Чепасов В.И., Рычко О.К., Никонорова О.А.
Оренбургский государственный университет

ОБУСЛОВЛЕННОСТЬ ТЕМПЕРАТУРЫ ВОЗДУХА ПРИЛИВНЫМИ ИЗМЕНЕНИЯМИ СИЛЫ ТЯЖЕСТИ

Детерминированными методами строится матрица исследования с усредненными по часовым сечениям приливными изменениями силы тяжести (поправками) и исследуемыми параметрами. Статистическими методами определяется качественная и количественная обусловленность параметров. Приводятся полиномиальные и гармонические модели прогноза.

Для определения обусловленности была построена матрица исследования со следующими параметрами-столбиками: 1 – месяцы, 2 – поправка в момент времени = 1, 3 – поправка в момент времени = 2

……………………………………

25 – поправка в момент времени = 24

26 – слой стока в Сакмаре (месяцы)

27 – осадки по Сакмаре (месяцы)

28 – Т воздуха (Цельсий) в бассейне Сакмары (месяцы)

29 – испарение а (Е, мм) в бассейне Сакмары (месяцы)

30 – расход воды (куб.м/сек) в Сакмаре (месяцы)

Строчки-наблюдения в матрице исследования – это средние месячные значения приливных изменений силы тяжести (поправки) по временным сечениям 1, 2…24 часа и месячные значения параметров исследования.

В нашем исследовании:

– слой стока в Сакмаре (месяцы); – осадки по Сакмаре (месяцы); – Т воздуха (Цельсий) в бассейне Сакмары (месяцы); – испарение а (Е, мм) в бассейне Сакмары (месяцы);

– расход воды (куб. м/сек) в Сакмаре (месяцы).

Рассматривались четыре матрицы исследования:

1. Месячные значения параметров за 1991 год (размер 12 ∗30).
2. Месячные значения параметров за 1992 год (размер 12 ∗30).
3. Месячные значения параметров за 1991, 1992 годы (размер 24 ∗30).
4. Месячные значения параметров за 1991-1995 годы (размер 60 ∗30).

Для определения парных обусловленностей был проведен корреляционный анализ для всех матриц исследования [1].

Предположим, что производится выборка из случайных величин X и У, в результате чего получается N пар их наблюденных значений. Коэффициент корреляции можно оценить по этим выборочным значениям следующим образом:

$$r_{x,y}=\frac{\sum_{i=1}^{N}(X_i-XS)\cdot(Y_i-YS)}{\left[\sum_{i=1}^{N}(X_i-XS)^2\cdot\sum_{i=1}^{N}(Y_i-YS)\right]}, \qquad (1)$$

где XS,YS – выборочные средние.

Выборочный коэффициент корреляции r_{xy} лежит в пределах между -1 и +1. Граничные значения достигаются только в том случае, когда наблюдения обнаруживают идеальную линейную зависимость. Если же зависимость отлична от линейной и (или) наблюдается разброс измеренных значений, то независимо от того, обусловлено ли это обстоятельство ошибками измерений или нелинейным характером связи исследуемых величин, коэффициент уменьшается.

Для того чтобы оценить точность полученной оценки r_{xy}, целесообразно ввести в рассмотрение следующую функцию коэффициента r_{xy}:

$$W=(1/2)\ln[(1+r_{xy})/(1-r_{xy})]. \qquad (2)$$

Как известно, случайная величина W приближенно подчиняется гауссовскому распределению со средним значением:

$$\mu_W=(1/2)\ln[(1+\rho_{xy})/(1-\rho_{xy})] \qquad (3)$$

и дисперсией:

$$\sigma^2_W=1/(N-3) \qquad (4)$$

Зная оценку r_{xy}, нетрудно найти на основании этих соотношений доверительные интервалы для коэффициента ρ_{xy}.

Область принятия гипотезы о равенстве коэффициента корреляции нулю определяется неравенством:

$$-Z_{\alpha/2}\le\sqrt{(N-3)}/2*\ln[(1+r_{xy})/(1-r_{xy})]\le Z_{\alpha/2}, \qquad (5)$$

где $Z_{\alpha/2}$ – величина, подчиняющаяся нормированному гауссовскому распределению. Если рассматриваемая величина W лежит вне приведенного интервала, то это означает наличие корреляционной зависимости при уровне значимости α.

Результаты корреляционного анализа для 1991 года

Таблица 1. Корреляционная матрица R (фрагмент). Параметр 28 - Т воздуха (Цельсий) (месяцы)

0,20	-0,93	-0,88	-0,82	-0,73	-0,59	-0,36	-0,01	0,43	0,78
0,95	0,99	0,98	0,93	0,87	0,78	0,68	0,54	0,33	0,02
-0,40	-0,76	-0,92	-0,97	-0,97	0,10	0,19	1,00	0,91	0,10

Значимые коэффициенты парной корреляции для параметра 28: с параметрами 2, 3, 4, 5, 10, 11, 12, 13, 14, 15, 16, 22, 23, 24, 25, 29.

Для определения групповых обусловленностей между параметрами исследования был использован факторный анализ [4, 5].

В факторном анализе основным предположением является равенство:

$$X_i = \sum_{r=1}^{k} a_{ir} \cdot F_r + e_i \text{ , } (i = 1, 2,...,p), \qquad (6)$$

где X – i-я переменная,

F_r – r-й фактор,

a_{ir} – факторная нагрузка,

k – количество факторов,

e_i – остатки, которые представляют источники отклонений, действующие только на X_i.

Матрицу факторных нагрузок A находим методом главных компонент.

Алгоритм метода главных компонент:

1. Расчет корреляционной матрицы R:

$$r_{jk} = S_{jk} / (\sqrt{S_{jj}} \cdot \sqrt{S_{kk}}), \qquad (7)$$

где

$$S_{jk} = \sum_{i=1}^{n} (X_{ij} - T_j) \cdot (X_{ik} - T_k) -$$

$$-1/n \sum_{i=1}^{n} (X_{ij} - T_j) \cdot \sum_{i=1}^{n} (X_{ik} - T_k), \; T_j = \left(\sum_{i-1}^{n} X_{ij} \right) / n$$

i = 1, 2,..., n – наблюдения,

j = 1, 2,..., m – переменные.

2. Вычисление собственных значений, собственных векторов корреляционной матрицы.

3. Вычисление накопленных отношений собственных значений корреляционной матрицы, больших или равных заданной пользователем константе.

4. Вычисление матрицы факторных нагрузок по собственным значениям и соответствующим собственным векторам корреляционной матрицы.

5. Ортогональное вращение матрицы факторов.

Результаты факторного анализа для 1991 года

Таблица 2. Сумма квадратов нагрузок по факторам

Номер фактора	Сумма квадратов нагрузок
1	16,841
2	9,141
3	2,254
4	1,004

В первом факторе объединились параметры под номерами (смотри выше): 2, 3, 4, 5, 6, 11, 12, 13, 14, 15, 16, 17, 18, 23, 24, 25, 28, 29.

Для определения количественных обусловленностей параметров исследования использовались полиномиальные модели, построенные по разработанному алгоритму упрощенного метода Брандона [3].

При решении задач, связанных с отысканием оптимальных условий протекания сложных многопараметрических процессов, широкое распространение получили полиномиальные математические модели процесса [2]

$$y = b_0 + \sum_{i=1}^{m} b_i \cdot X_i + \sum_{i<j}^{m} b_{ij} \cdot X_i \cdot X_j + ... \qquad (8)$$

где y – параметр оптимизации;

b_0, b_i, b_{ij}, b_{ii} – выборочные коэффициенты регрессии, полученные по результатам эксперимента;

X_i, X_iX_j – параметры и их взаимодействия, i, j = 1, 2…

Упрощенный метод определения коэффициентов уравнения регрессии (8) предложен в работе Д. Брандона [3].

Разработанный алгоритм упрощенного метода Д. Брандона

Пусть задан процесс в виде матрицы наблюдений:

$$\begin{pmatrix} y_1 X_{11} ... X_{j1} ... X_{m_1} \\ .... \\ y_n X_{1n} ... X_{j1} ... X_{mn} \end{pmatrix},$$

где m – число независимых параметров;

n – число наблюдений;

y – зависимый параметр;

X – матрица наблюдений независимых параметров.

1. Последовательно для всех пар массивов y-X_1, y-X_2, …, y-X_m вычислить остаточную дисперсию $D_{(k-1)1}$, $D_{(k-1)2}$, …, $D_{(k-1)m}$, характеризующую

степень влияния независимого параметра на зависимый y.

Для этого воспользуемся ортогональными многочленами Чебышева.

Искомая зависимость условного среднего y от X_j имеет вид:

$$y_k = \varphi_j(X_j) = \sum_{a=0}^{K} b_a \cdot S_a(X_j), \tag{9}$$

где $S_k(X_j)$ – ортогональный многочлен Чебышева k-го порядка.

$S_0(X_j) = 1$ – полином нулевой степени.

Общая дисперсия:

$$D_0 = 1/n\sum_{i=1}^{N}(y_i - b_0)^2, \tag{10}$$

где $b_0 = 1/n\sum_{i=1}^{n} y_i$, $S_1(X_j)=X_j+f_1$,

$$f_i = -1/n\sum_{i=1}^{n} X_{ij}$$

$$b_1 = \sum_{i=1}^{n} y_i \cdot S_1(X_{ji}) / \sum_{i=1}^{n}[S_1(X_{ji})]^2,$$

$$y_k=b_0 \cdot S_0(X_j)+b_1 \cdot S_1(X_j)$$

Остаточная дисперсия:

$$D_{kj} = 1/n\sum_{i=1}^{n}(y_i - y_{ki})^2,$$

где k = 1.

Если $D_{(k-1)j}/D_{kj} \le F_{kp}$, где F_{kp} – табличное значение для распределения Фишера, то осуществляется переход к следующему j.

Иначе k = k+1 и переход к пункту 2.

2. Вычисляем:

$$S_k(X_j)=(X_j+B_k) \cdot S_{k-1}(X_j)+g_k \cdot S_{k-2}(X_j),$$

где

$$B_k = -\sum_{i=1}^{n} X_{ji} [S_{k-1}(X_{ji})]^2 / \sum_{i=1}^{n}[S_{k-1}(X_{ji})]^2, \tag{11}$$

$$g_k = \sum_{i=1}^{n} X_{ji} \cdot S_{k-2}(X_{ji}) \cdot (X_{ji}) / \sum_{i=1}^{n}[S_{k-2}(X_{ji})]^2$$

$$y_k=y_{k-1}+b_k \cdot S_k(X_j),$$

$$b_k = \sum_{i=1}^{n} y_i \cdot S_k(X_{ji}) / \sum_{i=1}^{n}[S_k(X_{ji})]^2$$

Если $D_{(k-1)j}/D_{kj} \le F_{kp}$, то запоминается $D_{(k-1)j}$, иначе k = k+1 и возврат к пункту 2 и так до тех пор, пока $D_{(k-1)j} \le F_{kp} D_{kj}$ для всех j=1, 2,..., m.

3. Выбрать из всех $D_{(k-1)j}$ минимальное, где j=1, 2,..., m.

Если минимальных остаточных дисперсий получится несколько равных между собой, то выбрать последнюю при переборе.

4. Если $D_{(k-1)j} = D_n$, начальной промежуточной дисперсии, то конец решения. Иначе перейти к пункту 5.

5. Произвести как бы разглаживание поверхностей отклика в направлении переменной X_j, вычитая из выборочных значений y_i величины, рассчитанные по $\varphi_j(X_j)$.

Сформировать массив:

$$y_{1i}=y_i-\varphi_j(X_j)$$

6. Заменить в матрице наблюдений массив y_i на y_{1i}.

7. Повторить вычисления с пункта 1 до пункта 6 для матрицы наблюдений с учетом замененных массивов y_i на y_{1i} до тех пор, пока последняя остаточная дисперсия $D_{(k-1)j} = D/r$ уменьшится в r раз по сравнению с начальной общей дисперсией величины y.

Результаты регрессионного анализа

Модель – 28 – Т воздуха (Цельсий) (месяцы) за 1991 год: значимый по вкладу параметр – 12 – поправка в момент времени = 11

$$y = +(-24{,}26081)+(0{,}3432284) \cdot (x\ 12) \tag{12}$$

Таблица 3. Характеристики модели

ХАРАКТЕРИСТИКИ МОДЕЛИ	ЗНАЧЕНИЯ
Коэффициент детерминации	0,99
Средняя абсолютная ошибка	1,48
Средняя ошибка в процентах	11,23

По модели (12) был сделан прогноз температуры воздуха на 1992 год:

Таблица 4. Прогноз температуры

Месяцы 1992 года	Прогнозируемое значение	Истинное значение
Январь	-12.77	-8.9
Февраль	-11.68	-11.4
Март	-4.19	-7.1
Апрель	6.79	6.1
Май	17.86	13.6
Июнь	22.94	18.5
Июль	22.27	19
Август	17.98	17.1
Сентябрь	13.37	14.7
Октябрь	6.09	4.2
Ноябрь	-1.82	-1.6
Декабрь	-8.32	-10.8

Средняя абсолютная ошибка при прогнозе – 2,212185.

С целью проверки правильности полиномиальных моделей и с целью временного прогноза был использован спектральный анализ [1].

Пусть периодический процесс представлен функцией $f(t)$, кусочной или непрерывной на отрезке [-T/2, T/2] с периодом T.

Тогда процесс может быть представлен следующим тригонометрическим рядом:

$$f(t)=a_0/2+\sum_{k=1}^{\infty}(a_k*\cos(2\pi kt/T)+b_k*\sin(2\pi kt/T)),\quad(13)$$

где

$$A_k=(2/T)*\int_{-T/2}^{T/2}f(t)*\cos(2\pi kt/T)*dt,\ (k=0,1,2,3,...)\quad(14)$$

$$b_k=(2/T)\int_{-T/2}^{T/2}f(t)*\sin(2\pi kt/T)dt,\ (k=1,2,3,...)$$

Амплитудный спектр:

$$A_k=\sqrt{a_k^2+b_k^2},\ (k=0,1,2,3...)\quad(15)$$

Фазовый спектр:

$$\varphi_k=-\mathrm{arctg}(b_k/a_k),\ (k=1,2,...)\quad(16)$$

Фурье-анализ [1] подтвердил своими спектральными характеристиками правильность количественных оценок, полученных по регрессионным моделям.

Результаты спектрального анализа для 1991 года

Температура воздуха (Цельсий) имеет с поправкой в момент времени -11 максимальный коэффициент корреляции (модуль) -0,99 (истинное значение R= 0,99).

Таблица 5. Спектральные характеристики по 1991 году

Параметры	Период процесса	Максимальная по амплитуде гармоника	Период гармоники	Фаза гармоники
Т воздуха	12	1	12	-3,694
Поправка-11	12	1	12	-3,664

Модель по максимальной гармонике для температуры воздуха (Цельсий):

Y(t) = 6.9352+(16.3085)*cos((2*pi*t/T)*1+(-3.6939)) (17)

где pi = 3,14....T – период процесса = 12.00.

Модель по максимальной гармонике для поправки в момент времени -11:

Y(t) = 84.9698+(58.9308)*cos((2*pi*t/T)*1+
+(-3.6638)) (18)

где pi = 3,14....T – период процесса = 12.00.

Аналогичные результаты были получены по остальным матрицам исследования.

Таким образом, мы можем по приливным изменениям силы тяжести осуществлять прогноз температуры воздуха как по полиномиальным моделям, так и по гармоническим.

Список использованной литературы:
1. Бендат Д. Ж., Пирсол А. Измерение и анализ случайных процессов. – М.: Мир, 1974.
2. Драйпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Статистика, 1973.
3. Brandon D. B. Developing Mathematical Models for Computer Control, USA Journal, 1959, V.S, N7.
4. Харман Г. Современный факторный анализ. – М.: Сатистика, 1972.
5. Иберла К. Факторный анализ. – М.: Статистика, 1980.