

К вопросу о повышении достоверности оценки качества воздушной среды урбанизированной территории

Н.В. Мензелинцева, Н.Ю. Каратузова, И. М. Статюха, Е.В. Попова

Волгоградский государственный технический университет, Волгоград

Аннотация: Показана возможность повышения достоверности оценки качества атмосферы урбанизированных территорий за счет использования вероятностных способов. Оценка показателей качества воздуха по количественному признаку позволила найти ошибку первого и второго рода. На основе подхода Байеса получены зависимости для оценки и прогноза загрязнения атмосферы в районе города, не имеющего стационарного поста, на основе данных стационарных постов других районов.

Ключевые слова: качество атмосферы, достоверность оценки, вероятностные способы, количественный признак, ошибка первого и второго рода, подход Байеса, априорная вероятность, апостериорная вероятность, правдоподобие величины.

В последние годы требования к экологической безопасности городов, и прежде всего к атмосферному воздуху урбанизированных территорий, постоянно ужесточаются. Соответственно повышаются требования к точности оценки загрязнения воздуха [1,2].

Качество атмосферного воздуха города оценивается по замерам стационарных постов экологических служб, которые располагаются не во всех районах, поэтому такая оценка является достаточно усредненной. Значительная изменчивость концентрации пыли по районам города, и недостаточное количество замеров повышают вероятность ошибки при оценке загрязнения воздуха в населенном пункте в целом.

Достоверность оценки качества воздушной среды можно повысить, используя вероятностные способы [3-5].

Основной нормативной характеристикой, определяющей качество атмосферного воздуха является предельно допустимая концентрация (ПДК), которую при вероятностной оценке качества атмосферы можно рассматривать, как верхнюю предельную норму для сводных характеристик выборки, на-

пример, среднего арифметического значения фактической концентрации $C_{факт}$.

Возможен случай, когда по контролируемому веществу в целом для атмосферы города выполняется соотношение $C_{факт} \leq \text{ПДК}$, но в контрольных пробах (выборке) отмечается превышение ПДК. В теории управления качеством такое события (обозначим его событием Z) называют ошибкой первого рода [3], максимальную вероятность которого можно определить по формуле:

$$P(C_{выб} \geq C_{факт} + k\sigma_C) = \alpha, C_{факт} = \text{ПДК} - \Delta \quad \Delta \rightarrow 0 \quad (1)$$

где $C_{выб}$ – выборочное значение концентрации; k – коэффициент, зависящий от характера распределения $C_{выб}$ и заданной вероятности α , σ_C – среднее квадратическое отклонение, Δ – некоторая положительная величина.

Рассмотрим другой случай (событие Y), когда качество атмосферы в целом не соответствует нормативным требованиям, но результаты анализа выборочных проб оценены как не превышающие ПДК. Эта ошибка называется ошибкой второго рода [3]. В этом случае максимальную вероятность можно определить в соответствии с выражением (2):

$$P(C_{выб} \leq C_{факт} - k\sigma_C) = \beta, C_{факт} = \text{ПДК} + \Delta \quad \Delta \rightarrow 0 \quad (2)$$

где $C_{выб}$ – выборочное значение концентрации; k – коэффициент, зависящий от характера распределения $C_{выб}$ и заданной вероятности β , σ_C – среднее квадратическое отклонение, Δ – некоторая положительная величина.

При достаточно большом объеме выборки среднее арифметическое значение имеет нормальное распределение [4]. В этом случае вероятность ошибки первого рода $p = 1 - \alpha$ и ошибки второго рода $q = 1 - \beta$ можно определить по формулам (3), (4) соответственно:

$$[(\text{ПДК} - \bar{C}_{выб}) / \sqrt{n}] / \sigma_C \geq U_{p=1-\alpha} \quad (3)$$

$$[(\bar{C}_{выб} - \text{ПДК}) / \sqrt{n}] / \sigma_C \geq U_{q=1-\beta} \quad (4)$$

где $\bar{C}_{выб}$ – выборочное среднее значение, n – число испытаний в выборке, σ_C – среднее квадратическое отклонение, U – квантиль нормального распределения.

По формулам (3,4) можно построить зависимость, характеризующую вероятность ошибки 1 или 2 рода от величины квадратического отклонения, которая в общем виде представлена на рис.1.

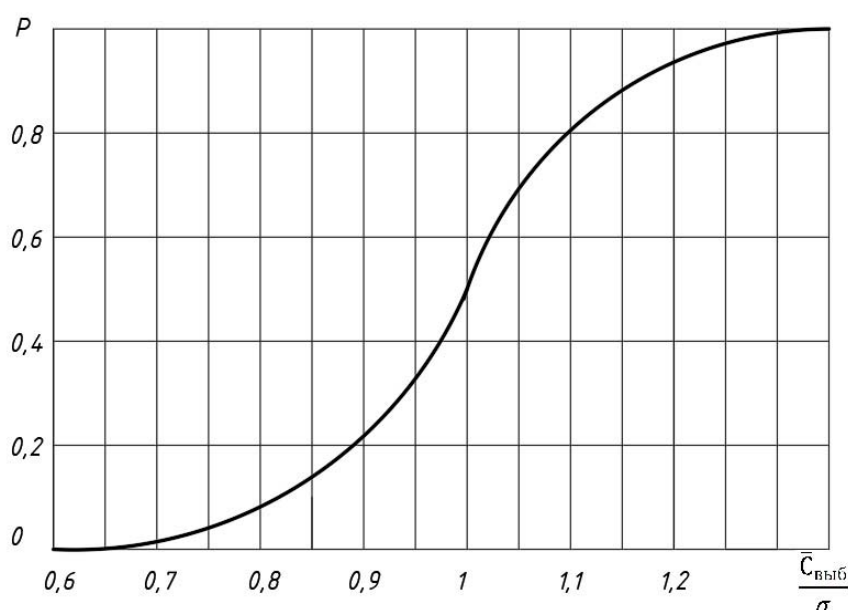


Рис.1. - Зависимость вероятности ошибки 1 или 2 рода от величины квадратического отклонения.

Для решения экологических задач по оценке качества воды, почвы известно использование подхода Байеса [6,9,12]. С целью повышения оценки качества атмосферы можно также использовать подход Байеса, позволяющий проводить анализ экспериментальных данных стационарных постов совместно с данными, накопленными при разовых замерах в районах, не имеющих таковых. Простое объединение подобных данных может быть неэффективным, так как текущие данные стационарного поста и данные разовых замеров в других районах города взаимно тем меньше коррелированы, чем больше временной интервал между замерами [8,10,11].

По теореме Байеса [4]:

$$P(\theta | D) = \frac{P(D | \theta)P(\theta)}{P(D)} = \frac{P(D | \theta)P(\theta)}{\int_{\psi} P(D | \theta)P(\theta)d\theta} \quad (5)$$

где $P(D | \theta)$ – функция правдоподобия, описывающая распределение случайной величины D при фиксированном значении θ ; $P(\theta | D)$ – апостериорная расчетная плотность вероятности, $P(D) = \int_{\theta} P(D | \theta)P(\theta)d\theta$ – безусловная плотность вероятности (нормирующий множитель), $P(\theta)$ – априорная вероятность гипотезы θ , ψ – область определения параметра θ .

В статистике Байеса вероятность отражает степень доверия событию, которая может быть основана на данных предыдущих экспериментов, личном опыте прошлых наблюдений, на основе модельных гипотез, т.е. на тех данных, которые получены до эксперимента, или априорных данных. Соответственно, начальная вероятность называется априорной.

По мере накопления новой информации вероятность может меняться. В ходе эксперимента появляются новые данные, которые называются апостериорными. Нормализованная величина после вычисления вероятности для каждого наблюдения называется апостериорной.

Для вычисления и обновления вероятности после получения новых данных используют теорему Байеса, которая описывает условную вероятность события на основе как данных эксперимента, так и априорной информации или доверия событию, или условий, связанных с событием. Условная вероятность, т.е. вероятность каждого события – функция правдоподобия.

Формулу Байеса можно применять последовательно после поступления очередной серии экспериментальных данных, при этом текущее распределение параметра считается априорным, а новые данные – апостериорными, за-

тем, при необходимости, они становятся априорными для следующего случая.

Предположим, что событие A состоит в том, что концентрация пыли в районе 1, на территории которого есть стационарный пост, превышает предельно допустимую концентрацию (ПДК), под событием B будем понимать превышение ПДК пыли в районе 2, где стационарный пост отсутствует.

Исходя из теоремы Байеса можно определить вероятность превышения ПДК в районе B при условии, что стационарный пост в районе A зарегистрировал такое превышение, при этом проводить экспериментальные замеры в районе B не требуется.

В этом случае выражение (1) запишется в виде

$$P(B/A) = \frac{P(A|B)P(B)}{P(A)} \quad (6)$$

где $P(A|B)$ – функция правдоподобия, описывающая распределение случайной величины A при фиксированном значении B , $P(B|A)$ – апостериорная расчетная плотность вероятности, $P(A) = \int_{\omega} P(A|B)P(B)dB$ – безусловная плотность вероятности (нормирующий множитель), ω – область определения параметра B , $P(B)$ – априорная вероятность наступления события B .

Предположим, что из общего числа n наблюдений за концентрацией пыли в воздухе s замеров превышают ПДК и $I = n - s$ замеров, соответственно, не превышают ПДК. В этом случае функцию правдоподобия рекомендуется представлять биномиальным распределением [5,6,7]:

$$P(s | n, R_{np}) = \frac{n!}{s!(n-s)!} R_n^s (1 - R_n)^{n-s} \quad (7)$$

где R_{np} – вероятность того, что концентрация пыли в воздухе превышает ПДК, индекс « np » соответствует правдоподобию величины.

К априорной плотности вероятности предъявляется целый ряд требований [6,9], наиболее полно которым отвечает бета-распределение [5,6,7]:

$$P_0(R_0) = B(x_0, t_0) R_0^{x_0} (1-R_0)^{t_0-1} \quad (8)$$

где $B(x_0, t_0)$ - бета-функция с параметрами $x_0 \geq 1, t_0 \geq 1$, индекс 0 соответствует априорному распределению.

Из выражений (6-8) можно найти апостериорную вероятность того, что концентрация пыли превысит ПДК:

$$P(R_B | s, n) = B(x, t) R_B^{x-1} (1-R_B)^{t-1} \quad (9)$$

где $x = x_0 + s, t = y_0 + n - s$, индекс Б обозначает значение величины по Байесу.

В качестве оценки параметра R можно принять апостериорное математическое ожидание [6,7,9]:

$$\hat{R}_B = [R_B | s, n] = \frac{x}{x+t} \quad (10)$$

Если параметры x_0, t_0 получены по результатам разовых замеров в районе 2 (без стационарного поста), то $x_0 = s_0, t_0 = n_0 - s_0$. Оценка вероятности превышения концентрации пыли в этом районе по частоте определяется по формуле:

$$\hat{R} = \frac{s_0}{n_0} \quad (11)$$

Оценка вероятности по частоте по объединенной выборке наблюдений в двух районах представляет собой байесовскую оценку:

$$\hat{R}_B = \frac{s_0 + s}{n_0 + n} \quad (12)$$

При объединении выборок экспериментальной и байесовской плотность вероятности $P_{об}$ может быть найдена по формуле:

$$P_{об}(R) = \alpha * P_B(R) + (1 - \alpha) * P_n(R) \quad (13)$$

$$\alpha^* = \frac{\alpha_0 - \alpha_{гр}}{1 - \alpha_{гр}} \quad (14)$$

где α_0 – наблюдаемый уровень значимости ; $\alpha_{гр}$ – уровень значимости на границе поля допуска , $\alpha_{гр} = [0,05 \div 0,1]$ [6].

Анализ формулы (10) показывает, что при $\alpha^* = 1$ используется апостериорная плотность вероятности $P_B(R)$ и для оценки используются данные объединенной выборки замеров , при $\alpha^* = 0$ оценка идет по функции правдоподобия экспериментальных данных стационарного поста $P_n(R)$.

Исходя из того, что, что байесовская оценка имеет вид (10), \hat{R}_B определяется по формуле (12):

$$\hat{R}_B = \frac{s_0 + s}{n_0 + n} = \frac{n_0}{n_0 + n} \hat{R}_0 + \left(1 - \frac{n_0}{n_0 + n}\right) \hat{R}_n \quad (15)$$

где $\hat{R}_n = \frac{s}{n}$ (16) .

Тогда с учетом (13), получаем:

$$\hat{R}_{OB} = \frac{n_0 \alpha^*}{n_0 + n} \hat{R}_0 + \left(1 - \frac{n_0 \alpha^*}{n_0 + n}\right) \hat{R}_n \quad (17)$$

Сопоставляя выражения (15) и(17) ,можно заключить о возможности перераспределения суммарного объема выборки (n_0+n) в сторону уменьшения относительного априорного объема ,т.е.относительного объема разовых замеров в районе $2\left(\frac{n_0 \alpha^*}{n_0 + n}$ вместо $\frac{n_0}{n_0 + n}$) и увеличения доли относительного объема замеров стационарного поста $\left(1 - \frac{n_0 \alpha^*}{n_0 + n}\right)$ вместо $\left(1 - \frac{n_0}{n_0 + n}\right)$).При необходимости объединения наблюдений в нескольких районах города полученная оценка (17) рассматривается как априорная для последующего этапа.

Для трех районов имеем $n_0, s_0, \hat{R}_0 = \frac{s_0}{n_0}, \alpha_1, n_1, s_1, \hat{R}_1 = \frac{s_1}{n_1}, \alpha_2, n_2, s_2,$

$\hat{R}_2 = \frac{s_2}{n_2}$. Следовательно, как априорная информация для третьего района

может быть использована оценка в виде:

$$\hat{R}_{OB_1} = \frac{n_0 \alpha_1}{n_0 + n_1} \hat{R}_0 + \left(1 - \frac{n_0 \alpha_1}{n_0 + n_1}\right) \hat{R}_{II} \quad (18)$$

Объединенная оценка \hat{R}_{OB_2} может быть найдена по формуле :

$$\hat{R}_{OB_2} = \frac{(n_0 + n_1) \alpha_2}{(n_0 + n_1 + n_2)} \hat{R}_{OB_1} + \left[1 - \frac{(n_0 + n_1) \alpha_2}{(n_0 + n_1 + n_2)}\right] \hat{R}_{II_2} \quad (19)$$

Выводы

1. Оценка качества атмосферного воздуха города по замерам стационарных постов экологических служб, которые располагаются не во всех районах, является достаточно усредненной, недостаточное количество замеров и значительная изменчивость концентрации вредных веществ по районам города повышают вероятность ошибки при оценке загрязнения атмосферы в населенном пункте в целом.
2. Показана возможность повышения достоверности оценки качества воздушной среды урбанизированной территории за счет вероятностных способов, в том числе оценки показателя качества по количественному признаку и подхода Байеса.

Литература

1. Азаров В.Н, Донцова Т.В., Сытник Л.Е. Оценка качества воздушной среды города на основе среднебалансового интегрального критерия загрязнения атмосферы // Инженерный вестник Дона, 2009, №1. URL:

- Ivdon.ru/ru/magazine/archive/N2y2017/4008.
2. Азаров В.Н, Кошкарев С.А, Николенко М.А. К определению фактических размеров частицы пыли выбросов стройиндустрии и строительства // Инженерный вестник Дона, 2015, №1. URL:Ivdon.ru/ru/magazine/archive/N2y 2015/2858.
 3. Ноулер Л. Статистические методы контроля качества продукции. М.: Издательство стандартов,1989. 96с.
 4. Гмурман В.Е. Теория вероятности и математическая статистика. М.: Высшая школа, 2003. 479 с
 5. Лимер. Э. Статистический анализ не экспериментальных данных. Выбор формы связи. М.: Финансы и статистика,1983. 381 с.
 6. Розенталь. О.М., Александровская Л.Н., Кириллин А.В. Байесовский подход к повышению достоверности контроля качества сточных вод // Аналитика и контроль. 2018.№3. С.334-340.
 7. Зайцева Т.В., Нестерова Е. В., Игрунова С.В. Байесовская стратегия оценки достоверности выводов // Научные ведомости Белгородского государственного университета.2012, №14(132), выпуск 23\1. С.180-184.
 8. Кожомбердиева Г.И., Бураков Д.П., Гарина М.И. Использование формулы Байеса при оценивании выполнения практик модели СММІ // Программные продукты и системы. 2017.т.30,№1.С. 67-74.
 9. Благовещенский Ю.Н., Самсонов В.П., Мешалкина Ю.Л.Байесовский подход при анализе структуры почвенного покрова // Вестник Московского университета. Серия 17. Почвоведение.2014. №4. С.13-20.
 10. Соколов В.А. Интервальные методы в Байесовском подходе при диагностировании технического состояния строительных конструкций зданий // Научно-технические ведомости СПбПУ. Наука и образование. Серия Естественные и инженерные науки. 2011.№4.С.227-235.
 11. Kass R.E., Raftery A.E Bayes Factors. // J.Amer. Statist. Assoc. 1995. Vol.
-

90.№430. pp.773-775.

12. Bogaert P., D'Or D. Estimating soil properties from thematic soil maps: the Bayesian maximum entropy. //Soil Sci.Soc. Amer.J. 2002.Vol.66.pp.1492-1500

References

1. Azarov V.N, Doncova T.V., Sytnik L.E. Inzhenernyj vestnik Dona, 2009, №1. URL:Ivdon.ru/ru/magazine/archive/N2y2017/4008.
 2. Azarov V.N, Koshkarev S.A, Nikolenko M.A. Inzhenernyj vestnik Dona, 2015, №1. URL:Ivdon.ru/ru/magazine/archive/N2y 2015/2858.
 3. Nouler L. Statisticheskie metody kontrolya kachestva produkcii [Statistical methods of product quality control]. M.: Izdatel'stvo standartov, 1989. 96p.
 4. Gmurman V.E. Teoriya veroyatnosti i matematicheskaya statistika [Probability Theory and Mathematical Statistics]. M.: Vysshaya shkola, 2003. 479 p.
 5. Limer. E. Statisticheskij analiz ne eksperimental'nyh dannyh. Vybor formy svyazi [Statistical analysis of non-experimental data. The choice of the form of communication]. M.: Finansy i statistika, 1983. 381 p.
 6. Rozental'. O.M., Aleksandrovskaya L.N., Kirillin A.V. Analitika i kontrol'. 2018.№3. pp.334-340.
 7. Zajceva T.V., Nesterova E.V., Igrunova S.V. Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta.2012, №14 (132),vypusk 23\1. pp.180-184.
 8. Kozhombdieva G.I., Burakov D.P., GarinaM.I . Programmnye produkty i sistemy. 2017.t.30., №1.pp.67-74.
 9. Blagoveshchenskij YU.N., Samsonov V.P., Meshalkina YU.L. Vestnik Moskovskogo universiteta. Seriya 17. Pochvovedenie.2014.№4.pp.13-20.
 10. Sokolov V.A. Nauchno-tehnicheskie vedomosti SPbPU. Nauka i obrazovanie.
-



- Seriya Estestvennyye i inzhenernyye nauki. 2011.№4.pp.227-235.
11. Kass R.E., Raftery A.E J. Amer.Statist.Assoc. 1995. Vol. 90.№430. pp.773-775.
 12. Bogaert P., D'Or D. Soil Sci.Soc. Amer.J. 2002.Vol.66. pp.1492-1500