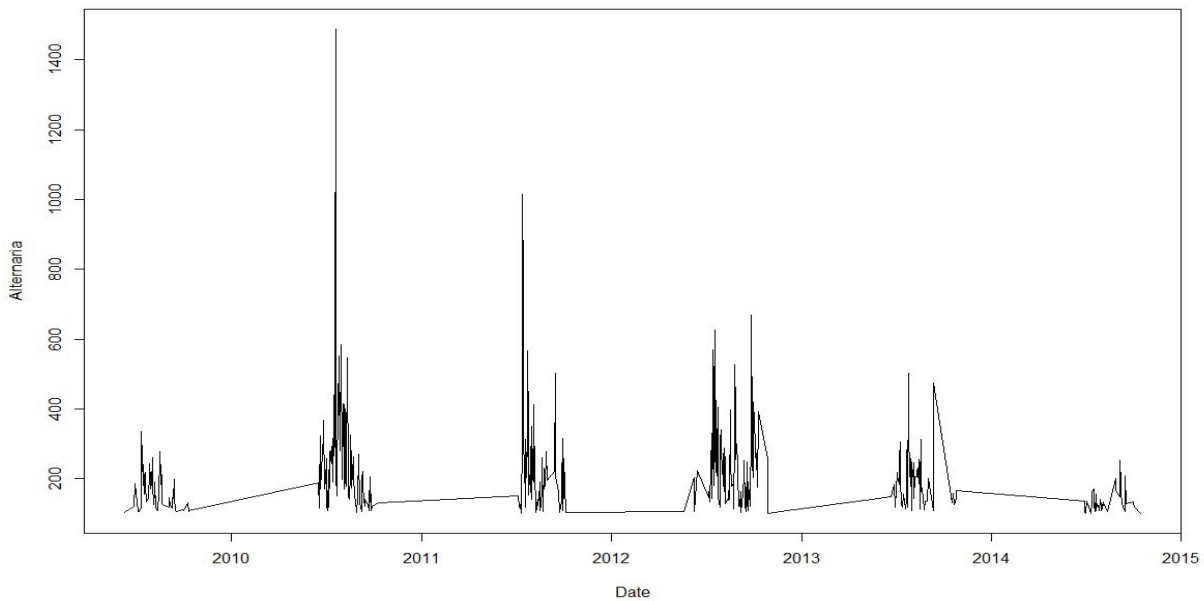


Рис. 3. Дані EAN про вміст спор грибів *Alternaria* у повітрі м. Вінниця за період 2009—2014 рр., обчислені за середньодобовими значеннями, які є не меншими 100 одиниць

#### Етап 4. Агрегування даних

Традиційно, аналіз закономірностей динаміки спор *Alternaria* у повітрі агрегують за період 2, 3, 4, 7, 10, 30 днів та по сезонах року [2—4], але отримані залежності сильно варіюються по кожному посту спостережень та роках. За даними статей [2—4] такий підхід теж дає можливість побудувати регресії, які пояснюють не більше 30...40 % даних ряду. Варто шукати закономірності іншим шляхом.

Відомо, що концентрація спор грибів *Alternaria* у повітрі залежить від багатьох метеофакторів. Зазвичай, дослідники аналізують залежність цієї концентрації від таких метеопказників: щодобова середня, мінімальна та максимальна температура повітря протягом доби, сонячні години, атмосферний тиск, точка роси, вологість повітря, кількість опадів, напрям та сила вітру, випаровування тощо [2—4]. При цьому, якщо намагались шукати залежності між необробленими даними по спорах і цими факторами, то, по-перше, такі залежності будуть дуже специфічними для кожного посту спостережень, по-друге, досить випадковими, а не системними, і це видно, наприклад, з ретельного аналізу багатьох джерел, проведеного у статті [4]. Причина цього у тому, що усі ці процеси (і кількість спор, і метеофактори) є нестационарними випадковими процесами. І вибірки у декілька років не дають можливості відтворити повну картину (наприклад, в гідрології для аналізу закономірностей з обсягу води у заданому пості прийнято брати ряди щодобових значень, не менше, ніж за 30 років підряд, а бажано – за 100). У сфері аеробіологічного моніторингу таких рядів немає. В Австралії є ряди за 20 років [3], в Україні 6—10 [4], але цього недостатньо.

Для того, щоб максимально уникнути специфічних особливостей років, сезонів, постів, тощо, для побудови регресії та виявлення основних закономірностей пропонуємо здійснювати агрегування, у першому наближенні — по роках, і брати до уваги метеофактори та значення концентрацій спор грибів *Alternaria* тільки у ті дні, коли ці концентрації перевищували 100 спор/м<sup>3</sup>.

Позначимо новий ряд концентрації спор грибів *Alternaria* у повітрі, утворений у такий спосіб, як  $S_{avg}$ . Але, для ретельнішого пошуку залежностей між  $S_{avg}$  та метеоданими під час агрегування останніх теж по роках пропонуємо шукати мінімальне, максимальне і середні значення цих метеоданих за рік.

Саме така вибірка з 6 років дозволяє коректно порівнювати сумарний вплив усіх метеофакторів та уникати впливу багатьох випадковостей. Недоліком такого підходу є те, що на малій вибірці не можна провести множинний кореляційний та регресійний аналіз з усіма метеофакторами одночасно. Як відомо, методи апроксимації експериментальних даних вимагають, щоб кількість коефіцієнтів регресії  $N$  була не меншою  $N + 1$  [7, 9]. А ще ж треба дотримуватись принципу перевірки адекватності регресійної моделі, коли коефіцієнти знаходяться за одними значеннями ряду даних, а адекватність регресії перевіряється за іншими. Для цієї перевірки потрібно мати, як мінімум, одне значення з цього ж ряду даних, яке для побудови регресії не використовувалось. Отже, щоразу

можна порівнювати кількість спор з не більше, ніж 3 метеофакторами. А для цього їх треба попередньо відібрати на основі кореляційного аналізу. А цьому аналізу, у свою чергу, повинен передувати етап висунення гіпотези про закони розподілу як агрегованих за рік даних, так і відповідних їм агрегованих метеоданих, що дещо ускладнюється дуже малою кількістю вхідних даних. І вже відповідно, до результатів цього аналізу, правильно вибирати математичний апарат і кореляційно-го, і регресійного аналізу. Пропонуємо аналізувати два типових варіанти:

– статистична гіпотеза 1 про те, що наявна вибірка  $S_{avg}$  з 6 агрегованих за рік даних, як і відповідні агреговані метеодані (2009—2014 рр.) — це частина генеральної сукупності, яка, в цілому розподілена за нормальним законом, що дозволило застосувати класичний кореляційний аналіз та провести множинний регресійний аналіз та ідентифікацію лінійної регресії з використанням оператора `lm` мови R.

– статистична гіпотеза 2 про те, що наявні вибірки розподілені за законом Пуасона (вважається, що саме закон Пуасона або експоненціальний найкраще описують дані по спорах [2—4]), а тому для їх аналізу слід було використовувати коефіцієнт рангової кореляції Спірмена та, за його результатами, було ідентифіковано нелінійну регресію (з використанням експоненти) за допомогою оператора `glm`.

Ще нагадаємо, що, як це прийнято, значення за якийсь рік (не найменше і не найбільше) не повинне братись до уваги при ідентифікації регресії [7, 9]. Воно має бути використане під час перевірки адекватності побудованої регресії.

### Застосування кореляційного та регресійного аналізу даних на прикладі

Застосуємо етап 4 для середньорічної кількості *Alternaria* у повітрі м. Вінниці за 2009—2014 рр. Для перевірки адекватності будемо використовувати рік 2011 р. А ідентифікацію регресії будемо проводити за даними решти 5 років.

#### Статистична гіпотеза 1.

	ALTE_avg
ALTE_avg	1.0000000
AAT_min	-0.5771366
AAT_max	0.5214923
AAT_avg	0.4163884
MAAT_min	-0.7357711
MAAT_max	0.4878287
MAAT_avg	0.3185724
MIAT_min	0.3479594
MIAT_max	0.8794898
MIAT_avg	0.6839532
DPT_min	0.2999916
DPT_max	0.6749285
DPT_avg	0.7232005
RH_min	0.3914866
RH_max	0.7239055
RH_avg	0.7380126
PRE_max	0.7897582
PRE_avg	0.7508044
AWS_max	-0.1059168
AWS_avg	-0.3435801

Рис. 4. Коефіцієнти кореляції між  $S_{avg}$  (ALTE\_avg) та мінімальними (min), максимальними (max) та середніми (avg) за рік значеннями таких метеопказників: середньодобові температури  $X_1$  (AAT\_min — мінімальні за рік),  $X_2$  (AAT\_max — максимальні за рік),  $X_3$  (AAT\_avg — середні за рік); максимальні за добу температури  $X_4$  (MAAT\_min),  $X_5$  (MAAT\_max),  $X_6$  (MAAT\_avg); мінімальні за добу температури  $X_7$  (MIAT\_min),  $X_8$  (MIAT\_max),  $X_9$  (MIAT\_avg); точка роси  $X_{10}$  (DPT\_min),  $X_{11}$  (DPT\_max),  $X_{12}$  (DPT\_avg); вологість повітря  $X_{13}$  (RH\_min),  $X_{14}$  (RH\_max),  $X_{15}$  (RH\_avg); опади  $X_{16}$  (PRE\_min),  $X_{17}$  (PRE\_max),  $X_{18}$  (PRE\_avg); швидкість вітру  $X_{19}$  (AWS\_min),  $X_{20}$  (AWS\_max),  $X_{21}$  (AWS\_avg)

З використанням класичної формули обчислення кореляційного коефіцієнта для аналізу випадкових величин, розподілених за нормальним законом, обчислюємо коефіцієнт між середньою за рік кількістю *Alternaria* та усіма агрегованими метеопказниками (рис. 4).

З рис. 4 видно, що потенційно найбільш впливовими є такі показники (для них коефіцієнт кореляції складає 0,7 і вище): MAAT\_min, MIAT\_max, DPT\_avg, RH\_max, RH\_avg, PRE\_max, PRE\_avg.

Далі засобами мови R з використанням функції `lm` проводимо множинний регресійний аналіз даних шляхом перебору різних комбінацій по 3 зі списку цих факторів і підбираємо оптимальну комбінацію з відповідною ідентифікацією регресії (рис. 5).

Residuals:				
1	2	3	4	5
-4.3352	2.0372	-0.3048	-0.7116	3.3145
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-560.878	75.675	-7.412	0.0854 .
MAAT_min	-4.083	1.105	-3.696	0.1682
MIAT_max	11.858	3.420	3.468	0.1787
RH_avg	8.497	1.054	8.063	0.0786 .
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 5.876 on 1 degrees of freedom				
Multiple R-squared: 0.9966, Adjusted R-squared: 0.9866				

Рис. 5. Результат множинного регресійного аналізу середньорічних концентрацій спор грибів *Alternaria* у повітрі м. Вінниці за 2009—2014 рр., обчислених за середньодобовими значеннями, які дорівнюють або перевищують 100 спор/м<sup>3</sup> на добу

На основі рис. 5 можна зробити такі висновки:

– ідентифікована регресія для обчислення середньорічних концентрацій  $Y$  спор грибів *Alternaria* у повітрі м. Вінниці за 2009—2014 рр. має вигляд

$$Y_1 = -560,878 - 4,083X_4 + 11,858X_8 + 8,497X_{15}; \quad (1)$$

– обидва  $R^2$ -критерії (Multiple R-squared та Adjusted R-squared) дуже наближені до 1, що означає, що уся вибірка даних за 5 років  $S_{avg}$  (на 99,7 %) дуже добре описується ідентифікованою регресією (1);

– усі 3 метеофактори є близькими до границі значущості 0,05 (тобто ймовірність того, що вони є значущими, дорівнює 0,8318, 0,8213, 0,9214, відповідно), але все ж таки є більшими (ймовірності є меншими 0,95), тобто є певна ймовірність того, що виявлені закономірності та коефіцієнти регресії (1) є дещо випадковими.

Результат ідентифікації, з урахуванням 2011 року, подано у таблиці 1.

Таблиця 1

Результат перевірки адекватності ідентифікованої регресії (1) за статистичною гіпотезою 1

Роки	2009	2010	2011	2012	2013	2014
Вхідні дані $S_{avg}$	164,5849	254,8442	225,375	241,1252	200,2820	134,2307
Обчислені значення $Y_1$	168,9201	252,8070	232,8498	241,4300	200,9936	130,9162
Відносна похибка, %	2,63	0,80	3,32	0,13	0,36	2,47

Гіпотеза щодо значущості відібраних факторів залишилась під питанням, однак, обчислена похибка у 3,3 % для контрольного 2011 року є прийнятною, отже, у першому наближенні, можна стверджувати, що регресія (1) добре описує реальні закономірності. Тому усереднена за рік кількість спор грибів *Alternaria* у повітрі м. Вінниці за 2009—2014 роки, обчислена за середньодобовими значеннями, які дорівнюють або перевищують 100 спор/м<sup>3</sup> на добу (ALTE\_avg):

– зменшується, у разі збільшення мінімальних за рік максимальних за добу (MAAT\_min) добових температур (MAAT\_min);

– збільшується, у разі збільшення максимальних за рік мінімальних за добу (MIAT\_max) температур;

– збільшується, у разі збільшення середньорічних значень вологості повітря (RH\_avg);

– практично не залежить від середньорічних значень точки роси, кількості опадів та швидкості вітру або ця залежність є явно нелінійною.

Виявлені закономірності можна, у першому наближенні, використати для аналізу первинного ряду значень, показаного на рис. 1, зокрема для ідентифікації регресії, яка описує тренд первинного ряду.

	ALTE_avg
ALTE_avg	1.0
AAT_min	-0.6
AAT_max	0.7
AAT_avg	0.1
MAAT_min	-0.9
MAAT_max	0.6
MAAT_avg	0.3
MIAT_min	0.2
MIAT_max	0.9
MIAT_avg	0.9
DPT_min	0.5
DPT_max	0.6
DPT_avg	0.9
RH_min	0.3
RH_max	0.6
RH_avg	0.9
PRE_max	0.9
PRE_avg	0.7
AWS_max	-0.1
AWS_avg	0.0

#### Статистична гіпотеза 2.

Обчислюємо коефіцієнти рангової кореляції Спірмена (рис. 6).

Як видно з рис. 6, ці коефіцієнти є більшими 0,7 для таких показників: AAT\_max, MAAT\_min, MIAT\_max, MIAT\_avg, DPT\_avg, RH\_avg, PRE\_max, PRE\_avg, тобто — такі ж показники, як і у випадку гіпотези 1, за винятком того, що додалися AAT\_max і MIAT\_avg та зник показник RH\_max.

Далі проводимо множинний регресійний аналіз даних на основі узагальненої функції glm (General Linear Model) для пуассонівського закону розподілу та регресії на основі експоненціальної функції (рис. 7).

Рис. 6. Коефіцієнти рангової кореляції Спірмена залежності ALTE\_avg від метеопказників

```

Deviance Residuals:
 1          2          3          4          5
0.064638 -0.024356  0.003749  0.009604 -0.054545

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.29366    0.87830   1.473 0.140773
MAAT_min     -0.02420    0.01277  -1.895 0.058053 .
MIAT_max      0.04938    0.03980   1.241 0.214789
RH_avg        0.04858    0.01335   3.640 0.000273 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 52.8984953 on 4 degrees of freedom
Residual deviance: 0.0078527 on 1 degrees of freedom

```

Рис. 7. Результат множинного регресійного аналізу середньорічних концентрацій спор грибів *Alternaria* у повітрі м. Вінниці за 2009-2014 рр., обчислених за середньодобовими значеннями, які дорівнюють або перевищують 100 одиниць на добу, у припущенні, що дані розподілені з пуасонівським законом, а регресія є експоненціальною функцією

На основі рис. 7 можна зробити такі висновки:

– ідентифікована регресія має вигляд

$$Y_2 = \exp(-1,29366 - 0,02420X_4 + 0,04938X_8 + 0,04858X_{15}); \quad (2)$$

– залишки Residuals зменшились у порівнянні з регресією (1), з чого можна зробити висновок про те, що регресія (2) ще краще описує вхідні дані;

– значущість вологості RH\_avg повністю підтверджена, показник MAAT\_min — на грані підтвердження значущості, а ось значущість показника MIAT\_max залишилась під питанням, отже, все одно є деяка ймовірність того, що виявлені закономірності та коефіцієнти регресії (2) є дещо випадковими.

Результат ідентифікації, з урахуванням року 2011 р., подано у таблиці 2.

Таблиця 2

Результат перевірки адекватності ідентифікованої регресії (2) за статистичною гіпотезою 2

Роки	2009	2010	2011	2012	2013	2014
Вхідні дані $S_{avg}$	164,5849	254,8442	225,375	241,1252	200,2820	134,2307
Обчислені значення $Y_2$	163,7571	255,2332	234,764	241,0670	200,1461	134,8637
Відносна похибка, %	0,50	0,15	4,17	0,02	0,07	0,47

Як видно, з табл. 2, сумарна відносна похибка стала меншою, ніж за регресією (1), але значущість показника MIAT\_max все одно залишилась під питанням. Обчислена похибка для контрольного 2011 р. у 4,17 % стала дещо більшою, ніж 3,32 % для регресії (1), але все одно є прийнятною, отже, у першому наближенні можна стверджувати, що регресія (2) теж добре описує реальні закономірності, а їх характер такий самий, як і для регресії (1). Але, їх подібність доводить, що, у першому наближенні, все ж таки можна використовувати лінійну регресію замість нелінійної, що значно спрощує подальше її використання.

Виявлені закономірності та побудовані регресії можна, у першому наближенні, використати для аналізу первинного ряду значень, показаного на рис. 1, зокрема для ідентифікації регресії, яка описує тренд цього первинного ряду.

Для усунення випадковості отриманих регресії рекомендуємо збільшувати ряди первинних даних, наприклад, за рахунок швидшого оброблення даних за 2015—2016 рр.

Запропонований алгоритм та програмна реалізація мовою програмування R є універсальними і їх можна застосовувати і для обробки даних інших постів Європейської системи аеробіологічного моніторингу.

## Висновки

Розглянуто задачу статистичного аналізу динаміки вмісту спор грибів *Alternaria* за даними Європейської системи аеробіологічного моніторингу на прикладі щодобових даних для м. Вінниця за 2009—2014 рр. та їх залежності від метеофакторів.

Проаналізовано типові проблеми такого аналізу та недоліки підходів, використаних іншими дослідниками, зокрема, нехтування перевіркою вибірок різних років за критерієм (тестом) Кохрена за їх об'єднання в один часовий ряд; побудова та аналіз регресійної моделі, яка описує менше 40 % даних, тобто не є коректною, та ін.

Запропоновано низку підходів щодо оброблення первинних даних спостережень, які дозволяють встановити закономірності залежності динаміки концентрації спор грибів *Alternaria* у повітрі від метеофакторів та за ними побудовано дві регресії — лінійну для статистичної гіпотези про нормальний закон розподілу для усіх величин та нелінійну на основі експоненти для статистичної гіпотези про пуасонівський закон розподілу. Перевірка адекватності цих регресій на контрольному році, яке не використовувалось для ідентифікації показала, що результати практично ідентичні (3,3 % та 4,2 %, відповідно), що дозволяє зробити висновок про те, що, у першому наближенні, все ж таки можна використовувати лінійну регресію замість нелінійної, що значно спрощує подальше її використання. Для усунення випадковості отриманих регресій рекомендовано збільшувати ряди первинних даних.

Виявлені закономірності та побудовані регресії можна використати для ідентифікації регресії, яка описує тренд первинного ряду спостережень. Запропонований алгоритм та програмну реалізацію мовою програмування R можна застосовувати і для оброблення даних інших постів Європейської системи аеробіологічного моніторингу.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Рекалова О. М. Динаміка деяких клініко-функціональних показників у хворих на бронхіальну астму на фоні проведення протигрибкової сублінгвальної алерген-специфічної імунотерапії / О. М. Рекалова, Л. В. Петренко // Астма та алергія. — 2017. — № 1. — С. 13—16.
2. A 10-year study of *Alternaria* and *Cladosporium* in two Polish cities (Szczecin and Cracow) and relationship with the meteorological parameters / [A. Grinn-Gofroń, A. Strzelczak, D. Stępańska et al.] // *Aerobiologia*. — 2016. — № 32, Issue 1. — Pp. 83—94.
3. Ianovici N. Atmospheric concentrations of selected allergenic fungal spores in relation to some meteorological factors, in Timișoara (Romania) / N. Ianovici // *Aerobiologia*. — 2016. — № 32, Issue 1. — Pp. 139—156.
4. Stennett P. J. *Alternaria* spores in the atmosphere of Sydney, Australia, and relationships with meteorological factors / P. J. Stennett, P. J. Beggs // *International Journal of Biometeorology*. — 2004. — № 49, Issue 2. — Pp. 98—105.
5. The improvement of the volumetric monitoring system to raise the analysis accuracy for the allergic pollen found in the city atmosphere [Електронне видання] / [Vitalii B. Mokin, Victoria V. Rodinkova, Tatiana Y. Vuzh, Waldemar Wójcik, Saltanat Sailarbek] // *Przegląd Elektrotechniczny*, ISSN 0033-2097. — R. 93 NR 5/2017. — 5 pages. — doi:10.15199/48.2017.05.17. — Режим доступу: <http://pe.org.pl/articles/2017/5/17.pdf>.
6. George E. P. Box. Time series analysis [Electronic resource] / George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel. — Fourth Edition. John Wiley & Sons, Inc., Hoboken, NJ, Inc., 2008. — doi: 10.1002/9781118619193.
7. Мокін Б. І. Методологія та організація наукових досліджень : навч. посіб. / Б. І. Мокін, О. Б. Мокін. — 2-ге вид., змін. та доп. — Вінниця : ВНТУ, 2015. — 317 с.
8. Kokoska S. Statistical Tables and Formulae / Stephen Kokoska, Christopher Nevison. — New York : Springer New York, 1989. — 93 p.
9. Моделювання та оптимізація систем : підручник / В. М. Дубової, Р. Н. Кветний, О. І. Михальов, А. В. Усов. — Вінниця : ПП ТД «Едельвейс», 2017. — 804 с.

Рекомендовано кафедрою системного аналізу, комп'ютерного моніторингу та інженерної графіки ВНТУ

Стаття надійшла до редакції 28.08. 2017

**Мокін Віталій Борисович** — д-р техн. наук, професор, завідувач кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки, e-mail: [vbmokin@gmail.com](mailto:vbmokin@gmail.com) ;

**Родінкова Вікторія Валеріївна** — д-р біол. наук, професор, професор кафедри фармації, e-mail: [vikarodi@gmail.com](mailto:vikarodi@gmail.com) ;

**Дратований Михайло Володимирович** — аспірант кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки, e-mail: [mishadratovany@gmail.com](mailto:mishadratovany@gmail.com) ;

**Білоус Олена Сергіївна** — здобувач наукового-дослідного центру



V. B. Mokin<sup>1</sup>  
V. V. Rodinkova<sup>2</sup>  
M. V. Dratovanyi<sup>1</sup>  
O. S. Bilous<sup>2</sup>

## Statistical Analysis of Dynamics of *Alternaria* Fungal Spores Using Data of the European System of Aerobiological Monitoring

<sup>1</sup>Vinnitsia National Technical University;

<sup>2</sup>National Pyrogov Memorial Medical University, Vinnitsia

*The article deals with the actual problem of statistical analysis of *Alternaria* fungal spores dynamics in the atmosphere and their dependence on meteorological factors according to the data of the European system of aerobiological monitoring. The typical problems of this analysis are discussed using the example of daily data for Vinnitsia for the years 2009—2014. The algorithm of the problem solution is proposed and its approbation with the use of the R<sup>1</sup> tool is made by means of constructing linear and nonlinear regressions of the dependence of the *Alternaria* data on three meteorological factors. New laws have been revealed and recommendations for more effective application of this algorithm have been made.*

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

**Mokin Vitalii B.** — Dr. Sc. (Eng.), Professor, Head of the Chair of Systems Analysis, Computer Monitoring and Engineering Graphic, e-mail: vbmokin@gmail.com ;

**Rodinkova Victoria V.** — Dr. Sc. (Biology), Professor of the Chair of Pharmacy, e-mail: vikaodi@gmail.com ;

**Dratovanyi Mykhailo V.** — Post-Graduate Student of the Chair of Systems Analysis, Computer Monitoring and Engineering Graphic, e-mail: mishadratovany@gmail.com ;

**Bilous Olena S.** — Researcher of the Scientific Research Center

В. Б. Мокин<sup>1</sup>  
В. В. Родинкова<sup>2</sup>  
М. В. Дратованый<sup>1</sup>  
Е. С. Билоус<sup>2</sup>

## Статистический анализ динамики спор грибов *Alternaria* по данным европейской системы аэробиологического мониторинга

<sup>1</sup>Винницкий национальный технический университет;

<sup>2</sup>Винницкий национальный медицинский университет им. Н. И. Пирогова

*Рассмотрена актуальная задача статистического анализа динамики содержания спор грибов *Alternaria* по данным Европейской системы аэробиологического мониторинга и зависимости этих данных от метеофакторов. Исследованы типичные проблемы этого анализа на примере ежесуточных данных для г. Винница за 2009—2014 гг. Предложен алгоритм их решения и проведена апробация этого алгоритма с использованием инструментария языка программирования R путем построения линейной и нелинейной регрессий зависимости этих данных от трех метеофакторов. Обнаружены новые закономерности и сделаны рекомендации по более эффективному применению этого метода.*

**Ключевые слова:** статистический анализ, моделирование аллергенных спор грибов в воздухе, *Alternaria*, временные ряды, корреляционный анализ, регрессия, язык R, аэробиологический мониторинг.

**Мокин Виталий Борисович** — д-р техн. наук, профессор, заведующий кафедрой системного анализа, компьютерного мониторинга и инженерной графики, e-mail: vbmokin@gmail.com ;

**Родинкова Виктория Валерьевна** — д-р биол. наук, профессор, профессор кафедры фармации, e-mail: vikaodi@gmail.com ;

**Дратованый Михаил Владимирович** — аспирант кафедры системного анализа, компьютерного мониторинга и инженерной графики; e-mail: mishadratovany@gmail.com ;

**Билоус Елена Сергеевна** — соискатель научно-исследовательского центра