

УДК 004.02

Домашовець О. В.

аспірант

кафедра загальної екології та екоінформаційних систем
Національний університет «Львівська політехніка»

Казимира І. Я.

кандидат технічних наук, доцент

кафедра загальної екології та екоінформаційних систем
Національний університет «Львівська політехніка»

Домашовець О. В.

аспірант

кафедра общей экологии и экоинформационных систем
Национальный университет «Львовская политехника»

Казимира И. Я.

кандидат технических наук, доцент

кафедра общей экологии и экоинформационных систем
Национальный университет «Львовская политехника»

Domashovets O.V.

Postgraduate Student

Department of General Ecology and Eco-Information Systems
Lviv Polytechnic National University

Kazymyra I. Y.

Ph.D., Associate Professor

Department of General Ecology and Eco-Information Systems
Lviv Polytechnic National University

**КЛАСТЕРНИЙ ПІДХІД У ДОСЛІДЖЕННІ ЗАБРУДНЕННЯ
АТМОСФЕРНОГО ПОВІТРЯ (НА ПРИКЛАДІ ЛЬВІВСЬКОЇ
ОБЛАСТІ)**

**КЛАСТЕРНЫЙ ПОДХОД В ИССЛЕДОВАНИИ ЗАГРЯЗНЕНИЯ
АТМОСФЕРНОГО ВОЗДУХА (НА ПРИМЕРЕ ЛЬВОВСКОЙ
ОБЛАСТИ)**

**THE CLUSTER APPROACH FOR THE INVESTIGATION OF AIR
POLLUTION (BY EXAMPLE OF LVIV REGION)**

Анотація. Пропонується застосувати двоетапний кластерний підхід до дослідження динаміки викидів найпоширеніших забруднюючих речовин в атмосферне повітря Львівської області. Описано математичну модель. Виконано програмну реалізацію запропонованого підходу у середовищі Matlab 7.10.0 та представлено результати.

Ключові слова: кластеризація, алгоритм k-середніх, карта Кохонена, що самоорганізується, атмосферне повітря, викиди забруднюючих речовин.

Аннотация. Предлагается использовать двухэтапный кластерный подход к исследованию динамики выбросов наиболее распространенных загрязняющих веществ в атмосферный воздух Львовской области. Описана математическая модель. Выполнена программная реализация предложенного подхода в среде Matlab 7.10.0 и представлены результаты.

Ключевые слова: кластеризация, алгоритм k-средних, самоорганизующаяся карта Кохонена, атмосферный воздух, выбросы загрязняющих веществ.

Summary. We propose to apply the two-stage cluster approach for the investigation of the dynamics in emissions of the most common pollutants into the atmosphere in Lviv region. The mathematical model was described. The implementation of the proposed approach was performed in the Matlab 7.10.0 environment and the results were presented.

Key words: clustering, k-means algorithm, Kohonen's self-organizing maps, the atmosphere air, emissions of pollutants.

Постановка проблеми. Екологічна ситуація на планеті з кожним роком ускладнюється. Це пов'язано з високою концентрацією промислового виробництва, розвиненою транспортною інфраструктурою, появою нових технологічних процесів, тощо. Все це призводить до значного забруднення довкілля. Це підкреслює актуальність проблематики розроблення та застосування алгоритмів, засобів та інформаційних

технологій для проведення ефективного, точного і достовірного оцінювання стану якості навколишнього середовища.

Атмосфера – одна з найважливіших складових довкілля. Вона надійно захищає живі організми від космічного й ультрафіолетового випромінювання, визначає загальний тепловий режим поверхні Землі, впливає на кліматичні умови, регулює кількість сонячної енергії, необхідної для життя. Спостереження за змінами стану атмосфери є надзвичайно важливим, оскільки допомагає оцінити можливі наслідки цих змін та вчасно реагувати на них [1].

У роботі представлено один із підходів до аналізу динаміки викидів найпоширеніших забруднюючих речовин в атмосферне повітря, який ґрунтується на застосуванні методу кластеризації до даних, що описують викиди від стаціонарних та пересувних джерел.

Аналіз останніх досліджень та публікацій. Мета кластерного аналізу полягає в пошуку наявних структур, що виражається в утворенні груп схожих між собою об'єктів – кластерів. Водночас його дія полягає й у привнесенні структури в досліджувані об'єкти. Отже, методи кластеризації необхідні для виявлення структури в даних, яку нелегко знайти при візуальному обстеженні або за допомогою експертів [6].

Теоретичним аспектам застосування кластерного аналізу присвячені наукові праці багатьох вітчизняних і закордонних вчених, зокрема Н. Бидголі, Н.А. Мікула, М.І. Мельник, М. Портер, Т.Харріс, С.Л. Шульц, Л. Янг та ін. ([2], [6], [8], [11]). Ці та інші автори сформували математичну базу для використання кластерного аналізу в різних галузях.

Аналіз літературних джерел показав, що в багатьох країнах світу вже існує позитивний досвід застосування можливостей кластерного аналізу при дослідженні параметрів довкілля, а в Україні він знаходиться на ранній стадії застосування.

Мета дослідження: застосування кластерного підходу для оцінювання тенденції зміни забруднення атмосферного повітря від

стаціонарних та пересувних джерел у Львівській області за даними 2013-2015 років. Вхідні дані отримано з офіційного сайту Регіональної статистики Львівщини [4] та Регіональної доповіді про стан навколишнього природного середовища у Львівській області в 2015 році [3]. У таблиці 1 наведено перелік використаних у дослідженні показників найпоширеніших забруднюючих речовин.

Таблиця 1

Показники викидів забруднюючих речовин

x_1^t – викиди забруднюючих речовин в атмосферу стаціонарними джерелами забруднення у розрахунку на 1 км^2 , т.
x_2^t - викиди діоксид сірки (SO_2) в атмосферу стаціонарними джерелами забруднення, т.
x_3^t - викиди оксид вуглецю (CO) в атмосферу стаціонарними джерелами забруднення, т
x_4^t - викиди діоксид азоту (NO_2) в атмосферу стаціонарними джерелами забруднення, т
x_5^t - викиди метану (CH_4) в атмосферу стаціонарними джерелами забруднення, т
x_6^t - викиди неметанових летких органічних сполук в атмосферу стаціонарними джерелами забруднення, т
x_7^t - викиди речовини у вигляді суспендових твердих частинок, т
x_8^t - щільність викидів забруднюючих речовин в атмосферу автотранспортом на 1 км^2 , т
x_9^t - викиди CO в атмосферу автотранспортом, т
x_{10}^t - викиди CH_4 в атмосферу автотранспортом, т
x_{11}^t - викиди NO_2 в атмосферу автотранспортом, т.

Опис кластерного підходу. Дослідження проводиться двоетапним методом, який передбачає послідовне застосування карт Кохонена, що самоорганізуються (KSOM), та алгоритму k-середніх (Рис.1). KSOM дає можливість зменшити вимірність багатовимірних вхідних даних та відобразити кластери на карті, тому його обрано для попередньої обробки даних – зменшення розмірності, перегрупування та візуалізації. Основними перевагами даного методу є економія часу в обчисленнях великих і складних наборів даних [9] та отримання точніших результатів кластеризації [11], [9].

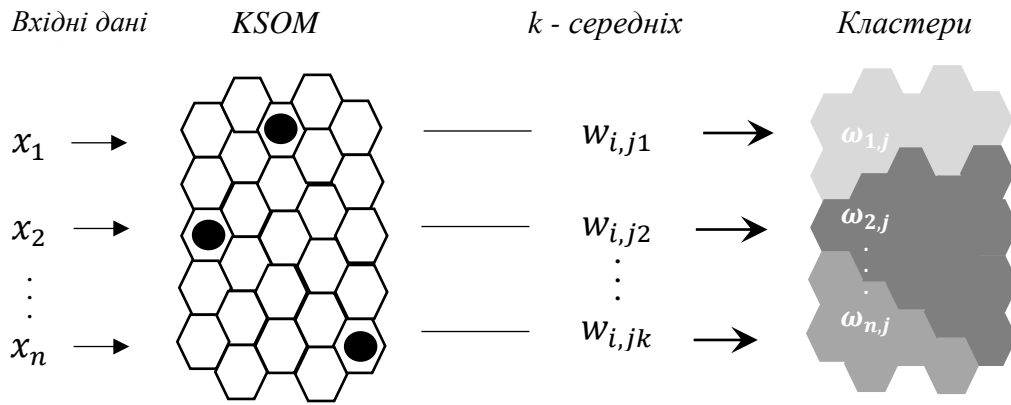


Рис.1. Застосування двоетапного підходу до кластеризації

(складено авторами на основі [7]).

Таким чином, пропонується застосувати двоетапний підхід для аналізу динаміки зміни забруднення атмосферного повітря.

На основі [8] розроблена така матриця вхідних даних P:

$$\mathbf{P} = \begin{matrix} o_1^t & \begin{matrix} x_1^t & \dots & x_k^t & \dots & x_m^t \\ x_{1,1}^t & \dots & x_{1,k}^t & \dots & x_{1,m}^t \\ \dots & \dots & \dots & \dots & \dots \\ o_i^t & \begin{matrix} x_{i,1}^t & \dots & x_{i,k}^t & \dots & x_{i,m}^t \\ \dots & \dots & \dots & \dots & \dots \\ o_n^t & \begin{matrix} x_{n,1}^t & \dots & x_{n,k}^t & \dots & x_{n,m}^t \end{matrix} \end{matrix} \end{matrix} \omega_{i,j}^t \\ \omega_{1,j}^t \\ \dots \\ \omega_{i,j}^t \\ \dots \\ \omega_{n,j}^t \end{matrix} \quad (1)$$

де $o_i^t \in O, O = \{o_1^t, o_2^t, \dots, o_i^t, \dots, o_n^t\}$ - об'єкти (райони) в часі t,

x_k^t - k-й параметр в часі t,

$x_{i,k}^t$ - значення параметра x_k^t для i-го об'єкту $o_i^t \in O$,

$\omega_{i,j}^t$ - j-й клас визначений для i-го об'єкту $o_i^t \in O$,

$p_i^t = (x_{i,1}^t, x_{i,2}^t, \dots, x_{i,k}^t, \dots, x_{i,m}^t)$ - i-й вхідний образ,

$x^t = (x_1^t, x_2^t, \dots, x_k^t, \dots, x_m^t)$ - вектор параметрів.

Коротко розглянемо принцип роботи даного методу. KSOM [11] базується на конкурентному принципі навчання, складається з вхідного шару (кількість нейронів відповідає кількості ознак об'єкта) та вихідного або конкурентного (кількість нейронів відповідає кількості кластерів). Вектор вхідних даних $p_i^t = (x_{i,1}^t, x_{i,2}^t, \dots, x_{i,k}^t, \dots, x_{i,m}^t)$ розподіляється на

вхідному шарі карти. Далі обчислюються відстані d_j між вхідним вектором p_i^t і векторами ваг $w_{i,j}$ для усіх нейронів конкурентного шару:

$$d_j = \sum_{i=1}^n (p_i^t - w_{i,j})^2 \quad (2)$$

де j проходить s нейронів конкурентного шару, $j=1,2,\dots,s$, p_i^t - i -й образ, $i=1,2,\dots, n$, $w_{i,j}$ - синаптичні ваги. Вибирається нейрон переможець j^* (ВМУ), для якого відстань d_j є мінімальною. Коли $w_{i,j}$ визначені, кожному вектору p_i^t присвоюється w_{i,j^*} нейрон переможець.

Після того, як усі ВМУ знайдені, відбувається адаптація синаптичних ваг $w_{i,j}$. Принцип алгоритму послідовного навчання полягає у тому, що w_{i,j^*} нейрони переможці та їх топологічні сусіди наближаються до фактичного вхідного вектора p_i відповідно до співвідношення:

$$w_{i,j}(t' + 1) = w_{i,j}(t') + \eta(t') \times h(j^*, j) \times (p_i^t(t') - w_{i,j}(t')), \quad (3)$$

де $\eta(t') \in (0,1)$ — швидкість навчання, функція Гауса, яка визначається як

$$h(j^*, j) = e^{-\left(\frac{d_E^2(j^*, j)}{\lambda^2(t')}\right)} \quad (4)$$

де $h(j^*, j)$ - функція сусідства, $d_E^2(j^*, j)$ - Евклідова відстань нейронів j^* та j на площині, $\lambda(t)$ - розмір околу в момент часу t' .

Навчання проводиться доки карта ознак стабілізується, тобто значення вагових коефіцієнтів перестануть змінюватися.

На другому етапі отримані топологічні координати KSOM передаються на вхід алгоритму k -середніх. Вхідна множина розбивається на K кластерів, та випадково обираються k об'єктів, в якості початкових центрів груп. Решту об'єктів присвоюються цим кластерам на основі мінімальної Евклідової відстані до центру:

$$d_j = \sqrt{\sum_{i=1}^n (k_i - w_{i,j^*})^2} \quad (5)$$

де k_i - значення центра кластера, w_{i,j^*} - значення нейронів переможців.

Далі розраховується середнє значення кожного кластера для знаходження нових центрів. Алгоритм працює доти, поки центри груп не перестануть змінюватися.

Перевагами такого підходу є простота та швидкість використання, зрозумілість і прозорість алгоритму. Недоліками є чутливість до вибору початкових значень центрів кластерів та труднощі, пов'язані з визначенням прогнозованої кількості кластерів.

Виклад основних результатів. Здійснено кластеризацію районів Львівської області за показниками викидів найпоширеніших забруднюючих речовин у атмосферне повітря. Для отримання результатів використано такі програмні засоби: інструментарій SOM Toolbox, середовище Matlab 7.10.0 [5].

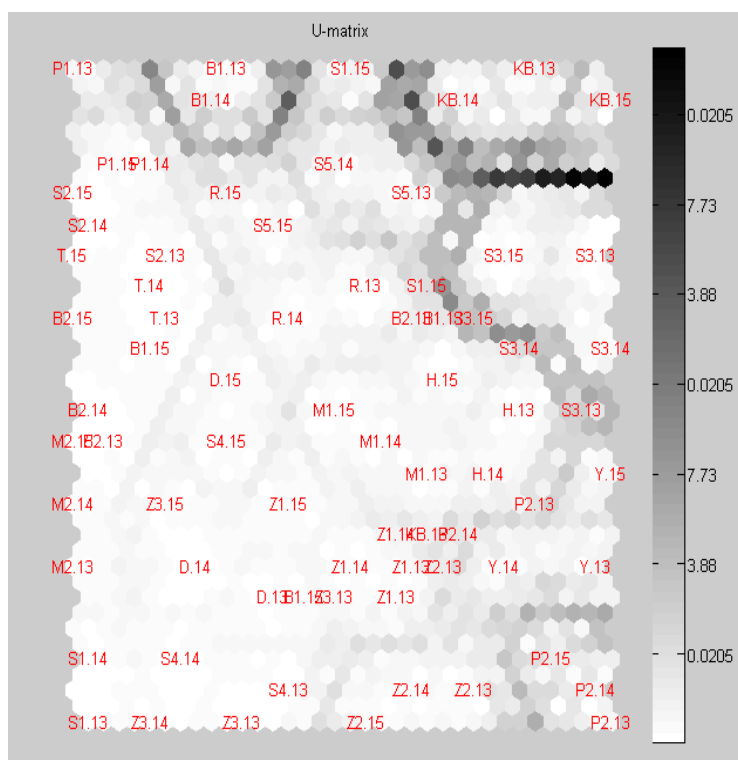
На рис.2 представлена KSOM у вигляді U-матриці, яка показує область Евклідових відстаней d_j між $w_{i,j}$ та відображає багатовимірні вхідні дані на двовимірній карті. На отриманій карті проводилась кластеризація районів алгоритмом k-середніх. Виявлено чотири групи з різними показниками стану якості атмосферного повітря (див. рис.3).

За розташуванням районів на карті можна оцінити динаміку зміни кількості викидів забруднюючих речовин в атмосферу. Ті що, знаходяться внизу карти, характеризуються найбільшою кількістю викидів, тобто гіршим станом якості повітря. Верхня частина карти, відповідно, показує райони з найменшою концентрацією шкідливих домішок.

Наприклад, Миколаївський район впродовж трьох років належить до трьох різних кластерів. Причому, в межах досліджуваного періоду, спостерігається тенденція до зниження кількості викидів (значно зменшилися викиди SO_2 та NO_2).

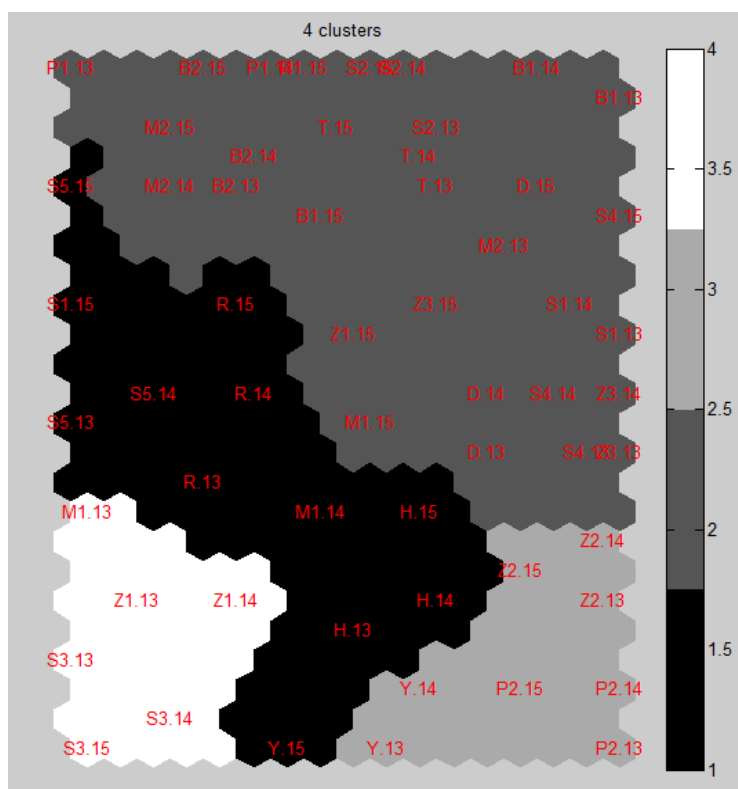
Найбільший відсоток забруднення атмосферного повітря області спостерігається у Сокальському, Жидачівському, Миколаївському, а також у Кам'янка-Бузькому районах. В одну групу їх об'єднує значне збільшення

у 2015 році викидів неметанових летких органічних сполук від стаціонарних джерел забруднення.



B1 - Бродівський
B2 - Буський
H - Городоцький
D - Дрогобицький
Z1 - Жидачівський
Z2 - Жовківський
Z3 - Золочівський
KB - Кам'янка-Бузький
M1 - Миколаївський
M2 - Мостиський
P1 - Перемишлянський
P2 - Пустомитівський
R - Радехівський,
S1 - Самбірський
S2 - Сколівський
S3 - Сокальський,
S4 - Старосамбірський
S5 - Стрийський
T - Турківський
Y - Яворівський
Позначки .13, .14, .15 –
відповідають 2013, 2014,
2015 рокам)

Рис.2. U-матриця Евклідових відстаней (1-ий етап).



B1 - Бродівський
B2 - Буський
H - Городоцький
D - Дрогобицький
Z1 - Жидачівський
Z2 - Жовківський
Z3 - Золочівський
KB - Кам'янка-Бузький
M1 - Миколаївський
M2 - Мостиський
P1 - Перемишлянський
P2 - Пустомитівський
R - Радехівський,
S1 - Самбірський
S2 - Сколівський
S3 - Сокальський,
S4 - Старосамбірський
S5 - Стрийський
T - Турківський
Y - Яворівський
Позначки .13, .14, .15 –
відповідають 2013, 2014,
2015 рокам)

Рис.3. Результати кластеризації алгоритмом k-середніх (2-ий етап).

Пустомитівський, Яворівський та Жовківський райони (кластер 3) характеризуються збільшенням викидів шкідливих речовин (SO₂ та CO) в атмосферу стаціонарними джерелами забруднення та зменшенням викидів від пересувних засобів.

У Перемишлянському, Турківському, Сколівському та інших районах (кластер 2) спостерігається найменша кількість викидів забруднюючих речовин і зміни стану атмосферного повітря є незначними.

Числовими характеристиками якості отриманих результатів є помилка квантування (QE) і топографічна помилка (TE) [9].

$$qe_j = \sum_{i=1}^{n_c} \left| |w_j - p_i^t| \right| \quad (6)$$

За параметром QE (6) оцінимо якість апроксимації вхідних даних нейронною мережею після навчання. Отримане значення QE = 0,044.

За значенням топографічної помилки оцінимо якість впорядкування нейронів у ґратці, тобто невідповідність між розташуванням нейронів нейромережевої ґратки та особливостями вхідних даних. Менше значення цього параметру свідчить про кращу впорядкованість нейронів навченої мережі. TE обчислимо за формулою:

$$TE = \frac{1}{n} \sum_{w_j \in R^n} \begin{cases} 1, & \left| |r_{j^*} - r_{sj^*}| \right| > 1 \\ 0, & \left| |r_{j^*} - r_{sj^*}| \right| \leq 1 \end{cases} \quad (7)$$

де r_{j^*} - координати нейрона найкращого наближення, r_{sj^*} - координати другого нейрона найкращого наближення. Розраховане значення TE = 0,018.

Одержані значення параметрів QE та TE свідчать про адекватність отриманих результатів.

Висновок. Для оцінювання тенденції зміни забруднення атмосферного повітря від стаціонарних та пересувних джерел у Львівській області застосовано двоетапний підхід до кластеризації, який передбачає послідовне застосування карт Кохонена, що самоорганізуються, та алгоритму k-середніх.

Отримані результати свідчать, що виявлено кластери із найвищими показниками викидів забруднюючих речовин і тенденцією до їх зростання. Тобто такі райони Львівської області як Кам'янка-Бузький, Яворівський, Сокальський та Жидачівський потребують особливої уваги у сфері екологічної політики та розроблення природоохоронних заходів.

Література:

- [1] Друзюк В. Динаміка викидів забруднювальних речовин в атмосферу у Львівській області / В. Друзюк, В. Погребенник, Ю. Сікач // *Вимірювальна техніка та метрологія*, 2009, №70, С.71-76
- [2] Кравців В.С. Розвиток природоексплуатуючих галузей господарства Львівської області: стан, проблеми, перспективи / В.С. Кравців, П.В. Жук, О.І. Гулич, І.А. Колодійчук, В.О. Полюга // *НАН України. Ін-т регіональних досліджень*. – Львів, 2011. – с.90
- [3] Регіональна доповідь про стан навколишнього природного середовища у Львівській області в 2015 році. : [Електронний ресурс]. – Режим доступу: http://www.ekologia.lviv.ua/file/monitoring/ND_2016.pdf
- [4] Регіональна статистика Львівщини. : [Електронний ресурс]. – Режим доступу: <http://database.ukrcensus.gov.ua>
- [5] *Laboratory of computer and information science*: [Електронний ресурс]. – Режим доступу: <http://www.cis.hut.fi/projects/somtoolbox/>
- [6] Chon T.S. *Self-organizing Maps Applied to Ecological Sciences* / T.S. Chon // *Ecological Informatics*, 2011, Vol.6(1), pp.50-61.
- [7] Laerhoven K. V. *Combining the self-organizing map and k-means clustering for on-line classification of sensor data* / K. V. Laerhoven // *Proc. International Conference on Artificial Neural Networks*, 2001, pp. 464-469
- [8] Olej V. *Air Quality Modelling by Kohonen's Neural Network* / V. Olej, P. Hájek, J. Křupka, I. Obršálová, // *Proc. of the the 5th WSEAS International Conference on Environment, Ecosystems and Development (EED'07)*, WSEAS Press, Tenerife, Canary Islands, Spain, 2007.
- [9] Si J. *Dynamic Topology Representing Networks* / J. Si, S. Lin, M. A. Vuong // *Neural Networks*, 2000, Vol.13, pp. 617-627.
- [10] Vesanto J. *Clustering of the self-organizing map* / J. Vesanto, E. Alhoniemi // *IEEE Transactions on Neural Networks*, 2000, Vol.11, pp. 586-600.
- [11] Wang J. H. *Two-stage clustering via neural networks* / J. H. Wang, J. D. Rau, and W. J. Liu // *IEEE Transactions on Neural Networks*, 2003, Vol. 14, pp. 606-615.