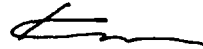


На правах рукописи



КОЛЕСНИК Алексей Николаевич

**МЕТОДОЛОГИЧЕСКИЕ СХЕМЫ
ОПРЕДЕЛЕНИЯ КАЧЕСТВЕННЫХ И КОЛИЧЕСТВЕННЫХ
ОБУСЛОВЛЕННОСТЕЙ ПАРАМЕТРОВ ТЕХНОГЕНЕЗА**

25.00.36 - Геоэкология

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Оренбург-2004

Работа выполнена на кафедре прикладной информатики Государственного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет»

Научные руководители: доктор технических наук, доцент
Чепасов Валерий Иванович,

доктор географических наук, профессор
Рычко Олег Константинович

Официальные оппоненты: доктор медицинских наук, профессор
Греков Иван Иванович,

кандидат технических наук, доцент
Матвейкин Игорь Витальевич

Ведущая организация: ГОУ ВПО «Оренбургская государственная медицинская академия Министерства здравоохранения РФ»

Защита диссертации состоится «27» мая 2004 в ¹¹⁰⁰ часов на заседании диссертационного совета КР212.181.13в Государственном образовательном учреждении высшего профессионального образования «Оренбургский государственный университет» по адресу: 460352, Оренбург, проспект Победы 13, ауд. *CZO\$*

С диссертацией можно ознакомиться в библиотеке Государственного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет»

Автореферат разослан «26» апреля 2004г.

Ученый секретарь
диссертационного совета



Тарасова Т.Ф.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Окружающая среда и человеческий организм — многопараметрические и многофакторные объекты, тесно связанные и взаимодействующие между собой. Такое взаимодействие в случаях нарушения человеком окружающей среды, в результате сверхнормативного техногенеза может иметь отрицательные последствия, в том числе и для жизнеобеспечения населения, предопределяя его заболеваемость.

В связи с этим необходимо определение обусловленности факторов человеческого организма биологически значимыми параметрами окружающей среды.

Поскольку параметры окружающей среды и факторы человеческого организма определяются посредством либо дискретного, либо непрерывного измерения, то для оценки обусловленности могут использоваться детерминированные подходы, связанные со спектральными методами; и статистические, связанные с корреляционным и факторным анализами и с моделями нелинейной регрессии. Исключительно важно определение параметров объектов исследования по их спектральным характеристикам, что связано с периодичностью изменчивости компонентов окружающей среды, и факторов человеческого организма.

В связи с чем, представляется актуальным исследование влияния слабых воздействий окружающей среды (солнечная активность, магнитная активность, приливные изменения силы тяжести и т.д.) на отдельные группы населения в условиях техногенеза.

Временной характер изменения заданных параметров предопределяет проблему локального и регионального прогноза по различным видам онкозаболеваний (новообразований), включая количественные оценки качественных показателей для осуществления классификаций заболеваемости. Кроме этого методически важно оценить результаты объединений параметров факторным анализом по физическим вкладам в регрессионных моделях этих параметров.

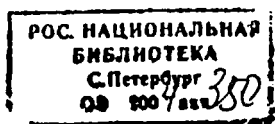
Решение указанных вопросов позволит значительно упростить определение обусловленности параметров, оптимизировать модели прогнозирования и сократить время проведения экспресс-анализа параметров техногенеза окружающей среды.

Целью диссертационной работы является геоэкологическое, математическое моделирование и разработка системы методов для анализа и прогноза техногенно обусловленных новообразований среди взрослого населения районов Оренбургской области.

Для достижения поставленной цели были решены следующие задачи:

- выполнен анализ состояния изученности обозначенной проблемы, в том числе компонентов окружающей среды и распространенности новообразований среди населения региона исследования;

- проведен выбор антропогенных параметров окружающей среды и факторов, характеризующих состояние здоровья населения города Оренбурга и районов Оренбургской области;



- разработаны математические модели, описывающие корреляционные связи между параметрами окружающей среды и распространенностью новообразований;
- предложена методика прогнозирования новообразований по основным природным и техногенным параметрам;
- выявлены качественные связи между параметрами измеряемого сигнала и его спектральными характеристиками, с использованием метода главных компонент;
- обоснована возможность применения упрощенного алгоритма ступенчатого регрессионного метода для построения полиномиальных моделей искомых параметров;
- разработан комплекс программ для реализации метода главных компонент и ступенчатого регрессионного метода;
- выполнены экспериментальные проверки моделей прогнозирования и оценка применимости разработанных методов в других субъектах РФ.

Объектом исследования является система: «техногенные параметры окружающей среды - новообразования».

Методика исследования - комплексная, содержащая методы: системного анализа, математической статистики, теории цифровой обработки исходных данных, математического моделирования и другие.

Научная новизна работы заключается в следующем:

- предложен упрощенный алгоритм ступенчатого регрессионного метода;
- дана оценка вкладов как результатов варимаксного вращения матрицы факторных нагрузок;...
- рассмотрено приближение геоэкологических процессов рациональными полиномами..

Наиболее существенные результаты, полученные лично автором:

- разработан метод оценок вкладов в полиномиальных моделях;
- предложена количественная оценка качественных признаков для факторов человеческого организма и параметров окружающей среды;
- разработан алгоритм уплотнения матрицы факторных нагрузок в случае ее вырождения.

Достоверность научных положений, выводов и рекомендаций работы подтверждается:

- теоретическими исследованиями обусловленности новообразований качественными и количественными параметрами техногенеза;
- сопоставлением результатов прогнозирования новообразований по полученным моделям с фактической диагностикой медицинскими учреждениями г.Оренбурга..

Практическую ценность работы составляют:

- разработанные методы, алгоритмы и программы для построения полиномиальных моделей;
- возможность более оперативного получения моделей заданных параметров по сравнению с использованием быстрого преобразования ряда Фурье;

- методика расчета критических точек приливных изменений силы тяжести;
- прогноз периодов проведения требуемых операций;
- получение моделей прогноза на множестве обусловленных параметров;
- проведение косинор-анализа на временных массивах показателей;
- методика проведения экспресс-анализа на базе разработанного программного комплекса.

Внедрение результатов исследования:

- в учреждениях государственной санитарно-эпидемиологической службы Оренбургской области;
- в Оренбургской областной онкологической больнице для прогнозирования радиорезистентного рака молочной железы.
- в Оренбургском государственном университете в учебных курсах: «Математическое моделирование», «Геоэкологический мониторинг».

Основные положения, выносимые на защиту :

- методика адаптированного использования корреляционного анализа в геоэкологии;
- модификация метода главных компонент для определения групповых обусловленностей параметров и факторов исследования;
- алгоритм упрощенного метода Брандона;
- оценка количественной обусловленности изучаемых параметров;
- модели прогнозирования новообразований по районам Оренбургской области.

Апробация работы. Основные результаты диссертационной работы доложены и обсуждались на: семинаре Оренбургской Государственной Медицинской Академии (г.Оренбург, 1999г.); научно-практической конференции «Новые подходы к оценке реактивности организмов: математическое моделирование, функциональная диагностика, клиника, влияние геофизических факторов» (г.Оренбург, 2001г.); научно-практической конференции «Корреляции биологических и физико-химических процессов с солнечной активностью и другими факторами окружающей среды» (г.Оренбург, 2000г.); полученные прогностические модели прогноза были использованы для прогноза новообразований в 2000-2002 годах, результаты которого не превысили 12% по сравнению с фактическими значениями.

Публикации. Основные результаты диссертационной работы изложены в 6 печатных работах, в том числе одном методическом пособии.

Структура и объем диссертации. Диссертация состоит из введения, пяти глав и заключения, изложенных на 160 страницах машинописного текста и включает 17 таблиц, 2 рисунка. Список литературы содержит **101** источник.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ.

В первой главе рассмотрено построение моделей факторов человеческого организма и параметров окружающей среды с использованием ряда Фурье, быстрого преобразования ряда Фурье, линейных регрессионных моделей.

Обосновано применение факторного анализа, ступенчатого регрессионного метода для построения моделей и определения вкладов в них спектральных характеристик.

Для ускорения процесса построения моделей более адекватных истинным, автором предлагается следующий алгоритм обработки данных:

- совместно аппаратно реализовать сигналы, соответственно их параметры, амплитудные и фазовые спектры.

- осуществить цифровую запись параметров и спектров.

- определить связи между параметрами и гармониками спектров с использованием факторного анализа, метода главных компонент.

- по найденным связям построить ступенчатым регрессионным методом полиномиальные модели с оптимальной степенью аппроксимирующих полиномов.

- определить по построенным моделям вклады параметров в амплитудные и фазовые спектры.

- по остаточной и начальной дисперсии найти коэффициент детерминации.

В представленном алгоритме обработки отсутствует использование быстрого преобразования ряда Фурье, что существенно снижает затраты машинного времени на обработку. Кроме этого, модели, полученные ступенчатым регрессионным методом, имеют более высокий коэффициент детерминации, чем линейные регрессионные модели.

Во второй главе рассмотрен один из основных методов факторного анализа - метод главных компонент. Он совпадает с методом расчленения ковариационной или корреляционной матрицы на совокупность ортогональных векторов (компонент) или направлений по числу рассматриваемых переменных.

Эти векторы соответствуют собственным векторам и собственным значениям корреляционной матрицы. По этому методу собственные значения выделяются в порядке убывания их величины, что становится существенным, если для описания данных используется лишь незначительное число компонент.

Векторы попарно ортогональны, и компоненты, полученные по ним, некоррелированы, хотя несколько компонент могут выделить большую часть суммарной дисперсии переменных, однако для точного воспроизведения корреляций между переменными требуются все компоненты.

В тех случаях, когда применяется метод главных компонент, не нужно выдвигать никаких гипотез о переменных, они не обязаны быть даже случайными.

Факторный анализ в противоположность методу главных компонент заранее объясняет матрицу ковариаций наличием минимального или по крайней мере небольшого числа гипотетических переменных или факторов. В то время, как метод главных компонент ориентирован на дисперсии, факторный анализ ориентирован на ковариаций (или на корреляционную связь).

В факторном анализе основным предположением является равенство:

$$X_i = \sum_{r=1}^k a_{ir} \cdot F_r + e_i, (i=1,2,\dots,p) \quad (1)$$

где X_i - переменная; F_r - r -ый фактор; a_{ir} - факторная нагрузка; k - количество факторов; e_i - остатки, которые представляют источники отклонений, действующие только на X_i .

Эти p случайных величин e_i предполагаются независимыми как между собой, так и с k величинами F_r .

Уравнение (1) нельзя проверить непосредственно, поскольку p переменных X_i выражены в них через $(p+k)$ ненаблюдаемых переменных.

Но эти уравнения заключают в себе гипотезу о ковариациях и дисперсиях X_i , которую можно проверить.

Когда число факторов $k > 1$, то ни факторы, ни нагрузки не определяются, однозначно, поскольку в уравнении (1) факторы F_r могут быть заменены любым ортогональным преобразованием их с соответствующим преобразованием нагрузок. Это свойство использовано для преобразования или вращения факторов, полученных в каком-либо практическом исследовании.

Вращение подбирается так, чтобы переменные, которые в большей или меньшей степени измеряют некоторые легко опознаваемые стороны, имели бы достаточно высокие нагрузки на один фактор и нулевые или почти нулевые на другие факторы.

Формализация метода главных компонент заключается в том, что если отправной точкой является корреляционная матрица R с единицами на главной диагонали, то говорят о компонентном анализе, чья модель отлична от модели классического факторного анализа и приводит к дескриптивным факторам. Если же в матрице R используют оценки общностей, то получают модель факторного анализа.

Классическая модель факторного анализа имеет вид:

$$R = A \cdot C \cdot A', \quad (2)$$

где R - корреляционная матрица, A - матрица факторных нагрузок, C - корреляционная матрица, отражающая связи между факторами, A' - транспонированная матрица факторных нагрузок.

Если наложить условие некоррелированности факторов, т. е. $C = I$, где I - единичная матрица, то в результате получим:

$$R = A \cdot A'. \quad (3)$$

Система уравнений, соответствующая (3), имеет однозначное решение с вводом дополнительных условий, а именно: сумма квадратов нагрузок первого фактора должна составлять максимум от полной дисперсии; сумма квадра-

тов нагрузок второго фактора должна составлять максимум оставшейся дисперсии и т. д., т. е. максимизирует функцию:

$$s_1 = \sum_{i=1}^m a_{i1} = \max \quad (4)$$

при $m(m-1)/2$ независимых друг от друга условиях

$$r_{ik} = a_{i1} * a_{k1} \quad (i, k = 1, 2, \dots, m, i < k), \quad (5)$$

где m - число переменных в матрице наблюдений.

Для максимизирования функции, связанной некоторым числом дополнительных условий, используем метод множителей Лагранжа.

В результате приходим к системе m однородных уравнений с m неизвестными a_{i1} :

$$\begin{aligned} (1-\lambda)a_{11} + r_{12} * a_{21} + \dots + r_{1m} * a_{m1} &= 0 \\ r_{21} * a_{11} + (1-\lambda)a_{21} + \dots + r_{2m} * a_{m1} &= 0 \\ \dots & \\ r_{m1} * a_{11} + r_{m2} * a_{21} + \dots + (1-\lambda)a_{m1} &= 0 \end{aligned} \quad (6)$$

Система равенств (6) составляет проблему собственных значений действительной симметричной матрицы.

В общем она записывается в следующем виде:

$$R\alpha_e = \lambda * \alpha_e, \quad (R - \lambda I)\alpha_e = 0, \quad (7)$$

где λ - собственные значения, они соответствуют собственным векторам α_e матрицы R .

Факторы пропорциональны собственным векторам матрицы R .

Путем нормирования получим искомые значения α_e матрицы A по компонентам собственных векторов матрицы R :

$$a_{ie} = \alpha_{ie} \cdot \frac{\sqrt{\lambda_e}}{\sqrt{\alpha_{1e}^2 + \alpha_{2e}^2 + \dots + \alpha_{me}^2}} \quad (8)$$

После определения матрицы факторных нагрузок A для лучшей интерпретации факторов используем вращение A в пространстве общих факторов.

В настоящее время наиболее наиболее распространен «метод варимакс» для осуществления вращения матрицы A . Метод варимакс был предложен Кайзером.

Алгоритм метода главных компонент.

1. Расчет корреляционной матрицы R :

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}} \cdot \sqrt{S_{kk}}}, \quad (9)$$

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

$i=1, 2, \dots, n$ - наблюдения,

$j=1, 2, \dots, m$ - переменные.

2. Вычисление собственных значений, собственных векторов корреляционной матрицы.

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

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

5. Ортогональное вращение матрицы факторов.

В третьей главе изложен ступенчатый регрессионный метод, упрощенный алгоритм этого метода, его программная реализация. Разработанный метод дает возможность определить автоматически оптимальную степень аппроксимирующего полинома, коэффициент детерминации модели, вклады параметров в модели для зависимой переменной, значение остаточной дисперсии.

В четвертой главе, содержатся результаты исследований по построению моделей прогнозирования онкозаболеваемости населения г. Оренбурга в связи с содержанием в воздухе, воде и почвах микроэлементов, пестицидов, а также некоторых неорганических веществ.

Схема определения обусловленности заболевания раком легкого параметрами техногенеза окружающей среды в городе Оренбурге за период 1970-1995г: на нормализованной матрице наблюдений с параметрами -столбиками (таблица 1) был проведен корреляционный анализ, результаты которого представлены в таблице 2.

Как видно из приведенных в таблице 2 значений коэффициентов парной корреляции, количество заболевших раком легкого наиболее сильно коррелирует с параметрами-столбиками под номерами 16,21,26,39,50(таблица 1).

Однако проведенный на той же матрице наблюдений факторный анализ (метод главных компонент с варимаксным вращением матрицы факторных нагрузок) дал следующее объединение по фактору новообразования и техногенных параметров окружающей среды (таблица 3).

Из таблицы 3 следует, что количество заболевших раком легкого обусловлено техногенными параметрами окружающей среды под номерами 10, 21, 31, 44 (таблица 1).

На основании объединения по фактору ступенчатым регрессионным методом была построена модель для прогноза количества заболевших раком легкого по параметрам техногенеза окружающей среды, объединившихся в одном факторе с новообразованием, а по модели были получены следующие вклады параметров-аргументов (таблица 4).

Модель для y (заболеваемость раком легкого)

$$\begin{aligned} y = & 6930,97 - 459191 * x_{10}^1 + 9643680 * x_{10}^2 - 62147400 * x_{10}^3 - 42408,7 + \\ & + 2256360 * x_{21}^1 - 39321600 * x_{21}^2 + 223909000 * x_{21}^3 - 70272,2 + \\ & + 210419 * x_{31}^1 - 208240 * x_{31}^2 + 68163,1 * x_{31}^3 + 180467 - 2604730 * x_{44}^1 + \\ & + 12506800 * x_{44}^2 - 19976200 * x_{44}^3 \end{aligned} \quad (10)$$

Достаточно высокий коэффициент детерминации (таблица 5) говорит о хорошей адекватности модели истинной зависимости на исследуемом поле параметров, в связи с чем модель можно использовать для прогноза.

Таблица 1 – Параметры исследования

Номер п/п	Название параметра	Единица измерения	Характер
1	Заболеласмость раком легкого		Прогнозируемый
2	Азот при сгорании бензина	% (объемный)	Параметр-аргумент
3	Кислород при сгорании бензина	% (объемный)	Параметр-аргумент
4	Пары воды при сгорании бензина	% (объемный)	Параметр-аргумент
5	Диоксид углерода при сгорании бензина	% (объемный)	Параметр-аргумент
6	Оксид углерода при сгорании бензина	% (объемный)	Параметр-аргумент
7	Оксиды азота при сгорании бензина	% (объемный)	Параметр-аргумент
8	Угледороды при сгорании бензина	% (объемный)	Параметр-аргумент
9	Альдегиды при сгорании бензина	% (объемный)	Параметр-аргумент
10	Сажа при сгорании бензина	% (объемный)	Параметр-аргумент
11	Прометрин (арборицид)	Мг/кг почвы	Параметр-аргумент
12	Хлорамп (арборицид)	Мг/кг почвы	Параметр-аргумент
13	Хлорофос (инсектицид)	Мг/кг почвы	Параметр-аргумент
14	Карбофос (инсектицид)	Мг/кг почвы	Параметр-аргумент
15	Гексахлорциклогексан (инсектицид)	Мг/кг почвы	Параметр-аргумент
16	Гамма-изомер гексахлорана (инсектицид)	Мг/кг почвы	Параметр-аргумент
17	Полихлорпинен (инсектицид)	Мг/кг почвы	Параметр-аргумент
18	Полихлоркалфен (инсектицид)	Мг/кг почвы	Параметр-аргумент
19	Алкилсульфонаты в воде(питьевой)	Мг/л	Параметр-аргумент
20	Акриловая кислота в воде(питьевой)	Мг/л	Параметр-аргумент
21	Анизол в воде(питьевой)	Мг/л	Параметр-аргумент
22	Анилин в воде (питьевой)	Мг/л	Параметр-аргумент
23	Аммиак в воде(питьевой)	Мг/л	Параметр-аргумент
24	Ацетон в воде(питьевой)	Мг/л	Параметр-аргумент
25	Ацетальдгид в воде(питьевой)	Мг/л	Параметр-аргумент
26	Бензол в воде (питьевой)	Мг/л	Параметр-аргумент
27	Железо в воде(питьевой)	Мг/л	Параметр-аргумент

Таблица 2 – Корреляционная матрица R (заболеваемость раком легкого)

1,00	0,30	0,40	-0,28	0,26	0,37	0,39	-0,09	0,18	-0,21
0,28	-0,05	0,03	-0,17	0,44	-0,52	-0,18	-0,11	-0,04	0,05
-0,57	-0,13	0,26	-0,33	0,27	0,66	-0,38	-0,22	-0,45	0,26
-0,28	0,07	-0,19	0,16	-0,21	-0,16	-0,16	0,03	-0,52	-0,12
0,12	0,28	-0,15	-0,38	0,06	0,08	-0,21	-0,11	-0,26	0,73
0,48	-0,34	-0,20	-0,08	0,53	0,02	-0,45	-0,43	0,06	-0,46
0,18									

Таблица 3 - Объединение по фактору 10

Номер	Название параметра	Нагрузка
1	(заболеваемость раком легкого)	0,6062
10	(сажа при сгорании бензина в % по объему)	-0,7315
21	(анизол в воде мг/л)	-0,8825
31	(медь в воде мг/л)	-0,6801
44	(аммиак в воздухе мг/м ³)	-0,5776

Схема определения обусловленности заболевания раком крови параметрами техногенеза окружающей среды в городе Оренбурге за период 1970-1995г.

На нормализованной матрице наблюдений с параметрами-столбиками (в таблице 1 под номером 1. - рак крови в количестве заболевших человек, далее те же параметры окружающей среды) был проведен корреляционный анализ, который дал следующие значения коэффициентов парной корреляции новообразования с представленными параметрами окружающей среды (таблица 6).

Факторный анализ дал следующее объединение новообразования с техногенными параметрами окружающей среды (таблица 7).

На основании объединения по фактору была построена регрессионная модель и по ней определены вклады параметров-аргументов (таблица 8).

Таблица 4 - Вклады параметров-аргументов в модели (10)

Номер	Название параметра	Вклад в модель
10	(сажа при сгорании бензина в % по объему)	0,5025774
21	(анизол в воде мг/л)	0,0759868
31	(медь в воде мг/л)	0,3098558
44	(аммиак в воздухе мг/м ³)	0,1115800

Таблица 5 - Характеристики модели (10)

Характеристики модели	Значения
Коэффициент детерминации	0,87
Средняя абсолютная ошибка	32,21
Средняя ошибка в процентах	19,47

Таблица 6 - Корреляционная матрица R (заболеваемость раком крови)

Номер	Название параметра	Нагрузка
1	(заболеваемость раком крови)	0,6212
7	(оксиды азота при сгорании бензина в % по объему)	0,9036
16	(гамма-изомер гексахлорана(инсектицид) мг/кг почвы)	-0,7306
34	(никель в воде мг/л)	0,5597
39	(хром в воде мг/л)	-0,8266
52	(кислота серная в воздухе мг/м ³)	-0,8923
55	(медь оксид в воздухе мг/м ³)	0,7127

Таблица 7 - Объединение по фактору 3

Номер	Название параметра	Нагрузка
1	(заболеваемость раком крови)	0,6212
7	(оксиды азота при сгорании бензина в % по объему)	0,9036
16	(гамма-изомер гексахлорана(инсектицид) мг/кг почвы)	-0,7306
34	(никель в воде мг/л)	0,5597
39	(хром в воде мг/л)	-0,8266
52	(кислота серная в воздухе мг/м ³)	-0,8923
55	(медь оксид в воздухе мг/м ³)	0,7127

- Модель для у(заболеваемость раком крови)

$$\begin{aligned}
 y = & 868,440 - 3970,24 * x_7^1 + 5782,57 * x_7^2 - 2716,88 * x_7^3 + \\
 & + 301,319 - 251,329 * x_{16}^1 - 906,564 + 12296,2 * x_{34}^1 + \\
 & + 30032,0 * x_{34}^2 - 622219 * x_{34}^3 + 455,251 - 7842,48 * x_{39}^1 + \\
 & + 35014,7 * x_{39}^2 - 24316,9 * x_{39}^3 - 623,064 + 10547,2 * x_{52}^1 - \\
 & - 42663,7 * x_{52}^2 - 327,117 + 307115 * x_{55}^1
 \end{aligned} \quad (11)$$

Высокий коэффициент детерминации (таблица 9) предполагает успешное использование модели для прогноза.

Схема определения обусловленности заболевания раком пищевода параметрами техногенеза окружающей среды в городе Оренбурге за период 1970-1995г.

На нормализованной матрице наблюдений с параметрами-столбиками (в таблице 1 под номером 1. - рак пищевода в количестве заболевших человек, далее те же параметры окружающей среды) был проведен корреляционный анализ, который дал следующие значения коэффициентов парной корреляции новообразования с представленными параметрами окружающей среды (таблица 10).

Факторный анализ выявил следующее объединение (таблица 11).

По построенной регрессионной модели определили вклады параметров-аргументов (таблица 12).

Модель для у(заболеваемость раком пищевода)

$$\begin{aligned}
 y = & 552,553 - 13935,0 * x_{30}^1 + 103713 * x_{30}^2 - 190412 * x_{30}^3 + \\
 & + 13005,2 - 39472500 * x_{37}^1 + 39803000000 * x_{37}^2 - 1331200000000 * x_{37}^3 + \\
 & + 127,195 - 291,343 * x_{41}^1 + 162,671 * x_{41}^2 + 5,16917 * x_{41}^3 + \\
 & + 8897,76 - 133594 * x_{44}^1 + 665871 * x_{44}^2 - 1102170 * x_{44}^3.
 \end{aligned} \quad (12)$$

В пятой главе содержатся результаты исследований по построению моделей для прогнозов онкозаболеваемости населения 5 сельских районов Оренбургской области (Александровский, Тоцкий, Красногвардейский, Шарлыкский и Сорочинский районы).

Схема определения обусловленности новообразований радиоактивностью и микроэлементами в Тоцком районе.

$$\text{Модель для } y \text{ (онкозаболеваемость)} \quad y = 312,949 + 0,0157012 * x_1. \quad (13)$$

$$\text{Модель для } y \text{ (оикосмертность)} \quad y = 233,949 + 0,0184518 * x_2. \quad (14)$$

Модель для у (заболеваемость раком) Щитовидная железа

$$\begin{aligned}
 y = & 7,92172 - 0,00252542 * x_1 + 0,000000191583 * x_1^2 + 10,9643 - 0,00623111 * x_2 + \\
 & + 0,000000964550 * x_2^2 - 187538 + 3616810000 * x_{37} - 23251300000000 * x_{37}^2 + \\
 & + 4982560000000000 * x_{37}^3.
 \end{aligned} \quad (15)$$

Таблица 8 - Вклады параметров-аргументов в модели (11)

Номер	Названия параметра	Вклад в модель
7	(оксиды азота при сгорании бензина в % по объему)	0,0075152
16	Гамма-изомер гексахлорана(инсектицид) мг/кг почвы)	1,0149660
34	(никель в воде мг/л)	0,2250772
39	(хром в воде мг/л)	0,2252426
52	(кислота серная в воздухе мг/м ³)	0,0324239
55	(медь оксид в воздухе мг/м ³)	0,0400399

Таблица 9 - Характеристики модели (11)

Характеристики модели	Значения
Коэффициент детерминации	0,95
Средняя абсолютная ошибка	7,94
Средняя ошибка в процентах	17,32

Таблица 10 - Корреляционная матрица (заболеваемость раком пищевода)

1,00	0,26	-0,09	-0,33	0,12	0,37	0,08	0,34	-0,04	-0,21
0,26	0,25	0,10	0,23	-0,04	-0,40	-0,27	-0,44	0,14	-0,20
0,06	-0,20	-0,17	-0,56	0,18	0,01	0,02	0,08	-0,49	0,31
-0,08	0,17	0,05	-0,05	-0,39	-0,39	0,65	-0,39	-0,24	-0,60
0,29	0,22	0,52	-0,47	0,36	-0,10	0,08	0,34	0,04	-0,04
-0,22	-0,03	-0,04	0,32	0,13	0,41	0,02	0,32	0,35	-0,17
0,22									

Таблица 11 - Объединение по фактору 12

Номер	Названия параметра	Нагрузка
1	(заболеваемость раком пищевода)	0,7043
30	(керосин в воде мг/л)	0,5693
37	(фенол в воде мг/л)	0,6453
41	(цинк в воде мг/л)	0,6990
44	(аммиак в воздухе мг/м ³)	-0,5146

Таблица 12 - Вклады параметров-аргументов в модели (12)

Номер	Названия параметра	Вклад в модель
30	(керосин в воде мг/л)	0,0967044
37	(фенол в воде мг/л)	0,8660473
41	(цинк в воде мг/л)	0,2361671
44	(аммиак в воздухе мг/м ³)	0,1989187

Таблица 13 - Характеристики модели (12)

Характеристики модели	Значения
Коэффициент детерминации	0,94
Средняя абсолютная ошибка	1,46
Средняя ошибка в процентах	6,08

Таблица 14 - Вклады параметров-аргументов в модели (13)

Номер	Названия параметра	Вклад в модель
1	Плутоний, в почве	0,9830
2	Цезий 137, в почве	0,0313
57	Ве-снег	0,0053
80	Sr-волосы	0,0091

Таблица 15 - Характеристики модели (13)

Характеристики модели	Значения
Коэффициент детерминации	0,95
Средняя абсолютная ошибка	5,54
Средняя ошибка в процентах	1,84

Таблица 16 - Вклады параметров-аргументов в модели (14)

Номер	Названия параметра	Вклад в модель
2	Цезий 137, в почве	1,0111
57	Ве-снег	0,0111

Таблица 17 - Характеристики модели(14)

Характеристики модели	Значения
Коэффициент детерминации	0,96
Средняя абсолютная ошибка	4,73
Средняя ошибка в процентах	2,67

Таблица 18 - Вклады параметров-аргументов в модели(15)

Номер	Названия параметра	Вклад в модель
1	Плутоний, в почве	0,9206
2	Цезий 137, в почве	0,3278
50	Ni-снег	0,0013
57	Ве-снег	0,2736
60	Ga-снег	0,0239
72	Zn-волосы	0,0003
80	Sr-волосы	0,0022
114	Zn-молоко	0,0001

Таблица 19-Характеристики модели(15)

Характеристики модели	Значения
Коэффициент детерминации	0,95
Средняя абсолютная ошибка	0,62
Средняя ошибка в процентах	10,26

ЗАКЛЮЧЕНИЕ

1. Учитывая, что факторы человеческого организма и параметры техногенеза окружающей среды могут определяться посредством измерений, предлагается новая методика исследования детерминированных процессов с использованием статистических методов для определения обусловленности заболеваний и смертности параметрами окружающей среды.

2. В результате анализа известных методов установлено, что обычная форма нахождения функциональных зависимостей параметров измерительных сигналов от их амплитудного и фазового спектра связана с использованием быстрого преобразования ряда Фурье и линейных регрессионных моделей. Этот путь достаточно трудоемок по машинным затратам и не приводит к моделям, адекватным истинным зависимостям, в случае нелинейности последних. Кроме этого, в ныне применяемом подходе остается открытым вопрос о гармониках амплитудного и фазового спектров, наиболее сильно связанных с исследуемым параметром.

3. В настоящей работе показано новое, более эффективное применение метода главных компонент для выделения гармоник амплитудного и фазового спектров, наиболее сильно влияющих на тот или иной параметр измерительного сигнала. Этот метод, несмотря на переход в пространство гипотетических переменных (факторов), позволил выделить связи между заданными параметрами и спектральными характеристиками.

По реализации параметров и связанных с ними гармоник спектров были построены с помощью ступенчатого регрессионного метода полиномиальные модели спектров и параметров, для чего автором был разработан алгоритм упрощенного ступенчатого регрессионного метода с автоматическим выбором оптимальной степени аппроксимирующего полинома.

Представление установленных спектров полиномиальными моделями дало возможность определить вклады параметров в рассматриваемые гармоники спектра, дать более точную количественную оценку связи между параметрами и выделенными гармониками по сравнению с методом главных компонент.

При построении исходных матриц для метода главных компонент и ступенчатого регрессионного метода не требуется применять быстрое преобразование ряда Фурье, что существенно снижает машинное время обработки (в 5-6 раз) по сравнению с существующей.

4. Программная реализация факторного анализа осуществлялась с использованием широко распространенного пакета SSP американской фирмы IBM. Автором разработана полезная подпрограмма по определению параметров, объединившихся в каждом факторе.

Достаточно просто построена подпрограмма по реализации ступенчатого регрессионного метода, использование которой позволяет получать все необходимые для практического использования параметры: коэффициент детерминации, коэффициенты разложения, ошибки и т. д.

На основе использования программных реализаций факторного анализа

и ступенчатого регрессионного метода, определены связи между параметрами часто используемых сигналов и их спектров, построены их полиномиальные модели и модели параметров.

5. Наряду с построением полиномиальных моделей параметров для изучения динамики их изменения от значений спектральных характеристик показано использование ступенчатого регрессионного метода для приближений сигналов во временной области с достаточно высоким коэффициентом детерминации.

Преимуществом полиномиальных моделей является их достаточно простая схемная реализация, что делает возможным использование таких моделей во многих электронных и радиотехнических приборах.

Указанная методика оказалась достаточно эффективной при исследовании малых воздействий приливных изменений силы тяжести на характер изменения контролируемых параметров в медицине, биофизике, геоэкологии, в частности при определении обусловленности новообразований техногенными

окружающей среды, по разработанной методике были найдены параметры окружающей среды, наиболее значимые для определенного вида заболеваемости и смертности от новообразований. Получены модели прогнозирования новообразований на основе использования наиболее связанных параметров окружающей среды по различным районам области, по городу Оренбургу и в целом по Оренбургской области .

Список основных публикаций:

1. Колесник А.Н., Смирных Т.А. Распределение больных злокачественными новообразованиями по полу и возрасту//Оренб. гос. уни-т, Оренбург, 2001 бс./Рукопись деп. в ВИНТИ 16.08.2002 г. № 1500-B2002/(Доля автора 50%).

2. Колесник А.Н., Смирных Т.А. Число больных с впервые в жизни установленным диагнозом злокачественного новообразования в г.Оренбурге (на 100000 жителей)//Оренб. гос. уни-т, Оренбург, 20014с./Рукопись деп. в ВИНТИ 16.08.2002 г. № 1501-B2002/(Доля автора 50%).

3. Колесник А.Н., Смирных Т.А. Прогноз заболеваемости населения Оренбургской области злокачественными новообразованиями.//Оренб. гос. унит, Оренбург, 2001 7с./Рукопись деп. в ВИНТИ 16.08.2002 г. № 1502-B2002 / (Доля автора 50%).

4. Колесник А.Н., Смирных Т.А. Прогноз смертности от злокачественных новообразований населения Оренбургской области//Оренб. гос. уни-т, Оренбург, 2001 4с./Рукопись деп. в ВИНТИ 16.08.2002 г. № 1503-2002 / (Доля автора 50%).

5. Чепасов В.И., Колесник А.Н., Смирных Т.А. Новообразования и окружающая среда УОренбург: ИПК ГОУ ОГУ, 2001 - 150с, Методическое пособие. (Доля автора 30%).

6. Рычко О.К., Чепасов В.И., Колесник А.Н. Алгоритм минимизации количества параметров исследования в геоэкологии //Вестник Оренбургского государственного университета. №3, 2004. - С. 13-18 (Доля автора 30%).

Формат 60x84 $\frac{1}{16}$, гарнитура «Тайме»
Усл. печ. листов 1,0. Тираж 100 экз, Заказ 129.

РИК ОГУ
460352, г. Оренбург, пр. Победы 13,
Оренбургский государственный университет.

8737