

ГЕОИНФОРМАЦИОННЫЙ КОМПЛЕКС АНАЛИЗА СОСТОЯНИЯ ОКРУЖАЮЩЕЙ СРЕДЫ НА ОСНОВЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Ю.М. Полищук

Институт химии нефти СО РАН

Т.О. ПЕРЕМИТИНА

Институт химии нефти СО РАН

8 января 2004 г.

The methodological problems of geoinformation approach to environmental analysis and its algorithmic and program developments are considered. Geoinformation program complex, which have been realized with CIS-technologies, is described. Two samples of this approach application to environmental problems solving are shown.

1. Введение

При изучении свойств и поведения природных объектов приходится использовать большие объемы разнородной информации, в связи с чем в задачах анализа состояния окружающей среды возникает потребность в сжатии информации. Одним из перспективных подходов можно считать подход, основанный на методе главных компонент (МГК), позволяющий описывать объекты меньшим числом обобщенных показателей - главными компонентами. Последние являются удобными интегральными показателями, так как они отражают внутренние закономерности, которые не поддаются непосредственному наблюдению. При использовании метода главных компонент (МГК), корреляционная матрица используется как исходная ступень для дальнейшего анализа и появляется возможность извлечения дополнительной информации об исследуемом объекте, проведения причинного анализа взаимосвязей характеристик и определения их стохастической связи с главными компонентами [1].

Существует широкий класс природных объектов, при анализе которых необходимо учитывать пространственные свойства, связанные с заданием формы, взаиморасположения объектов и др. Включение пространственных свойств в анализ вносит разнородность в информационный массив и требует проведения комплексного анализа данных на основе методов пространственного анализа, в реализации которых широкое распространение получили геоинформационные системы (ГИС) [2,3], позволяющие манипулировать и управлять пространственными данными, хранящимися в виде тематических слоев, географически определенных относительно карты-основы.

В наших работах [4,6-8] рассмотрены методические вопросы комплексного анализа многомерных данных о природных объектах на основе сочетания методов статистического анализа и методов пространственного анализа. Разработаны алгоритмы и программные средства анализа таких данных на основе метода главных компонент с учетом пространственных свойств объектов. Однако в большинство работ по данной тематике опубликованы в трудах научных конференций и других малодоступных изданиях и рассматривают частные вопросы геоинформационного подхода к анализу данных. Это делает необходимым обобщение методических разработок, что и явилось целью данной работы.

2. Методические вопросы реализации геоинформационного подхода

Геоинформационный подход к анализу многомерных данных о пространственно-распределенных объектах, основанный на сочетании компонентного и пространственного анализов, реализуется в виде двух схем (рис. 1).

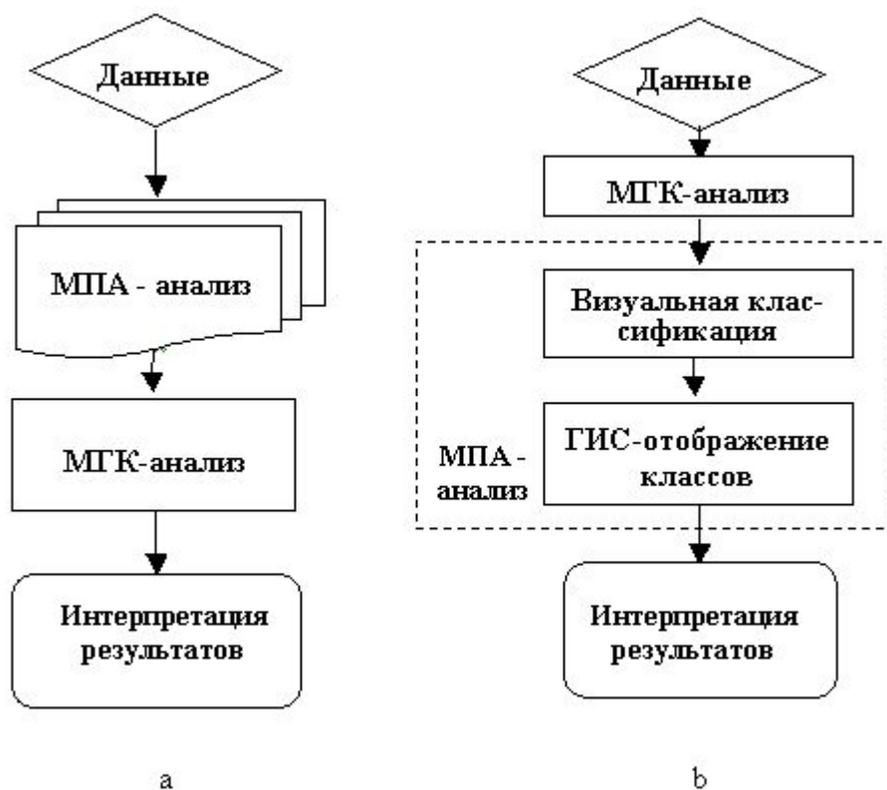


Рис. 1. Схемы комплексного анализа многомерных данных; а — схема П-ГК; б — схема ГК-П

Схема П-ГК (рис. 1а) комплексного анализа предполагает выполнение пространственного анализа на первом этапе и анализ данных методом главных компонент на втором. Результатом пространственного анализа является разделение всех объектов на однородные группы с применением ГИС и цифровых тематических карт. Методом главных компонент анализируются выделенные однородные группы объектов.

Схема ГК-П (рис. 1б) предполагает проведение анализа многомерных данных методом главных компонент на первом этапе и пространственный анализ на втором. В пространстве главных компонент проводится анализ групп объектов исследования.

На рис. 2 приведена схема разработанного алгоритма комплексного анализа многомерных данных, где приняты следующие обозначения: ЦК - цифровая карта; Кг - критерий отбора главных компонент, ГК1 и ГК2 - первая и вторая главные компоненты соответственно. Алгоритм предусматривает возможность проведения анализа по любому из двух направлений, представленных схемами на рис. 1а и 1б.

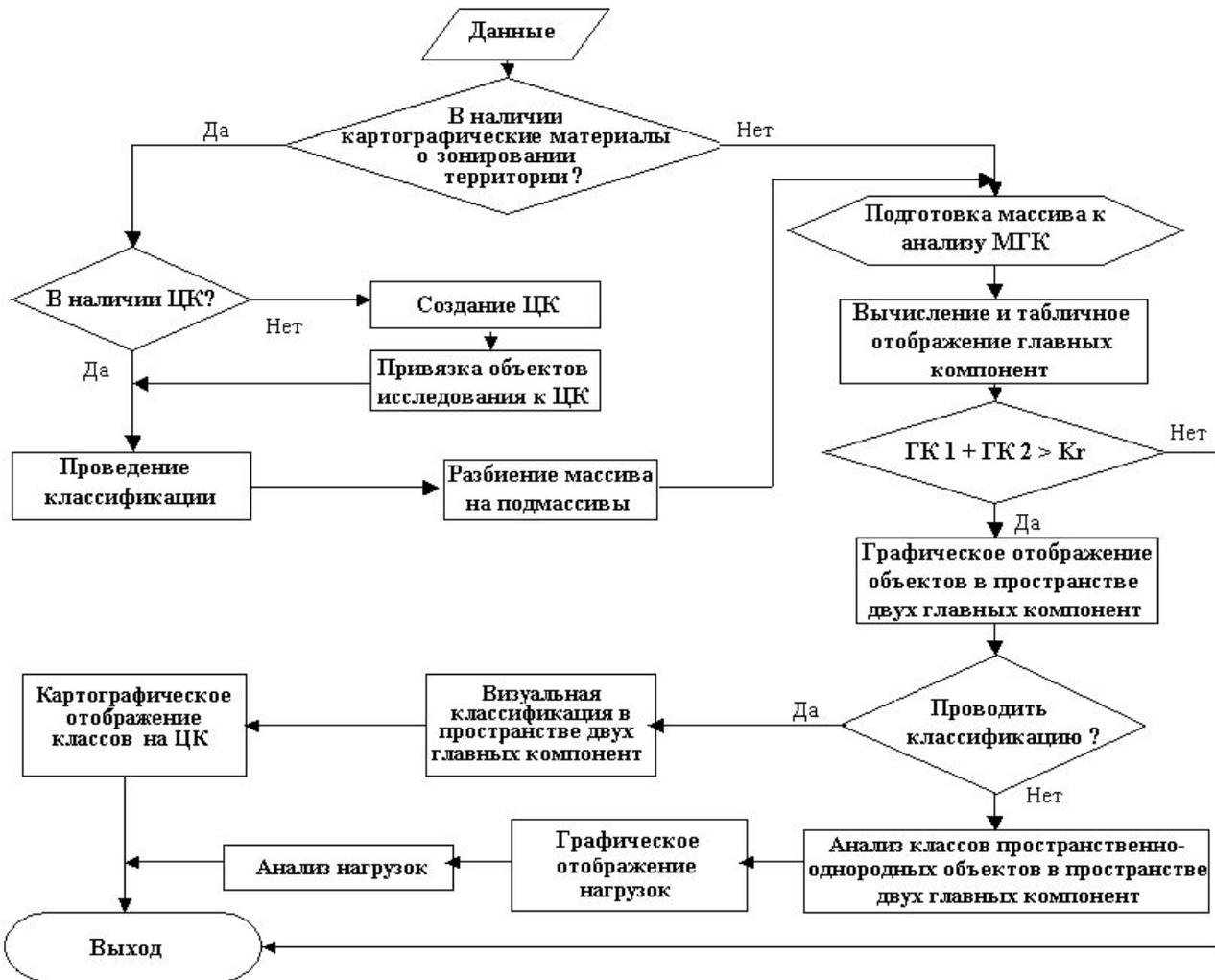


Рис. 2. Алгоритм комплексного анализа

Основные процедуры, используемые в алгоритме, показаны на рис. 2. Наиболее важными из них являются процедуры графического отображения. Предусмотренная в алгоритме процедура «Графическое отображение объектов в пространстве двух главных компонент» предполагает представление результатов МГК-анализа в виде точечных объектов в двумерном пространстве главных компонент ГК1 и ГК2 (рис. 3а). Как правило, свойства природных объектов являются статистически зависимыми между собой, что приводит к наложению областей рассеивания объектов, относящихся к разным группам, что и иллюстрирует рис. 3а. Наложение областей рассеивания затрудняет восприятие и осложняет интерпретацию результатов анализа многомерных данных.

В связи с этим предлагается отображать результаты МГК- анализа с помощью представления среднего значения и доверительной области для каждой группы объектов в пространстве двух главных компонент (рис. 3б). Такое представление результатов позволяет достоверно определить, имеется ли статистически значимое различие между выявленными группами объектов с доверительной вероятностью 90%–95% (ГОСТ 11.001-73: «Прикладная статистика. Ряды предпочтительных значений статистических характеристик»).

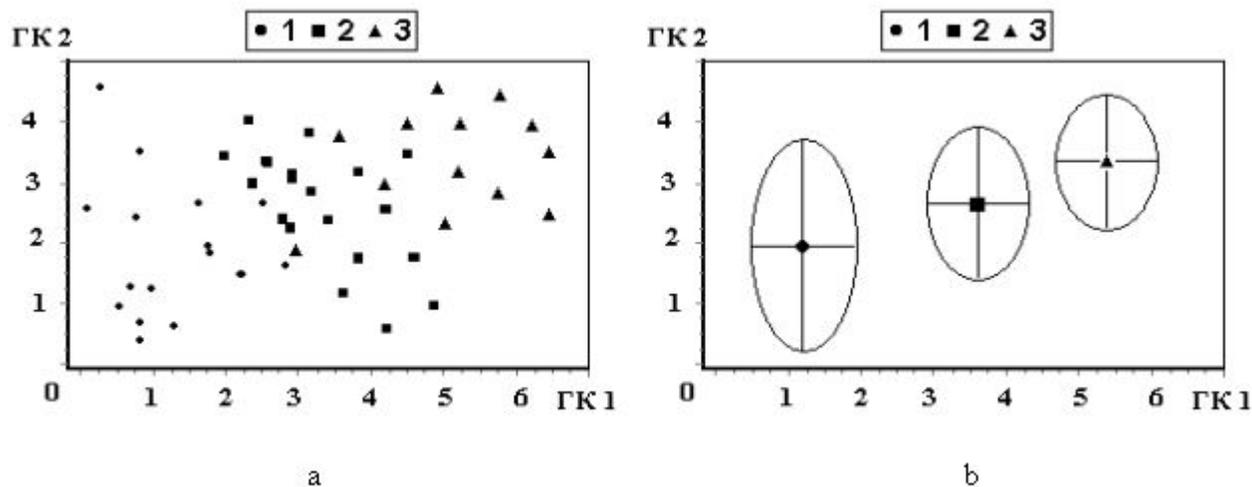


Рис. 3. Графическое отображение результатов комплексного анализа в пространстве двух главных компонент

Применение МГК позволяет не только перейти в пространство меньшей размерности, но и исследовать значения параметров объектов путем изучения нагрузок на главные компоненты. Нагрузками являются координаты собственных векторов, определяемые как проекции наблюдаемых переменных в пространстве главных компонент [5].

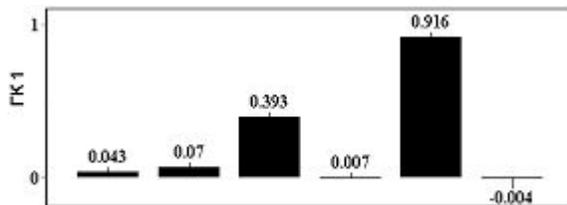


Рис. 4. Графическое представление нагрузок на первую главную компоненту

В рассматриваемом алгоритме (рис. 2) предусмотрена процедура графического представления нагрузок на главные компоненты, что позволяет наглядно и оперативно оценить вклад каждого показателя (характеристики) состояния объекта в значение той или иной главной компоненты. Пример графического представления нагрузок дан на рис. 4, из которого видно, что анализ проводился по шести показателям, пятый из которых дает максимальный вклад (91,6%) в ГК1.

3. Программный комплекс анализа многомерных данных о природных объектах

Описанный выше алгоритм реализован в программном комплексе анализа многомерных данных [6], который включает модуль интерфейсов, модуль пространственного анализа данных, использующий средства ArcView 3.x, модуль анализа методом главных компонент и модуль графического отображения, реализованные в среде Delphi 6. Общая схема программного комплекса представлена на рис. 5.

Модуль анализа методом главных компонент включает программные средства, предназначенные для выполнения следующих операций и процедур:

- диалоговую процедуру открытия данных (таблицы типа dBase);
- процедуры центрирования и стандартизации исходного массива данных;
- вычисление ковариационной матрицы;
- нахождение собственных значений и собственных векторов ковариационной матрицы методом вращения Якоби;
- вычисление главных компонент;
- вычисление вклада каждой главной компоненты в общую дисперсию;
- анализ вкладов главных компонент в общую дисперсию и отбор главных компонент для проведения дальнейшего анализа;
- вычисление проекций исходных наблюдений (нагрузок) на отобранные главные компоненты;
- табличное представление нагрузок на отобранные главные компоненты;

Модуль ПА (модуль пространственного анализа) данных предназначен для выполнения следующих операций:

- подготовка атрибутивной таблицы - создание массива исследуемых данных средствами ArcView 3.x;
- отображение объектов на цифровых картах - создание цифрового слоя исследуемых объектов;
- выявление пространственных группирований объектов - в атрибутивной таблице каждому объекту присписывается номер класса, к которому он отнесен.

Модуль графического отображения включает программные средства, реализующие следующие процедуры:

- графическое отображение объектов исследования в пространстве двух главных компонент;
- графическое представление нагрузок на две главные компоненты;

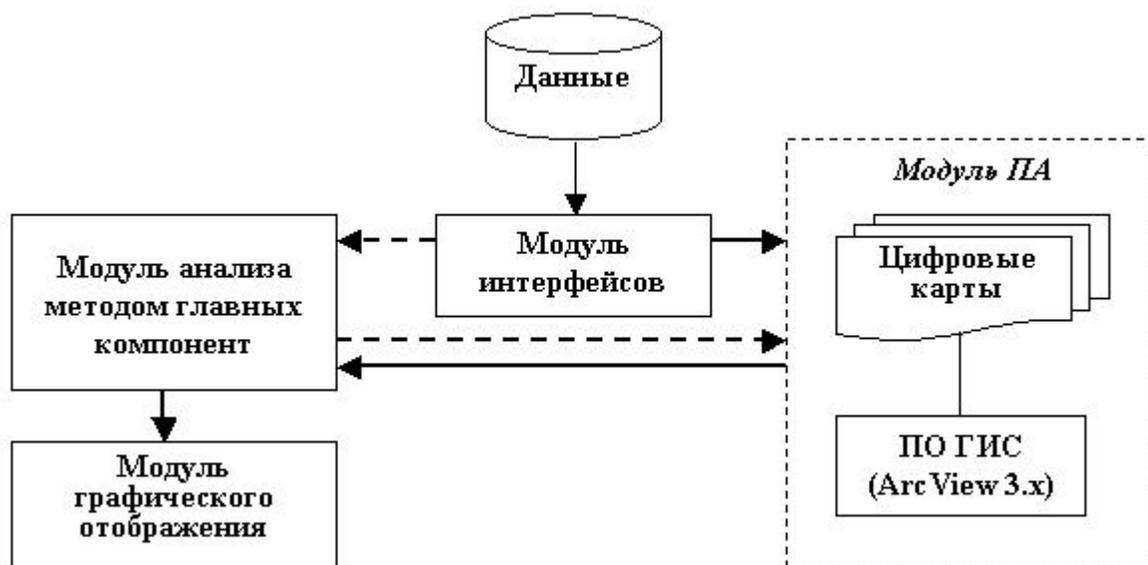


Рис. 5. Структура программного комплекса анализа многомерных данных

- векторное представление нагрузок в пространстве двух главных компонент;
- графическое представление результатов анализа с отображением средних значений и доверительных областей.

Модуль интерфейсов предназначен для осуществления взаимодействия между другими модулями программного комплекса.

4. Результаты комплексного анализа

В данной работе рассматривается применение разработанного подхода и программного комплекса на двух конкретных примерах.

а) Анализ природно-климатического состояния территории юго-востока Западно-Сибирской равнины.

В качестве объекта исследования выбрана территория юга Западно-Сибирской низменности в границах трёх административных областей: Томской, Новосибирской и Омской (рис. 6). Центральную часть этой территории занимает Васюганская равнина, где располагается одна из крупнейших в мире болотных систем - Большое Васюганское болото, которое в последние годы стало объектом комплексных междисциплинарных исследований как российских, так и зарубежных ученых [7].

Для анализа климатических особенностей территории Васюганской равнины были использованы временные ряды среднегодовых значений температуры воздуха и количества атмосферных осадков за период 1955–1992 гг. (по 49 станциям), за период 1955–1985 гг. на 49 станциях Томской, Новосибирской и Омской областях. Сформированный для проведения исследований файл данных среднегодовых значений содержит 1 862 записи.

В настоящей работе пространственный анализ проводился путем наложения границ административных образований и условной границы территории Васюганской равнины на

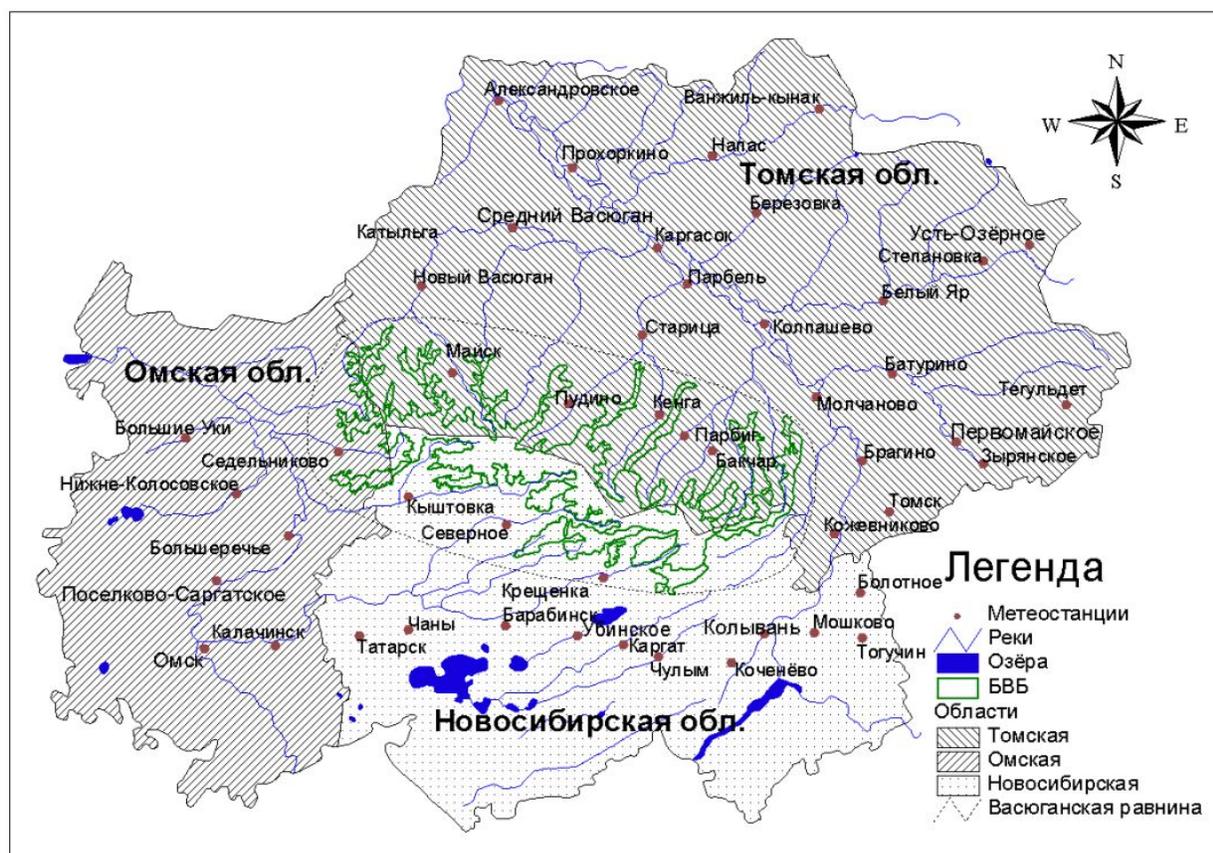


Рис. 6. Расположение метеостанций на исследуемой территории

карту расположения метеостанций (рис. 6). Результатом этого стало разбиение метеостанций на четыре группы в зависимости от их расположения на территориях Васюганской равнины и трёх указанных выше административных областей.

В анализе климатических особенностей территории Васюганской равнины по сравнению с примыкающими к ней территориями Омской, Томской и Новосибирской областей были использованы шесть параметров, полученных по данным наблюдений на 49 метеостанциях, относящихся к территории юго-востока Западно- Сибирской равнины:

1. временные ряды среднегодовых значений температуры воздуха;
2. временные ряды среднегодовых значений количества атмосферных осадков;
3. максимальные годовые значения высоты снежного покрова;
4. максимальные годовые значения плотность снежного покрова;
5. максимальные годовые значения запаса воды в снежном покрове;
6. значения их высоты станции над уровнем моря.

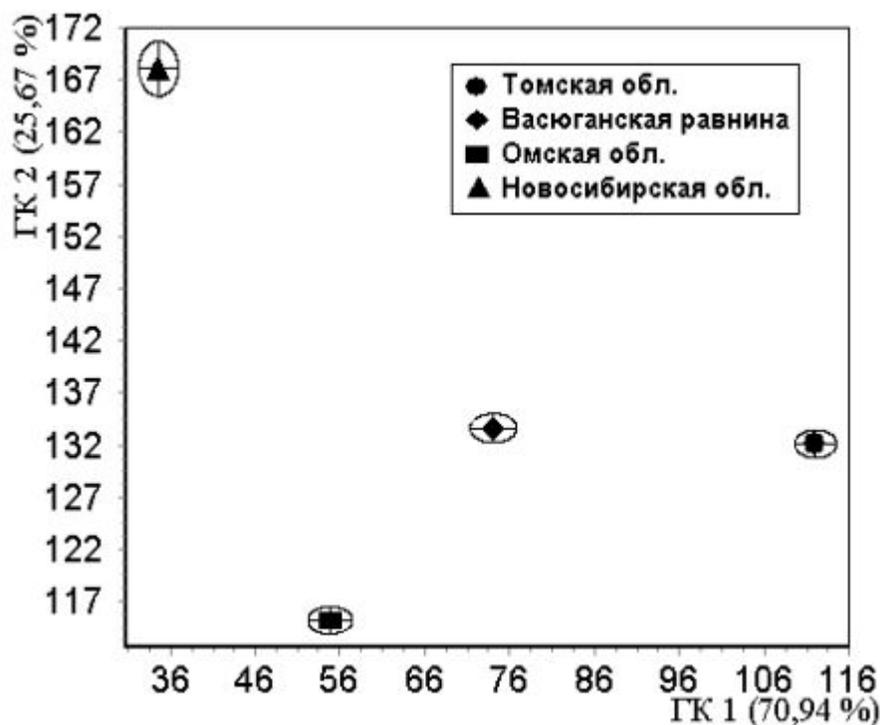


Рис. 7. Отображение доверительных интервалов и средних значений классов объектов исследования в пространстве двух главных компонент ($= 90\%$)

Результаты комплексного анализа данных, отображенных в виде средних значений и доверительных интервалов, показали статистически значимое различие между выявленными четырьмя зонами (рис. 7).

Анализ нагрузок на главные компоненты (рис. 8) позволил выявить особенности в климатическом состоянии Васюганской равнины по сравнению с сопредельными территориями. На первую главную компоненту для Томской, Омской областей и Васюганской

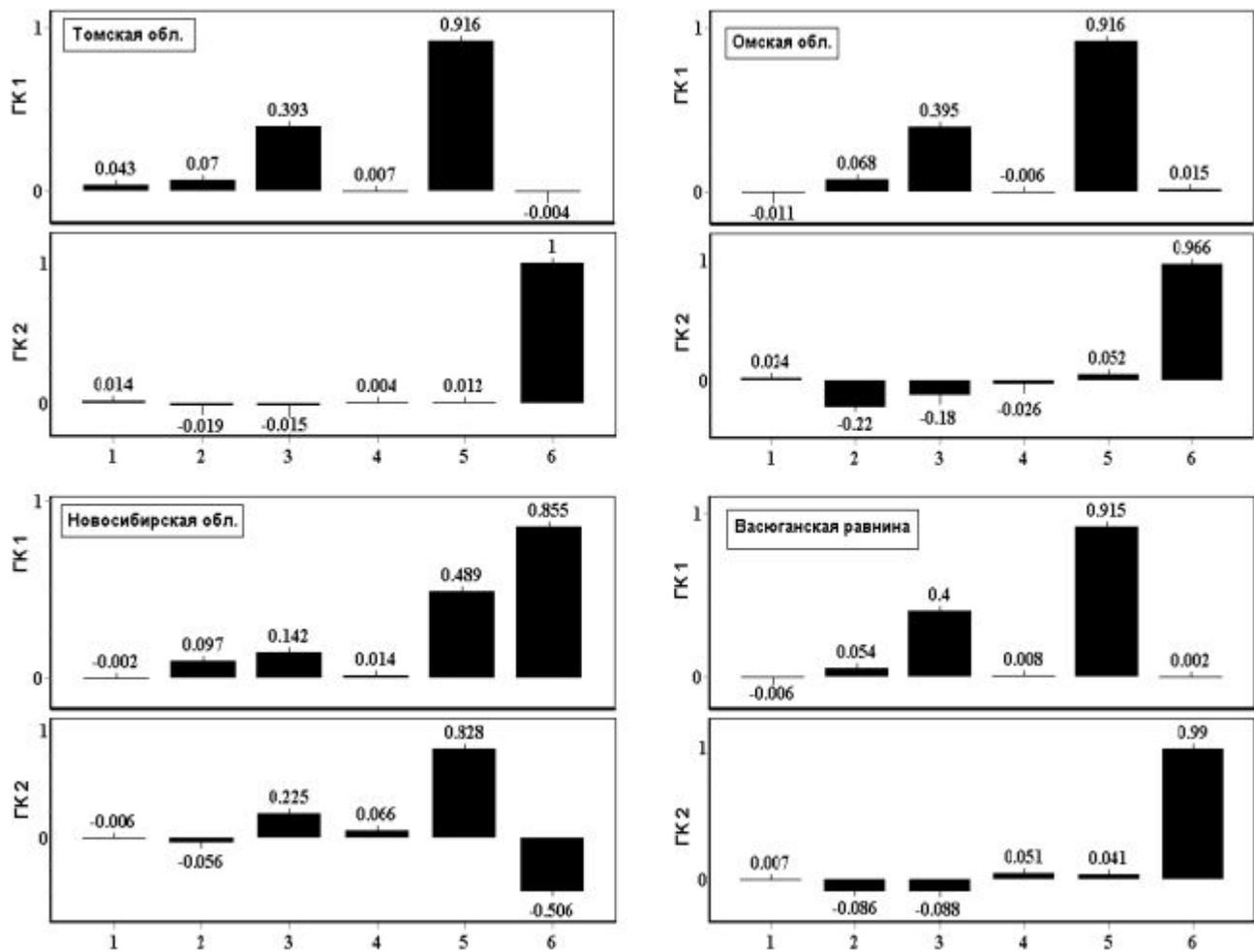


Рис. 8. Нагрузки на главные компоненты для исследуемых территорий

равнины наибольшее значение нагрузки имеет показатель 5 — запас воды в снежном покрове. Однако, хотя для Новосибирской области наибольший вклад даёт показатель 6 — высота станции над уровнем моря, запас воды в снежном покрове также оказывает существенное влияние. Поэтому этот показатель следует рассматривать как существенный для всех исследуемых территорий, однако их отличия в климатическом состоянии определяются различием значений показателя водозаписа в снежном покрове.

б) Анализ данных о радиационном загрязнении территории

Рассматривается территория в окрестностях г. Томска, находящаяся в зоне воздействия крупного предприятия ядерного цикла — Сибирского химического комбината (СХК), крупнейшего в России предприятия по производству оружейного плутония [9]. Анализ радиационного загрязнения территории проводится по данным о содержании радиоактивных веществ в годичных кольцах деревьев [10], произрастающих в зоне воздействия СХК. Целью анализа является выявление закономерностей радиоактивного загрязнения окружающей среды в зависимости от расстояния до источника загрязнения.

Для анализа радиационной обстановки использовались данные за 40-летний период по удельной активности радиоуглерода, трития и цезия в годичных кольцах деревьев, произрастающих в 30-ти км зоне СХК и за её пределами (рис. 9). Деревья из деревень Георгиевка и Наумовка включены в анализ вследствие того, что они оказались «накрыты» радиоактивным облаком, выброшенным СХК во время известной производственной аварии 6 апреля 1993 г. [11].

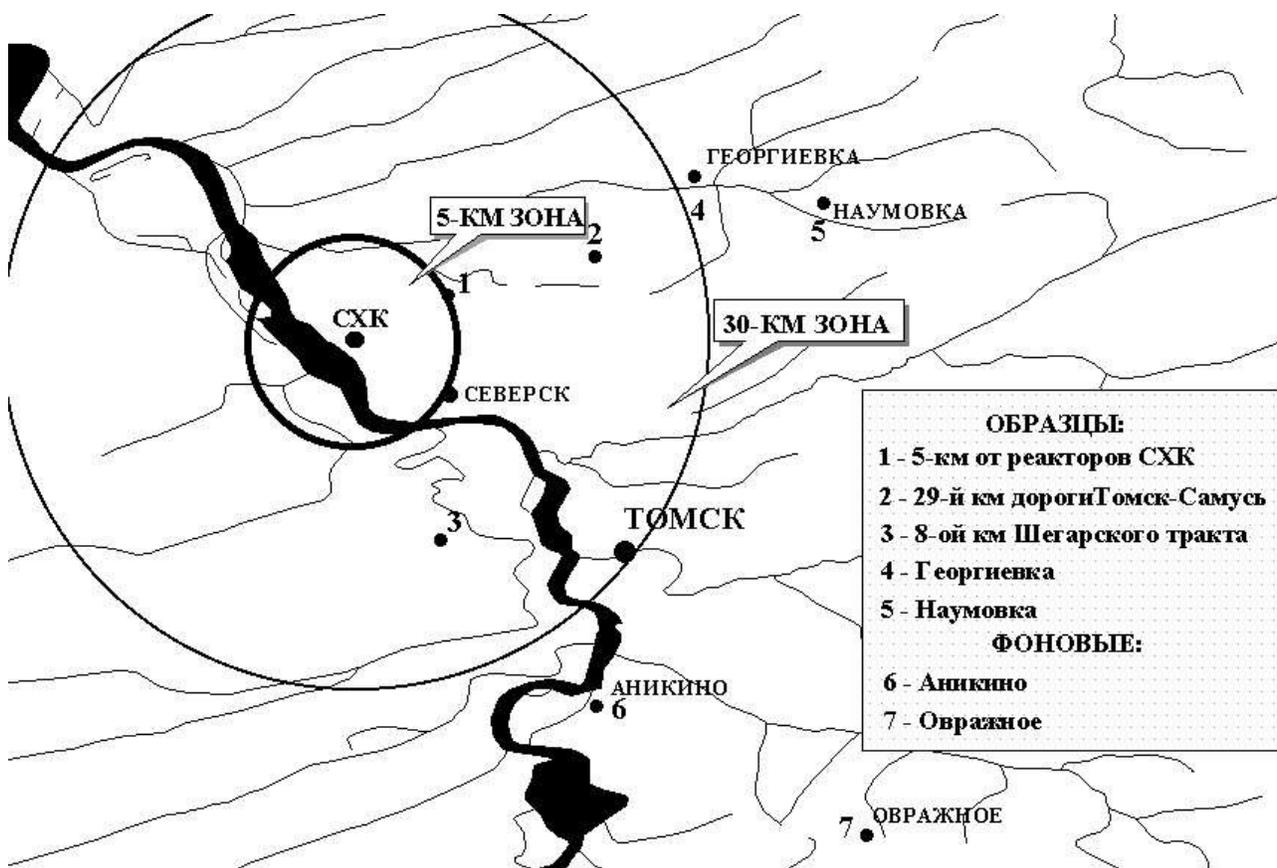


Рис. 9. Пространственное отображение объектов исследования

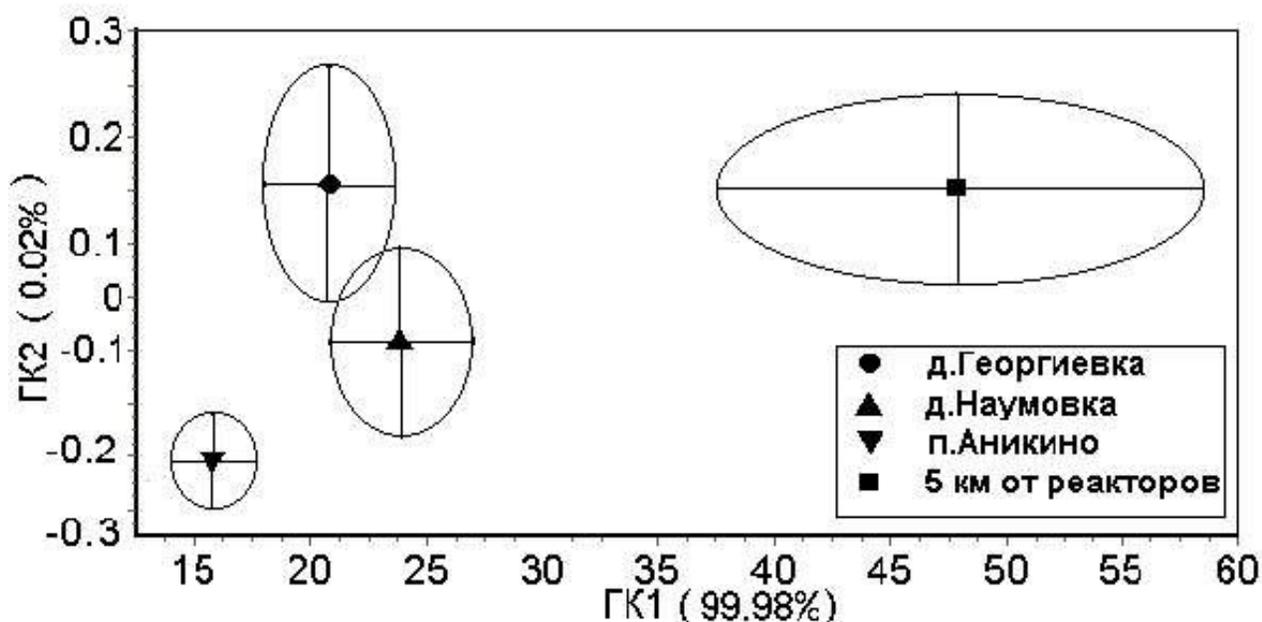


Рис. 10. Средние значения и доверительные интервалы (= 90%) активности трития в образцах древесины годовичных колец деревьев

Результаты проведенного анализа показали (рис. 10), что между значениями концентраций трития в годовичных кольцах исследуемых деревьев имеется статистически значимое различие. Максимальную активность трития имеет сосна, выросшая в 5-км зоне от реакторов СХК. Меньшее количество трития накапливают деревья в районах деревень Георгиевка и Наумовка (за пределами 30-км зоны), а сосна на окраине пос. Аникино (контрольная точка за пределами 30-км зоны) имеет активность трития меньшую, чем в предыдущих случаях.

5. Заключение

При описании природных объектов в задачах окружающей среды и природопользования, как правило, используется большой объем пространственной информации. Традиционный метод главных компонент не позволяет учитывать пространственные свойства объектов, что требует разработки методических, алгоритмических и программных средств его использования в сочетании с методами пространственного анализа, реализуемыми с помощью ГИС-технологиями. В настоящей статье разработан геоинформационный подход к анализу природных объектов на основе сочетания методов пространственного и компонентного анализов. Данный подход проиллюстрирован в работе на примерах анализа данных о природно-климатическом состоянии региона и о радиационном загрязнении территории. При этом были выявлены закономерности уровня концентрации радиоактивных веществ в зависимости от расстояния до источника загрязнения. А в задачах анализа климатических особенностей территорий данный подход позволил выявить пространственные особенности различных территорий, при этом анализ нагрузок переменных на главные компоненты дал возможность установить наиболее значимые климатические характеристики для каждой исследуемой территории.

Геоинформационный подход применим к анализу широкого класса природных объектов. В наших работах рассмотрено применение этого подхода в задачах геологии и геохимии органического вещества осадочных пород [12] и в задачах анализа физико-химических свойств нефтей Западной Сибири [13,14].

Работа выполнена в рамках интеграционных проектов СО РАН № 137, 138, 167 и 168, при поддержке Программы ЕС ИНКО Коперникус 2 (проект ИСИРЕММб контракт ICA2-ST-2000-10024), Программы ИНТАС (проект АТМОС, контракт INTAS-00-189), Программы «Интеграция» (проект Ф0009/1248) и гранта РФФИ — Югра № 03-05-96829.

Список литературы

- [1] ДУБРОВ А.М. Обработка статистических данных методом главных компонент. М.: <Статистика>, 1978. 135 с.
- [2] КОШКАРЕВ А.В., ТИКУНОВ В.С. Геоинформатика. –М.:Картгеоцентр-Геодезиздат, 1993. 213 с.
- [3] ArcView GIS. Environmental Systems Research Institute. Inc. N.Y., 1997. 316 p.
- [4] ПОЛИЩУК Ю.М., ПЕРЕМИТИНА Т.О. Геоинформационный подход к анализу многомерных данных о пространственно-распределенных объектах // Геоинформатика. 2003., № 1. С. 18–21.
- [5] АНДЕРСОН Т. Введение в многомерный статистический анализ. М.: Физматгиз, 1963., 500 с.
- [6] ПЕРЕМИТИНА Т.О., ПОЛИЩУК Ю.М. Комплексный анализ многомерных данных на основе метода главных компонент // Официальный бюллетень Российского агентства по патентам и товарным знакам. М.: 2002., № 3, С. 109–110.
- [7] ПЕРЕМИТИНА Т.О. Программный комплекс обработки многомерных данных с применением метода главных компонент и геоинформационных технологий // Электронный «Журнал радиоэлектроники». 2003, № 3., <http://jre.cplire.ru/jre/mar03/6/text.html>.
- [8] ЗАДДЕ Г.О. ИРОВОДА В.П., ПЕРЕМИТИНА Т.О., ПОЛИЩУК Ю.М. Геоинформационный анализ климатических особенностей территории Васюганского болота на основе метода главных компонент // Большое Васюганское болото. Современное состояние и процессы развития: Сб. науч. трудов., Томск, 2002., С. 134–138.
- [9] Состояние окружающей среды и здоровье населения в зоне влияния Сибирского химического комбината. – Аналитический обзор. Томск: Госкомитет экологии и природных ресурсов Томской области, 1994., С. 83.
- [10] НЕСВЕТАЙЛО В.Д. Дендрохроноиндикация как метод ретроспективного мониторинга. – В кн.: Проблемы исследования и преодоления экологической опасности в промышленном регионе., Кемерово, 1990., С. 108–109.

- [11] Израэль Ю., Артемов Е., Пахомов В., Чириков В. и др. Радиоактивное загрязнение местности в результате аварии на радиохимическом заводе Томске-7 // Метеорология и гидрология, 1993., № 6., С. 5–8.
- [12] ПЕРМИТИНА Т.О., ПЕВНЕВА Г.С., ПОЛИЩУК Ю.М., ГОЛОВКО А.К. Компьютерные средства обработки данных по составу углеводородов на основе метода главных компонент. // Матер. 4-й межд. конф. «химия нефти и газа» – В 2-х т., Томск: Изд-во STT, 2000. – Т. 2, С. 389–392.
- [13] ПОЛИЩУК Ю.М., ПЕРМИТИНА Т.О. Геоинформационный подход к анализу состава нефтей нижнесреднеюрских отложений Западной Сибири // Материалы 6 Межд. конференции «Новые идеи в геологии и геохимии нефти и газа. К созданию общей теории нефтегазоносности недр». В 2-х т., М.: ГЕОС, 2002., Т. 2., С. 100–103.
- [14] ПОЛИЩУК Ю.М., ПЕРМИТИНА Т.О. Геоинформационный подход к анализу данных о физико-химических свойствах нефтей нижней и средней юры Западной Сибири // Труды межд. конф. «Генезис нефти и газа», М.: ГЕОС, 2003., С. 252–254.