

ИСПОЛЬЗОВАНИЕ ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ РЕШЕНИЯ ЗАДАЧ РАЙОНИРОВАНИЯ (на примере годовых атмосферных осадков Беларуси)

А. А. Волчек¹, Н. В. Санюк², О. В. Безсилко²

¹ Отдел проблем Полесья НАН Беларуси

Брест, Беларусь

E-mail: Volchak@tut.by

² Брестский государственный университет имени А. С. Пушкина

Брест, Беларусь

E-mail: lapan@mail.ru

Метод дисперсионного анализа применяется к решению проблемы районирования территорий по климатическим показателям. Как пример использована территория Беларуси и режим выпадения годовых атмосферных осадков. Установлено, что режим выпадения годовых атмосферных осадков на территориях Минской и Могилевской областей идентичен.

Ключевые слова: районирование территорий, атмосферные осадки, метеостанции, дисперсионный анализ.

Для решения ряда научных и прикладных задач необходимо проведение климатического районирования территорий. Традиционно такие задачи решаются путем выделения доминирующих факторов либо на основании разных метрик, общим недостатком которых являются жесткие требования к исходным данным и трудности интерпретации результатов.

В последние годы более широкое распространение получили математические методы районирования, такие как факторный анализ, теория распознавания образов, метод пространственного объединения временных рядов и др. [1]. Преимущество математических методов заключается в их объективности, возможности использования компьютерных технологий, что позволяет привлечь к анализу большие объемы информации, а конечный результат в любом удобном для пользователя виде. Однако на этом пути исследователя ждет ряд трудностей, а именно: выделение объективных критериев районирования, обоснование методов анализа, интерпретации полученных результатов.

Нами предпринята попытка использования дисперсионного анализа для решения задач районирования территорий по климатическим показателям. С этой целью были отобраны временные ряды годовых величин атмосферных осадков равномерно расположенных на территории Беларуси 37 метеостанций, имеющих период наблюдения с 1949 по 2000 г., т. е. 52 года.

Для проведения районирования методом дисперсионного анализа [2, 3] имеющиеся 37 временных рядов длиной 52 года ($p = 52$) предварительно разбиты на

6 групп ($J = 6$) по областям (группа 1 метеостанции Брестской области, группа 2 – Витебской, группа 3 – Гомельской, группа 4 – Гродненской, группа 5 – Минской, группа 6 – Могилевской областей).

Векторы наблюдений размерности p , относящиеся к j -й группе ($j = 1, 2, \dots, 6$), должны иметь нормальное распределение $N(\mu_j, \Sigma)$, где μ_j – вектор средних значений группы j и Σ – ковариационная матрица, одинаковая для всех групп. Векторы μ_j и матрица Σ считаются неизвестными.

Предполагается, что для каждой группы указана выборка p -мерных векторов наблюдений, т. е. для каждой группы должна существовать выборка соответственно из n_j векторов наблюдений. Отдельные векторы наблюдений j -й группы обозначаются:

$$y_{jk} = \begin{pmatrix} y_{1jk} \\ y_{2jk} \\ y_{52jk} \end{pmatrix} \quad (j = 1, \dots, 6; k = 1, \dots, n_j), \text{ где } \begin{pmatrix} n_1 = 7 \\ n_2 = 9 \\ n_3 = 5 \\ n_4 = 4 \\ n_5 = 7 \\ n_6 = 5 \end{pmatrix}.$$

Дальнейшие вычисления основываются на векторах средних значений отдельных групп:

$$y_j = \frac{1}{n_j} \sum_{k=1}^{n_j} y_{jk} \quad (j = 1, \dots, 6),$$

и векторе средних значений всех групп (общее среднее)

$$y_{..} = \frac{1}{37} \sum_{j=1}^J \sum_{k=1}^{n_j} y_{jk} = \frac{1}{37} \sum_{j=1}^J n_j y_j,$$

оценках ковариационных матриц отдельных групп:

$$\begin{aligned} S_j &= \frac{1}{n_j - 1} \sum_{k=1}^{n_j} (y_{jk} - y_j)(y_{jk} - y_j)' = \\ &= \frac{1}{n_j - 1} \left(\sum_{k=1}^{n_j} y_{jk} y_{jk}' - n_j y_j y_j' \right), \end{aligned} \quad (j = 1, \dots, 6)$$

а также усредненной ковариационной матрице

$$S = \frac{1}{31} \sum_{j=1}^J (n_j - 1) S_j.$$

Векторы y_j и матрица S – несмещенные оценки μ_j и Σ .

При классификации использовали элементарные дискриминантные признаки v_1, \dots, v_t . Так как не все v_i линейно независимы один от другого, то для классификации возьмем вектор v^* только первых t дискриминантных признаков v_1, \dots, v_t . Если $J \leq p + 1$, то $t = J - 1$. В этом случае используем все дискриминантные признаки, кроме последнего.

Вектор v вычислим по формуле

$$v = D'y,$$

где $D = (d_1, d_2, \dots, d_t) = S^{-1}A$.

$$A = (y_1 - y_{..}, y_2 - y_{..}, \dots, y_J - y_{..}).$$

Признак v_j по формуле

$$v_j = d'_j y = \sum_{i=1}^p d_{ij} y_i,$$

$$d_j = S^{-1}(y_j - y_{..}).$$

Здесь классификация основана на критерии значимости. С его помощью проверяется принадлежность вектора v^* к группе j . При построении критерия вектор v^* будем рассматривать как представителя группы с номером 0 наряду с группами 1, 2, ..., J . Тогда получаем статистику:

$$\tilde{F}_{0/j} = k_j = \frac{n - J - t + 1}{t(n - J)} \frac{n_j}{n_j + 1} (v^* - v_{j.}^*)' S_{v^*}^{-1} (v^* - v_{j.}^*).$$

В соответствии с этой процедурой мы относим вектор v^* , т. е. соответствующую ему станцию наблюдения, к группе j , если

$$k_j \leq F_{t, n-J+1; \alpha}.$$

При использовании этого метода станция наблюдения может быть одновременно отнесена к нескольким группам. Но может случиться, что она не будет причислена ни к одной из групп. В этом случае выбирают группы с наименьшим k_j .

Классификацию можно провести точнее, если дополнительно учитывать априорные вероятности p_j каждой группы. Если с самого начала учитывать, каковы вероятности попадания станций в отдельные группы, то для классификации выборок большого объема надо применять величины

$$m_j = \frac{1}{2} (v^* - v_{j.}^*)' S_{v^*}^{-1} (v^* - v_{j.}^*) - \log p_j.$$

Станцию относят к группе с наименьшим m_j .

Классификация по 52 признакам с применением 37 данных векторов наблюдений приводит к результатам, отраженным в таблице.

Результаты классификации

Заданные группы	Конечные группы					
	Группа 1	Группа 2	Группа 3	Группа 4	Группа 5	Группа 6
Группа 1	7	0	0	0	0	0
Группа 2	0	9	0	0	0	0
Группа 3	0	0	0	0	0	5
Группа 4	0	0	0	4	0	0
Группа 5	0	0	7	0	0	0
Группа 6	0	0	5	0	0	0

В данной таблице числа, отличные от нуля, обозначают количество метеостанций, относящихся к какой-то конкретной группе. Строка, в которой находится число, обозначает, к какой группе относились метеостанции изначально, а столбец – к какой группе стали относиться эти же метеостанции после проведения дисперсионного анализа.

В результате дисперсионного анализа установлена однородность формирования режима годовых атмосферных осадков Минской и Могилевской областей, следовательно, эти территории можно объединить в одну группу. Остальные группы станций можно оставить без изменения.

Схематическое районирование территории Республики Беларусь по уровню годовых атмосферных осадков представлено на рисунке. Как видно из рисунка, отличия в формировании годовых атмосферных осадков имеют отношение к повышенным участкам местности, в основном к Белорусской гряде. Обращает на себя внимание стабильностью выпадения осадков северо-запад территории Беларуси, что ярко выделяет зону формирования максимальных модулей стока.



Схематическое представление районирования территории Беларуси по режиму выпадения годовых атмосферных осадков

ЛИТЕРАТУРА

1. Исследование и расчеты речного стока / Под ред. В. Д. Быкова. М.: Изд-во МГУ, 1981. 228 с.
2. Арнс Х., Лейтер Ю. Многомерный дисперсионный анализ. М., 1985. С. 101–157.
3. Хальд А. Математическая статистика с техническими приложениями. М., 1956. С. 351–415.