

На правах рукописи

Катаев Сергей Григорьевич

**Алгоритмические и программные средства аналитического и
неразрушающего контроля параметров природной среды на основе метода
выделения структур**

05.11.13

Приборы и методы контроля природной среды, веществ,
материалов и изделий

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора технических наук

Томск - 2012

Работа выполнена в Томском государственном педагогическом университете и
Томском государственном университете систем управления и
радиоэлектронике

Научный консультант: **Кориков Анатолий Михайлович**, доктор
технических наук, профессор, заведующий кафедрой АСУ ТУСУР

**Официальные
оппоненты:**

Кулешов Валерий Константинович, доктор
технических наук, профессор, ИНК при ФГБОУ ВПО
«Национальный исследовательский Томский
политехнический университет», профессор кафедры
физических методов и приборов контроля качества

Кулижников Александр Михайлович, доктор
технических наук, профессор, ФГУП «Российский
дорожный научно-исследовательский институт»,
заведующий отделением методов проектирования и
экспертизы проектной документации

Немирович-Данченко Михаил Михайлович,
доктор физико-математических наук, Томский филиал
Института нефтегазовой геологии и геофизики СО
РАН, ведущий научный сотрудник

**Ведущая
организация:**

государственное учреждение Федеральной службы
России по гидрометеорологии и мониторингу
окружающей среды «Центральная аэрологическая
обсерватория», г. Долгопрудный, Московская область.

Защита состоится 12.02.2013 г. в 15-00 на заседании совета по защите
докторских и кандидатских диссертаций Д 212.269.09 при ФГБОУ ВПО
«Национальный исследовательский Томский политехнический университет» по
адресу: Россия, 634028, г. Томск, ул. Савиных, 7, ауд. 215 (актовый зал).

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО
«Национальный исследовательский Томский политехнический университет».

Автореферат разослан _____ 2012 г.

Ученый секретарь
диссертационного совета
к.т.н., доцент

Винокуров Б.Б.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Для решения глобальных проблем, стоящих перед человечеством, необходимо иметь глубокое понимание процессов, происходящих на нашей планете. Построение реалистичных моделей, описывающих различные явления в такой сложной системе, как наша планета, требует знания множества разнообразных параметров. Именно поэтому в настоящее время проводится большое количество измерений в режиме мониторинга, что порождает стремительный рост объемов получаемых данных. Эти данные несут в себе много скрытой информации, поскольку они, как правило, отражают совместное действие большого количества влияющих факторов. И здесь исследователи сталкиваются с серьезными трудностями, связанными, в частности, с отсутствием универсального подхода, который позволял бы получать необходимую информацию. К каждой сложной системе применяется свой набор методов исследований, зависящий от специфики рассматриваемой системы, различного рода ограничений, типа и объема имеющихся данных и, конечно, целей исследования.

Научное исследование любой системы начинается со сбора фактов и их анализа. Если отбор данных о системе проведен на основе корректных теоретических воззрений, то массив таких данных будет внутренне структурирован. И уже на этом, первом этапе, особенно при изучении сложных систем, структуризация предоставляет дополнительные возможности в осмыслении того, что предстоит изучать. Выявление информации о существующих связях, т.е., обнаружение определенных закономерностей в наборе данных – это одна из задач обработки данных; очень полезный этап, позволяющий, в частности, сжать первоначальную информацию и представить задачу в обозримом виде, что дает возможность строить гипотезы и модели для дальнейших более глубоких исследований. Для решения подобных задач в настоящее время широко используют, наряду с обычными статистическими методами, методы многомерной статистики: факторный, дискриминантный и кластерный анализы.

Данная диссертационная работа посвящена разработке подходов, созданию методов и алгоритмов на основе предлагаемого автором **метода выявления структур** (МВС) для изучения сложных систем различной природы. МВС относится к классу кластерных методов, т.е., является процедурой, позволяющей либо провести разложение данных на структуры, либо обнаружить в них заранее заданные структуры. Структуры, если они существуют в данных наблюдений – это математические структуры. Они отражают совокупное проявление связей между переменными системы, которые не всегда очевидны, и по этой причине часто являются скрытыми.

Фактически МВС реализует скрытый алгоритм, присутствующий, согласно математической энциклопедии, в самом определении понятия математической структуры: *«Структуры математические – родовое название, объединяющее понятия, общей чертой которых является то, что они применимы к множествам, природа элементов которых не определена. Чтобы определить*

структуру, задают отношения, в которых находятся элементы множества, затем постулируют, что данные отношения удовлетворяют условиям – аксиомам структуры» (Математическая энциклопедия. – М.: Советская энциклопедия, 1985. - с. 250).

Эта формулировка и определила название метода, которое показалось автору довольно удачным, поскольку оно, кроме всего прочего, отражает главную цель кластерного анализа – поиск существующих структур.

Кластерный анализ – это обобщенное название ряда процедур, целью которых является создание классификации, т.е. распределение объектов на небольшое число групп, которые называются кластерами, используя при этом тот или иной принцип схожести объектов по выбранной системе признаков. Кластерные методы нашли особенно широкое применение в социологии, практических геофизических исследованиях, биологии, археологии. В литературе можно встретить разнообразные названия, характеризующие, в общем, одну и ту же задачу – разбиение множества объектов, обладающих общей системой признаков, на классы. Общепринятого определения понятия кластер нет, в каждом конкретном случае исследователь вкладывает свой смысл, который зависит от специфики задачи и используемого алгоритма. Но обычно под кластером понимается совокупность точек, объединяемая в единое целое, исходя *только* из их положения в пространстве признаков.

Результаты, полученные автором с помощью МВС, показывают, что использование понятия структура, как совокупности точек, объединяемые в один класс не только по степени близости, но и с использованием дополнительного набора правил на отношения между ними, позволяет существенно расширить круг решаемых задач.

Согласно М.С. Олдендерферу и Р.К. Блэшфилду (Факторный, дискриминантный и кластерный анализ: пер с англ. / Дж.–О. Ким [и др.]; под ред. И. С. Енюкова. – М.: Финансы и статистика, 1989. –с.143) , различные приложения кластерного анализа можно свести к четырем основным задачам:

- 1) разработка типологии или классификации;
- 2) исследование полезных концептуальных схем группирования объектов;
- 3) порождение гипотез на основе исследования данных;
- 4) проверка гипотез или исследования для определения, действительно ли типы (группы), выделенные тем или иным способом, присутствуют в имеющихся данных.

Задачи, рассмотренные в данной диссертационной работе, можно отнести, главным образом, к третьему и четвертому типам и в меньшей степени ко второму. Эти задачи относятся к классу обратных задач, которые, как известно, не имеют однозначного решения за редким исключением. Каждая задача обладает своими особенностями, языком описания и областью практического применения. Естественно, что при исследовании и интерпретации полученных результатов использовались и другие методы, но МВС в той или иной модификации в качестве одного из этапов присутствует в каждой задаче. Адаптация МВС к конкретной проблеме заключается, главным образом, в

выборе объектов, системы признаков для них и определении понятия структуры.

Цель работы

Разработка и исследование методов и алгоритмов аналитического и неразрушающего контроля параметров природной среды на основе метода выделения структур и построение с использованием МВС программных средств для решения задач обнаружения закономерностей в экспериментальных данных геофизической и иной природы.

Основные задачи исследований

1. Разработать алгоритмы, создать программы, реализующие выделения структур из эмпирических данных произвольной природы.
2. Разработать подход на основе МВС к исследованию периодических и непериодических полей данных.
3. Изучить с помощью предлагаемого подхода различные метеорологические поля (температуры, осадков, O_3 , CO_2 , ветра и т.д.).
4. Разработать алгоритмы и создать программы, позволяющие исследовать тепловлажностные и диэлектрические свойства пористых сред (почв и строительных материалов) при разнообразных внешних воздействиях, а также моделировать поведение отраженного от такой среды электромагнитного импульса.
5. Создать алгоритмы и программы, позволяющие обнаруживать структуру подповерхностного слоя и объекты, находящиеся в нем, а также объекты, располагающиеся за преградой из данных по электромагнитному зондированию сверхширокополосным сигналом.
6. Создать алгоритмы и программы на основе МВС, позволяющие аппроксимировать неизвестный поток событий МС-поток.
7. Разработать на основе МВС метод, позволяющий по потоку изображений находить контур движущегося объекта.

Методологическими и теоретическими основами послужили работы:

- зарубежных ученых: У. Гренандера, Н. Винера, Р. Сокэла, П. Снита, Дж.–О. Кима, Ч.У. Мьюллера, У.Р. Клекка, М. Кендала, Э. Майники, Р. Ними, А. Стюарта, К. Фу, Э.П. Фридмана, С.Г. Голдстейна, Дж. Д. Линдена и других;
- российских ученых: С.А. Айвазяна, Э.М. Браверманна, В.М. Бухштабера, И.С. Енюкова, Н.Г. Загоруйко, В.Н. Елкиной, Г.С. Лбова, Л.Д. Мешалкина, И.Б. Мучник, Г.В. Груза, Р.Г. Рейтенбаха, Ю.А. Будыко, М.С. Израэля, Е.С. Рубинштейна, С.П. Перова, А.Х. Хргиана, Л.С. Гандина, Р.Л. Каган и других.

Обозначенные в работе цели и задачи диссертационного исследования ставились и решались с применением разных методов, включая методы многомерной и обычной статистики.

Научная новизна

Научной новизной обладают следующие результаты:

1. Создан метод выявления структур (МВС), позволяющий отыскивать структуры в массивах данных по заданному правилу объединения элементов в структуру. Метод и его модификации, в отличие от многих существующих, не требует априорной информации о природе исследуемых экспериментальных данных, способах их получения, а также о числе отыскиваемых структур.
2. Предложен и использован при решении задач геофизики комплексный подход к исследованию природных периодических полей, заключающийся в согласованном выполнении ряда этапов: разложения геофизических полей на составляющие, их кластеризации и компонентного анализа.
3. В задачах определения структуры подповерхностного слоя и обнаружения находящихся в нем объектов по данным электромагнитного сверхширокополосного зондирования впервые предложены оригинальные алгоритмы, являющиеся модификациями МВС.
4. Предложена алгоритмическая разновидность МВС, адаптированная к задаче построения радиоизображения объектов, находящихся за преградой, с использованием данных электромагнитного сверхширокополосного зондирования.
5. Впервые предложен подход, использующий совместное решение тепловлажностной задачи и задачи обнаружения, к решению проблемы распознавания подповерхностных объектов по данным электромагнитного сверхширокополосного зондирования. Этот подход позволяет более точно интерпретировать отраженный сигнал, поскольку учитывает зависимость формы отраженных пиков от профиля влажности в среде.
6. Впервые предложен метод, использующий идеологию выделения структур для выявления участков стационарности во входящем потоке событий. Это позволяет решать задачу аппроксимации входящего потока событий дважды стохастическим марковским потоком.
7. Для решения задачи отбора модельных деревьев впервые применен метод, в основе которого лежат базисные положения МВС, позволяющий описывать пространственно-временные характеристики окружающей среды.
8. Впервые предложен алгоритм локализации фигуры человека, на основе МВС, обладающий значительно большей устойчивостью и скоростью обнаружения, при сохранении точности существующих методов.

Практическая ценность работы и реализация результатов исследования диссертационной работы определяется возможностью широкого применения методов выделения структур в различных модификациях при решении практических задач неразрушающего контроля состояния сложных систем. Они могут быть использованы при создании математического обеспечения для решения задач мониторинга окружающей среды, для геолокаторов, предназначенных для решения разнообразных задач

инженерного и экологического плана, а также при разработке и исследовании систем массового обслуживания.

Созданы программы:

1. Разложения временных рядов на составляющие с последующей классификацией и компонентным анализом каждой составляющей (*ICCC*).
2. Для моделирования при заданных граничных условиях профилей температуры, влажности в слоистых пористых материалах (почвы, строительные материалы), построения для этих профилей модельной функции комплексной диэлектрической проницаемости, и расчета отраженного от такой структуры электромагнитного сигнала (*TWES*).
3. Решения задачи профилирования и обнаружения подповерхностных объектов из данных по электромагнитному зондированию сверхширокополосным сигналом (*GEO*).
4. Решения задачи аппроксимации входящего потока событий дважды стохастическим потоком (*APFLOW*).

Разработанный программный продукт *ICCC* используется на кафедре метеорологии ТГУ (г. Томск), экологии и природопользовании ЮГУ (г. Ханты-Мансийск) и в ИМЭКС СО РАН (г. Томск) при изучении различных метеорологических полей. В ОАО Томгипротранс (г. Томск), ОАО Фундаментпроект (г. Москва) и РГУПС (г. Ростов-на-Дону) при решении георадиолокационных задач применяются программы интерпретации георадарных данных комплекса *GEO_TWES*.

Разделы работы, посвященные изучению поведения природных полей, использовались при чтении лекций студентам гуманитарных специальностей по курсу «Концепции современного естествознания» Томского государственного педагогического университета.

Апробация работы Основное содержание диссертационной работы докладывалось на: на Сибирском Совещании по климатозэкологическому мониторингу (I, Томск 1995, III 1999, IV 2001, VI 2005), 4-ой Международной научно-технической конференции «Распространение и дифракция электромагнитных волн в неоднородных средах» (Москва, 1994), Progress in Electromagnetic Research Symposium (PIERS), (1996, Innsbruck, Austria), 3-ем Сибирском Конгрессе по прикладной и индустриальной математике, ИНПРИМ-98 (Новосибирск, 1998), Научно-практической конференции «Актуальные вопросы неврологии и нейрохирургии» (Рязань, 1999), VII-ом Международном симпозиуме «Оптика атмосферы и океана», KOREA-RUSSIA International Symposium on Science and Technology (Tomsk, 2001, 2004), Сибирском поляризационном семинаре, СибПол (Сургут, 2004), Международных конференциях по измерениям, моделированию и информационным системам для изучения окружающей среды, ENVIROMIS (Томск 2004, 2007, 2011), Conference on spatial interpolation in climatology and meteorology (Budapest, 2004), Всероссийской научно-практической конференции «Проблемы информационной безопасности государства, общества и личности», (Томск, 2005), Всероссийских симпозиумах

«Радиолокационное исследование природных сред» (XXV,XXVI Санкт Петербург, 2007), 11-th international conference on ground penetrating radar (USA, Columbus Ohio, 2006), 3-ей, 5-ой, 7-ой и 8-ой международных научно-практических конференциях «Инженерная и рудная геофизика» (Геленджик, 2007, 2009, 2012; Москва 2011), Международной конференции Ultrawideband and Ultrashort impulse signals, (UWBUSIS, Sevastopol, 2008, 2009), 2-ой Международной геолого-геофизической конференции «Тюмень – 2009» (Тюмень, 2009), 29th Annual EARSeL Symposium 2009, (Chania, Crete, Greece).

Основные защищаемые положения

1. Метод выявления структур (МВС), дающий возможность отыскивать структуры в массивах данных по заданному правилу объединения элементов в структуру, и не требующий вычисления функционала качества.
2. Комплексный подход к исследованию периодических и непериодических геофизических полей естественного и антропогенного происхождения, заключающийся в выполнении последовательности известных алгоритмов по определенным правилам.
3. Методика обработки пространственно-временных наборов данных электромагнитного сверхширокополосного зондирования, основанная на МВС, позволяющая выявлять структуры изучаемого подповерхностного слоя.
4. Методика обработки данных электромагнитного сверхширокополосного зондирования, представляющая собой алгоритмическую разновидность МВС, позволяющая проводить обнаружение объектов неизвестной природы, в том числе находящихся за преградой.
5. Методика обработки информационных потоков, базирующаяся на принципах МВС, позволяющая только по временам прихода событий аппроксимировать входящий поток МС-поток.
6. Основанная на МВС методика обработки и анализа потока изображений, позволяющая выделять движущиеся объекты в режиме, близком к реальному времени.

Публикации По теме диссертации опубликовано 76 работ, в том числе 2 монографии, 15 статей в изданиях, входящих в перечень ведущих рецензируемых научных журналов и изданий из списка ВАК, получено 2 патента и 3 свидетельства о регистрации программ для ЭВМ.

Структура диссертации Диссертация состоит из введения, шести глав, заключения, списка использованной литературы, включающего 284 наименования, приложения. Содержит 355 страниц текста, 145 рисунков, 37 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность и новизна предлагаемых методов и исследований, сформулирована основная цель и решаемые задачи, общая характеристика работы, приводятся основные положения, выносимые на защиту.

В первой главе излагаются основные положения метода выделения структур, сравнение этого метода с другими методами многомерной статистики. Приводится краткий обзор методов кластерного анализа.

Метод выделения структур

Пусть имеется множество объектов $a = \{a_1, a_2, \dots, a_M\}$, каждый из которых характеризуется общей системой признаков $X_\Omega = \{X, \Omega\}$, где $X = \{x_1, x_2, \dots, x_N\}$ – множество основных, а $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ – множество вспомогательных или дополнительных признаков. Исходная матрица для основных признаков: $T = \|t_{ij}\|_{M \times N}$. Пусть $D = \|d_{ij}\|_{M \times M}$ – матрица расстояний между объектами, построенная с использованием только основных признаков, задаваемых T . Можно сформулировать задачу отыскания на множестве a структур, удовлетворяющих определенному набору аксиом. Аксиомы структуры – это набор требований (ограничений), которые вытекают из цели исследования, и их можно разбить на 3 типа.

1. Тип A_D . Аксиомы, сформулированные в терминах только матрицы D .
2. Тип A_X . Аксиомы, являющиеся ограничениями, записанными через множество основных признаков X .
3. Тип A_Ω . Аксиомы, определенные с помощью множества вспомогательных признаков.

В этих обозначениях возможный набор аксиом можно представить в виде упорядоченной тройки следующим образом: $A = \{A_\Omega, A_X, A_D\}$. Ясно, что в рамках такого представления можно выделить 8 классов аксиом, в зависимости от присутствия того или иного типа.

Алгоритм выделения структур *REVPAT* наиболее просто может быть описан на языке теории графов. Соотнесем с каждым объектом $a_i, i = 1, \dots, N$ вершину x_i некоторого графа $G = (X, E)$, а каждому ребру $e_{ij} = (x_i, x_j) \in E, i, j = 1, \dots, N$, связывающему вершины x_i и x_j , в качестве стоимости c_{ij} припишем величину, равную расстоянию d_{ij} между соответствующими объектами в пространстве признаков: $c_{ij} = d_{ij}$. Выделение структур возможно в двух вариантах.

Вариант 1.

Начальное состояние – полный граф. Будем последовательно удалять из графа ребра e_{ij} со значениями, большими некоторого порога d_p^t , получая на каждом шаге $G^t = (X^t, E^t)$ – граф на уровне t , множество ребер которого $E^t = \{e_{ij} \in E : c_{ij} \leq d_p^t\}$. Будем уменьшать величину d_p^t , и тогда, начиная с некоторого значения d_p^{t*} , исходный граф станет несвязным, появится k_t подграфов (компонент связности) $G_1^t, G_2^t, \dots, G_{k_t}^t$ с различным числом вершин, среди которых могут быть подграфы, состоящие из одной вершины.

Повторение процедуры дальнейшего уменьшения порогового значения d_p^t приводит к выделению все более связанных подграфов.

Вариант 2.

Начальное состояние – граф без ребер. Будем последовательно добавлять в граф ребра e_{ij} со значениями, меньшими некоторого порога d_p^t , получая на каждом шаге $G^t = (X^t, E^t)$ – граф на уровне t , множество ребер которого $E^t = \{e_{ij} \in E : c_{ij} \leq d_p^t\}$, состоящий из k_t компонент связности $G_1^t, G_2^t, \dots, G_{k_t}^t$ с различным числом вершин.

Будем увеличивать величину d_p^t . При этом число компонент связности и число изолированных вершин будет уменьшаться и, начиная с некоторого значения d_p^{t*} , исходный граф станет связным. Назовем **структурой** множество вершин графа, образующее на некотором шаге t компоненту связности и удовлетворяющее при этом наперед заданным аксиомам структуры $(A): S = \{x : x \in X_\alpha^t, G_\alpha^t = (X_\alpha^t, E_\alpha^t); c_\alpha^t \leq d_p^t, (A)\}$. Аксиомы отражают и цель исследования, и специфику решаемой задачи. Например, при изучении поля какой-либо метеорологической величины можно сформулировать следующее понятие структуры:

$$S = \{X_\alpha^t : G_\alpha^t = (X_\alpha^t, E_\alpha^t); \bar{R}(X_\alpha^t) \geq R_0, D(X_\alpha^t) \leq D_0, \forall (x_j, x_k) \in X_\alpha^t, \max |r_k - r_j| \leq r_0\}$$

Т.е., средняя связность поля в каждой выделяемой структуре должна быть не меньше, чем R_0 , среднее значение дисперсии не больше, чем D_0 и при этом максимальное расстояние между объектами из одной структуры в реальном пространстве не должно превышать r_0 . Введение понятия структуры и поиск именно структур отличает МВС от традиционных кластерных методов. На самом деле, любой метод кластерного анализа основывается на предположении о существовании оптимального разбиения, т.е. такого разбиения множества объектов на классы, которое удовлетворяет некоторому критерию качества, сформулированному для всего множества. Иными словами кластерные методы осуществляют разложение системы на структуры. В МВС каждая структура фиксируется по отдельности в соответствии с внешними требованиями к структуре; полагается, что в массиве экспериментальных данных структуры могут существовать, но не обязательно массив данных должен раскладываться на структуры без остатка.

В сравнении с кластерным анализом такой подход имеет ряд преимуществ, а именно, нет проблемы выбора функционала качества, не требуется задания количества классов, кластеры могут иметь любую форму и т.д. В качестве примера работы МВС на рис. 1 и 2 отображены начальная, промежуточные и конечная стадии классификации аномальной составляющей средней месячной температуры 90 станций на азиатской территории России.

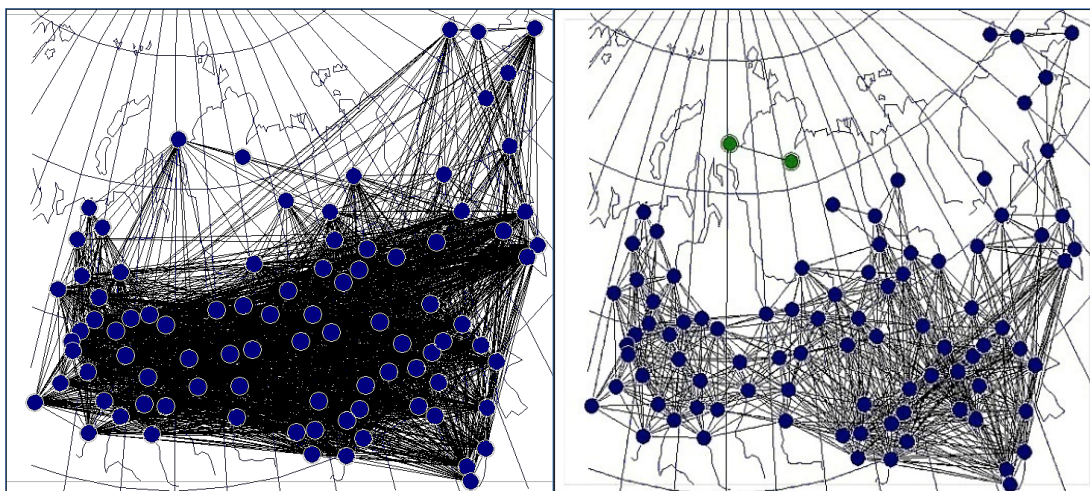


Рис. 1– Начальная стадия (слева) и появление первого класса (справа)

Левая часть рисунка 1 иллюстрирует начальную стадию – практически полный граф, а в правой части - ситуация после удаления ребер с большими весами; граф распался на 2 компоненты связности, на Севере АТР выделился 1 класс, состоящий из 2 станций. В левой части рисунка 2 помещена практически конечная стадия выделения, связей уже очень мало. На этой стадии видно присутствие одновременно 6-и классов, (пункты, окрашенные в разные цвета), и окончательный результат - 11 классов приведен в правой части рисунка 2.

Далее следует описание алгоритма *BINREL*, являющегося разновидностью МВС и метод структурного анализа *MSA*.

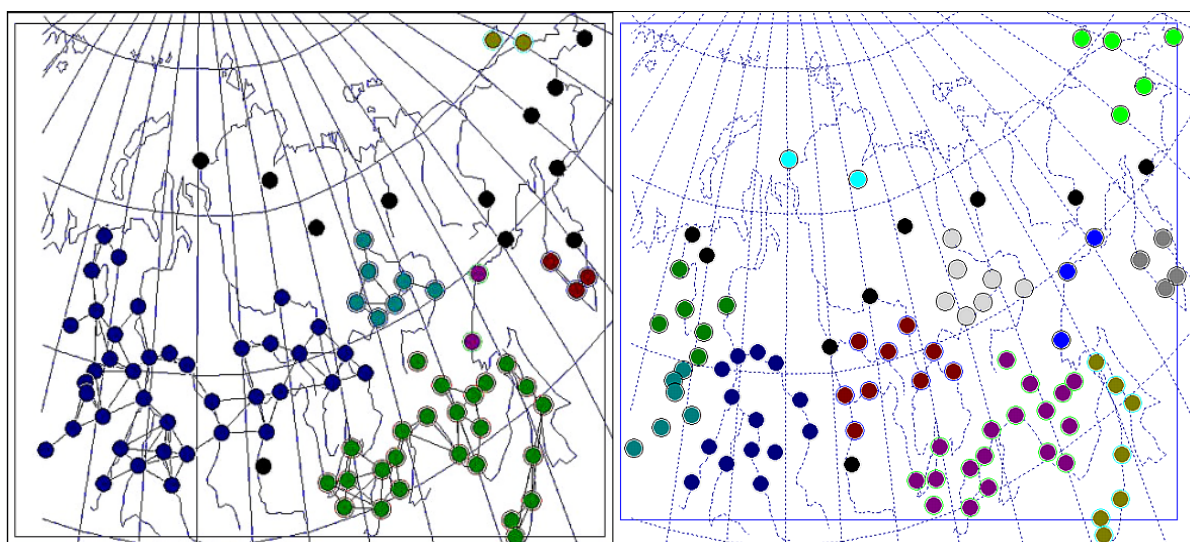


Рис. 2 – Конечная стадия классификации

Вторая глава посвящена решению задачи аппроксимации входящего потока событий дважды стохастическим потоком.

В теории массового обслуживания широко исследуются системы массового обслуживания с входящим МС-потоком заявок. Обычно его исследование осуществляют в предположении, что математическая модель потока известна с

точностью до определенных параметров. Как правило, относительно исследуемого потока известны лишь времена наступления событий. В связи с этим при изучении реального потока можно поставить задачу об его аппроксимации МС-поток для дальнейшего исследования СМО уже разработанными методами. МС-поток называется поток, интенсивность которого представляет собой кусочно-постоянный марковский процесс $\lambda(t)$, принимающий значения из конечного множества констант $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, называемых состояниями процесса $\lambda(t)$. Длительности участков стационарности, т.е. участков, где $\lambda = const$, есть случайные величины, распределенные по экспоненциальному закону с параметрами $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ соответственно. На каждом таком интервале времени поток событий ведет себя как пуассоновский поток событий с интенсивностью λ_i (простейший поток событий с интенсивностью λ_i). Переход из состояния i в состояние j осуществляется по вероятности p_{ij} из матрицы вероятностей переходов, которая обладает свойством: $\sum_j p_{ij} = 1, p_{ii} = 0$. Очевидно, для аппроксимации наблюдаемого потока событий МС-поток необходимо выделить в реализации наблюдаемого потока событий интервалы времени, на которых поток ведет себя как простейший, т.е. события поступают с примерно одинаковой интенсивностью. Тем самым будут определены и моменты смены состояний. После обнаружения таких интервалов можно оценить интенсивности потока на этих интервалах, а при наличии в потоке большого числа смены состояний и интенсивности переходов, и вероятности перехода из состояния в состояние.

В данной главе предлагается метод, использующий идеологию МВС, определения параметров МС-потока, который мог бы аппроксимировать реальный поток по наблюдениям за моментами наступления событий потока. Основная характеристика потока событий – это его интенсивность $\hat{\lambda}$, которая для простейшего потока оценивается по формуле:

$$\hat{\lambda} = k / (t_{k+1} - t_0). \quad (1)$$

где k – это число событий, наступивших на интервале времени от t_0 до t_{k+1} .

Таким образом, под структурой реализации МС-потока с s состояниями можно понимать последовательность времен t_m переключения с одного состояния на другое. И тогда задача аппроксимации произвольного потока из N событий МС-поток сводится к отысканию его структуры, т.е., таких t_m , чтобы поток событий, приходящих за каждый интервал времени (t_m, t_{m+1}) , являлся бы приблизительно пуассоновским. Для нахождения такой структуры предлагается воспользоваться МВС, построив матрицу расстояний D . Входящий поток можно рассматривать как совокупность объектов (событий), которые характеризуются только одним признаком – временем их наступления. В этом случае матрица T будет представлять собой столбец, размерностью N . В качестве расстояния между объектами i и j можно выбрать разность $\Delta t_{ij} = t_i - t_j$,

либо какую-нибудь иную величину d_{ij} , связанную с Δt_{ij} . Поскольку в данной задаче в качестве характеристики состояния МС-потока выступает интенсивность, определяемая согласно (1), то для отношения d_{ij} , удобно использовать подобную оценку.

$$d_{ij} = \begin{cases} \frac{j-i+1}{t_{j+1}-t_{i-1}}, & i \leq j; \quad i, j = 1, \dots, N \\ 0, & i > j; \quad i, j = 1, \dots, N \end{cases} \quad (2)$$

Определенные таким образом элементы d_{ij} , образуют верхнетреугольную матрицу $D = [d_{ij}]_{N \times N}$. Тогда формализация МВС для данной задачи может быть изложена в следующей редакции. Сопоставим каждому i -му событию потока, $i = 1, N$, вершину с тем же номером i некоторого графа $G(V, E)$, где V – множество вершин графа мощности N , а E – множество пар вершин, т.е. множество ребер графа. Стоимостью каждого ребра графа, связывающего i -ю и j -ю ($i < j$) вершины, будем считать величину d_{ij} , т.е. соответствующий элемент матрицы D . Определение структуры можно сформулировать, исходя из понятия интервала стационарности. А именно, если группа событий с номерами $i_1^p, i_2^p, \dots, i_k^p$ составляет участок стационарности потока с интенсивностью $\lambda = \lambda_p$, то это означает, что

а) среднее значение всех ребер, связывающих между собой вершины с номерами $i_1^p, i_2^p, \dots, i_k^p$ примерно равно интенсивности λ_p :

$$\frac{2}{k(k-1)} \sum_{(i,j) \in I} d_{ij} \cong \lambda_p, \quad (3)$$

где I – множество всевозможных упорядоченных по номерам пар из $\{i_1^p, i_2^p, \dots, i_k^p\}$;

б) номера $i_j^p, j = 1, k$, идут подряд, т.е. $i_j^p = i_{j-1}^p + 1$;

в) все $d_{ij}, i, j \in I$, принадлежат, в основном, некоторому интервалу $(\lambda_p - \Delta\lambda, \lambda_p + \Delta\lambda)$. Необходимо отметить, что оценка интенсивности, полученная по всей группе событий, т.е. (i_1^p, i_k^p) -я компонента матрицы D , должна быть наиболее близкой к истинному λ_p . Изложенные соображения определяют идею алгоритма выделения участков стационарности потока. При выборе удачного интервала $(\lambda_p - \Delta\lambda, \lambda_p + \Delta\lambda)$ и удалении из графа G всех ребер, стоимость которых не попадает в указанный интервал, в графе выделяются компоненты связности $G_k = (V_k, E_k)$, соответствующие участкам стационарности, относящимся к p -му состоянию МС-потока (G_k является подграфом графа G). Разумно ожидать, что участку стационарности в наибольшей степени отвечает такой подграф, в котором, во-первых, максимальное число ребер, а во-вторых, каждая вершина должна быть соединена с наибольшим числом вершин, т.е. степени вершин должны быть примерно одинаковыми. Эти требования могут быть учтены

введением следующих характеристик для выделенных компонент связности G_k :

а) Степень компактности компоненты:

$$COMP[k] = \frac{2 \sum_{i=1}^{N_k} s_i}{N_k (N_k - 1)}, \quad (4)$$

где N_k – число вершин G_k , а s_i – степень i -ой вершины; величина $COMP[k]$ характеризует близость подграфа G_k к полному;

б) степень однородности компоненты:

$$ENTR[k] = \frac{-\sum_{i=1}^{N_k} p_i \log p_i}{\log N_k}, \quad (5)$$

где p_k есть отношение степени i -ой вершины компоненты G_k к величине $(N_k - 1)$; характеристика $ENTR[k]$ по виду напоминает формулу энтропии S случайной величины с распределением p_k , $i = 1, N_k$:

$$S = -\sum_{k=1}^{N_k} p_k \log p_k, \quad \sum_{k=1}^{N_k} p_k = 1. \quad (6)$$

Величина энтропии, как известно, характеризует степень однородности множества, и для заданного N_k $S \Rightarrow \max$ при $p_k = \frac{1}{N_k}$.

Будем называть элемент матрицы $D = \{\hat{\lambda}_{ij}\}$ чистой оценкой интенсивности потока на интервале стационарности, если все события, наступившие в моменты $\{t_i, t_{i+1}, \dots, t_{j+1}\}$, принадлежат одному и тому же интервалу стационарности потока. Остальные элементы матрицы назовем ложными оценками или шумом. Если исследуемый поток пуассоновский, то все элементы матрицы D являются чистыми оценками его интенсивности. В случае МС-потока матрица оценок имеет скрытую блочную структуру: на диагонали компактно в виде треугольных блоков расположены чистые оценки интенсивности. Каждый блок соответствует участку стационарности. Размеры блоков совпадают с количеством событий в соответствующем отрезке пуассоновского потока. Остальные элементы матрицы – шум. Далее сгруппируем данные, содержащиеся в матрице оценок, в гистограмму:

$$\left\{ kh; \frac{n_k}{N}; N = \sum_k n_k \right\} \quad (7)$$

Здесь: h – шаг гистограммы; n_k – частоты, т.е. количество элементов матрицы D , значения которых принадлежат интервалу $((k-1)h, kh]$; N – количество элементов матрицы D , значения которых принадлежат интервалу $(0, mh]$; m – количество шагов гистограммы. Из свойств оценки $\hat{\lambda}$ вытекает, что для пуассоновского потока полигон частот, при достаточной длине реализации, представляет собой δ -образную кривую, максимум которой совпадает со значением интенсивности потока λ с точностью до шага гистограммы (левая

часть рис.3). Если бы для МС-потока удалось выделить из матрицы оценок только те элементы, которые являются чистыми оценками, полигон частот выглядел бы как сумма δ -образных кривых, максимумы которых совпадают с интенсивностями потока на интервалах стационарности. Наличие шумовых оценок в МС-потоке приводит к усложнению вида гистограммы (правая часть рис.3).

Процедура выделения связанных компонент графа может быть описана следующим образом.

1. Задан интервал $[\lambda', \lambda'']$. Из графа удаляются все ребра, стоимости которых не попадают в данный интервал. Граф G разбивается на компоненты связности $G_k, k = 1, 2, \dots$, (в том числе изолированные вершины).
2. Интервал $[\lambda', \lambda'']$ увеличивается. Получаются новые разбиения графа G , в которых компоненты изменяются: расширяются за счет изолированных вершин, поглощают другие компоненты. Наконец, при достижении интервалом некоторой величины граф G не разобьется на компоненты, останется связным. На множестве разбиений k -я компонента характеризуется последовательностями характеристик $COMP[k]$ и $ENTR[k]$.

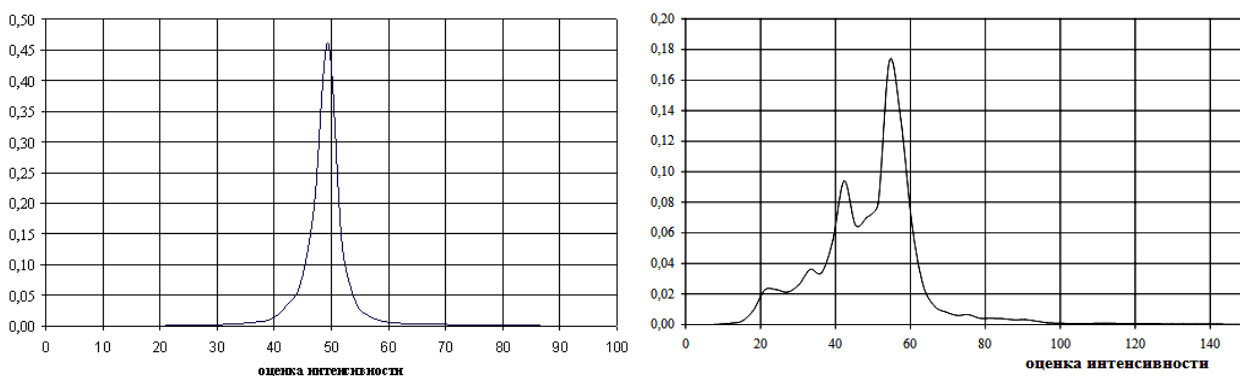


Рис. 3 – Полигон частот для пуассоновского потока с интенсивностью $\lambda=50$, состоящего из 500 событий (слева) и для МС-потока с 3-мя состояниями с интенсивностью $\lambda=20, 50, 90$ (справа)

Выбирается то разбиение, для которого $COMP[k]$ достигает своего локального максимума, но при этом и функция $ENTR[k]$ близка к своему локальному максимуму. Те компоненты, которые этому условию удовлетворяют, считаются перспективными в смысле соответствия интервалам стационарности. Средняя стоимость ребра полученной компоненты, которая рассчитывается с учетом ранее отброшенных ребер, будет оценкой интенсивности потока для одного из состояний. Если эти оценки для разных компонент близки, то это интервалы стационарности одного состояния МС-потока.

Для выбора перспективных интервалов $[\lambda', \lambda'']$ можно воспользоваться информацией о местоположении пиков огибающей гистограммы. Ясно, что, если события с номерами $i_1, i_2, \dots, i_k$, относятся к одному интервалу

стационарности, значит (i, j) -ые компоненты матрицы D при $i \in \{1, \dots, i_1 - 1\}$, а $j \in \{i_1, N\}$, и при $i \in \{i_1, i_k\}$, а $j \in \{i_k + 1, N\}$, являются заведомо шумом. Если при удалении из матрицы D ложных оценок и повторения процедуры выделения связанных компонент графа состав компонент не изменяется, то интервалы стационарности для данных состояний считаются определенными. Для устранения неоднозначности при определении границ интервалов стационарности и для оценки качества проведенной аппроксимации разработаны соответствующие критерии, основанные на критерии Пирсона.

Эффективность метода иллюстрируется рисунком 4, где приводятся «экспериментальная» матрица чистых оценок МС-потока с тремя состояниями, полученного методом имитационного моделирования, (слева) и «теоретическая» матрица оценок того же потока, построенная в результате работы алгоритма (справа).

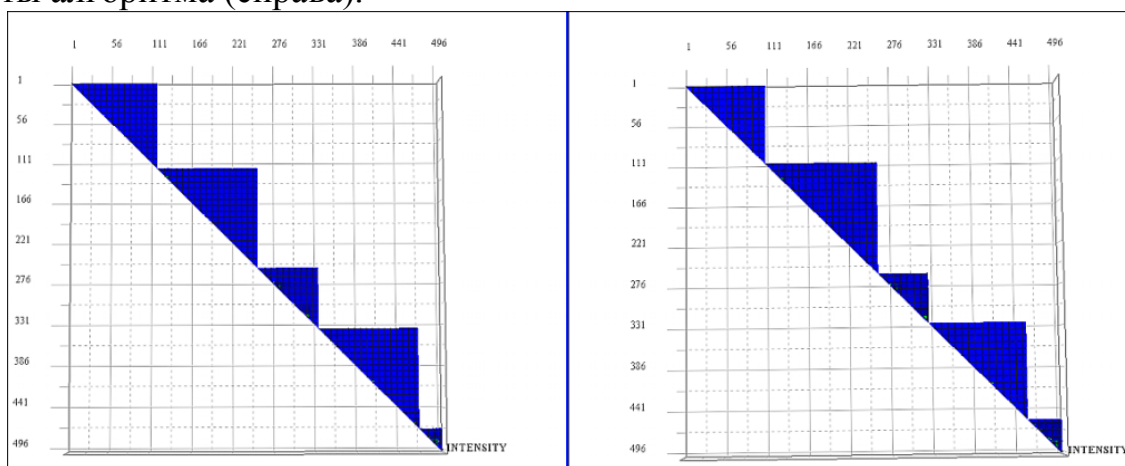


Рис. 4 Структуры «экспериментальной» и «теоретической» матриц оценок потока с тремя состояниями

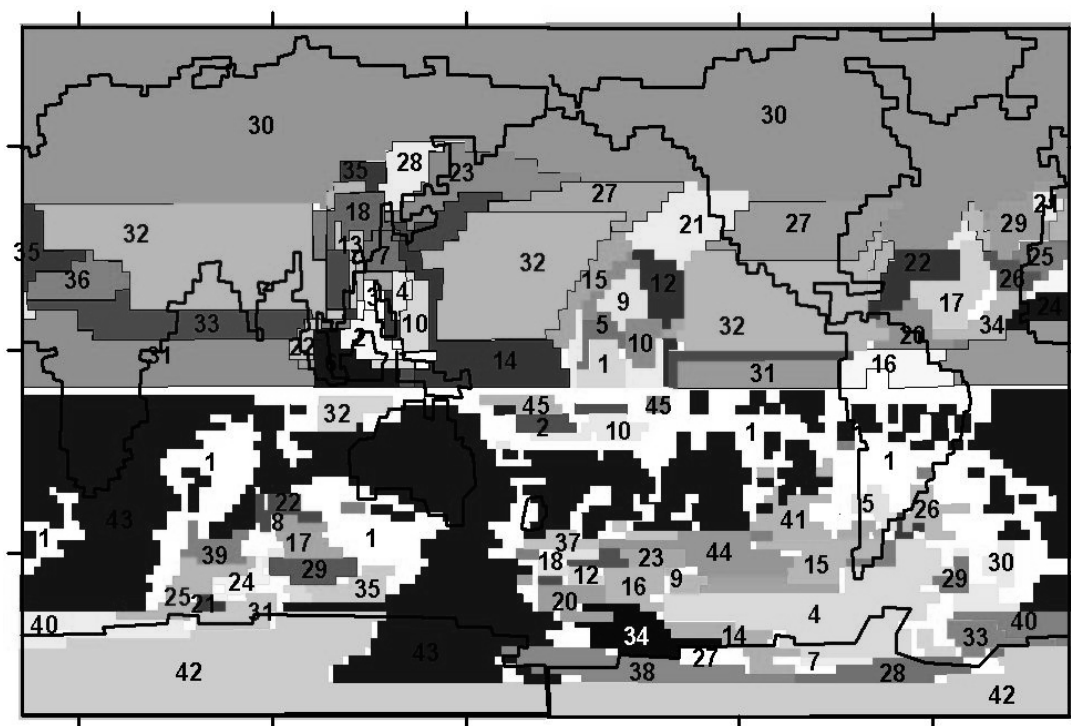
Проведенное большое количество численных экспериментов и исследование реального потока событий позволяют сделать вывод о возможности аппроксимации исходного потока МС-поток с небольшим числом состояний с помощью разработанной авторами процедуры.

В третьей главе приводятся конкретные примеры применения МВС при исследовании неперіодических геофизических полей, а также излагается комплексный подход к исследованию периодических и неперіодических природных полей (ПП) (температура, содержание озона, ветер и т.д.).

В начале главы рассматривается применение МВС для исследования профилей концентрации CO_2 в различных регионах Земного шара с целью оценивания возможности этого метода для нахождения областей с близкими характеристиками поведения. В качестве исходных данных для моделирования были взяты 3-х часовые значения профилей концентрации CO_2 , на пространственной сетке $[0.5^\circ \times 0.5^\circ]$ на 15 стандартных высотах. Были отобраны несколько однородных и неоднородных по свойствам подстилающей поверхности и

равных по географическим размерам ареалов, находящихся на разных континентах. В каждой области располагалось по 377 точек сетки; каждой точке был поставлен в соответствие временной ряд, составленный из суточных временных рядов (8 значений) на каждой высоте. Таким образом, в данной задаче 377 объектов и 120 признаков.

В результате применения МВС в каждой области было выделено от 20 до 40 структур (классов) с близким поведением профилей. Каждой точке, относящейся к определенному классу, приписывался типовой (средний по структуре) профиль. Средняя квадратичная ошибка такой аппроксимации по однородным районам не превышает 1%, а по неоднородным – 2% от $\sqrt{D}$, где D – исходная дисперсия всей области. Таким образом, предлагаемый подход позволяет обеспечивать необходимую (заданную) точность при аппроксимации временных рядов концентрации CO_2 . Разбив всю поверхность Земли на несколько сотен регионов, мы тем самым сокращаем необходимый объем вычислений при восстановлении, например, спутникового сигнала в случае сетки $[2.5^\circ \times 2.5^\circ]$, по крайней мере, на 2 порядка. На рис.5 приводится результат выделения структур для приповерхностного поля концентрации CO_2 . (2001 год).



*Рис. 5 – Результирующее положение структур поля концентрации CO_2 .
Нумерация ведется отдельно для каждого полушария*

Далее излагается подход к исследованию периодических природных полей. Значение конкретного природного поля в пространственной точке x представляет собой непрерывно изменяющуюся величину, которая состоит как из детерминированных, так и случайных составляющих. Выбор временного интервала, по которому проводится усреднение, зависит, в первую очередь, от

цели исследования. При этом уровень усреднения должен позволять, с одной стороны, получать достаточно подробную информацию, а с другой стороны, давать возможность строить обобщенную картину эволюции ПП. Использование среднемесячных значений для многих ПП вполне удовлетворяет подобным требованиям. Этот уровень усреднения, с одной стороны, позволяет выявлять достаточно тонкие тенденции изменения ПП внутри года по исследуемой территории, которые исчезают при использовании среднегодовых значений, а с другой стороны, дает возможность избежать анализа возможных артефактов, неизбежно возникающих при выборе более короткого временного периода. Предполагается, что для описания ПП следует использовать некоторую промежуточную стадию – результаты обработки экспериментальных данных, представленных в виде набора параметров, характеризующих временные ряды в каждой пространственной точке x_i . Необходима математическая модель, представляющая собой набор эмпирических уравнений эволюции этих параметров и соотношений, связывающих их друг с другом. Эта модель может играть роль эксперимента при построении физических моделей, объясняющих изменения природных полей.

Предлагаемый подход к исследованию периодических ПП заключается в последовательном применении известных и используемых при анализе временных рядов, как правило, по отдельности процедур:

- 1) выделение из временных рядов в каждой точке пространства составляющих: периодической, тренда и аномалии;
- 2) пространственную и (или) временную классификацию каждой из составляющих;
- 3) выявление сопряженности, как исходного поля, так и его составляющих для различных классов.

Первый этап позволяет сформировать набор понятий для сжатого описания временных рядов: среднее значение, дисперсия, фаза, параметры тренда и периодической составляющей. При этом предполагается, что каждая из составляющих временного ряда обусловлена действием своего природного механизма в своем временном масштабе. Т.е., что значения ПП, порожденные процессами меньшего масштаба, колеблются около величин, сформированных процессами большего масштаба. Выбор вида тренда зависит от характера наблюдаемого долговременного изменения температуры. В нашем подходе в качестве тренда можно выбирать либо степенную, либо полиномиальную функцию, для оценивания параметров которой используется метод наименьших квадратов. При указанном способе разложения, составляющие исходного временного ряда не коррелированы относительно друг друга. Дисперсия исходного ряда D_T (общая дисперсия) равна сумме дисперсией составляющих.

Второй этап дает возможность определять структуру рассматриваемого поля, выделяя пространственные или временные области (классы) с близким характером эволюции. Причем, применяя процедуру классификации для каждой из составляющих, мы тем самым можем оценить пространственные

(временные) области действия каждого из природных механизмов. В данном случае классификация конкретного ПП преследует несколько целей. Первая – разбиение исследуемого пространства на отдельные непересекающиеся области – дает возможность проводить анализ как внутри каждого класса, так и изучать взаимодействие между классами. В частности, привязка классов к ландшафтам позволяет понять роль подстилающей поверхности при формировании климата. Вторая цель – нахождение временных структур в метеоданных. Под временной структурой понимается период времени, в течение которого на всех пунктах наблюдения искомая величина изменялась примерно одинаково. Обнаружение таких временных структур является очень важной задачей; границы выделенных временных интервалов указывают на моменты изменений климата, и в зависимости от выбора исследуемой области можно говорить либо о региональном, либо о глобальном изменении. Разбиение объектов на классы позволяет вычислить статистические характеристики исследуемого признака или признаков для каждого кластера: среднее значение, дисперсия, и т.д. Но кроме этого, каждому кластеру можно приписать и определенные структурные характеристики, зависящие только от отношений между элементами: среднее значение длины связи, ее дисперсия, изолированность и компактность кластера и т.д. Это, безусловно, полезно, т.к. дает возможность количественно сравнивать различные кластеры друг с другом по этим структурным параметрам и оценивать все разбиение в целом по результатам анализа удаленности кластеров друг от друга. В терминах длины связи обычно формулируется и критерий качества всего разбиения, который обычно называется функционалом качества.

Третий этап позволяет из анализа поведения собственных значений каждой из компонент строить типовые поля, характерные для различных сезонов, определять степень связности поля для каждого класса.

Использование предложенного подхода к описанию различных периодических (общее содержание озона (ОСО), температура) и непериодических (ветер) метеорологических полей позволило получить представление о структуре и динамике исследуемых полей. Так, зависимость амплитуды от среднего значения ОСО для периода 1997-2004 гг., изображенная на рис. 6, свидетельствует о том, что с ростом среднего значения амплитуда колебаний возрастает, причем скорость возрастания амплитуды ОСО – параметр экспоненциальной интерполирующей функции – практически не зависит от выбора временного интервала и способа наблюдения.

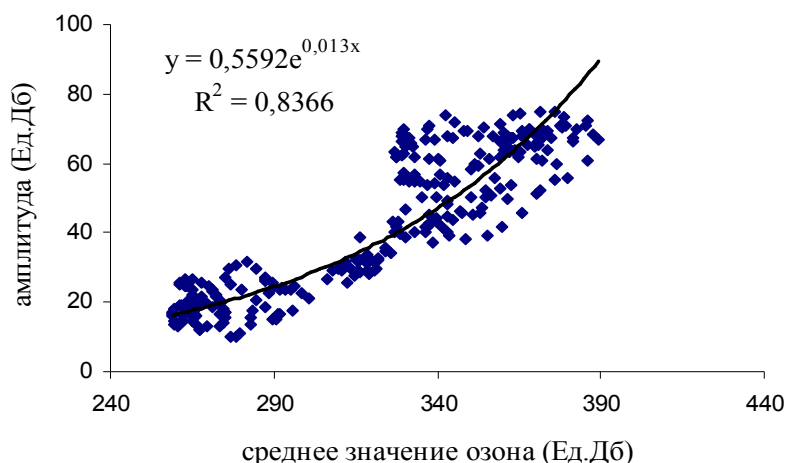


Рис.6. – Зависимость амплитуды от среднего значения ОСО

Классификация и анализ поля ОСО в северном полушарии выявили существование хорошо локализованных областей с повышенным и пониженным содержанием озона относительно среднеширотных значений, картографическое изображение которых приводится на рис.7.

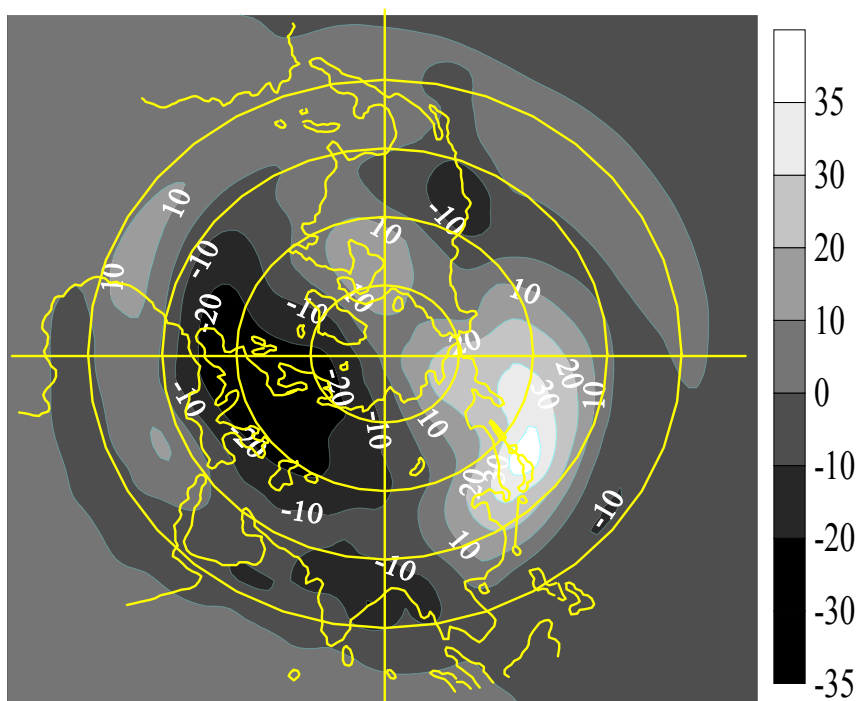


Рис. 7 – Отклонения усредненных значений ОСО в каждой точке наблюдения от среднеширотных значений

Появление таких областей и степень контраста, т.е., разность между максимальным и минимальным значениями в указанных областях, можно связать с временным градиентом ОСО. Так, когда ОСО нарастает, а это, в среднем, наблюдается в период с октября по февраль, контраст непрерывно возрастает, а в процессе уменьшения содержания ОСО (март-сентябрь) контраст между этими областями постепенно уменьшается и, начиная с мая,

практически исчезает. Максимальное значение контраста в декабре достигает довольно большой величины, порядка 160 ед. Дб, что составляет примерно 50% от среднего значения ОСО по всему Северному полушарию.

Результаты проведенных исследований позволяют говорить о существовании устойчивой структуры, состоящей из областей как повышенного, так и пониженного значения ОСО. Надо отметить, что эта закономерность выявлена и для средних значений за все годы, и для каждого года в отдельности, что позволяет высказать предположение о существовании сложного комплекса, скорее всего, природных механизмов, управляющих этим процессом.

В следующем пункте рассматриваются результаты, полученные в ходе исследования структуры ветра, при решении задачи нахождения на основе данных глобальной сети NCEP наиболее информативных районов для измерения в то или иное время года с помощью лидарной установки концентраций вулканических газов около вулкана Этна (о. Сицилия, южная Италия). Решение данной задачи связано с установлением преимущественного или наиболее вероятного направления ветра в районе вулкана на высотах максимального содержания изучаемых газов в разное время года. Фактически, это задача нахождения структуры поля ветра, усредненного за каждый месяц, по данным наблюдения. Сеть NCEP содержит 6-часовые данные основных метеорологических параметров на 17 стандартных высотах в достаточно редкой сетке. Для решения данной задачи использовался подход, сочетающий процедуру выделения структур с интерполяцией. С применением интерполяционной процедуры Гандина по базовой сетке из 25 точек было сгенерировано дополнительно 625 точек, равномерно распределенных по исследуемой области сети, которая охватывала район примерно $10^0 \times 10^0$. Исследование проводилось для данных, относящихся к стандартной высоте $P = 700$ мбар, измеряемой в единицах давления, что соответствует обычной высоте, равной примерно 3.5 км (высота вулкана Этна составляет 3340 м). Результат интерполяции для 2003 года иллюстрируется на левом фрагменте рис. 8, где

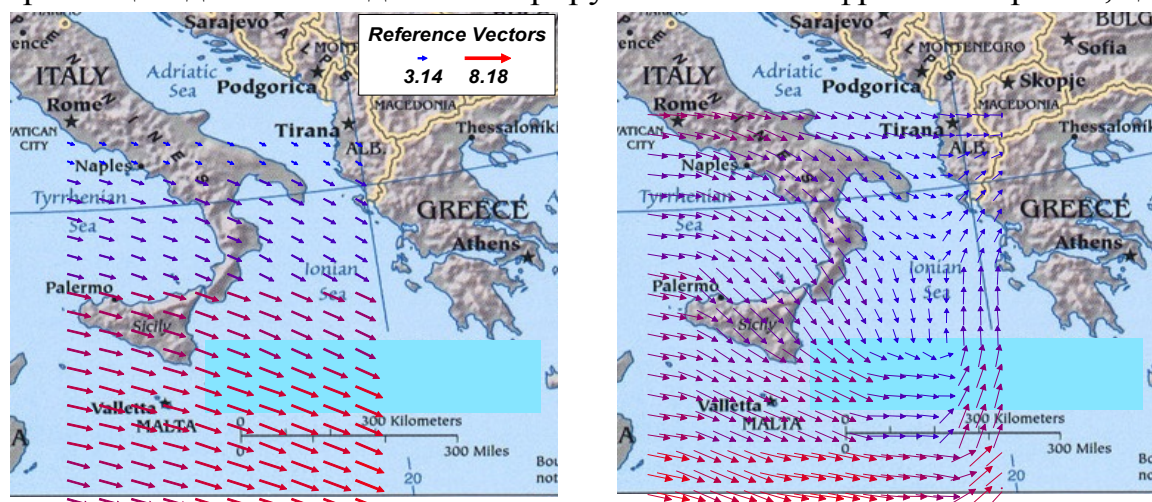


Рис. 8– Поле, усредненного за год, полного ветра на высоте $P=700$ мбар (слева) и на высоте $P=1000$ мбар для апреля 2003 г. (справа)

векторы отображают направление и силу ветра в каждой точке получившейся сети. Для сравнения в правой части рис. 8 приведено поведение приземного ветра. Далее, с применением МВС были выделены районы с приблизительно одинаковым поведением U и V компонент ветра для каждого календарного месяца и в целом для всего года. Для того чтобы выяснить наиболее вероятное направление ветра в то или иное время года были рассчитаны характеристики поля ветра отдельно для каждого месяца. На рис.9 в осях U V помещены средние значения для классов, на территории которых находится Этна, полученных в результате кластерного анализа отдельно U - и V -компонент ветра для каждого месяца (номера точек на рис.9 соответствуют номеру месяца) и для года в целом (точка №13).

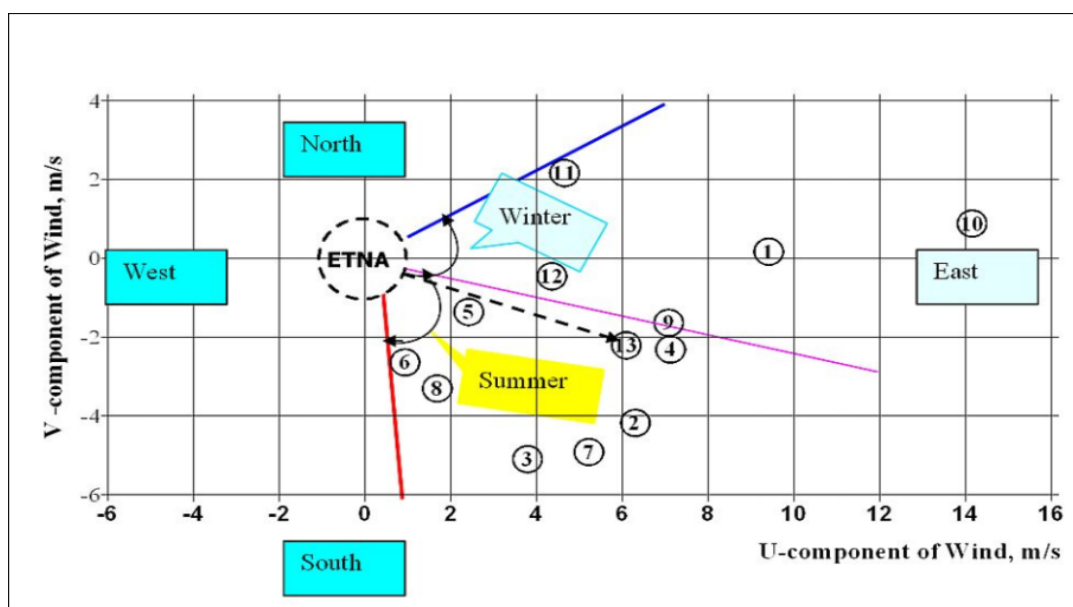


Рис.9 – Положение средних значений по классам за каждый месяц в осях U и V компонент ветра

Из рис. 9 видно, что в локализации точек в фазовой плоскости наблюдается определенная сезонная закономерность, а именно, точки, относящиеся к осенне-зимнему периоду (с сентября по январь) располагаются широтно в узком секторе 38° со средним значением угла, равным 265° . (За ноль считается направление на север и угол отсчитывается против часовой стрелки). Все остальные месяцы попадают во второй сезон (с февраля по август). В этот, условно «летний» сезон, ветер ориентирован больше к югу, сектор тоже достаточно узкий и составляет 52° . Среднее за год значение угла, под которым направлен ветер, таким образом, равно 250.1° .

Четвертая глава посвящена рассмотрению задачи интерпретации сверхширокополосного (СШП) электромагнитного сигнала, отраженного от многослойной диэлектрической структуры. Георадары или геолокаторы – получили за последние 30 лет широкое распространение в различных сферах

деятельности: геологические и инженерные изыскания, дорожные, в том числе и железнодорожные, строительные работы, археологические исследования. Схематически все георадары работают по одному принципу. Одна антенна (передающая) испускает радиосигнал, другая (приемная) принимает сигнал, отраженный от исследуемой среды. Таким образом, отраженный сигнал несет в себе информацию как о структуре подповерхностного слоя, объектах или неоднородностях находящихся в нем, так и о состоянии самой среды.

Фиксируются снятые радаром сигналы в виде радарограммы – совокупности записей сигналов, отраженных от среды и пришедших к приемной антенне в интервале времени от посылки зондирующего сигнала до конца интервала записи. На рис.10 изображена типичная радарограмма, полученная геолокатором ОКО-2 с антенной АБ-400. На ней представлены 279 отраженных сигналов, каждый длительностью 100 нс с шагом 10 см.

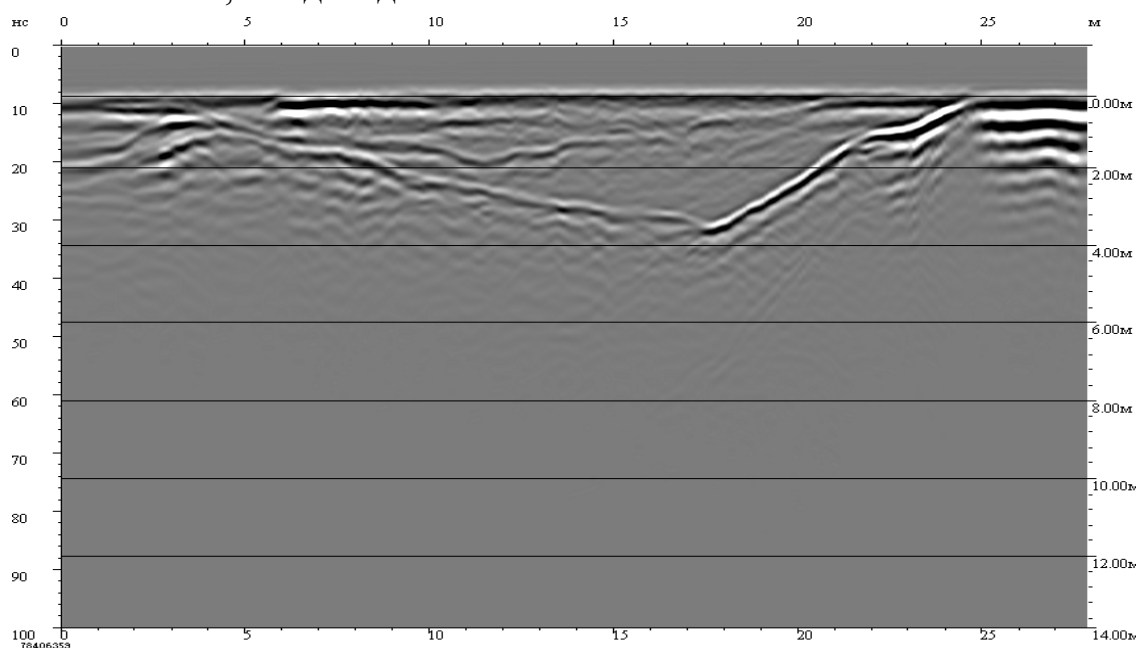


Рис.10 – Пример радарограммы

В идеале интерпретация отраженного сигнала должна позволить восстановить в каждой точке исследуемой среды значения комплексной диэлектрической проницаемости. А зная эту функцию, можно затем высказывать предположение о происхождении той или иной неоднородности. Если создание георадаров сейчас уже не вызывает принципиальных технических трудностей, то интерпретация полученных сигналов по-прежнему является актуальной задачей. Процедура обработки радарограмм практически не автоматизирована и осуществляется интерпретатором. Главная проблема при интерпретации СШП сигнала в общем случае состоит в неоднозначности определения как положения объекта в подповерхностном слое, так и его диэлектрических свойств. Как правило, априорная информация о пространственном поведении комплексной функции диэлектрической проницаемости вмещающей среды отсутствует, и именно с этим связана основная сложность задачи корректной интерпретации отраженного сигнала. В

диссертационной работе в этом случае предлагается использовать комплексный самосогласованный подход, включающий в себя:

1. Определение пространственного распределения температуры и влажности $T(z), W(z)$ в среде с предполагаемыми свойствами при известных внешних климатических условиях и вычисление для этих распределений модельной функции комплексной диэлектрической проницаемости $\varepsilon(z, \omega)$. Разработанный с этой целью пакет *TWES* рассчитывает $T(z), W(z)$ путем совместного решения одномерной системы нелинейных уравнений теплопроводности и диффузии для любых реальных сочетаний температуры и влажности.

2. Решение задачи обнаружения локальных подповерхностных объектов или задачи профилирования по измеренным отраженным СШП-сигналам. 3. Расчет модельного отраженного сигнала на основе полученной выше информации и сравнение его с реальным; коррекция интерпретации с учетом всех имеющихся данных. Интерпретация основывается на закономерностях поведения отраженного сигнала (искажение формы и размытие пиков), в зависимости от степени неоднородности распределения влажности.

Для решения задачи обнаружения в случае, когда известны георадарные сигналы в узлах сетки, автором предлагается использовать алгоритмы *REVPAT* и *BINREL*.

Наличие под поверхностью объекта с другой, нежели у среды, диэлектрической проницаемостью приводит к искажению отраженного сигнала в области нахождения объекта. Если эти точки расположены компактно или демонстрируют иную пространственную закономерность, то сигналы от этой совокупности точек требуют дальнейшего исследования. Для реализации указанных алгоритмов каждая пространственная точка $x_i, i = 1, \dots, N$, в которой фиксировался сигнал, рассматривается как *объект*, а последовательность значений сигнала при временах $t_1, t_2, \dots, t_M$ – совокупность *признаков*. Таким образом, отраженный сигнал представляется в виде вектора размерности M , где M – число временных отсчетов сигнала. В данной задаче под структурой надо понимать совокупность точек, удовлетворяющих следующим требованиям: сигналы в этих точках, во-первых, должны быть удалены от точек фона, а во-вторых, быть близкими друг другу. Иными словами, при классификации точки, относящиеся к объекту, должны попасть в один класс.

Алгоритм *REVPAT* реализует МВС в варианте, описанном выше, т.е., разбивает все множество точек на классы, используя процедуру выделения компонент связности графа. Последующий анализ внутриклассовых характеристик и расстояний между классами позволяет ранжировать все полученные кластеры по степени удаленности от фона.

В алгоритме *BINREL* не применяется процедура выделения компонент связности. В нем для отыскания индивидуальных характеристик объектов используются только бинарные отношения между ними. В результате анализа совокупности отношений каждому объекту k присваивается определенное значение функции c_k , которая характеризует степень относительного контраста

этого объекта. Данный алгоритм удобно применять, когда сканирование ведется по площади и есть необходимость получить представление о зондируемой среде в целом. Результаты работы *BINREL* представлены на рис. 11. Каждая клетка таблицы соответствует пространственной точке изучаемой среды (кирпич в сухом песке), а цифры в ячейке указывают значения c_k данной точки. Георадарные измерения проводились на прямоугольном стенде, поэтому значения контраста для точек, находящихся на границах стенда, были положены равными нулю. Расстояние между точками в обоих направлениях равно 10 см. На рис. 11 серым помечены девять, компактно расположенных точек, обладающих наибольшими значениями c_k . Положение этих точек соответствуют расположению кирпича на стенде.

Рис. 12 иллюстрирует применение алгоритма *BINREL* в модельной задаче построения «радиоизображения» объектов, находящихся за стеной. Определение положения объектов в плоскости, т.е. получение своеобразной фотографии, осуществляется при обработке либо полного сигнала, либо какой-нибудь его части. На этом рисунке изображены контрастные точки, полученные при анализе алгоритмом различных частей сигнала, связанных с отражением от различных объектов, разнесенных в комнате по глубине.

0	0	0	0	0	0	0	0	0	0	0
0	11	16	16	5	0	6	9	15	12	0
0	8	12	10	4	0	5	6	15	15	0
0	25	8	14	29	30	23	5	5	9	0
0	25	3	33	75	109	77	35	4	7	0
0	25	7	44	97	128	89	36	6	8	0
0	21	17	36	72	98	68	32	9	14	0
0	49	17	18	29	24	14	8	15	19	0
0	58	13	8	8	7	4	8	15	17	0
0	42	3	8	8	7	3	7	12	16	0
0	0	0	0	0	0	0	0	0	0	0

Рис. 11 – Значения функции контраста c_k для точек стенда.
(Кирпич на поверхности сухого песка)

Надо подчеркнуть универсальность предлагаемого подхода к обнаружению объектов и построению их «радиоизображений». При использовании данного метода в задаче обнаружения не требуется никакой информации об объектах, среде, форме и природе излучаемого сигнала.

В задаче профилирования предлагается использовать *метод структурного анализа (MSA)*, который базируется на следующей идее. Отраженный от объекта сигнал несет в себе *количественную* информацию о пространственном поведении диэлектрической проницаемости и среды и объекта.

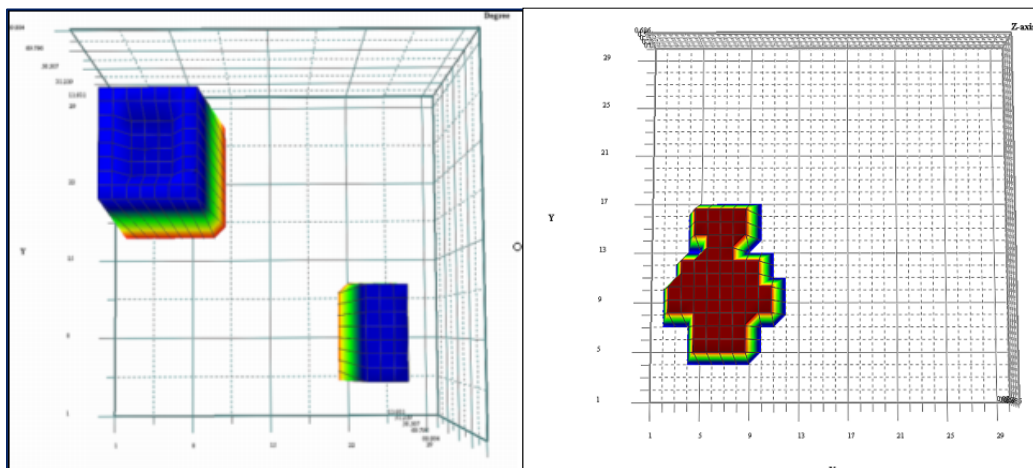


Рис 12 - «Радиоизображение» объектов, находящихся в комнате за стеной.

Все это содержится в численных значениях функции, описывающей отраженный сигнал. При этом структура той части сигнала, которая определяется отражением от объекта или слоя, часто остается неизменной, несмотря на громадное количество возможных сочетаний характеристик среды и объекта. Структура несет в себе качественные отличия объекта и среды. Эти соображения, идеологически близкие к структурным подходам, которые используются в распознавании образов и анализе сцен, лежат в основе предлагаемого метода: отыскивать в сигналах характерные, заранее определенные структуры. В диссертационной работе структура функции задается последовательностью ее максимумов, минимумов и нулей. Для передачи структуры осуществлялась процедура кодирования. Каждой особой точке по определенному правилу ставился в соответствие некоторый символ, что позволяет записывать функцию-сигнал в виде определенного слова – последовательности букв. Несмотря на разнообразие реальных ситуаций, количество качественных вариантов (своеобразных «генотипов»), отражающих соотношения диэлектрических проницаемостей не так велико. Например, для случая, когда среда может рассматриваться как трехслойная структура, генотипов всего четыре: $\epsilon_1 > \epsilon_2 > \epsilon_3$, $\epsilon_1 < \epsilon_2 < \epsilon_3$, $\epsilon_1 > \epsilon_2 < \epsilon_3$, $\epsilon_1 < \epsilon_2 > \epsilon_3$, где ϵ_1 , ϵ_2 и ϵ_3 – диэлектрические проницаемости соответственно 1-ой, 2-ой и 3-ей сред. На рис. 13 изображены фрагменты рассчитанного сигнала, отраженного от объекта с большей диэлектрической проницаемостью, для различных вариантов распределения влажности в первом слое. На нем хорошо просматриваются отражения от передней и задней границ объекта. Ясно, что для малоразмерных объектов, при достаточно большой длительности импульса отражения от границ объекта могут слиться, что приведет к исчезновению некоторых промежуточных особенностей. Это необходимо учитывать. Но самый главный вывод – сигнал, имеющий структуру, как на рис. 13, характеризует отражение от объекта с большей диэлектрической проницаемостью. Следовательно, для широкого круга реальных ситуаций, код отраженного сигнала должен быть неизменным.

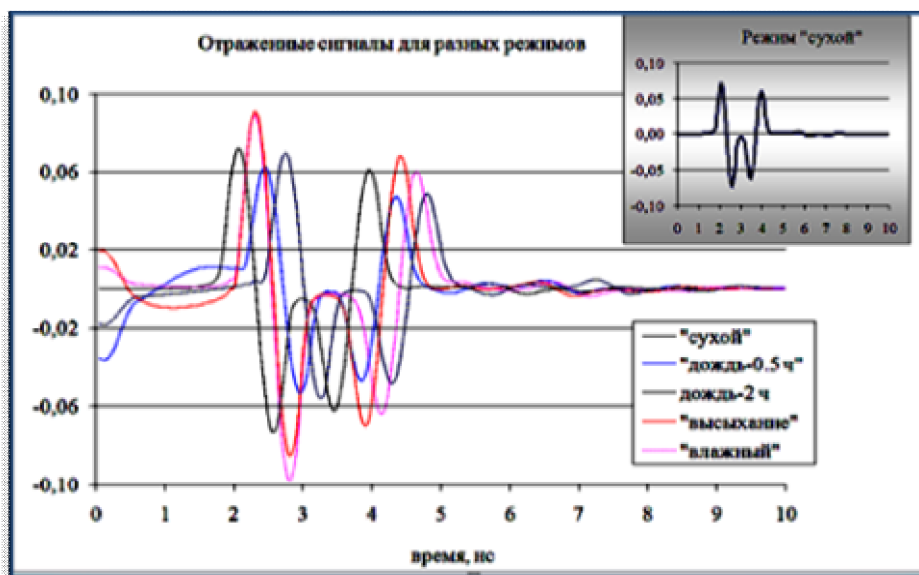


Рис. 13 – Отражение от границы с большей диэлектрической проницаемостью. Теоретический модельный расчет.

В результате применения данного алгоритма интерпретатор получает в автоматическом режиме качественный образ (последовательность «генотипов») исследуемого участка. При этом:

- 1) «гены», т.е. кодовые слова, могут иметь произвольную длину (количество символов);
- 2) можно отыскивать только определенный «ген», соответствующий, например, промоине или подкопу;
- 3) можно закладывать в банк данных «гены» сложных многослойных конструкций.

Особенно удобно использовать данный подход в случае обнаружения вполне определенной неоднородности, например, пустоты. В этом случае обработка данных может проводиться без участия оператора с большой скоростью, что является весьма существенным, т.к. объемы выполняемых работ с помощью георадаров быстро растут, и, тем самым, задача автоматизации процесса обработки сигналов является сейчас необычайно актуальной. Причем, есть задачи, которые не требуют подробного и тщательного изучения исследуемой среды во всех пространственных точках; достаточно получить только качественный результат. Примером такого типа работ могут служить радиолокационные исследования железнодорожного полотна, проводящиеся в рамках инженерно-геологических изысканий с целью обнаружения «больных мест». Описанный выше самосогласованный подход с использованием *MSA* был применен для обработки результатов георадарного исследования участка железной дороги, которые проводили сотрудники отдела инженерной геологии (начальник отдела Сайковский А.В.) ОАО Томгипротранс Брокон С.А. и Скворцов В.С. Распределение реальной части диэлектрической проницаемости в точках среды относительно поверхности приведено на рис. 14.

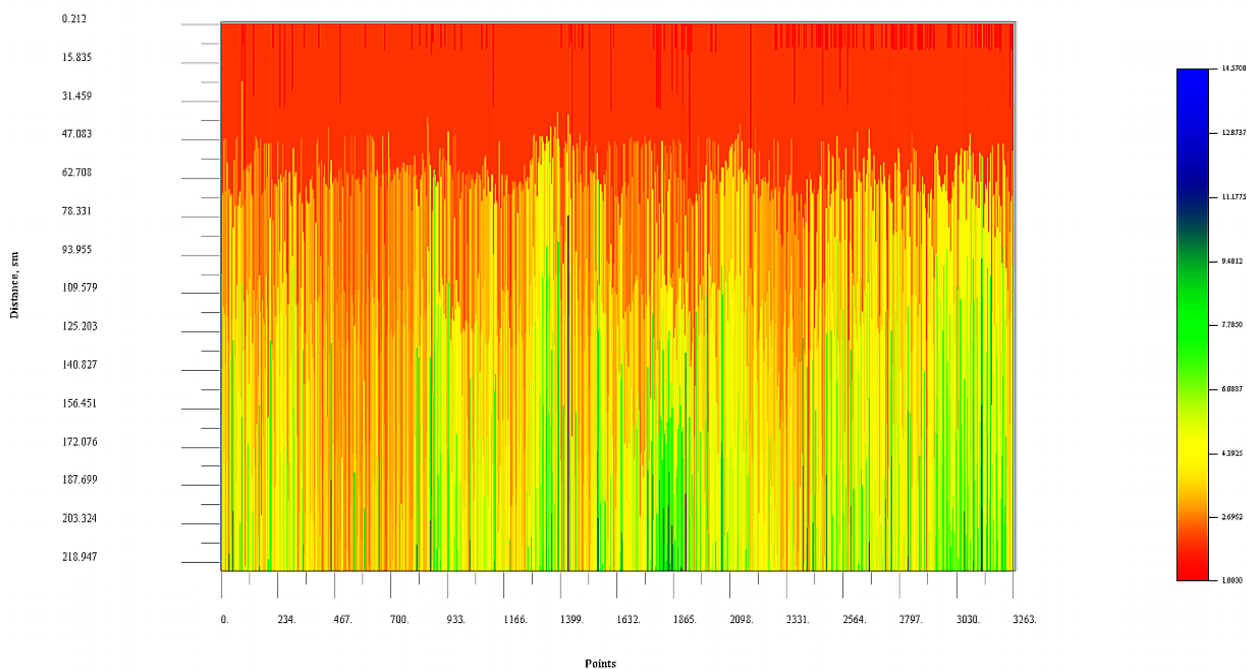


Рис. 14 – Значение реальной части диэлектрической проницаемости в точках среды относительно поверхности, пересчитанные в предположении постоянства диэлектрических проницаемостей в каждом слое

Надо заметить, что применение *MSA* не ограничивается только задачами геолокации. Этот метод универсален. Он дает возможность находить качественно одинаковое поведение в «количественно» различных процессах. Его можно применять либо для поиска заранее заданных структур в массиве данных, либо при классификации; каждый класс будет содержать сигналы, обладающие одинаковым качественным поведением. Так, например, временные ряды метеопараметров в различных точках Земли могут сильно отличаться друг от друга, однако, если в какой-то период времени структуры этих «сигналов» будут одинаковыми, то имеет смысл рассматривать гипотезу о существовании общего влияющего фактора.

В пятой главе приводятся результаты исследования биологических объектов. Живые системы необычайно динамичны, обладают высокой степенью индивидуальности, и это обычно проявляется в данных измерений. Обнаруженные закономерности, особенно, если исследования касаются людей, имеют, в основном, характер только качественных тенденций.

Одна из наиболее важных задач современной науки – исследование динамики климатических параметров и прогноз состояния составляющих физико-географической среды. Прогноз климата тесно связан с реконструкцией прошлого состояния биосферы Земли. Наибольшую ценность для такого рода прогнозов имеют объекты, характеристики которых обладают тесной связью с состоянием окружающей среды. В этом плане деревья являются, наверное, наиболее подходящими объектами. Несмотря на сильную зависимость прироста от локальных особенностей, дерево является индикатором

ретроспективной информации об изменениях в окружающей среде, и одной из задач дендроклиматологии является получение связи индексов прироста деревьев с климатическими переменными. В связи со значительной изменчивостью ширины годичных колец у разных деревьев выбор модельных деревьев для их дальнейшего исследования достаточно сложен и противоречив.

Обычной процедурой для этого является осреднение ширины годичных колец по нескольким деревьям, но при этом невозможно с уверенностью сказать, что «осредненное модельное дерево» отражает общие черты изменения ширины годичных колец всей совокупности деревьев данного ландшафта. По этой причине разработка метода выбора модельных деревьев для различных микроландшафтов является очень важной задачей дендроклиматологии.

Сущность предлагаемого подхода заключается в классификации с помощью МВС всей совокупности выбранных деревьев, а затем в выявлении характера сопряженности внутри каждого из выделенных классов. Материалом для исследования послужили керны сосны болотной в трех различных микроландшафтах Большого Васюганского болота. Для проведения классификации по принципу выбора наиболее похожих кернов временные ряды ширины годичных колец деревьев с разных микроландшафтов объединялись в одну совокупность (исходный массив), затем производилась стандартизация исходной совокупности. Характеристика проведенной классификации для двух массивов представлена на рисунке 15, где по оси абсцисс отложены названия кернов, по оси ординат – номер класса.

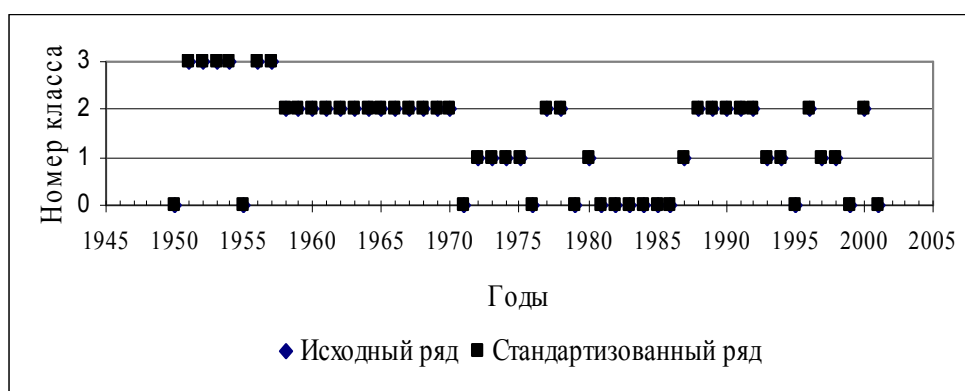


Рис. 15 – Распределение кернов по классам в разных микроландшафтах

Была также проведена классификация, где в качестве объекта использовались годы, а в качестве признаков – керны всех деревьев. В этом варианте выявляются временные структуры, т.е., отрезки времени за период с 1945 по 2005 гг. с близким по характеру радиальным приростом сосны. Классификации подвергались два массива – исходный и стандартизированный, и, согласно рисунку 15, эти периоды полностью совпали. При этом выделились три класса; первый класс отмечался в течение 10 лет, второй – 22 года, при этом в течение 13 лет непрерывно, а третий класс наблюдался в течение 6 лет, и то в начале временных рядов.

Реализация данного подхода, т.е., последовательного применения кластерного и компонентного анализов, позволяет устанавливать группы деревьев, характеризующихся близким по характеру развитием, а также выявлять периоды квазидинакового роста для больших групп деревьев.

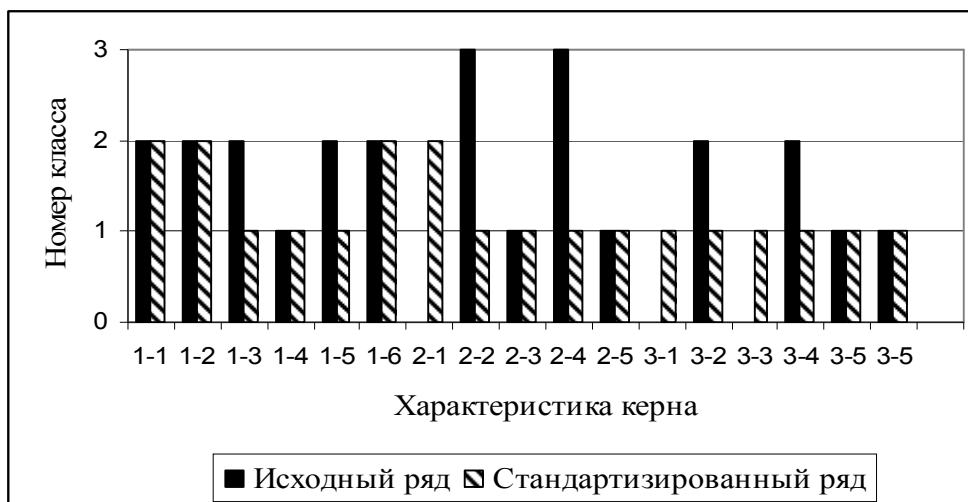


Рис.16 – Распределение по времени периодов с характерным изменением ширины годичных колец сосны болотной

Следующая часть пятой главы диссертации посвящена решению задачи определения дифференциально-диагностических клинико-лабораторных критериев у больных, перенесших инсульт и имеющих депрессивное расстройство или органическое поражение центральной нервной системы (ЦНС) с психическими нарушениями. Подобные критерии позволят получить более достоверные сведения, что, в свою очередь, облегчит ранний выбор адекватной терапии и прогнозирование течения инсульта.

Инсульты являются важнейшей медико-социальной проблемой, что обусловлено их высокой долей в структуре заболеваемости и смертности населения. Инсульт рассматривается как катастрофическая реакция мозга, неконтролируемый стресс, при котором из-за нарушения взаимодействия нейротрансмиттеров могут развиваться депрессивное расстройство и психические нарушения. По данным ВОЗ постинсультная депрессия развивается у 50% больных. Эмоциональные нарушения, возникающие у больного в период инсульта, серьезно сказываются на его физическом и социальном статусе, снижают качество жизни, ведут к суицидальному поведению. Известен способ дифференциальной диагностики депрессивного и психоорганического синдромов у постинсультных больных, основанный на клинической картине, который заключается в сборе анамнестических и клинических признаков и описании их в соответствии с международной классификацией. Однако данные методы основаны на обработке клинических симптомов и признаков, таких как острое начало болезни, неравномерное течение болезни, острые психотические эпизоды. Тестирование же больных из группы с органическим поражением

ЦНС провести просто невозможно, а адекватная оценка психопатологического состояния зависит во многом от субъективного фактора (знания и опыта врача). По этой причине отсутствуют достаточно достоверные и надежные критерии диагноза степени тяжести состояния больного.

В работе предложен дополнительный способ дифференциальной диагностики депрессивных расстройств и органического поражения ЦНС у постинсультных больных. Этот способ основан на результатах кластерного анализа и статистической обработки данных о содержании форменных элементов крови в зависимости от степени тяжести состояния больных и наличия или отсутствия у них сопутствующей депрессии. Всего было обследовано около 420 человек, разбитых на 4 группы: Органическое поражение ЦНС (**П**) с психическими нарушениями наблюдалось у 37 человек. Больные с разной степенью депрессии составили группу из 195 человек, а в летальную группу (**Л**) вошли 50 человек, умерших в острый период инсульта. Контрольная группа состояла из 137 человек. Средние значения и доверительные интервалы для каждой группы и исследуемых показателей белой крови приведены в таблице. Результаты исследования позволили сформулировать дополнительный составной критерий к оценке психического состояния постинсультных больных. Критерий основан на определении в периферической крови количества эозинофилов, базофилов и палочкоядерных нейтрофилов. При количестве эозинофилов от 1.97 до 2.52 %, базофилов от 0.12 до 0.14%, палочкоядерных нейтрофилов от 0.64 до 0.91% диагностируют депрессивное расстройство, а при количестве эозинофилов менее 1.32%, базофилов от 0.03% и меньше, палочкоядерных нейтрофилов от 1.27% и выше диагностируют органическое поражение ЦНС. Кроме этого, выявленные тенденции в поведении средних значений показателей в зависимости от степени тяжести состояния больного можно использовать для прогнозирования течения инсульта.

Таблица. Средние значения показателей белой крови и доверительные интервалы на уровне 95% надежности

Группа	Лейкоциты $\times 10^9$	Сегменто- ядерные нейтрофилы %	Лимфоциты %	Базофилы. %	Палочко- ядерные нейтрофилы %.	Эозино- филы %
К	5,59 $\pm$ 0,40	67,73 $\pm$ 1,80	23,95 $\pm$ 1,51	0,39 $\pm$ 0,17	0,33 $\pm$ 0,11	2,11 $\pm$ 0,31
Д	6,51 $\pm$ 0,58	68,73 $\pm$ 2,53	22,35 $\pm$ 2,15	0,12 $\pm$ 0,11	0,75 $\pm$ 0,33	2,23 $\pm$ 0,55
П	7,85 $\pm$ 1,46	68,68 $\pm$ 5,26	22,03 $\pm$ 5,29	0,03 $\pm$ 0,06	1,27 $\pm$ 0,73	1,32 $\pm$ 0,58
Л	12,53 $\pm$ 1,71	78,68 $\pm$ 3,79	9,13 $\pm$ 1,51	0,04 $\pm$ 0,06	6,26 $\pm$ 3,25	0,3 $\pm$ 0,27

Далее рассматриваются результаты изучения минерального гомеостаза у больных с острой и хронической сосудистой патологией головного мозга. Цель этого исследования – определить область гомеостаза у здоровых людей и найти, какие изменения этой области происходят у перенесших инсульт больных с депрессией и без нее. Существующие методы диагностики являются

либо недостаточно информативными и специфическими для диагностики депрессивных расстройств, а также технически сложными и дорогостоящими, либо с ограниченной областью применения. Как известно, ЦНС оказывает регулирующее действие на содержание и распределение микроэлементов в тканях организма, и при возбуждении ЦНС содержание микроэлементов в крови повышается (вследствие распада белков), а при торможении происходит снижение содержания микроэлементов в крови и накопление их в тканях головного мозга. Были использованы данные обследования содержания некоторых микро- и макроэлементов в крови примерно 200 человек. Из них около 100 с депрессией (после инсульта), примерно 30 без депрессии, и в контрольную группу практически здоровых людей вошло около 40 человек. Проводилось изучение содержания в крови у людей таких микроэлементов, как алюминий, магний, свинец, цинк, медь, марганец, железо. Было получено, что у более чем половины больных с цереброваскулярной патологией отмечается снижение количества алюминия в крови: пределы изменения от 6.67 до 29.6 мкмоль/л, тогда как в контрольной группе эти значения составляют от 20 до 55 мкмоль/л. При наличии сопутствующей депрессии уровень алюминия оказывается еще ниже: от 6.67 до 21.6 мкмоль/л. Этот факт послужил основанием для предложения дополнительного способа диагностики депрессивных расстройств у больных с острой и хронической сосудистой патологией головного мозга. В совокупности с другими методами этот способ позволяет значительно повысить информативность, что дает возможность быстро провести обследование у больных с разной степенью вербальных нарушений, выделить группы риска, назначить адекватную медицинскую коррекцию депрессивных расстройств.

В заключение главы рассматривается метод нахождения контура движущегося объекта из анализа потока изображений и определения положения центра тяжести. Это важная задача, поскольку параметры ходьбы являются индикаторами, позволяющими определить состояния пациента. Метод базируется на МВС и может быть реализован в зависимости от цели с использованием как алгоритма *REVPAT*, так и *BINREL* без какой-либо предварительной обработки данных. Результаты определения центра тяжести обоими алгоритмами разнятся незначительно. На рис.17 представлен результат обработки последовательных изображений, полученный с применением алгоритма *REVPAT*.

Шестая глава содержит описание структуры и основных компонент созданных программ. Все программы написаны на языке Фортран.

Целью программного обеспечения *ICCC* является создание удобного, практически полностью автоматизированного пакета для проведения процедур многомерного анализа, описанных в главе 1, с массивом экспериментальных данных произвольной природы, представленным в виде таблицы типа «объект-свойство». *ICCC* (блок-схема на рис. 18) включает в себя следующие основные процедуры.

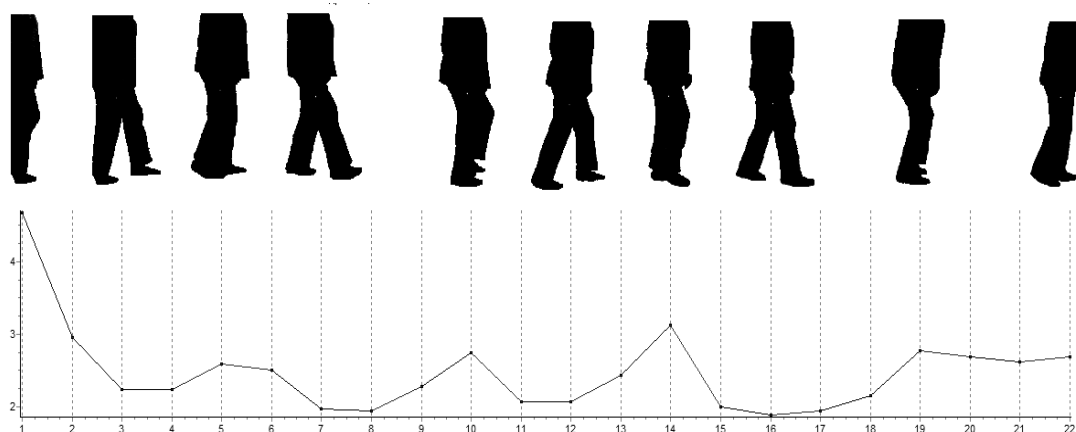


Рис. 17 – Оценки положения центра тяжести движущегося человека, полученные с помощью алгоритма REVPAT

1. *Предварительный анализ исходных данных.* Выбирается вариант решения задачи, формируются аксиомы структуры.
2. *Интерполяция.* Для каждой заданной пользователем пространственной точки проводится интерполяция. Контроль качества интерполяции осуществляется с помощью процедуры взаимной интерполяции базисных станций.

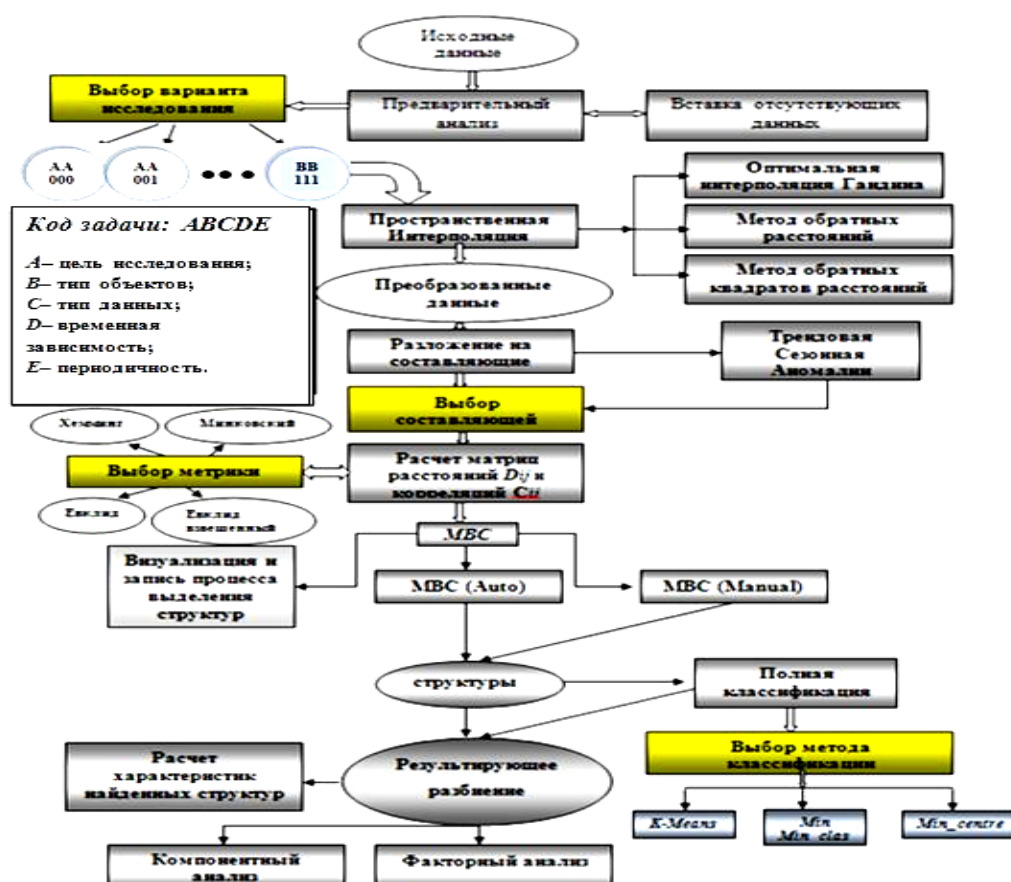


Рис.18 – Структурная блок-схема пакета ICCS

3. *Разложение на составляющие.* На этом этапе выполняются следующие процедуры:
- выделение тренда, заданного полиномом l -ой степени и удаление тренда из исходных временных рядов;
 - выделение циклической составляющей заданного периода;
 - образование аномальной составляющей путем удаления из исходного ряда тренда и циклической составляющих.
4. *Нахождение структур.* С использованием МВС отыскиваются структуры, формируется вариант «оптимального» разбиения.
5. *Нахождение сопряженности полей.* На этом этапе для получения количественных характеристик связности полей в каждом классе и для всего исходного поля используется метод главных компонент.

В следующем пункте главы описывается комплекс программ для анализа георадиолокационных данных *TWES_GEO* (рис. 19). На этом рисунке ОИ – объект исследований, БДПС – база данных параметров сред, СТРУКТУРА – структурная модель среды, АНАЛИЗ – сравнение наблюдаемого и теоретического сигналов. Фактически, *TWES_GEO* состоит из двух программных продуктов: *TWES* и *GEO*, которые могут применяться как совместно, так и по отдельности.

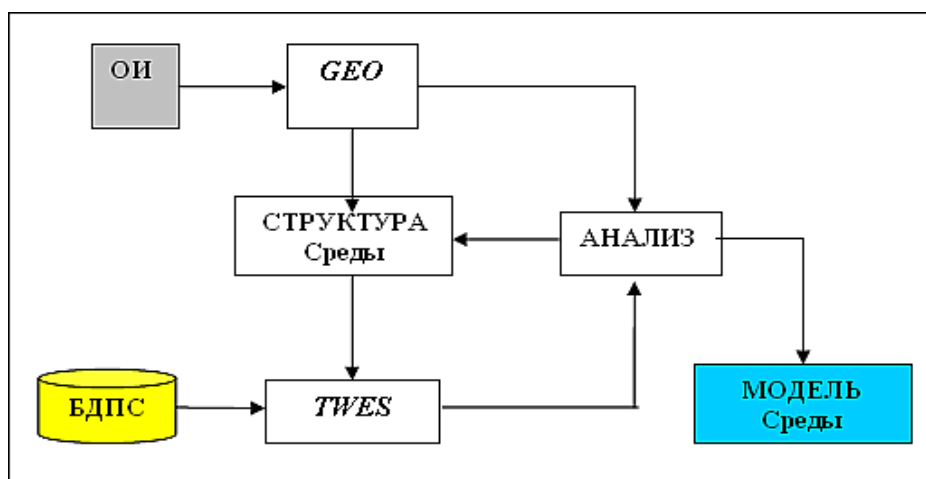


Рис. 19 – Обобщенная структурная блок-схема обработки и интерпретации георадарных данных с использованием *TWES* и *GEO*

TWES предназначена для исследования поведения электромагнитного импульса в слоистой среде с произвольным распределением температуры и влажности, а *GEO* позволяет решать ряд задач, возникающих при георадиолокационном зондировании. *TWES* для заданной одномерной среды и определенных граничных условий рассчитывает путем совместного решения нелинейных уравнений теплопроводности и влажопроводности пространственные распределения температуры, влажности, комплексной функции диэлектрической проницаемости, а также отраженного от такой среды электромагнитного импульса. Пакет *GEO* реализует различные модификации метода выделения структур и метода структурного анализа.

Пакет программ *APFLOW* обеспечивает выполнение алгоритмов, описанных в главе 2, которые позволяют аппроксимировать входящий поток событий МС-поток, с использованием только времен прихода событий. Блок-схема пакета *APFLOW* помещена на рис. 20.

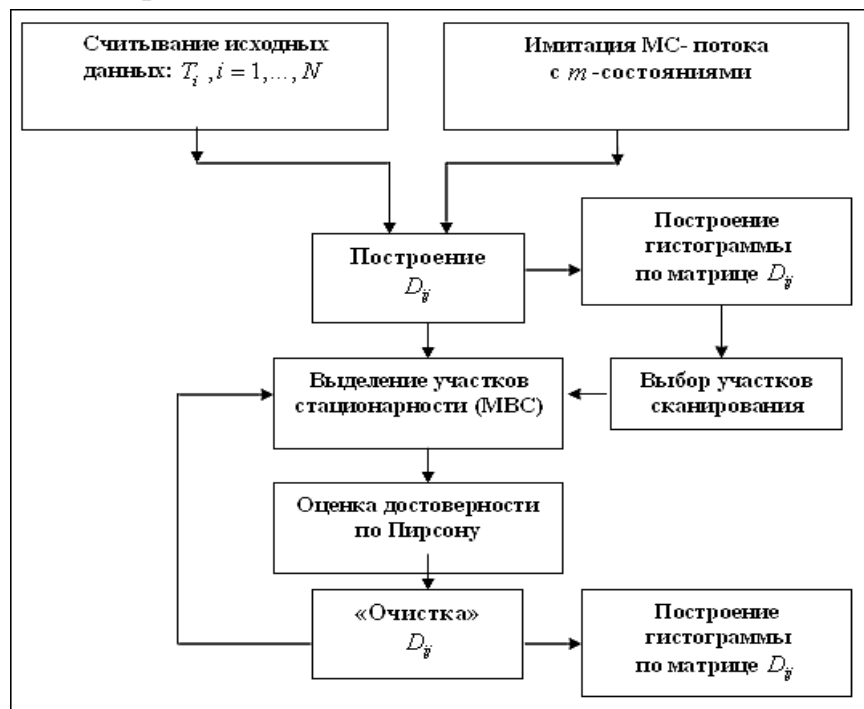


Рис. 20 – Структурная блок-схема процедуры выделения интервалов стационарности во входящем потоке событий

Пакет позволяет работать как с экспериментальным потоком, задаваемым массивом данных времен прихода событий, так и проводить исследования потока, построенного с использованием имитационной модели, генерирующей МС-поток с m – состояниями ($m \leq 5$).

ОБЩИЕ ВЫВОодЫ

Диссертация посвящена разработке и исследованию методов и алгоритмов, которые могут быть использованы для обнаружения закономерностей в экспериментальных данных геофизической и иной природы, что при изучении и контроле состояний сложных систем является одним из начальных этапов, а также построению на их основе соответствующих программных средств. Проведенные автором исследования позволили сформулировать следующие выводы.

1. Разработанная методика выделения структур (МВС) исходного поля или его компонент, по пространственным (или временным) характеристикам позволяет разбить исследуемое пространство (или период времени) на определенное число областей (интервалов времени), в которых временные ряды обладают схожим поведением. Предложенные на основе МВС методы анализа многомерных временных рядов обеспечивают необходимую (наперед

заданную) точность при их аппроксимации. Это дает возможность построить структурную модель поля метеорологического параметра и, тем самым, сжать первичную информацию без существенной потери точности. Так, структурная модель поля концентрации глобально распределенного CO_2 (найденная на основе МВС) включает в себя около сотни регионов, что сокращает необходимый объем дальнейших вычислений в случае сетки $[2.5^0 \times 2.5^0]$, по крайней мере на 2 порядка.

2. Предлагаемый подход к исследованию периодических и непериодических геофизических полей, состоящий из разложения временных рядов (тренд, сезонная и аномалии) и последующей структуризации каждой из этих составляющих, дает возможность получить разнообразную информацию о системе в виде набора параметров, характеризующих результаты разложения, и параметров, связанных с описанием структуры системы. Такая информация вместе с эмпирическими функциональными зависимостями между параметрами может рассматриваться в качестве обобщенных экспериментальных данных о системе. Кроме того, получаемые закономерности могут служить основой для создания гипотез и моделей о факторах, влияющих на исследуемую величину

3. Показано, что пространственная интерполяция повышает эффективность подхода МВС, при изучении геофизических полей на редкой сетке точек. Определен минимально возможный набор вычислительных процедур (интерполяция, выделение составляющих, структуризация и компонентный анализ) для анализа периодических геофизических данных. Получаемый в результате набор характеристик изучаемого поля обладает минимально-достаточными свойствами для его описания.

4. Создан подход, на основе МВС, позволяющий решать задачи обнаружения закономерностей с непериодическими полями. Созданные в рамках подхода алгоритмы, позволяют обнаруживать структуры в исследуемой среде (отдельные объекты, находящиеся в изучаемом слое, либо располагающиеся за преградой) при георадарном зондировании. Результаты обработки реальных геолокационных данных при автоматической обработке радарограмм показали эффективность подхода.

5. Предложенный подход выделения структур позволяет аппроксимировать неизвестный поток событий МС-поток с применением метода нахождения интервалов стационарности, который, в отличие от существующих подходов, использует только времена поступления событий и не требует другой исходной информации о входящем потоке.

6. Показана универсальность МВС при исследовании биологических объектов, которые обладают высокой степенью индивидуальности, что сказывается на пределах изменения каждого изучаемого параметра. Для решения задачи отбора модельных деревьев и описания пространственно-временных характеристик окружающей среды применена методика, в основе которой лежат базисные положения МВС. При нахождении характеристик двигательной активности человека из потока изображений, разработанные

алгоритмы позволяют перейти к режиму обработки, близкому к real-time, в отличие от известных, матричных подходов обработки изображений.

Результаты диссертационной работы показывают широкие возможности применения метода выделения структур в различных модификациях и в сочетании с другими известными методами для решения разнообразных практических задач неразрушающего контроля состояний сложных систем. Надо подчеркнуть, что все используемые варианты МВС, в отличие от многих существующих, не требуют априорной информации о природе исследуемых экспериментальных данных, способах их получения, а также о числе отыскиваемых структур. Они могут быть использованы как самостоятельный исследовательский инструмент, так и в качестве одного из блоков при создании математического обеспечения для решения задач мониторинга окружающей среды, медицины, для геолокаторов и другой геофизической аппаратуры, а также при разработке и исследовании систем массового обслуживания.

Основное содержание работы полностью раскрывается в следующих публикациях

Публикации в периодических изданиях из списка ВАК

1. Загоскин В.В., **Катаев С.Г.**, Тюльков Г.И., Чернышов В.Н. Исследование динамики пространственного распределения влажности, температуры и диэлектрических характеристик в капиллярно- пористых средах // Изв. Вузов "Физика". 1994. № 14. С. 10-16.
2. Загоскин В.В., **Катаев С.Г.**, Тюльков Г.И., Чернышов В.Н., Лукьянов С.П., Ильюшенко В.Н. Влияние тепловлажностных факторов на пространственное распределение диэлектрических характеристик в капиллярно-пористых средах с проводящим компонентом на сверхвысоких частотах // Изв. вузов "Физика". 1996. № 10. С. 55-62.
3. **Катаев С.Г.**, Кусков А.И. Исследование озоновых полей над территорией России и сопредельных государств. I. Составляющие полей озона и их структура // Вестник ТГПУ. Сер.: Естественные и точные науки. 1998. Вып. 5. С. 3-9.
4. **Катаев С.Г.**, Кусков А.И. Исследование озоновых полей над территорией России и сопредельных государств. II. Классификация составляющих полей озона // Вестник ТГПУ. Сер.: Естественные и точные науки. 1998. Вып. 5. С. 10-17.
5. **Катаев С.Г.**, Кусков А.И. Проблемы исследования геофизических полей // Вестник ТГПУ. Сер.: Естественные и точные науки. 2000. Вып. 2 (15). С. 21-27.
6. Беккерман Е.Н., **Катаев С.Г.**, Катаева С.С. Об одном алгоритме распознавания МС-потoka // Вестник ТГУ. 2000. № 271. С. 45-49.
7. **Катаев С.Г.**, Кусков А.И. Внутренняя структура полей температуры в Томской области // Вестник ТГПУ. Сер.: Естественные и точные науки. 2000. Вып. 2 (15). С. 33-39.

8. Кусков А.И., **Катаев С.Г.** Закономерности современных изменений теплового поля в приземном слое атмосферы Сибири и на Дальнем Востоке //Изв. Вузов «Физика». 2004. №11. С. 81-92.
9. Катаева Н.Г., Корнетов Н.А., **Катаев С.Г.**, Левина А.Ю. Клинико-лабораторная диагностика органических и депрессивных психических расстройств у постинсультных больных //Российский психиатрический журнал. 2005. №4. С. 9 – 12.
10. Катаев М.Ю., **Катаев С.Г.**, Андреев А.Г., Базелюк С.А., Лукьянов А.К. Непараметрические математические методы восстановления общего содержания CO₂ из данных спутникового мониторинга//Доклады ТУСУР. 2011. Т.2(24), ч. 3, с.181-186.
11. Катаев М.Ю., **Катаев С.Г.** Эффективный вычислительный метод выявления пространственных структур профилей концентрации углекислого газа// Доклады ТУСУР. 2011. Т.2(24), ч. 3, с.195-200.
12. Катаев М.Ю., **Катаев С.Г.** Численный метод и алгоритм определения центра тяжести движущегося человека из анализа потока изображений//Доклады ТУСУР. 2011. Т.2(24), ч. 3, с.201-205.
13. Катаев М.Ю., **Катаев С.Г.**, Максюттов Ш., Андреев А.Г., Базелюк С.А., Лукьянов А.К. Математические алгоритмы обработки и анализа данных спутникового фурье-спектрометра в ближней ИК-области спектра.//Изв. Вузов «Физика». 2012, №3. С. 84-90.
14. **Катаев С.Г.**, Катаев М.Ю. Математический метод и алгоритм сжатия многомерных временных рядов//Доклады ТУСУР. 2012. Т.1(25), ч. 2, с.208-2013.
15. **Катаев С.Г.**, Катаев М.Ю. Математический формализм и алгоритм структурирования многомерных данных //Ползуновский вестник № 2/1, 2012, с.99-103.

Монографии

16. Региональный мониторинг атмосферы: Коллективная монография/Под ред. М.В. Кабанова. Томск: МГП «РАСКО», 2000. Ч. 4: Природно-климатические изменения, гл. 1: Региональные климатические изменения метеорологических полей /[Задде Г.О., **Катаев С.Г.**, Кусков А.И.]. С. 11.
17. Кусков А.И., **Катаев С.Г.** Структура и динамика приземного температурного поля над азиатской территорией России. Томск: Изд-во ТГПУ, 2006. 176 с.

Патенты

18. Катаева Н.Г., Корнетов Н.А., Баранова О.В., Иванова Е.В., **Катаев С.Г.**, Левина А.Ю., Таткина Е.Г. Способ диагностики депрессивных расстройств у больных с острой и хронической сосудистой патологией головного мозга:

патент 2210778 Рос. Федерация. № 2002100489/14; заявл.08.01.02; опубл. 20.08.2003, Бюл. № 23. 11 с.

19. Катаева Н.Г., **Катаев С.Г.**, Левина А.Ю., Корнетов Н.А. Способ дифференциальной диагностики депрессивных расстройств и органического поражения ЦНС у постинсультных больных: патент 2254571 Рос. Федерация. № 2004102637/15; заявл. 29.01.04; опубл. 20.06.05, Бюл. №17. 11 с.

Научные статьи в других изданиях

20. Комаров А.И., Кусков А.И., **Катаев С.Г.** Классификация составляющих полей температуры в районе Большого Васюганского болота //Современные проблемы физики и технологии: сб. статей молодых ученых. Томск: Изд-во ТГУ, 2002. С. 204-207.
21. Беккерман Е.Н., **Катаев С.Г.**, Катаева С.С., Кузнецов Д.Ю. Аппроксимация МС-потокм реального потока событий // Вестник ТГУ. 2005. Приложение №14. С. 248-252.
22. Кусков А.И., Комаров С.И., Волкова М.А., **Катаев С.Г.** Циркуляционные условия формирования полей осадков теплого периода года на территории Томской области // Вестник ТГПУ. Сер.: Естественные и точные науки. 2006. № 6 (57). С. 87-92.
23. Кусков А.И., Волкова М.А., **Катаев С.Г.** Формирование поля долговременного тренда осадков на азиатской территории России //Вестник ТГПУ. Сер.: Естественные и точные науки. 2006. № 6 (57). С. 92-98.
24. **Катаев С.Г.**, Чернышов В.Н., Загоскин В.В. Исследование влияния профиля влажности на электромагнитный сигнал, отраженный от подповерхностного объекта //Изв. Вузов «Физика». 2008. № 6. С. 66-72.
25. **Катаев С.Г.** Об одном методе обнаружения подповерхностных объектов из данных сверхширокополосного зондирования //Изв. Вузов «Физика». 2008. № 11. С. 93-100.
26. Инишева Л. И., **Катаев С.Г.**, Иваницкий А. Е. и др. Оценка биохимической активности болотных экосистем с использованием кластерно-аналитического метода //Вестник ТГПУ. Биология 2009. № 6 (84). С. 112-116.

Свидетельства о регистрации программ для ЭВМ

27. Свидетельство о государственной регистрации программы для ЭВМ №2011612845. «MRS». Программа выделения структур из массива данных/**Катаев С.Г.** Зарегистрировано в Реестре программ для ЭВМ 8 апреля 2011 г.
28. Свидетельство о государственной регистрации программы для ЭВМ №2011612847. «BINREL». Программа выделения структур по данным зондирования на регулярной сетке/ **Катаев С.Г.** Зарегистрировано в Реестре программ для ЭВМ 8 апреля 2011 г.

29. Свидетельство о государственной регистрации программы для ЭВМ №2011612846. «APFLOW». Программа аппроксимации случайного потока событий марковским потоком /**Катаев С.Г.** Зарегистрировано в Реестре программ для ЭВМ 8 апреля 2011 г.

Материалы научно-технических конференций

30. Загоскин В.В., **Катаев С.Г.**, Тюльков Г.И., Чернышов В.Н. Влияние тепловлажностных факторов на распространение широкополосного сигнала в капиллярно-пористых средах // Тезисы докладов 4-ой Междунар. науч.-техн. конф. «Распространение и дифракция электромагнитных волн в неоднородных средах». Москва, 1994. С. 132.
31. Zagoskin V.V., **Kataev S.G.**, Tulkov G.I., Chernyshov V.N., Lukyanov S.P., Ilushenko V.N. Methods and models for radiowave investigation of the soils // Targets superresolution in subsurface radar. Progress in Electromagnetics Research Symposium (PIERS), Proceedings. July 8-12, 1996. University of Innsbruck, Austria, 1996. P. 340.
32. **Катаев С.Г.**, Кусков А.И. Классификация полей суммарного озона над Россией и сопредельными странами // Третий Сибирский Конгресс по прикладной и индустриальной математике (ИНПРИМ-98). Тезисы докладов, часть IV. Новосибирск: Изд-во Института математики СО РАН, 1998. С. 100.
33. **Катаев С.Г.**, Чернышов В.Н. Обнаружение и идентификация малоконтрастных и малоразмерных объектов из данных подповерхностного зондирования // Третий Сибирский Конгресс по прикладной и индустриальной математике (ИНПРИМ-98). Тезисы докладов, часть IV. Новосибирск: Изд-во Института математики СО РАН, 1998. С. 96.
34. Кусков А.И., **Катаев С.Г.**, Комаров А.И., Чередыко Н.Н. Мониторинг концентрации загрязнения атмосферы // Материалы международной конференции «Экология и рациональное природопользование на рубеже веков. Итоги и перспективы». Томск, 2000. Т. 2. С. 56-59.
35. Кабанов М.В., Задде Г.О., Ипполитов И.И., Кусков А.И., **Катаев С.Г.** Пространственно-временные изменения температурного режима Сибири // Материалы VII Межд. Симп. «Оптика Атмосферы и Океана», 2000. С. 44.
36. Загоскин В.В., **Катаев С.Г.**, Ильюшенко В.Н., Чернышов В.Н. Алгоритмы обнаружения малоконтрастных объектов из данных подповерхностного зондирования // Доклады Сибирского поляризационного семинара СибПол 2004. Сургут, 7-9 сентября, 2004. С. 232-237.
37. **Kataev S.G.**, Kuskov A.I. An approach for the investigation of structure and dynamic of the climatic field // Conference on spatial interpolation in climatology and meteorology. Budapest, October 24-29, 2004. P. 31.
38. **Катаев С.Г.**, Чернышов В.Н., Загоскин В.В., Ильюшенко В.Н. Об одном методе обнаружения подповерхностных объектов из данных сверхширокополосного зондирования // Труды XXIV Всероссийского

- симпозиума «Радиолокационное исследование природных сред». СПб., 2007. Вып. 6. С. 205-210.
39. Чернышов В.Н., **Катаев С.Г.**, Загоскин В.В., Ильюшенко В.Н. О влиянии профиля влажности на форму отраженного от поверхности электромагнитного сигнала // Труды XXIV Всероссийского симпозиума «Радиолокационное исследование природных сред». СПб. 2007. Вып. 6. С. 211-220.
 40. **Катаев С.Г.**, Чернышов В.Н., Загоскин В.В., Ильюшенко В.Н. Выявление информации из данных электромагнитного мониторинга // Труды XXIV Всероссийского симпозиума «Радиолокационное исследование природных сред». СПб. 2007. Вып. 6. С. 315-327.
 41. Chernyshov V.N., **Kataev S.G.**, Zagoskin V.V., Iljushenko V.N., Ligthart L.P., Yarovoy A.G. The GPR response of a subsurface objects and its use for target detection // Ultrawideband and Ultrashort Impulse Signals, (UWBUSIS'08) Conference, Proceedings. Sevastopol, Ukraine, September 18-22, 2006. P. 143.
 42. Zagoskin V.V., Iljushenko V.N., Ligthart L.P., **Kataev S.G.**, Yarovoy S.G. Dielectric properties of moist soils // 11th International Conference on Ground Penetrating Radar. Columbus Ohio, USA, June 19-22, 2006.
 43. **Катаев С.Г.**, Чернышов В.Н. Использование теоретических и эмпирических закономерностей при анализе GPR-сигнала // Тезисы докладов 3-ей международной научно-практической конференции «Инженерная и рудная геофизика-2007». Геленджик, 23-27 апреля, 2007. С. 78-80.
 44. Кусков А.И., **Катаев С.Г.**, Прейс Ю.И., Несветайло В.Д. Классификация прироста сосны болотной в различных микроландшафтах Большого Васюганского болота // Материалы Всероссийской научной конференции с международным участием. Иркутск: Изд-во Института географии СО РАН, 2007. С. 111-113.
 45. Kataev M.Yu., **Kataev S.G.**, Kuskov A.I. Interpolation of the meteorological information of database NCEP on the any spatial point // International Conference on Environmental Observations, Modeling and Information Systems ENVIROMIS-2008. Tomsk, Russia, 28 June - 6 July 2008. P. 43.
 46. L. Fiorani, A Aiuppa, **Kataev S.G.** et al. Chemicomorphological characterization of Etna's plume by atmospheric lidar, satellite radiometers and in-situ measurements. 29th Annual EARSeL Symposium 2009 - Chania, 15th - 18th June, Crete, Greece. - pp.105
 47. **Kataev S.G.**, Kataev M.Yu. Research of the spatial structure of CO₂ concentration profiles on territories of Western Siberia // International conference on computational information technologies for environmental studies CITES 2011. Tomsk, 3-13 July, 2011. P. 126-127.