

РУССКОЕ ГЕОГРАФИЧЕСКОЕ ОБЩЕСТВО
Томский отдел
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ВОПРОСЫ ГЕОГРАФИИ
СИБИРИ

Выпуск 25

Томск
2003

**ПРОСТРАНСТВЕННО-ВРЕМЕННЫЕ ИЗМЕНЕНИЯ
СОСТАВЛЯЮЩИХ ПОЛЕЙ КОНЦЕНТРАЦИЙ ПРИМЕСИ
В АТМОСФЕРЕ г. ТОМСКА**

Для выявления пространственной общности изменения составляющих временных рядов концентрации загрязнителей, а, следовательно, и причин, которые их формируют, воспользуемся методом разложения полей трендов на ортогональные составляющие. Традиционный способ определения взаимосвязи между переменными, которую также называют структурой зависимости, заключается в определении ковариаций, что эквивалентно получению дисперсий и корреляций между исходными переменными (в данном случае в качестве переменных выступают значения составляющих временных рядов концентрации загрязнителей в различных точках города Q_j). Однако в некоторых случаях удобнее найти линейные комбинации

Y_1, \dots, Y_m переменных Q_1, \dots, Q_p ($m \leq p$), по которым можно получить структуру зависимости между Q_1, \dots, Q_p .

Суть метода состоит в том, что находятся такие линейные комбинации исходных переменных

$$Y_i = \sum_{j=1}^p a_{ij} Q_j, \dots, Y_p = \sum_{j=1}^p a_{pj} Q_j, \quad (1)$$

что ковариации между новыми переменными (компонентами) равны нулю, т.е. новые переменные не коррелированы между собой:

$$\text{cov}(Y_i, Y_j) = 0, \quad i, j = 1, \dots, p, \quad i \neq j, \quad (2)$$

дисперсии компонент ранжированы, дисперсия предыдущей компоненты всегда больше или равна дисперсии последующей:

$$V(Y_1) \geq V(Y_2) \geq \dots \geq V(Y_p). \quad (3)$$

Сумма дисперсий компонент (общая дисперсия) равна сумме дисперсий переменных:

$$\sum_{i=1}^p V(Y_i) = \sum_{i=1}^p \sigma_i^2. \quad (4)$$

Таким образом, получается сжатое описание структуры зависимости, несущее почти всю информацию, которая содержится в исходных переменных. Этот метод носит название метода анализа главных компонент [1–7]

Для получения линейных комбинаций (1) необходимо определять коэффициенты a_{kj} , где k – номер компоненты, j – номер переменной. Чтобы определить коэффициенты, необходимо иметь ковариационную матрицу исходных переменных v_{ij} , однако удобнее перед получением v_{ij} исходные переменные центрировать. Тогда диагональные элементы v_{ij} будут представлять собой дисперсии переменных $\sigma_{ij}^2 = v_{ij}$. При этом Q_j в выражении (1) будут являться отклонениями от своих средних значений

При известной ковариационной матрице σ_{ij}^2 и линейном преобразовании $Y_1 = a_{11} Q_1 + \dots + a_{1p} Q_p$ требуется найти такие a_{11}, \dots, a_{1p} , чтобы величина $V(Y_1) = \sum_{i=1}^p \sum_{j=1}^p a_{1i} a_{1j} \sigma_{ij}^2$ была максимальной при $\sum_{j=1}^p a_{1j}^2 = 1$. Это обеспечивает единственность решения. Решение $\alpha_1 = (a_{11}, \dots, a_{1p})$ называется собственным вектором и соответствует максимальному собственному значению матрицы σ_{ij}^2 . Это собственное значение равно дисперсии $V(Y_1)$. Линейная комбинация $Y_1 = a_{11} Q_1 + \dots + a_{1p} Q_p$ называется первой главной компонентой переменных Q_1, \dots, Q_p . Она объясняет $100 V(Y_1)/V$ процентов общей дисперсии. По аналогии определяются собственные векторы и собственное значение второй главной компоненты и т.д.

Можно преобразовать систему (1) так, чтобы каждая из исходных переменных была выражена как линейная комбинация главных компонент. Это преобразование записывается в виде следующей формулы

$$Q_1 = \sum_{j=1}^p a_{j1} Y_j, \dots, Q_p = \sum_{j=1}^p a_{jp} Y_j \quad (5)$$

Связь между исходными переменными и главными компонентами можно определить традиционным методом, рассчитав коэффициент корреляции, или воспользоваться формулой:

$$\text{corr}(Q_i, Y_j) = [a_{ji} \sqrt{V(Y_j)}] / \sigma_i^2 \quad (6)$$

Учитывая некоррелированность компонент, можно рассчитать долю дисперсии i -ой переменной, выраженной m компонентами:

$$\frac{\sum_{j=1}^m V(Y_j)}{\sigma_i^2} = [a_{i1}^2 V(Y_1) + \dots + a_{im}^2 V(Y_m)] / \sigma_i^2. \quad (7)$$

Применение описанного метода определяется:

- 1) необходимостью наглядного представления (визуализации) данных;
- 2) упрощением интерпретации полученных статистических выводов;
- 3) необходимостью построения иерархической классификации переменных, которую не всегда можно применить к исходным данным

Если для поиска решения $\alpha_i = (a_{i1}, \dots, a_{ip})$ использовалась корреляционная матрица, тогда выражение (6) переписывается в виде.

$$r_{i,j} = \text{corr} [a_{ij} \sqrt{V(Y_j)}] \quad (8)$$

Из (8) следует, что корреляция между главными компонентами и уровнями загрязнения в различных пунктах города определяется собственным значением j -ой компоненты и коэффициентами $a_{i,j}$. Учитывая, что величина $V(Y_j)$ для всех пунктов города одинакова, значение $r_{i,j}$ будет определяться величиной и знаком $a_{i,j}$.

Материалом для исследования послужили данные наблюдений за концентрацией диоксида азота, двуокиси серы, угарного газа и пыли (основных ингредиентов) в г. Томске за период с 1981 по 2000 гг. в четырех пунктах: пл. Ленина (ПНЗ-2), ул. Герцена (ПНЗ-5), пр. Мира (ПНЗ-7) и завод измерительной аппаратуры (ПНЗ-11). Данные были предоставлены Томским центром по гидрометеорологии и мониторингу загрязнения окружающей природной среды (ЦГММ)

Рассмотрим связь между внутригодовой составляющей концентраций оксида углерода по территории города и главными компонентами. Результаты расчета приведены в табл. 1. Аналогичные результаты получены для других временных составляющих и для других ингредиентов.

Из анализа табл. 1 вытекает, что только первая компонента имеет наибольшее по абсолютной величине коэффициенты корреляции, кроме этого, она имеет и одинаковые знаки для всех пунктов. Из этого следует, что коэффициенты $a_{i,j}$ также имеют одинаковый знак. Для других компонент коэффициенты $a_{i,j}$ в зависимости от пункта отличаются не только величиной, но и знаком. Отсюда следует вывод, что первая главная компонента определяет синхронные колебания концентрации загрязнителей по территории города. При этом величину этих колебаний можно оценить по формуле

$$Q_j^o = a_{j,1} Y_1 = r_{j,1} \sqrt{V(Y_1)} Y_1. \quad (9)$$

Другие главные компоненты ответственны за локальные изменения концентрации в каждом конкретном пункте.

В табл 2–4 приведены коэффициенты корреляции между первой главной компонентой и различными составляющими временных рядов загрязнения по террито-

Таблица 1

Коэффициенты корреляции между внутригодовой составляющей концентрации оксида углерода в различных пунктах города и главными компонентами

Пункт	Номер компоненты			
	1	2	3	4
ПНЗ 2	0,947	-0,011	-0 268	0,177
ПНЗ-5	0,906	0,375	0 199	0,010
ПНЗ 7	0,910	-0 319	0,264	0,032
ПНЗ 11	0,961	0,039	-0 172	-0,214

рии города, а также собственные значения первой главной компоненты для разных загрязнителей. Квадраты коэффициентов корреляции, приведенные в табл 2–4, представляют собой долю вклада синхронных изменений в дисперсию каждой составляющей в различных пунктах города

Таблица 2

Коэффициенты корреляции между первой главной компонентой и внутригодовой составляющей для различных загрязнителей по территории города, а также собственные ($\sqrt{V(Y_i)}$) значения первой главной компоненты

	Загрязнители			
	CO	NO ₂	SO ₂	пыль
$\sqrt{V(Y_i)}$	0,931	0,688	0,926	0,837
ПНЗ-2	0,947	0,639	0,944	0,906
ПНЗ-5	0,906	0,669	0,893	0,860
ПНЗ-7	0,910	0,784	0,920	0,671
ПНЗ-11	0,961	0,903	0 947	0,888

Таблица 3

Коэффициенты корреляции между первой главной компонентой и внутримесячной составляющей для различных загрязнителей по территории города, а также собственные значения ($\sqrt{V(Y_i)}$) первой главной компоненты

	Загрязнители			
	CO	NO ₂	SO ₂	пыль
$\sqrt{V(Y_i)}$	0 623	0,592	0,746	0,630
ПНЗ-2	0,664	0,592	0,676	0,619
ПНЗ-5	0,604	0,510	0,770	0,740
ПНЗ 7	0 595	0,654	0,747	0,650
ПНЗ 11	0,625	0,606	0,791	0,711

Коэффициенты корреляции между первой главной компонентой и внутрисуточной составляющей для различных загрязнителей по территории города, а также собственные значения ($\sqrt{V(Y_i)}$) первой главной компоненты

	Загрязнители			
	CO	NO ₂	SO ₂	пыль
$\sqrt{V(Y_i)}$	0,609	0,549	0,571	0,555
ПНЗ-2	0,648	0,538	0,705	0,514
ПНЗ-5	0,532	0,468	0,396	0,598
ПНЗ-7	0,614	0,609	0,435	0,633
ПНЗ-11	0,635	0,572	0,680	0,454

Собственные значения $V(Y_i)$ первой главной компоненты можно интерпретировать как осредненную по территории города долю вклада синхронных изменений в изменчивость составляющих временных рядов. Для удобства анализа в табл. 2-4 приведен квадратный корень из собственных значений. При этом появляется возможность сравнения коэффициентов корреляции для конкретных пунктов и конкретных загрязнителей с некоторым средним уровнем по городу.

Характерной особенностью для всех загрязнителей является увеличение синхронности изменений по мере увеличения временного масштаба составляющей. Наибольшей связью по территории города обладают поля оксида углерода, диоксида серы и пыли, за исключением изменения концентрации пыли на ПНЗ-7, что можно объяснить влиянием местных особенностей.

Большая связность между первыми компонентами и полями концентраций загрязнителей по территории города, особенно для составляющих больших временных масштабов, указывает на формирование структурной зависимости загрязнителей атмосферы г. Томска и возможность прогноза составляющих загрязнения. Значения коэффициентов корреляции указывают предельный уровень надежности, который могли бы иметь прогнозы составляющих концентрации по городу в целом.

Литература

- 1 Афифи А., Айзис С. Статистический анализ. Подход с использованием ЭВМ. М. Мир, 1982. 488 с.
- 2 Груза Г.В., Рейтенбах Р.Г. Статистика и анализ гидрометеорологических данных. Л. Гидрометеоиздат, 1982. 216 с.
- 3 Дэвис Дж. Статистический анализ данных в геологии. Кн. 2. М. Недра, 1990. 428 с.
- 4 Каждан А.Б., Гуськов О.И. Математические методы в геологии. М. Недра, 1990. 252 с.
- 5 Казакевич Д.И. Основы теории случайных функций в задачах гидрометеорологии. Л. Гидрометеоиздат, 1989. 232 с.
- 6 Кендалл М. Дж., Стюарт А. Многомерный статистический анализ и временные ряды. М. Наука, 1976. 736 с.
- 7 Уданова Е.С., Забелин Б.Н. Методы корреляционного и регрессионного анализа в агрометеорологии. Л. Гидрометеоиздат, 1990. 208 с.