

УДК 504.349

В.М. ТРИСНЮК

МАТЕМАТИЧНІ МОДЕЛІ ВЗАЄМОЗВ'ЯЗКІВ МІЖ ПОКАЗНИКАМИ ЕКОЛОГІЧНОГО СЕРЕДОВИЩА І ЗДОРОВ'ЯМ НАСЕЛЕННЯ

***Анотація.** Розв'язано важливу науково-практичну задачу, яка полягає у визначенні закономірностей впливу окремих хімічних елементів на навколишнє середовище та здоров'я населення.*

***Ключові слова:** моніторинг, екологічна безпека, захворюваність, комплексний аналіз, здоров'я населення, техноприродна система.*

Вступ

Методологія оцінки екологічної ситуації, принципи застосування і розроблення інструментальних методів та математичних моделей для виявлення та ідентифікації аномалій параметрів стану екосистем за даними космічного і наземного моніторингу ґрунтуються на одному з пріоритетних напрямків – забезпечення екологічно та техногенно-безпечних умов життєдіяльності громадян і суспільства, збереження навколишнього природного середовища та раціональне використання природних ресурсів. Рівень небезпеки технічного об'єкта для довкілля і здоров'я людини може бути різним – від найнезначнішого до критичного і навіть катастрофічного. При цьому сам рівень є досить невизначеним і, як правило, відповідає первинному екологічному стану довкілля. Тому дуже важливим є створення системи екологічної безпеки, яка б дозволяла керувати прогнозними впливами, стежити за змінами екологічної ситуації, впроваджувати заходи мінімізації впливу та оцінювати їх [1].

Працями багатьох дослідників, зокрема О.М. Адаменка, Г.І. Рудька, М.В. Крихівського та інших [1–3], доведена залежність між показниками забруднення зовнішнього середовища і захворюваністю населення.

Метою роботи є створення нових математичних моделей взаємозв'язків між показниками екологічного середовища та методів, спрямованих на покращення здоров'я населення.

1. Загальна постановка задачі

Дослідження зв'язків між показниками забруднення зовнішнього середовища і здоров'ям жителів є пріоритетним завданням з точки зору екологічної безпеки. Інформативність і наглядність результатам такого аналізу надають кількісні характеристики. Їх подальше використання в управлінських рішеннях має найбільш широкі можливості. Для вимірних величин стану екологічної ситуації SES і досліджуваних змін здоров'я Z зв'язок між ними може бути представлений у вигляді функціональної залежності:

$$Z = f(SES) + \varepsilon, \quad (1)$$

де ε – відхилення від однозначно визначеної функції, що не пов'язане з впливом екологічних чинників і похибкою вимірювання.

Визначення в (1) є регресійною залежністю Z від SES , або регресією Z на SES , а функція f – регресійною функцією. Змінну SES вважатимемо незалежною, а змінну Z – залежною. Задачею кореляційно-регресійного аналізу є встановлення факту наявності зв'язку між Z від SES і пошук функції f , що адекватно описує цей зв'язок.

Якщо будуть знайдені статистично значущі зв'язки між досліджуваними даними та вдасться з незначним відхиленням виразити залежну змінну у вигляді аналітичної функції від незалежної, то отриману функціональну залежність можна буде використати для прогнозування значень залежної змінної у діапазоні значень, які детально не досліджувались. Такий прогноз міститиме певну неточність, що залежить від точності апроксимації, величини відхилення ε та інтервалу, для якого знайдено аналітичну залежність. У різних діапазонах значень зміни незалежної змінної аналітична функція може бути не однаковою, що відповідає апроксимації кусково-неперервними залежностями. Щодо захворюваності населення, то зміні рівня чинника ризику в інтервалі між пороговими значеннями відповідають зміни показників здоров'я, тоді як в області нижче від порога чутливості і вище від порога насичення (ці порогові значення можуть бути різними для різних показників здоров'я) такого впливу не спостерігається [4].

Формалізуючи задачу аналізу екологічної ситуації певної території та вплив на здоров'я концентрацій окремих хімічних елементів, введемо позначення для концентрацій та типів хвороб. Концентрації забруднювачів в ґрунтах і поверхневих водах екологічних ділянок опишемо як $kz_{i,j}$, де i змінюється від 1 до 17, відповідаючи номерам екологічних ділянок (райони Тернопільської області, рис. 1), а j змінюється від 1 до 9, відповідаючи хімічним елементам (1 – Pb, 2 – Zn, 3 – Cd, 4 – Ca, 5 – Co, 6 – Mn, 7 – Mg, 8 – Cu, 9 – Ni). Ці змінні у формулі (8) визначають SES , тобто описують екологічну обстановку і є незалежними змінними. Їх в математичних моделях інколи називають незалежними, факторами-аргументами, екзогенними, предикаторними або пояснючими.

Змінні, що описують вплив стану навколишнього середовища, а саме захворюваність або смертність жителів певної території, будемо позначати з використанням міжнародної статистичної класифікації хвороб, травм і причин смерті. Позначимо: 1_x – хвороби крові та кровотворних органів; 2_x – хвороби органів дихання, серед яких 3_x – хронічний бронхіт, 4_x – бронхіальна астма; 5_x – хвороби органів травлення; 6_x – онкологічні захворювання.

Поширеність захворювань запишемо як $zx_{i,j}$. Індекс i задає номер екологічної ділянки (змінюється від 1 до 17), а j – номер типу захворюваності: 1 – всього, 2 – 1_x , 3 – 2_x , 4 – 3_x , 5 – 4_x , 6 – 5_x , 7 – 6_x . Інші типи: первинна захворюваність – $zxI_{i,j}$, $sm_{i,j}$ – смертність, $zzx_{i,j}$ – зміни в часі поширеності захворювань, $zzxI_{i,j}$ – зміни в часі первинної захворюваності. Індеси i, j мають інтерпретацію, аналогічну до $zx_{i,j}$. Ці змінні в математичних моделях іноді називають відгукками, ендогенними, результуючими або пояснюваними.



Рисунок 1 – Картохсхема районних екологічних дільниць Тернопільської області

Ще один тип змінних, які безпосередньо не вимірюються, позначимо як nek_i , де i – номер екологічної дільниці, що змінюється від 1 до 17. Ці латентні змінні відображають вплив рівня доходів населення, рівень медичного обслуговування, спадковість та похибки вимірювання.

Тепер можна сформулювати задачу аналізу даних екологічної ситуації території та медичного обслуговування населення. Необхідно побудувати таку векторну функцію:

$$fex(kz_{i,j}) = \begin{pmatrix} fex_1(kz_{i,j}) \\ fex_2(kz_{i,j}) \\ \dots \\ fex_m(kz_{i,j}) \end{pmatrix}, \quad (2)$$

яку можна використовувати для розрахунку (прогнозування) в залежності від екологічних характеристик $(kz_{i,j})$ значення вихідних змінних $zx_{i,j}$, $zxI_{i,j}$, $sm_{i,j}$, $zzx_{i,j}$, $zzxI_{i,j}$. Функцію fex потрібно знаходити математичними методами, наприклад, методом найменших квадратів. Розв'язуючи цю задачу, необхідно починати зі з'ясування існування або відсутності статистично значущого зв'язку між $kz_{i,j}$ та nz_i $zx_{i,j}$, $zxI_{i,j}$, $sm_{i,j}$, $zzx_{i,j}$, $zzxI_{i,j}$. Шукати адекватну аналітичну залежність $fex_k(kz_{i,j})$ слід тільки у випадку існування такого зв'язку.

Тісноту статистичного зв'язку між випадковими змінними η (вихідна) і ξ (вхідна) показує [5] індекс кореляції $IK_{\eta,\xi}$:

$$IK_{\eta,\xi}^2 = \frac{\sigma_f^2}{\sigma_\eta^2} = 1 - \frac{\sigma_{f(x)}^2}{\sigma_\eta^2}, \quad (3)$$

де $\sigma_f^2 = Df(\xi)$ – дисперсія функції регресії $y_{cp}=f(x)$, $\sigma_\eta^2 = D\eta$ характеризує повну варіацію (дисперсію) результуючої змінної η , а $\sigma_{f(x)}^2$ характеризує середню величину неконтрольованого залишкового випадкового компоненту. У формулі (3) змінна η відповідає одному зі стовпців матриць nz_i $zx_{i,j}$, $zxI_{i,j}$, $sm_{i,j}$, $zzx_{i,j}$, $zzxI_{i,j}$ для фіксованого значення j , а саме одному з типів захворюваності або смертності, а змінна ξ відповідає одному з 7, тобто одному з заданого набору концентрацій хімічних елементів.

У випадку нормального розподілу випадкових величин η і ξ індекс кореляції перетворюється на коефіцієнт кореляції:

$$r = \frac{E[(\xi - E\xi)(\eta - E\eta)]}{\sqrt{D\xi \cdot D\eta}} = \frac{\text{cov}(\xi, \eta)}{\sigma_\xi \cdot \sigma_\eta}, \quad (4)$$

де коваріація $\text{cov}(\xi, \eta)$ – другий центральний змішаний елемент двовимірної випадкової величини (ξ, η) , а σ_ξ і σ_η – середньоквадратичні (безумовні) відхилення відповідних компонент. Значення коефіцієнта кореляції деякої вибірки r' (статистична оцінка r' невідомого значення r) розраховується за відомими парами даних вхідної і вихідної змінної (ses_1, z_1) , $(ses_2, z_2), \dots, (ses_n, z_n)$ як:

$$r' = \frac{\sum_{i=1}^n (ses_i - \overline{ses})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^n (ses_i - \overline{ses})^2 \sum_{i=1}^n (z_i - \bar{z})^2}}, \quad (5)$$

де

$$\overline{ses} = \frac{1}{n} \sum_{i=1}^n ses_i \quad \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i. \quad (6)$$

Для вибірок невеликого об'єму ($n < 15$) точнішою оцінкою коефіцієнта кореляції є r^* :

$$r^* = r' \left[1 + \frac{1 - r'^2}{2(n - 3)} \right]. \quad (7)$$

Кожен з розрахованих коефіцієнтів кореляції слід перевіряти на значущість статистичної гіпотези $r' = 0$, тобто гіпотези про відсутність кореляційного зв'язку між досліджуваними змінними. Для цього застосовують характеристики розподілу Стьюдента [6]:

$$r^* = r' \left[1 + \frac{1 - r'^2}{2(n - 3)} \right]. \quad (8)$$

де $t_{0.05}(n-2)$ – 5-відсоткова точка розподілу Стьюдента з $n-2$ степенями свободи. Якщо рівність $r' = 0$ справедлива, тоді з ймовірністю 0,05 гіпотеза про відсутність статистичного зв'язку приймається.

Діапазон застосування визначають довірчі інтервали для істинного значення коефіцієнта кореляції r , що будуються, виходячи із нормальності розподілу r' . Відрізки $[r_1, r_2]$ обчислюються як:

$$r_{1,2} \approx r' + \frac{r'(1 - r'^2)}{2n} \pm u_{\frac{\alpha}{2}} \frac{1 - r'^2}{\sqrt{n}}, \quad (9)$$

де $u_{\frac{\alpha}{2}}$ – $100 \frac{\alpha}{2}$ %-ва точка стандартного нормального розподілу. Можна вважати, що істинне значення коефіцієнта кореляції r з імовірністю $1-\alpha$ належить відрізку $[r_1, r_2]$.

Досліджуючи нелінійну залежність випадкових величин η і ξ , коефіцієнт кореляції r не вважається основною характеристикою тісноти зв'язку. Використовуючи набори даних $(ses_1, z_1), (ses_2, z_2), \dots, (ses_n, z_n)$, слід скористатися розрахунком за допомогою формули (10) індексу кореляції $IK_{\eta, \xi}$. Спочатку необхідно задатися типом функціональної залежності. Вибравши залежність:

$$y_{\eta\theta} = E(\eta | \xi = x) = f(x; \theta_0, \theta_1, \dots, \theta_p), \quad (10)$$

де f – фіксована відома функція з аргументом x , що залежить від $(p + 1)$ невідомих параметрів $\theta_0, \theta_1, \dots, \theta_p$. У цьому випадку оцінка $I_{\eta, \xi}^{r^2}$ індексу кореляції $I_{\eta, \xi}^2$ буде:

$$I_{\eta, \xi}^{\prime 2} = 1 - \frac{\frac{1}{n-p-1} \sum_{i=1}^n (y_i - f(x_i, \theta'_0, \theta'_1, \dots, \theta'_p))^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (11)$$

де $\theta'_0, \theta'_1, \dots, \theta'_p$ – побудовані за допомогою апріорної інформації оцінки невідомих параметрів, що входять в опис функції регресії.

Тісноту зв'язку між кількома випадковими змінними можна визначити множинним коефіцієнтом кореляції $R_{\eta, \xi}$. Для випадкових величин η (ендогенна ознака) і багатовимірною величиною $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(r)})$ (екзогенна змінна) він визначається формулою:

$$R_{\eta, \xi}^2 = 1 - \frac{\bar{\sigma}_{\eta(X)}^2}{\sigma_{\eta}^2}, \quad (12)$$

де варіація результуючого показника $\sigma_{\eta}^2 = D\eta$, а середня величина умовної дисперсії:

$$\bar{\sigma}_{\eta(X)}^2 = E_x D[\eta | \xi = X]. \quad (13)$$

Якщо розподіли випадкових змінних ses і z не можна вважати нормальними, використаємо непараметричні статистичні методи, наприклад метод рангових (порядкових) кореляцій. Він базується на характеристиці тісноти зв'язку коефіцієнтом Спірмена або Кендала. У випадку, коли у вибірках немає однакових значень, коефіцієнт рангової кореляції Спірмена розраховується за допомогою формули [7]:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}, \quad (14)$$

де D – різниця відповідних пар рангів, а n – об'єм вибірки. В іншому випадку використовуються формули [5]:

$$r_{S,B} = 1 - \frac{6 \sum D^2}{(n^3 - n)(T_{x'} + T_{y'})}, \quad (15)$$

$$T_{x'} = \frac{1}{2} \cdot \sum (t_{x'}^3 - t_{x'}); \quad T_{y'} = \frac{1}{2} \cdot \sum (t_{y'}^3 - t_{y'}), \quad (16)$$

де $t_{x'}$ – кількість однакових значень вхідної змінної, $t_{y'}$ – кількість однакових значень вихідної змінної, між якими виявляється залежність.

Множинний статистичний зв'язок між кількома порядковими змінними показує коефіцієнт конкордації (узгодженості), який розраховується за допомогою формули:

$$W'(m) = \frac{12}{m^2(n^3 - n)} \cdot \sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(k_j)} - \frac{m(n+1)}{2} \right)^2, \quad (17)$$

де k_1, k_2, \dots, k_m – номери відібраних порядкових змінних із сукупності $x^{(0)}, x^{(1)}, \dots, x^{(p)}$, m – кількість змінних (порівнюваних впорядкованих послідовностей), n – кількість елементів впорядкованостей (об'єм вибірки).

Висновки

Вирішено важливу науково-практичну задачу, яка полягає у визначенні закономірностей впливу окремих хімічних елементів на навколишнє середовище та здоров'я населення.

Рівень небезпеки технічного об'єкта для довкілля і здоров'я людини може бути різним – від найнезначнішого до критичного і навіть катастрофічного. Тому дуже важливим є створення системи екологічної безпеки, яка б дозволяла керувати прогнозними впливами, стежити за змінами екологічної ситуації, впроваджувати заходи мінімізації впливу та оцінювати їх.

СПИСОК ЛІТЕРАТУРИ

1. Trofymchuk O. Geo-information Technologies for Decision Issues of Municipal Solid Waste / O. Trofymchuk, V. Trysnyuk, N. Novokhatska, I. Radchuk // Journal of Environmental Science and Engineering A 3 (2014) s. 183–187.
2. Рудько Г. І. Конструктивна геоекологія : наукові основи та практичне втілення / Г. І. Рудько, О. М. Адаменко [за ред. Г. І. Рудька]. – К. : ТОВ «Маклаут», 2008. – 320 с.
3. Рудько Г. І. Землелогія. Еколого-ресурсна безпека Землі / Г. І. Рудько, О. М. Адаменко [за ред. Г. І. Рудька]. – К. : Академпрес, 2009. – 512 с.
4. Самойленко В. Г. Умови існування розв'язку неоднорідного диференціального рівняння другого порядку в просторі швидко спадних функцій / В. Г. Самойленко, Ю. І. Самойленко // Наукова публікація. 2012 р. С. 71–80.
5. Довгий С.О. Методи прогнозування в системах підтримки прийняття рішень : наук.-навч. вид. / С. О. Довгий, П. І. Бідюк, О. М. Трофимчук, О. І. Савенков; НАН України, Ін-т телекомунікацій і глобал. інформ. простору. – К. : Азимут-Україна, 2011. – 608 с. – Бібліогр.: – С. 600–607.
6. Stefanyshyn D.V. A method of decision making at risk in natural resources use by pairwise comparison of alternatives with taking account of risks of lost opportunities / D.V. Stefanyshyn, Yu.D. Stefanyshyna // Proc. of Int. Scientific School «Modelling and Analysis of Safety and Risk in Complex Systems». July 7-11, 2009. – S.-Petersburg, Russia. – P.P. 435–439.
7. King E. G. Identfyng Linkages Amond Conceptual Models of Ecosystem Degradation and Restoration : Towards an integrative Framework / E. G. King and R. J. Hobbs // Restoration Ecology, 2006. – Volume 14, Issue 3. – PP. 369–378.

Стаття надійшла до редакції 18.11.2017.