

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ОЦІНКИ ПАРАМЕТРІВ  
СПЛАЙН-РЕГРЕСІЇ ПРИ ОБРОБЦІ ДАНИХ ЗАБРУДНЕННЯ  
АТМОСФЕРНОГО ПОВІТРЯ В ГІС «AIRNORM»**

*Анотація. Задача моніторингу повітряного середовища промисловими підприємствами полягає в зборі, обробці параметрів забруднення атмосфери на локальній території та напрацюванні за результатами висновків стосовно прийняття рішень про екологічний стан як промайданчика, так і прилеглих житлових територій. Наведено удосконалений метод апроксимації даних за використанням кусково-поліноміальної регресії. Запропонований алгоритм надає можливість підвищити адекватність визначення границь при різкій зміні концентрації шкідливої речовини. Представлено результати проведення обчислювального експерименту проведеного за показниками рівня вмісту викиду речовини NH<sub>3</sub> для джерела V-109 на Державному науково-виробничому підприємстві (ДНВП) «Цирконій» м. Кам'янське.*

*Ключові слова: сплайн-регресія, моделі розсіювання речовини у повітрі, геоінформаційні системи.*

**Постановка проблеми.** Техногенне навантаження на повітряне середовище – досить складний процес, який формує різноманіття факторів впливу. Забруднення атмосфери – зміна хімічного складу атмосфери в результаті потрапляння в неї домішок. Домішки в атмосфері – це розсіяна в атмосфері речовина, що не міститься в її постійному складі.

При обробці результатів спостереження за забрудненням повітря необхідна розробка моделі переносу шкідливих речовин та визначення в цих процесах метеорологічних та техногенних факторів. Оскільки положення джерел викидів, значення їх параметрів та метеорологічні параметри системи відомі, за допомогою математичних та фізичних моделей розраховується поля концентрацій шкідливих речовин в атмосферному повітрі для будь яких ситуацій.

Поряд з математичними моделями при оцінці рівня забруднення використовуються геоінформаційні системи (ГІС), як потужний інструмент, що забезпечу-

ють: збір, збереження, обробку, відображення даних, аналіз, оцінку, прогноз стану навколишнього середовища територіальної громади.

Одним з напрямів застосування ГІС є вирішення задачі моніторингу забруднення довкілля шкідливими викидами для певних локальних ділянок, що є територіями техногенно навантаженого регіону.

**Аналіз останніх досліджень і публікацій.** Актуальні проблеми моделювання забруднення повітря шкідливими речовинами наведено в роботах [2–4, 6, 8].

При обробці експериментальних даних виникає задача достовірного відображення фізико-хімічних процесів, що, як правило, є неоднорідними. Застосування сплайнів в задачах екстраполяції, інтерполяції та згладжування представлено в роботах [1, 5, 7].

**Мета роботи.** З врахуванням стану сучасних розробок, полягає в розробці інформаційної технології, алгоритмів та програмного забезпечення моделювання та візуалізації даних забруднення атмосферного повітря, для прийняття рішень про екологічне забруднення локальної ділянки.

**Основний матеріал.** Для моделювання забруднення атмосферного повітря викидами промислового підприємства локальної ділянки використано моделі, що базуються на процедурах методики ОНД-86.

Одним з недоліків методики ОНД-86 є її детермінований характер. Значення певних коефіцієнтів які використовуються методикою є строго затабульованими постійними величинами. Однак їх природної характер, який вони в собі несуть, є досить мінливим.

Задача інтерполювання сплайнами  $S(x)$  початкового масиву  $\{x_i, y_i, i = \overline{1, N}\}$  формулюється як задача знаходження серед усіх  $S(x) \in W_2^m[a, b]$  такої  $S(x_i) = y_i$ . Знаходження згладжуючого сплайну виконується із умови

$$\min_{S \in W_2^m} \left\{ \sum_{i=1}^N (y_i - S(x_i, \vec{\theta}))^2 \right\} \quad (1)$$

Вважаючи, що мають місце нелінійні моделі регресії, що зводяться до лінійного типу, подальша задача полягає як до знаходження оцінок параметрів  $\vec{\theta}$ , так і у знаходженні вузлів точності наближення емпіричних даних лінійними сплайнами з реалізацією інтерполяційної процедури знаходження мінімуму залишкової дисперсії.

Обчислювальні схеми відтворення одновимірних сплайн-регресійних моделей нелінійного типу, які наведено нижче, базуються на перетворенні остан-

ніх до лінійного типу. Процедура відтворення нелінійних сплайн-регресійних залежностей зводиться до відтворення лінійних сплайн-регресійних моделей с подальшим визначенням коефіцієнтів нелінійної регресії.

Нехай початковий масив даних  $\{x_i, y_i, i = \overline{1, N}\}$  перетворено в масив  $\{t_i, z_i, i = \overline{1, N}\}$ .

Лінійна сплайн-регресія визначена у наступному вигляді:

$$S(t) = \begin{cases} b_1 t + b_0, & t^{(0)} \leq t \leq t^{(1)}; \\ b_2 t + (b_1 - b_2)t^{(1)} + b_0, & t^{(1)} \leq t \leq t^{(2)}; \\ \dots \\ b_i t + \sum_{s=1}^{i-1} (b_s - b_{s+1})t^{(s)} + b_0, & t^{(i-1)} \leq t \leq t^{(i)}; \\ \dots \\ b_k t + \sum_{s=1}^{k-1} (b_s - b_{s+1})t^{(s)} + b_0, & t^{(k-1)} \leq t \leq t^{(k)}; \end{cases} \quad (2)$$

Якщо накласти умову, що  $S(t) = z(i)$ , то

$$b_i = \frac{z^{(i)} - z^{(i-1)}}{t^{(i)} - t^{(i-1)}} \quad (3)$$

$$S(t) = \frac{z^{(i)} - z^{(i-1)}}{t^{(i)} - t^{(i-1)}} (t - t^{(i-1)}) + z^{(i-1)}, \quad t^{(i-1)} \leq t \leq t^{(i)} \quad (4)$$

Тим самим, задача відтворення лінійної сплайн-регресії зводиться до інтерполювання лінійними сплайнами.

У загальному випадку  $E\{S(t_i)\} = z_i, \quad i = \overline{1, N}$ .

Якщо  $N > k$  та для  $t^{(i-1)} \leq t_s \leq t^{(i)}$  значення  $t_{p_i}, t_{q_i}$  нижньою та верхньою межами статистичних даних відрізка  $[t^{(i-1)}, t^{(i)}]$ , то з умови

$$\min_{z^{(i)}, i=1, k} \sum_{i=1}^k \sum_{e=p_i}^{q_i} (S(t_e) - z_e)^2 \quad (5)$$

З урахуванням виразу (3) отримуємо систему лінійних рівнянь:

$$\left\{ \begin{aligned} & \sum_{s=p_1}^{q_1} \left( \frac{t_s - t^{(0)}}{h_1} z^{(1)} + \frac{t^{(0)} - t_s}{h_1} z^{(0)} + z^{(0)} - z_s \right) \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right) = 0; \\ & \sum_{s=p_1}^{q_1} \left( \frac{t_s - t^{(0)}}{h_1} z^{(1)} + \frac{t^{(0)} - t_s}{h_1} z^{(0)} + z^{(0)} - z_s \right) \left( \frac{t_s - t^{(0)}}{h_1} + 1 \right) + \\ & \quad + \sum_{s=p_2}^{q_2} \left( \frac{t_s - t^{(1)}}{h_2} z^{(2)} + \frac{t^{(1)} - t_s}{h_2} z^{(1)} + z^{(1)} - z_s \right) \left( \frac{t^{(1)} - t_s}{h_2} + 1 \right) = 0; \\ & \dots \\ & \sum_{s=p_i}^{q_i} \left( \frac{t_s - t^{(i-1)}}{h_i} z^{(i)} + \frac{t^{(i-1)} - t_s}{h_i} z^{(i-1)} + z^{(i-1)} - z_s \right) \left( \frac{t_s - t^{(i-1)}}{h_i} + 1 \right) + \\ & \quad + \sum_{s=p_{i+1}}^{q_{i+1}} \left( \frac{t_s - t^{(i)}}{h_{i+1}} z^{(i+1)} + \frac{t^{(i)} - t_s}{h_{i+1}} z^{(i)} + z^{(i)} - z_s \right) \left( \frac{t^{(i)} - t_s}{h_{i+1}} + 1 \right) = 0; \\ & \dots \\ & \sum_{s=p_k}^{q_k} \left( \frac{t_s - t^{(k-1)}}{h_k} z^{(k)} + \frac{t^{(k-1)} - t_s}{h_k} z^{(k-1)} + z^{(k-1)} - z_s \right) \left( \frac{t_s - t^{(k-1)}}{h_k} + 1 \right) = 0; \end{aligned} \right. \quad (6)$$

В матричному представленні система має вигляд

$$\begin{bmatrix} a_{00} & a_{01} & \dots & a_{0k} \\ a_{10} & a_{11} & \dots & a_{1k} \\ \dots & \dots & \dots & \dots \\ a_{i0} & a_{i1} & \dots & a_{ik} \\ \dots & \dots & \dots & \dots \\ a_{k0} & a_{k1} & \dots & a_{kk} \end{bmatrix} \begin{bmatrix} z^{(0)} \\ z^{(1)} \\ \vdots \\ z^{(i)} \\ \vdots \\ z^{(k)} \end{bmatrix} = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_i \\ \vdots \\ g_k \end{bmatrix} \quad (7)$$

або

$$A \times Z = G, \quad (8)$$

де

$$a_{e,e} = \sum_{s=p_e}^{q_e} \left( \frac{t_s - t^{(e-1)}}{h_e} \right)^2 + \sum_{s=p_{e+1}}^{q_{e+1}} \left( \frac{t^{(e)} - t_s}{h_{e+1}} + 1 \right)^2, \quad e = \overline{1, k-1}$$

$$a_{0,0} = \sum_{s=p_1}^{q_1} \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right)^2; \quad a_{k,k} = \sum_{s=p_k}^{q_k} \left( \frac{t_s - t^{(k-1)}}{h_k} \right)^2;$$

$$a_{e+1,e} = \sum_{s=p_{e+1}}^{q_{e+1}} \left( \frac{t^{(e)} - t_s}{h_{e+1}} + 1 \right) \left( \frac{t_s - t^{(e)}}{h_{e+1}} \right), \quad e = \overline{0, k-1};$$

$$a_{e-1,e} = \sum_{s=p_e}^{q_e} \left( \frac{t_s - t^{(e-1)}}{h_e} \right) \left( \frac{t^{(e-1)} - t_s}{h_e} + 1 \right), \quad e = \overline{1, k};$$

$$a_{r,e} = 0, \quad \text{якщо } r \geq e + 2, \quad e \geq r + 2;$$

$$g_e = \sum_{s=p_e}^{q_e} z_s \left( \frac{t_s - t^{(e-1)}}{h_e} \right) + \sum_{s=p_{e+1}}^{q_{e+1}} z_s \left( \frac{t^{(e)} - t_s}{h_{e+1}} + 1 \right), \quad e = \overline{1, k-1};$$

$$g_0 = \sum_{s=p_1}^{q_1} z_s \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right); \quad g_k = \sum_{s=p_k}^{q_k} z_s \left( \frac{t_s - t^{(k-1)}}{h_k} \right); \quad h_e = t^{(e)} - t^{(e-1)}.$$

Окремим випадком є  $h_e = h = const$ .

З урахуванням випадкових складових значення  $z_s$  на проміжку  $[t^{(e-1)}, t^{(e)}]$  можна представити у вигляді

$$z_s = S(t_s) + \varepsilon_s = \frac{t_s - t^{(e-1)}}{h_e} z^{(e)} + \frac{t^{(e-1)} - t_s}{h_e} z^{(e-1)} + z^{(e-1)} + \varepsilon_s, \quad (9)$$

де  $E\{\varepsilon_s\} = 0, D\{\varepsilon_s\} = \sigma^2$

З урахуванням виразу (9) умова (5) прийме наступний вигляд:

$$\min_{z^{(e)}} \sum_{i=1}^k \sum_{s=p_i}^{q_i} \left( z_s - \frac{t_s - t^{(e-1)}}{h_e} z^{(e)} - \frac{t^{(e-1)} - t_s}{h_e} z^{(e-1)} - z^{(e-1)} \right)^2 = S_{\text{заял}}^2 \quad (10)$$

Це еквівалентно розв'язку системи

$$A \times \hat{Z} = G, \quad (11)$$

де  $\hat{Z}^T = (\hat{z}^{(0)}, \hat{z}^{(1)}, \dots, \hat{z}^{(k)})$ .

Отримавши розв'язок системи (11), визначають вектор оцінок параметрів лінійної сплайн-регресійної залежності  $\vec{Z}$  та з урахуванням (3) – вектор  $B$  оцінок  $\vec{B}$ . Спираючись на отримані оцінки  $\vec{Z}$ ,  $\vec{B}$ , проводять обчислення оцінок параметрів нелінійної сплайн-регресійної залежності з урахуванням формул перерахунку.

З системи (11) випливає, що

$$E\{z_s\} = \frac{t_s - t^{(e-1)}}{h_e} z^{(e)} + \left( \frac{t^{(e-1)} - t_s}{h_e} + 1 \right) z^{(e-1)} \quad (12)$$

та леми, що наведено нижче.

*Лема 1*

$$E\{g_r\} = \sum_{s=0}^k a_{rs} z^{(s)}.$$

*Доведення.*

Дійсно,

$$\begin{aligned} E\{g_0\} &= \sum_{s=p_1}^{q_1} \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right) E\{z_s\} = \\ &= z^{(0)} \sum_{s=p_1}^{q_1} \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right)^2 + z^{(1)} \left( \frac{t_s - t^{(0)}}{h_1} \right) \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right) = \sum_{s=0}^k a_{0s} z^{(s)}; \\ E\{g_r\} &= \sum_{s=p_r}^{q_r} \frac{t_s - t^{(r-1)}}{h_r} E\{z_s\} + \sum_{s=p_{r+1}}^{q_{r+1}} \left( \frac{t^{(r)} - t_s}{h_{r+1}} + 1 \right) E\{z_s\} = \\ &= z^{(r)} \sum_{s=p_r}^{q_r} \left( \frac{t_s - t^{(r-1)}}{h_r} \right)^2 + z^{(r-1)} \sum_{s=p_r}^{q_r} \left( \frac{t_s - t^{(r-1)}}{h_r} \right) \left( \frac{t^{(r-1)} - t_s}{h_r} + 1 \right) + \\ &+ z^{(r+1)} \sum_{s=p_{r+1}}^{q_{r+1}} \left( \frac{t_s - t^{(r)}}{h_{r+1}} \right) \left( \frac{t^{(r)} - t_s}{h_{r+1}} + 1 \right) + z^{(r)} \sum_{s=p_{r+1}}^{q_{r+1}} \left( \frac{t^{(r)} - t_s}{h_{r+1}} + 1 \right)^2, \end{aligned}$$

або

$$\begin{aligned} E\{g_r\} &= z^{(r-1)} \sum_{s=p_r}^{q_r} \left( \frac{t_s - t^{(r-1)}}{h_r} \right) \left( \frac{t^{(r-1)} - t_s}{h_r} + 1 \right) + \\ &+ z^{(r)} \left\{ \sum_{s=p_r}^{q_r} \left( \frac{t_s - t^{(r-1)}}{h_r} \right)^2 + \sum_{s=p_{r+1}}^{q_{r+1}} \left( \frac{t^{(r)} - t_s}{h_{r+1}} + 1 \right)^2 \right\} + \end{aligned} \quad (13)$$

$$+ z^{(r+1)} \sum_{s=p_{r+1}}^{q_{r+1}} \left( \frac{t^{(r)} - t_s}{h_{r+1}} + 1 \right) \left( \frac{t_s - t^{(r-1)}}{h_{r+1}} \right) = \sum_{s=0}^k a_{rs} z^{(s)}, \quad 0 < r < k$$

$$\begin{aligned} E\{g_k\} &= \sum_{s=p_k}^{q_k} \left( \frac{t_s - t^{(k-1)}}{h_k} \right) E\{z_s\} = z^{(k-1)} \sum_{s=p_k}^{q_k} \left( \frac{t_s - t^{(k-1)}}{h_k} \right) \left( \frac{t^{(k-1)} - t_s}{h_k} + 1 \right) + \\ &+ z^{(k)} \sum_{s=p_k}^{q_k} \left( \frac{t_s - t^{(k-1)}}{h_k} \right)^2 = \sum_{s=0}^k a_{ks} z^{(s)}. \end{aligned}$$

Тим самим лему доведено.

Отже,  $E\{g_r\} = \sum_{s=0}^k a_{rs} z^{(s)}$ ,  $0 \leq r \leq s$ . (14)

З урахуванням виразів (13), (14), має місце система

$$T \times Z = E\{G\} \tag{15}$$

з якої випливає

$$Z = T^{-1} \times E\{G\}.$$

З іншого боку, з системи (11) з урахуванням  $A=T$ , отримують

$$\hat{Z} = T^{-1} \times G.$$

Тоді математичне сподівання вектору  $\hat{Z}$  має вигляд

$$E\{\hat{Z}\} = E\{T^{-1} \times G\} = T^{-1} \times E\{G\} = TT^{-1} \times Z = Z,$$

звідки випливає  $E\{\hat{z}^{(r)}\} = z^{(r)}$ ,  $r = 0, 1, 2, \dots, k$ .

Лема 2. Оцінкою ковариації вільних членів матриці  $G$  є

$$\text{cov}[g_r, g_e] = a_{re} \sigma^2.$$

Доведення.

Базуючись на визначенні ковариації, отримують

$$\begin{aligned} \text{cov}[g_r, g_e] &= E\{(g_r - E\{g_r\})(g_e - E\{g_e\})\} = E\left\{\sum_{s=p_r}^{q_r} z_s \frac{t_s - t^{(r-1)}}{h_r} + \right. \\ &+ \sum_{s=p_{r+1}}^{q_{r+1}} (z_s) \left(\frac{t^{(r)} - t_s}{h_{r+1}} + 1\right) - \sum_{s=p_r}^{q_r} (z_s - \varepsilon_s) \left(\frac{t_s - t^{(r-1)}}{h_r}\right) - \\ &- \sum_{s=p_{r+1}}^{q_{r+1}} (z_s - \varepsilon_s) \left(\frac{t^{(r)} - t_s}{h_{r+1}} + 1\right) \times E\left\{\sum_{s=p_e}^{q_e} z_s \frac{t_s - t^{(e-1)}}{h_e} + \right. \\ &+ \sum_{s=p_{e+1}}^{q_{e+1}} z_s \left(\frac{t^{(e)} - t_s}{h_{e+1}} + 1\right) - \sum_{s=p_e}^{q_e} (z_s - \varepsilon_s) \frac{t_s - t^{(e-1)}}{h_e} - \\ &\left. - \sum_{s=p_{e+1}}^{q_{e+1}} (z_s - \varepsilon_s) \left(\frac{t^{(e)} - t_s}{h_{e+1}} + 1\right)\right\} = E\left\{\left[\sum_{s=p_r}^{q_r} \varepsilon_s \frac{t_s - t^{(r-1)}}{h_r} + \sum_{s=p_{r+1}}^{q_{r+1}} \varepsilon_s \left(\frac{t^{(r)} - t_s}{h_{r+1}} + 1\right)\right] \times \right. \\ &\left. \times \left[\sum_{s=p_e}^{q_e} \varepsilon_s \frac{t_s - t^{(e-1)}}{h_e} + \sum_{s=p_{e+1}}^{q_{e+1}} \varepsilon_s \left(\frac{t^{(e)} - t_s}{h_{e+1}} + 1\right)\right]\right\} \end{aligned} \tag{16}$$

З виразу (16) випливають наступні випадки:

1) якщо  $r \neq e$ ,  $r \neq e + 1$ ,  $r \neq e - 1$ , то  $\text{cov}[g_r, g_e] = a_{re} \sigma^2 = 0$ ;

2) якщо  $r = e$ , то

$$\text{cov}[g_e, g_e] = E\left\{\left[\sum_{s=p_e}^{q_e} \varepsilon_s \frac{t_s - t^{(e-1)}}{h_e}\right]^2 + \left[\sum_{s=p_{e+1}}^{q_{e+1}} \varepsilon_s \left(\frac{t^{(e)} - t_s}{h_{e+1}} + 1\right)\right]^2\right\} = a_{ee} \sigma^2;$$

3) якщо  $r = e + 1$ , то  $\text{cov}[g_{e+1}, g_e] = a_{e+1,e} \sigma^2$ ;

4) якщо  $r = e - 1$ , то  $\text{cov}[g_{e-1}, g_e] = a_{e-1,e} \sigma^2$ .

Це доводить лему.

Теорема.

Якщо  $S_0^2$  - мінімум суми відхилень оцінки лінійного сплайну відносно емпіричних даних,  $N$  - загальна кількість варіант,  $k$  - кількість вузлів, тоді незсунена оцінка дисперсії  $\sigma^2$  має вигляд

$$D = E\{S_0^2\} / (N - k - 1).$$

Доведення.

Вважаємо, що значення  $S_0^2$  визначається за виразом

$$\begin{aligned} S_0^2 &= \sum_{i=1}^k \sum_{s=p_i}^{q_i} \left[ z_s - \frac{t_s - t^{(i-1)}}{h_i} z^{(i)} - \frac{t^{(i-1)} - t_s}{h_i} z^{(i-1)} - z^{(i-1)} \right]^2 = \\ &= \sum_{i=1}^k \sum_{s=p_i}^{q_i} z_s^2 - 2 \sum_{r=0}^k g_r z^{(r)} + \sum_{r=0}^k \sum_{s=0}^k a_{rs} z^{(r)} z^{(s)} = \\ &= \sum_{i=1}^k \sum_{s=p_i}^{q_i} z_s^2 - \sum_{r=0}^k g_r z^{(r)} \end{aligned} \quad (17)$$

Подальше доведення зводиться до знаходження математичних сподівань елементів правої частини виразу. При цьому реалізується оцінка

$$\begin{aligned} \{\text{cov}[z^{(r)}, z^{(s)}]\} &= \{[(z^{(r)} - z^{(r)})(z^{(s)} - z^{(s)})]\} = \{(\hat{Z} - Z)(\hat{Z} - Z)^T\} = \\ &= E\{T^{-1}(g - E\{g\})(g - E\{g\})^T T^{-1}\} = T^{-1} E\{(g - E\{g\})(g - E\{g\})^T\} T^{-1} = \\ &= T^{-1} (T\sigma^2) T^{-1} = \{d_{r,s}\} \sigma^2, \end{aligned}$$

де  $\{d_{r,s}\} = D = T^{-1}$ .

З урахуванням лем 1 та 2 визначають шукані математичні сподівання:

$$\begin{aligned} E \left\{ \sum_{i=1}^k \sum_{s=p_i}^{q_i} z_s^2 \right\} &= \sum_{i=1}^k \sum_{s=p_i}^{q_i} [D\{z_s\} + (E\{z_s\})^2] = \sum_{i=1}^k \sum_{s=p_i}^{q_i} \left[ \sigma^2 + \left( \frac{t_s - t^{(i-1)}}{h_i} z^{(i)} + \right. \right. \\ &+ \left. \left. \left( \frac{t^{(i-1)} - t_s}{h_i} + 1 \right) z^{(i-1)} \right)^2 \right] = N\sigma^2 + \sum_{i=1}^k \sum_{s=p_i}^{q_i} \left[ z^{(i)2} \left( \frac{t_s - t^{(i-1)}}{h_i} \right)^2 + z^{(i-1)2} \times \right. \\ &\times \left. \left( \frac{t^{(i-1)} - t_s}{h_i} + 1 \right)^2 + 2z^{(i)} z^{(i-1)} \left( \frac{t_s - t^{(i-1)}}{h_i} \right) \left( \frac{t^{(i-1)} - t_s}{h_i} + 1 \right) \right] = N\sigma^2 + \\ &+ z^{(0)2} \sum_{s=p_1}^{q_1} \left( \frac{t^{(1)} - t_s}{h_1} + 1 \right)^2 + z^{(0)} z^{(1)} \sum_{s=p_1}^{q_1} \left( \frac{t_s - t^{(0)}}{h_1} \right) \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right) + \\ &+ z^{(0)} z^{(2)} \cdot 0 + \dots + z^{(0)} z^{(k)} \cdot 0 + z^{(1)} z^{(0)} \sum_{s=p_1}^{q_1} \left( \frac{t^{(0)} - t_s}{h_1} + 1 \right) \left( \frac{t_s - t^{(0)}}{h_1} \right) + z^{(1)2} \times \\ &\times \left[ \sum_{s=p_1}^{q_1} \left( \frac{t_s - t^{(0)}}{h_1} \right)^2 + \sum_{s=p_2}^{q_2} \left( \frac{t^{(1)} - t_s}{h_2} + 1 \right)^2 \right] + z^{(1)} z^{(2)} \sum_{s=p_2}^{q_2} \left( \frac{t_s - t^{(1)}}{h_2} \right) \times \\ &\times \left( \frac{t^{(1)} - t_s}{h_2} + 1 \right) + z^{(1)} z^{(3)} \cdot 0 + \dots + z^{(1)} z^{(k)} \cdot 0 + \dots = N\sigma^2 + \\ &+ \sum_{r=0}^k \sum_{s=0}^k a_{rs} z^{(r)} z^{(s)}; \end{aligned}$$

$$\begin{aligned}
 E\left\{\sum_{r=0}^k g_r \hat{z}^{(r)}\right\} &= E\{\hat{Z}^T G\} = E\{\hat{Z}^T T \hat{Z}\} = E\left\{\sum_{r=0}^k \sum_{\vartheta=0}^k \hat{z}^{(r)} a_{r\vartheta} \hat{z}^{(\vartheta)}\right\} = \sum_{r=0}^k \sum_{\vartheta=0}^k a_{r\vartheta} E\{\hat{z}^{(r)} \hat{z}^{(\vartheta)}\} = \\
 &= \sum_{r=0}^k \sum_{\vartheta=0}^k a_{r\vartheta} [z^{(r)} z^{(\vartheta)} + \text{cov}[z^{(r)} z^{(\vartheta)}]] = \sum_{r=0}^k \sum_{\vartheta=0}^k a_{r\vartheta} z^{(r)} z^{(\vartheta)} + \sum_{r=0}^k \sum_{\vartheta=0}^k a_{r\vartheta} a_{r\vartheta} \sigma^2 = \\
 &= \sum_{r=0}^k \sum_{\vartheta=0}^k a_{r\vartheta} z^{(r)} z^{(\vartheta)} + (k+1)\sigma^2
 \end{aligned}$$

Отже, має місце результат

$$E\{S_0^2\} = E\left\{\sum_{i=1}^k \sum_{s=p_i}^{q_i} z_s^2\right\} - E\left\{\sum_{r=0}^k g_r \hat{z}^{(r)}\right\} = (N - k - 1)\sigma^2.$$

Тим самим теорему доведено.

*Наслідок.*

При заданому математичному сподіванні точність наближенні  $E\{S_2^0\}$  та дисперсії похибки заміни  $\sigma^2$  емпіричних даних лінійними сплайнами кількість вузлів визначається як

$$k = \frac{(N+1)\sigma^2 - E\{S_2^0\}}{\sigma^2}$$

або

$$k = (N + 1) - E\{S_2^0\} / \sigma^2$$

Більш складні обчислювальні схеми мають місце для поліноміальних сплайн-регресійних моделей степеню  $n \geq 2$ .

Проведемо сплайн-регресійний аналіз даних моделювання викиду для джерела V-109, розташованого на цій території промислового майданчика ДНВП «Цирконій» м. Дніпродзержинська за віссю факела викиду в ГІС «AirNorm». Розглянемо випадок використання сплайн-регресійної залежності типу «пряма-пряма» та «пряма-парабола». Вузол склеювання відомий.

Візуально бачимо, що дані розділились навпіл, тобто точка склеювання буде лежати десь ближче до середини. Тому обрано значення  $x_0=1$  та побудуємо залежність вигляду «пряма-пряма» (рис. 1, 2).



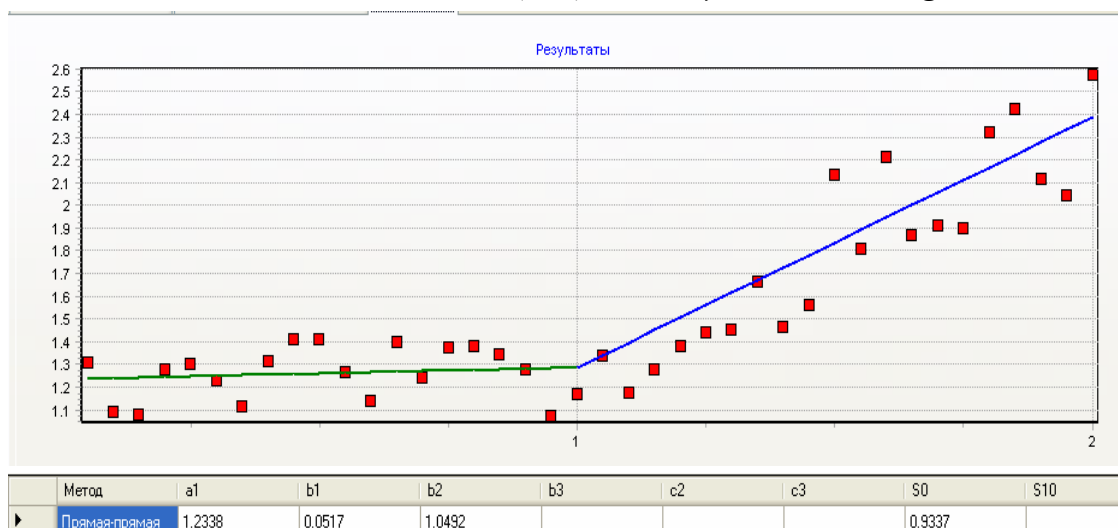


Рисунок 1 – Дані побудови сплайн-регресійної залежності «пряма-пряма» викиду  $NH_3$  для джерела V-109

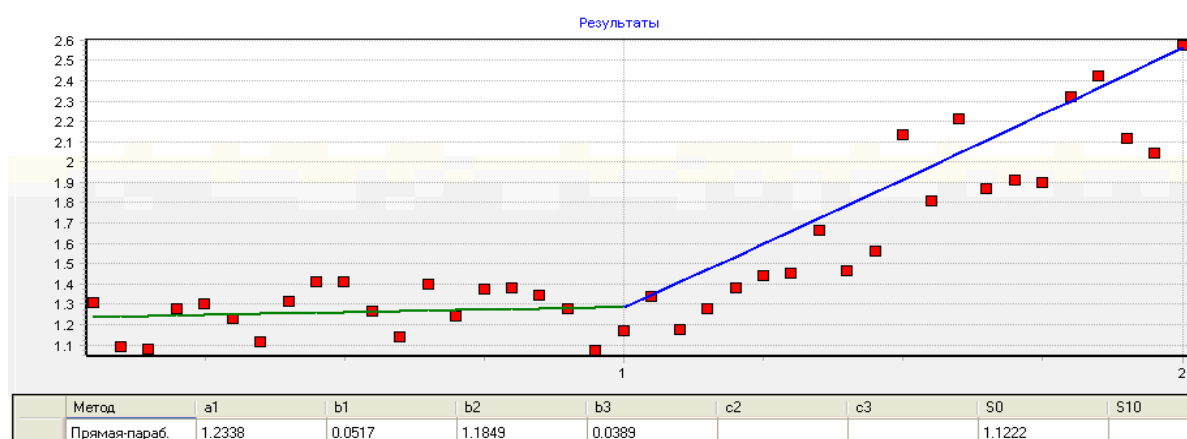


Рисунок 2 – Дані побудови сплайн-регресійної залежності «пряма-парабола» викиду  $NH_3$  для джерела V-109

Для з'ясування точності візуального пошуку точки склеювання, побудуємо залежності «пряма-пряма» та «пряма-парабола», коли точка склеювання  $x_0$  – невідома. Для типу «пряма-пряма» знайдено точку  $x_0=1.1$ , для типу «пряма-парабола» –  $x_0=1.15$ .

**Оцінка результатів.** Проведено порівняння результатів автоматичного пошуку точки склеювання та візуальних (рис.3).

Ґрунтуючись на значеннях залишкової дисперсії, можна зазначити, що результати автоматичного пошуку більш адекватні, причому в обох випадках.

Метод	a1	b1	b2	b3	c2	c3	S0	S10
▶ Прямая-прямая	1.2338	0.0517	1.0492				0.9337	
Прямая-параб.	1.2338	0.0517	1.1849	0.0389			1.1222	
Прямая-прямая	1.2403	0.0353	1.2421				0.8311	
Прямая-параб.	1.2405	0.0347	1.6125	-0.3992			0.8182	

Рисунок 3 – Дані порівняння результатів автоматичного та візуального пошуку точки склеювання

### ЛІТЕРАТУРА / LITERATURE

1. Бабак В.П. Статистична обробка даних / В.П. Бабак, А.Я. Білецький, О.П. Приставка, П.О. Приставка – Київ: МІВВЦ, 2001. – 312 с.
2. Белов И.В. Сравнение моделей распространения загрязнений в атмосфере / И.В. Белов, М.С. Беспалов, Л.В. Клочкова, Н.К. Павлова, Д.В. Сузан, В.Ф. Тишкин – М: Математическое моделирование, 1999. – Т.11, №8.– С.52–64.
3. Беляев Н.Н. Защита атмосферы от загрязнения при миграции токсических веществ / Н.Н. Беляев, В.М. Лисняк – Д: РВВ ДНУ, 2006. – 105с.
4. Берлянд М. Е. Прогноз и регулирование загрязнения атмосферы / М.Е. Берлянд – Л.: Гидрометеоиздат, 1985. – 273.
5. Полонська А.Є. Сплайн-регресія при обробці експериментальних даних / А.Є. Полонська, О.П. Приставка // Математичне моделювання. – Дніпродзержинськ, 2007. – Т. 2(17). – С. 46 – 49.
6. Митько Л.А. Компьютерное моделирование физико-химических процессов в приземном слое атмосферы – необходимый компонент информационного обеспечения в системе эколого-гигиенической безопасности [Электронный ресурс] / А.Л. Митько – Киев: Институт проблем моделирования в энергетике НАН Украины. – Режим доступа до ресурсу: <http://www.health.gov.ua/publ/conf.nsf/d2ffb20b75bd2a6cc2256d8f0029469d/08450493dfb77a1ac2256dc600401e3e?OpenDocument>
7. Приставка П.О. Поліноміальні сплайни при обробці даних: Монографія. / П.О. Приставка. – Д.: Вид-во Дніпропетр. ун-ту, 2004. – 236 с.
8. Сорока А.И. Стохастическая модель расчета рассеивания загрязняющих веществ в атмосферном воздухе / А.И. Сорока, А.Н. Тетельбаум // Инженерная экология. – М.: "Инженерная экология", 2001. – №2. – С.51–56.

### REFERENCES

1. Babak V.P. Statistical data processing / V.P. Babak, A.Y. Biletsky, O.P. Pristavka, P.O. Pristavka – Kyiv: MIVVC, 2001. – 312 p.

2. Belov I.V. Comparison of models of pollution distribution in the atmosphere / I.V. Belov, M.S. Bepalov, L.V. Klochkova, N.K. Pavlova, D.V. Susan, W.F. Tishkin – M: Mathematical modeling, 1999. – Vol.11, №8.–P.52–64.
3. Belyaev N.N. Protection of the atmosphere from pollution during the migration of toxic substances / NN Belyaev, B.M. Lisnyak – D: RVV DNU, 2006. – 105с.
4. Berland M.E. Forecast and regulation of atmospheric pollution / M.E. Berland – L. : Gidrometeoizdat, 1985. – 273 p.
5. Polonska A.Y. Spline regression in experimental data processing / A.Y. Polonska, O.P. Pristavka // Mathematical modeling. - Dniprodzerzhynsk, 2007. - Vol. 2 (17). - P. 46 - 49.
6. Mitko L.A. Computer modeling of physicochemical processes in the surface layer of the atmosphere - a necessary component of information support in the system of ecological and hygienic safety [Electronic resource] / L.A. Mitko – Kiev: Institute of Modeling Problems in Energy of the National Academy of Sciences of Ukraine. – Resource access mode:  
<http://www.health.gov.ua/publ/conf.nsf/d2ffb20b75bd2a6cc2256d8f0029469d/08450493dfb77a1ac2256dc600401e3e?OpenDocument>
7. Pristavka P.O. Polynomial splines in data processing: Monograph. / P.O. Pristavka – D. : Dnipropetrovsk Publishing House. University, 2004. – 236 p.
8. Soroka A.I. Stochastic model of calculation of scattering of pollutants in atmospheric air / A.I. Soroka, A.N. Tetelbaum // Engineering ecology. - M. : "Engineering Ecology", 2001. - №2. - P.51–56.

Received 29.03.2021.

Accepted 31.03.2021.

***Информационная технология оценки параметров сплайн-регрессии  
при обработке данных загрязнения атмосферного воздуха в ГИС «AIRNORM»***

*Особенностью задач, связанных с процессом распространения вредных веществ в природной среде, является наличие большого количества параметров, которые влияют на процесс миграции примесей. Для описания таких процессов используются математические модели различной сложности. Задача мониторинга воздушной среды промышленными предприятиями заключается в сборе, обработке параметров загрязнения атмосферы на локальной территории и выработке по результатам выводов относительно принятия решений об экологическом состоянии как промплощадки, так и прилегающих жилых территорий. Приведен усовершенствованный метод аппроксимации данных при использовании кусочно-полиномиальной регрессии. Предложенный алгоритм позволяет повысить адекватность определения границ при резком изменении концентрации вредного вещества, позволяет повысить точность построенных моделей выброса вещества в атмосферный воздух и более точно оценить масштабы загрязнения. Разработанные методы, позволяющие провести визуализацию полученных данных для повышения информативности при оценке загрязнения определенной территории.*

Наряду с математическими моделями при оценке уровня загрязнения используются геоинформационные системы (ГИС), как мощный инструмент, который обеспечивает: сбор, хранение, обработку, отображение данных, анализ, оценку, прогноз состояния окружающей среды территориальной общины. Описанные в статье методы и алгоритмы положены в основу информационной технологии и программного обеспечения решения научно-технической задачи визуально-аналитического анализа данных загрязнения атмосферного воздуха в ГИС «AirNorm». С помощью данной ГИС получены и представлены результаты проведения вычислительного эксперимента проведенного по показателям уровня содержания выброса вещества  $NH_3$  для источника V-109 на Государственном научно-производственном предприятии (ГНПП) «Цирконий» м.Каменское.

***Information technology for evaluating spline regression parameters  
when processing data on air pollution in the GIS «AIRNORM»***

*A feature of the tasks associated with the spread of harmful substances in the natural environment is the presence of a large number of parameters that affect the migration of impurities. Mathematical models of varying complexity are used to describe such processes. The task of air monitoring by industrial enterprises is to collect, process the parameters of atmospheric pollution in the local area and develop, based on the results, conclusions regarding decision-making on the ecological state of both the industrial site and adjacent residential areas. An improved method of data approximation using piecewise polynomial regression is presented. The proposed algorithm makes it possible to increase the adequacy of determining the boundaries with a sharp change in the concentration of a harmful substance, improves the accuracy of the constructed models of the release of a substance into the atmospheric air and more accurately assess the scale of pollution. The developed methods allow visualization of the obtained data to increase the information content in assessing the pollution of a certain area.*

*Along with mathematical models, when assessing the level of pollution, geographic information systems (GIS) are used as a powerful tool that provides: collection, storage, processing, display of data, analysis, assessment, forecast of the state of the environment of a territorial community. The methods and algorithms described in the article form the basis of information technology and software for solving the scientific and technical problem of visual and analytical analysis of atmospheric air pollution data in the GIS «AirNorm». With the help of this GIS, the results of a computational experiment carried out on the indicators of the level of emission of  $NH_3$  for the source V-109 at the State Research and Production Enterprise (GNPP) «Zirkoniy» m. Kamenskoye were obtained and presented.*

**Полонська Анна Євгенівна** – асистент, Дніпровський національний університет імені Олеся Гончара.

**Полонская Анна Евгеньевна** – ассистент, Днепровский национальный университет имени Олеся Гончара.

**Polonska Anna** – assistant, Oles Honchar Dnipro National University.