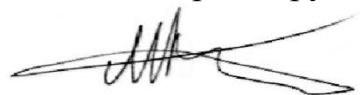


На правах рукописи



Копнов Максим Валериевич

**Модели, алгоритмы и программные средства
для анализа геополей**

05.13.01 — Системный анализ, управление и обработка информации
(отрасль: промышленность)

А в т о р е ф е р а т
диссертации на соискание ученой степени
кандидата технических наук

Томск — 2010

Работа выполнена на кафедре вычислительной техники ГОУ ВПО Национальный исследовательский Томский политехнический университет

Научный руководитель:

доктор технических наук, профессор,
Заслуженный деятель науки РФ
Марков Николай Григорьевич

Официальные оппоненты:

доктор технических наук, профессор
Кочегуров Владимир Александрович

кандидат технических наук
Сарайкин Андрей Витальевич

Ведущая организация:

**Томский государственный
университет систем управления и
радиоэлектроники**
(г. Томск)

Защита состоится 22 декабря 2010 г. в 17 часов на заседании Совета по защите докторских и кандидатских диссертаций Д 212.269.06 при Национальном исследовательском Томском политехническом университете по адресу: 634034, г. Томск, ул. Советская, 84/3, Институт кибернетики ТПУ, ауд. 214.

С диссертацией можно ознакомиться в библиотеке Национального исследовательского Томского политехнического университета по адресу: 634034, г. Томск, ул. Белинского, 55.

Автореферат разослан «___» ноября 2010 г.

Ученый секретарь
Совета по защите докторских
и кандидатских диссертаций Д 212.269.06
кандидат технических наук, доцент



М.А. Сонькин

Общая характеристика работы

Актуальность темы. Одним из наиболее актуальных и динамично развивающихся сегодня направлений в геоинформатике является пространственный анализ объектов, представляющих собой протяженные непрерывные поверхности. Поверхности, однозначно описываемые функцией, зависящей в общем случае от n пространственных координат, получили название n -мерные *геополя* (далее — геополя). Примерами геополей служат распределение в пространстве температуры и давления, высот рельефа местности над уровнем моря, распределения химических элементов в почвах и т.д.

Особым направлением, ставшим актуальным в последние несколько лет, является разработка методов анализа пространственно-временных геополей, когда наряду с зависимостью значений уровня геополя (далее — значений геополя) от пространственных координат имеется зависимость значений от времени.

Под анализом геополей понимают пространственный (в последнем случае и пространственно-временной) анализ поверхностей, включающий также их визуализацию, позволяющий сопоставлять значения геополя и выявлять взаимосвязи между ними. Анализ предполагает использование сложных методов и алгоритмов обработки геополей с целью выявления различных статистических зависимостей и корреляции значений геополя. Важным и практически значимым классом задач анализа геополей являются задачи восстановления геополей. Решение их сводится к поиску значений геополя в точках, где измерения не проводились. Решению задач восстановления геополей посвящено значительное число работ отечественных и зарубежных ученых: В.И. Аронова, В.В. Демьянова, М.Ф. Каневского, О.Р. Мусина, Г. Акима, Д. Дугласа, В. Франклина, Д. Ватсона и других. Однако вследствие некорректности задач восстановления, многие развитые ими детерминистические методы и алгоритмы оказываются неэффективными при решении практически важных задач восстановления геополей, особенно при решении задач восстановления пространственно-временных геополей. Поэтому актуальной продолжает оставаться проблема создания новых методов, алгоритмов и программных средств для восстановления геополей.

В настоящее время на рынке программного обеспечения (ПО) имеется несколько продуктов, предназначенных для анализа геополей (как правило, только двумерных, реже трехмерных). Наиболее полнофункциональный их представитель — пакет программ Surfer компании Golden Software. Каждый из них позволяет использовать тот или иной, часто небольшой, набор функций пространственного анализа. При обработке больших массивов пространственных данных многим из этих продуктов присущи недостатки, связанные с отсутствием в них средств работы с n -мерными данными, а также с отсутствием средств пространственного анализа таких данных. На наш взгляд, более перспективным направлением исследований и разработок является использование для такого анализа геоинформационных систем (ГИС), в которых уже изначально реализо-

ван базовый набор функций пространственного анализа элементарных объектов, и систем управления базами данных (СУБД), позволяющих довольно эффективно решать проблему хранения и обработки больших объемов информации. Проведенные исследования функций современных ГИС показали, что многие из них, к сожалению, имеют небольшой набор средств для решения указанных задач. Среди наиболее успешных разработок ГИС можно отметить продукты Spatial Analyst и Geostatistical Analyst фирмы ESRI (США), причем они не являются самостоятельными системами, а представляют собой дополнительное ПО к ГИС ArcGIS той же фирмы. Однако эти продукты позволяют анализировать только двумерные геополы и у них отсутствуют средства для анализа пространственно-временных геополей.

Учитывая изложенное выше, можно сделать вывод об актуальности проблемы создания моделей, алгоритмов и программных средств для полноценного анализа n -мерных геополей, включая большой их класс — пространственно-временные геополы.

Исследования и разработки по теме были поддержаны грантами РФФИ № 09-08-00309, № 06-08-00840, № 06-0596945-р_офи, №06-07-890, а также финансировалась по НТП Федерального агентства по образованию «Развитие научного потенциала высшей школы» (проект №432, направление 226, раздел 3.3).

Целью диссертационной работы является создание моделей n -мерных геополей, алгоритмов и программных средств для их анализа методами геостатистики, включая анализ пространственно-временных геополей. Для достижения этой цели необходимо решить следующие задачи.

1. Сформировать концепцию анализа геополей с учетом функциональных возможностей современных ГИС и СУБД.

2. Разработать цифровые модели геополей.

3. Разработать алгоритмы для анализа геополей на основе методов геостатистики. Решение этой задачи предполагает модификацию существующих и разработку новых алгоритмов анализа геополей и исследование их эффективности.

4. Создать с учетом разработанных моделей геополей и алгоритмов программные средства для анализа геополей. Эти средства должны интегрироваться с ПО универсальных ГИС и СУБД.

5. Апробировать разработанные модели, алгоритмы и программные средства для анализа геополей путем создания соответствующих информационных технологий для решения практически важных задач в нефтегазовой отрасли.

Методы исследований. В работе использованы методы математического моделирования, теории алгоритмов, методы математической статистики и объектно-ориентированного программирования.

Научную новизну полученных в работе результатов определяют:

1. Многоэтапный способ восстановления геополей по исходным точечным данным, включающий этапы предварительной обработки и вариограммный анализ исходных данных.

2. Алгоритм обработки совпадающих точечных данных, использующий

пространственное индексирование исходного набора точек для повышения вычислительной эффективности многоэтапного способа восстановления геополья.

3. Адаптивный алгоритм поиска ближайших точек к восстанавливаемой точке, отличающийся от известных тем, что позволяет исключить появление неоцененных зон.

4. Алгоритм локализации исходных точечных данных в задаче восстановления геополья.

5. Комплексный подход к восстановлению пространственно-временных геопольей, основанный на использовании при расчете расстояний между точками в n -мерном пространстве псевдоевклидова пространства Минковского или анизотропии.

6. Результаты исследований эффективности предложенных алгоритмов.

Практическая ценность и реализация результатов работы. Практически значимыми являются созданные модели представления геопольей, алгоритмы и программные средства для их анализа. Программные средства функционируют на персональных компьютерах типа IBM PC под управлением операционной системы Microsoft Windows. Объем исходного кода разработанного ПО составляет более 15 000 строк на языке Managed C++.

На основе предложенных моделей, разработанных алгоритмов и программных средств для анализа геопольей созданы информационные технологии для решения ряда практически важных задач в нефтегазовой отрасли, относящихся к классу задач восстановления пространственно-временных геопольей. Среди них задача восстановления недостающих значений потенциала при построении диаграмм потенциалов для установок катодной защиты газо- и нефтепроводов и задача получения отсутствующих значений буферных давлений в скважинах при добыче газа и газового конденсата. Эти информационные технологии реализованы на основе созданного комплекса программ в составе корпоративной геоинформационной системы управления производством (КГСУ) «Магистраль–Восток» и внедрены в ОАО «Томскгазпром» и ОАО «Востокгазпром». Внедрения подтверждены соответствующими актами.

Основные положения, выносимые на защиту:

1. Разработанный алгоритм локализации исходных данных позволяет эффективно решать задачу локализации значений геополья при восстановлении его по точечным данным.

2. Предложенный многоэтапный способ восстановления геопольей по точечным данным и реализующие его алгоритмы, разработанные в том числе на основе методов геостатистики, позволяют более точно и эффективно в вычислительном плане, чем известные детерминистические методы и алгоритмы, решать практически значимые задачи восстановления геопольей.

3. Разработанные на основе методов геостатистики алгоритмы анализа геопольей, результаты исследования этих алгоритмов и разработанные программные средства позволяют создавать информационные технологии для решения практически значимых задач в нефтегазовой отрасли.

Апробация работы. Основные результаты работы докладывались и обсуждались: на III Межрегиональной научно-практической конференции «Газораспределительные системы. АГНКС. АГЗС» (г.Томск, 2003г.), Международном российско-корейском симпозиуме по науке и технологиям KORUS'2004 (г.Томск, 2004г.), X, XI, XVI Международной научно-практической конференции студентов, аспирантов и молодых ученых «Современные техника и технологии» (г.Томск, 2004г., 2005г., 2010г.), Всероссийском конкурсе инновационных проектов аспирантов и студентов по приоритетному направлению развития науки и техники «Информационно-телекоммуникационные системы» (г.Москва, 2005г.), II, III, VI, VII Всероссийской научно-практической конференции-конкурсе «Технологии Microsoft в теории и практике программирования» (г.Новосибирск 2005г., 2006г.; г.Томск 2009г., 2010г.), Международной научно-практической конференции «Интеллектуальные информационно-телекоммуникационные системы для подвижных и труднодоступных объектов» (г.Томск, 2010г.).

По результатам исследований опубликовано 12 работ, в том числе 9 статей, из них одна в издании из перечня ВАК.

Личный вклад:

1. Постановка задач исследования и разработка концепции анализа полевых геополей, а также постановки задач исследования эффективности предложенных автором алгоритмов выполнены совместно с Н.Г. Марковым.

2. Модели представления геополей разработаны лично автором.

3. Алгоритмы восстановления геополей методами геостатистики, адаптивный алгоритм поиска ближайших точек и программная реализация этих алгоритмов, а также исследования эффективности алгоритмов выполнены автором совместно с Р.В. Ковиным.

4. Многоэтапный способ восстановления геополей, алгоритмы локализации исходных данных, вариограммного анализа и автоматизированного подбора параметров для восстановления геополей методом кригинг, их программная реализация разработаны автором, исследования их эффективности проведены автором лично.

5. Информационная технология для восстановления значений потенциала при построении диаграмм потенциалов газо- и нефтепроводов, информационная технология получения отсутствующих значений буферных давлений в газовых скважинах и результаты исследований этих технологий получены автором.

Объем и структура работы. Диссертация состоит из введения, четырех глав, заключения, списка использованных источников из 102 наименований и одного приложения. Объем основного текста диссертации составляет 135 страниц машинописного текста, иллюстрированного 52 рисунками и 2 таблицами.

Содержание работы

Во введении обосновывается актуальность темы диссертационной работы, формулируются цель и задачи исследования.

В первой главе приведен аналитический обзор состояния проблемы анализа n -мерных геополей.

N -мерное геополе представляет собой поверхность, в общем случае описываемую однозначной функцией $f = F(\vec{s})$, где $\vec{s}(s_1, s_2, \dots, s_n)$ — координаты точки в n -мерном пространстве. Частным случаем n -мерного геополя является пространственно-временное геополе, когда одной из координат является время t . Такое геополе помимо пространственного распределения его значений также содержит информацию о том, как поверхность (явление) изменялась с течением времени.

Рассматривается классификация задач, возникающих при анализе геополей. Значительное внимание уделяется цифровым моделям геополей и их визуализации. Приводятся результаты анализа современного состояния проблемы применения ГИС и ГИС-технологий в решении задач анализа геополей. Показывается, что на сегодняшний день существует несколько подходов к решению обозначенной проблемы, однако наиболее перспективным является путь создания алгоритмических и программных средств для анализа геополей на основе современной универсальной векторной ГИС.

Описывается постановка одной из основных задач анализа геополей — задачи восстановления n -мерного геополя, указывается, что эта задача является некорректной. Приводится обзор существующих методов и алгоритмов восстановления значений таких геополей в тех точках n -мерного пространства, где измерения геополя не проводились. Показываются основные сложности, возникающие при использовании этих методов, их главные недостатки. Делается вывод о необходимости модификации существующих методов и алгоритмов восстановления геополей и разработки новых алгоритмов восстановления геополей по точечным данным на основе геостатистических методов.

Проводится анализ возможностей разработки моделей геополей на основе моделей данных современных ГИС и СУБД. Делается вывод о том, что существующие модели данных универсальных векторных ГИС недостаточно эффективны для описания и хранения n -мерных геополей.

На основе результатов проведенного анализа проблемы формулируются цель исследования и задачи, решаемые в диссертационной работе.

Во второй главе изложены предлагаемые концепция анализа геополей и обобщенная структура комплекса программ для такого анализа.

Определены требования к технологии (включает модели, алгоритмы, программные средства и методики их применения) для анализа геополей. На основе этих требований сформулированы следующие принципы, составляющие основу концепции анализа геополей.

1. Известные методы геостатистики должны быть обобщены с целью применения в анализе n -мерных геополей, причем одной из координат может быть координата времени (анализируются пространственно-временные геополя).

2. Предлагаемые модели геополей должны опираться на возможности СУБД, поддерживающих реляционные модели данных, создаваемые алгоритмические и программные средства для анализа геополей должны быть макси-

мально абстрагированными от особенностей решения конкретных прикладных задач.

3. Учитывая необходимость в перечне процедур для анализа геополей функций пространственного анализа, в состав комплекса программ для анализа геополей должно быть включено ядро универсальной векторной ГИС.

4. Особенности конкретной прикладной задачи анализа геополей (формат исходных данных, возможность корректировки результатов обработки, способ представления результатов анализа геополя и т.д.) должны быть легко учитываемы при создании на основе разработанных алгоритмов и программных средств информационной технологии для решения этой прикладной задачи (задач).

5. Перед восстановлением геополя должен проводиться предварительный вариограммный анализ исходных данных, который позволит определить теоретическую модель вариограммы, а также параметры для восстановления геополя методом кригинг. Процесс определения теоретической модели вариограммы и параметров метода кригинг должен быть автоматизирован.

Показаны особенности восстановления методами геостатистики n -мерных геополей и одного из их классов — пространственно-временных геополей. При расчете восстанавливаемого значения геополя методом кригинг необходимо рассчитать расстояния (лаги h) между парами точек исходного набора $V(\vec{s}_i)$ и $V(\vec{s}_j)$ и расстояния между точкой $V(\vec{s})$, где необходимо восстановить значение геополя, и точками исходного набора данных $V(\vec{s}_i)$. Один из способов приведения всех координат пространства к единой системе расчетов базируется на использовании понятия анизотропии. В случае n -мерного геополя анизотропия — это различное распределение значений геополя в каждом из n направлений, при этом лаг h вычисляется следующим образом:

$$h = \sqrt{\frac{(s_{i1} - s_{j1})^2}{r_1^2} + \frac{(s_{i2} - s_{j2})^2}{r_2^2} + \dots + \frac{(s_{in} - s_{jn})^2}{r_n^2}},$$
 где $s_{i1}, s_{i2}, \dots, s_{in}$ — координаты $V(\vec{s}_i)$, $s_{j1}, s_{j2}, \dots, s_{jn}$ — координаты $V(\vec{s}_j)$, $r_1, r_2, \dots, r_n$ — радиусы влияния анизотропии в направлениях $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n$.

При восстановлении значений пространственно-временного геополя возможно использование понятия анизотропии. Также можно применять псевдоевклидово пространство — пространство Минковского. В последнем случае связь между пространственными расстояниями и промежутками времени, разделяющими события, характеризуется интервалом d . Например, в случае трехмерного пространства и времени t интервал d рассчитывается следующим образом:

$$d = \sqrt{c^2(t_i - t_j)^2 - (x_i - x_j)^2 - (y_i - y_j)^2 - (z_i - z_j)^2},$$
 где $x_i, x_j, y_i, y_j, z_i, z_j$ — координаты точек в трехмерном пространстве, t_i, t_j — временные координаты, c — скорость света.

Рассмотрены основные форматы двух и трехмерных моделей геополей. Обоснована необходимость разработки форматов для регулярных сетей, используемых в качестве цифровых моделей геополей. Приводится предлагаемый

формат для хранения значений n -мерных, в том числе и пространственно-временных геополей.

Предложена обобщенная структура программного комплекса для анализа геополей. В соответствии с результатами анализа существующего ПО для исследования геополей, включая анализ современных ГИС, показано, что основу программного комплекса должна составлять универсальная векторная ГИС и должна использоваться коммерческая СУБД, поддерживающая реляционную модель данных. В соответствии с выдвинутыми требованиями программный комплекс реализуется по модульному принципу. Каждый из разрабатываемых программных модулей (далее — модули) должен расширять базовые функциональные возможности универсальной векторной ГИС и являться основой для реализации ПО информационных технологий при решении конкретных прикладных задач.

В третьей главе рассматриваются предложенные алгоритмы анализа геополей и приводятся результаты их исследования.

Показано, что одной из основных проблем, возникающей при восстановлении геополей по точечным данным, является низкая вычислительная эффективность существующих алгоритмов интерполяции. Для решения этой проблемы предложен многоэтапный способ восстановления геополя по точечным данным, позволяющий повысить вычислительную эффективность процесса восстановления с помощью выполнения предварительной фильтрации исходных данных и применения методов пространственной индексации этих данных. Предложенный способ также повышает точность восстановления геополей за счет предварительной фильтрации исходных данных, позволяющей исключить из расчета недостоверные данные, а также за счет автоматизированного подбора параметров для метода кригинг, основанного на предварительном анализе исходных данных с помощью вариограмм.

Кратко рассмотрим основные этапы при реализации этого способа.

Этап 1. Загрузка исходных данных. Он предполагает преобразование исходных данных, представленных в различных форматах, к единому формату. В таком формате каждый элемент исходного набора данных представляется в виде пары $(\vec{s}, f)$. Атрибут $\vec{s}$ — координаты точки $(s_1, s_2, \dots, s_n)$ в n -мерном пространстве. Также $\vec{s}$ может содержать дополнительную координату t — время, показывающее когда было измерено значение геополя в этой точке. Атрибут f — значение геополя в этой точке.

Этап 2. Обработка и анализ исходных данных. Из набора данных, сформированных на этапе 1, удаляются точки, атрибут f которых не соответствует заданным условиям. Производится фильтрация исходных данных в пространстве координат, при этом из набора данных удаляются точки, положение которых не соответствует заданным условиям (точки, лежащие вне исследуемой области). Обработываются совпадающие точки в пространстве координат, находятся точки, расстояние между которыми меньше заданной величины. При этом группа близко расположенных точек заменяется одной точкой, а значение атрибута f в ней устанавливается по заданному правилу.

Этап 3. Вариограммный анализ. Используя набор данных, сформированный на этапе 2, строится экспериментальная вариограмма и подбирается теоретическая модель вариограммы. Далее производится автоматизированный подбор параметров теоретической модели вариограммы для последующего восстановления геополя методом кригинг.

Этап 4. Восстановление геополя. Параметры теоретической модели вариограммы, полученные на этапе 3, являются входными данными при восстановлении геополя. Набор данных, сформированный как результат этапа 2, подвергается соответствующей обработке методом простого кригинга.

Рассмотрим основные алгоритмы, реализующие предложенный многоэтапный способ. На этапе 2 в целях минимизации временных затрат целесообразно разрядить исходную сеть точек без существенного ущерба для точности восстановления геополя. Такая обработка особенно необходима в случае кластерного расположения исходных точек, где измерялось геополе. Для этого можно удалить точки, расстояние между которыми превышает заданное значение. Суть предложенного алгоритма удаления совпадающих точек: для каждой точки исходного набора выполняется поиск точек, попадающих в окружность с центром в этой точке и заданным радиусом. В исходный набор точек вносится новая точка с усредненными координатами и значением геополя, вычисленным по выбранному правилу. Текущая точка и найденная группа точек из исходного набора исключаются.

Наиболее трудоемкой частью этого алгоритма является поиск ближайших точек. Действительно, наши исследования показали, что при числе исходных точек $2 \cdot 10^4$ время работы алгоритма становится достаточно ощутимым — около 5 с., а при $2 \cdot 10^5$ точек время работы составляет около 100 с (на рис. 1 график, изображенный пунктирной линией). Исследование этого и описанных ниже алгоритмов проводилось на ПК с процессором AMD Athlon XP 1700+, ОЗУ 512 МБ с ОС Windows 2000. Учитывая, что реальные данные могут включать несколько миллионов точек, необходим более эффективный алгоритм.

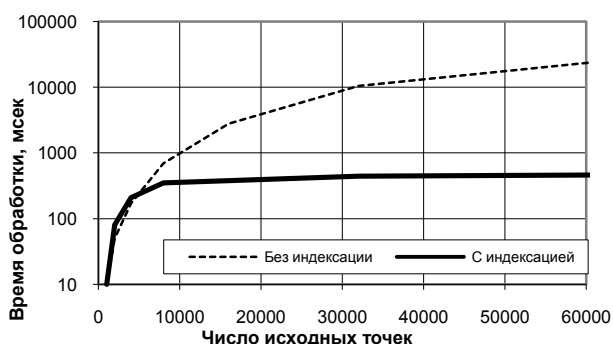


Рис. 1. Зависимость времени обработки совпадающих точек от числа исходных точек

Для повышения вычислительной эффективности исходного алгоритма предложено использовать пространственную индексацию. Показано, что наиболее подходящим методом, не требующим полной перестройки индексной системы при хранении пространственных данных, является *метод сетки*. Этот метод предполагает разбиение исходной области на сетку малых квадратов (ячеек). Создается матрица индексации, в которой каждая ячейка содержит индексы соот-

ветствующих ей точек. При обработке совпадающих точек их индексы удаляются из соответствующих ячеек, а при формировании новой — добавляются в нужную ячейку.

Первым шагом алгоритма является формирование индексной структуры. Для каждой точки определяются индексы ячейки, в которую эта точка попадает. В список этой ячейки заносится указатель на текущую точку. По завершению первого шага каждая ячейка вспомогательной сетки будет содержать список попадающих в нее точек.

На втором шаге алгоритма выполняется поиск точек, удаленных от текущей не далее заданного расстояния. Этот шаг выполняется для всех необработанных исходных точек. Очевидно что, размер ячейки влияет на вычислительную эффективность алгоритма. Проведенное исследование показало, что наименьшее время обработки для практически значимых случаев (когда отношение радиуса совпадения точек к максимальному расстоянию между точками меньше 15 %) получается, когда размер ячейки составляет $4R$, где R — радиус совпадения точек. Так, например, для $8 \cdot 10^3$ точек время обработки в этом случае сокращается в 2 раза, а для числа точек $6 \cdot 10^4$ — почти в 60 раз.

На этапе восстановления геополя часто необходимо решать задачу поиска точек, ближайших к восстанавливаемой. Задача осложняется тем, что пользователем для поиска могут задаваться разные условия: поиск в круге заданного радиуса или поиск заданного числа ближайших точек.

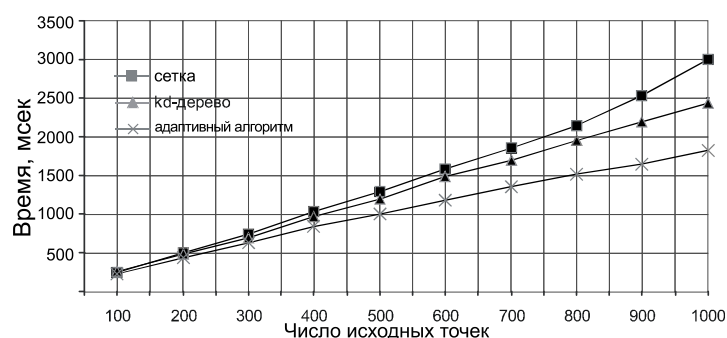
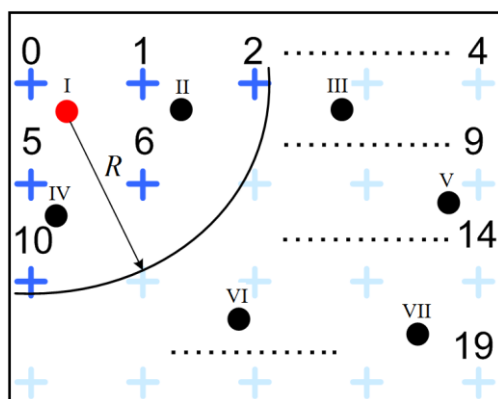


Рис.2 Зависимость времени поиска ближайших точек от числа исходных точек

при заданном радиусе поиска используется индексирование методом kd-дерева, а при заданном числе ближайших точек — метод сетки. Модификация метода kd-дерева заключается в использовании вспомогательной индексной сетки для случаев, когда необходимо расширять радиус поиска для исключения появления неоцененных зон. Адаптивность алгоритма также проявляется в том, что эта вспомогательная структура строится только при появлении первой неоцененной зоны. На рис. 2 приведены результаты исследования вычислительной эффективности этого алгоритма. Анализ результатов показал, что в тех случаях, когда возникают неоцененные зоны, время поиска сокращается в среднем на 15%-20% по сравнению с использованием метода kd-дерева и до 40% с использованием метода сетки.

В первом случае для исключения появления неоцененных зон дополнительно должно быть наложено ограничение на минимальное число ближайших точек. Предложен адаптивный алгоритм поиска ближайших точек, использующий пространственное индексирование методами kd-дерева и сетки. Адаптивность алгоритма проявляется в том, что



а)

	0	1	2	3	4	5	6	7	8	9	10	11	..	19	Номера точек
I	1	1	1	0	0	1	1	0	0	0	1	0	..	0	
II	1	1	1	1	0	0	0	1	0	0	1	1	..	0	
III	0	1	1	1	1	0	0	1	1	1	0	0	..	0	
IV	1	1	0	0	0	1	1	0	0	0	1	1	..	0	
V	0	0	0	0	0	0	0	0	1	1	0	0	..	1	
VI	0	0	0	0	0	0	1	1	1	0	0	0	..	0	
VII	0	0	0	0	0	0	0	0	0	1	0	0	..	1	
Номера ячеек															

б)

Рис.3 Схема индексации исходных точек (а)
и структура таблицы для хранения индексов (б)

Предложен алгоритм локализации исходных точечных данных для восстановления регулярной сети. Суть алгоритма заключается в том, что для каждой точки исходного набора данных с помощью алгоритма Брезенхейма определяются все ячейки регулярной сети, для которых точка исходного набора данных удалена на заданный радиус R . Для хранения информации о том, будет ли точка исходного набора данных участвовать в расчете значения ячейки сети (находится в круге с заданным радиусом), строится двумерная матрица размерностью, равной размерности восстанавливаемой регулярной сети. Каждая ячейка такой матрицы соответствует ячейке регулярной сети и содержит битовый вектор, число элементов которого равно количеству исходных данных (рис. 3). Каждый элемент такого вектора содержит информацию о том, будет ли точка исходного набора данных с соответствующим индексом участвовать в расчетах или нет.

Начало

Шаг 1. Вычисляются индексы узлов регулярной сети лежащие на окружности поиска с заданным радиусом R и центром в точке I.

Шаг 2. Из исходного набора выбирается точка.

Шаг 3. Определяются индексы узлов восстанавливаемой регулярной сети, удаленные от этой точки на заданный радиус R .

Шаг 4. Для каждого узла запоминается порядковый номер текущей точки из исходного набора.

Шаг 5. Если не пройдены все исходные точки, то на шаг 2.

Шаг 6. Для каждого узла регулярной сети вычисляется значение геополя f по всем локализованным точкам.

Конец

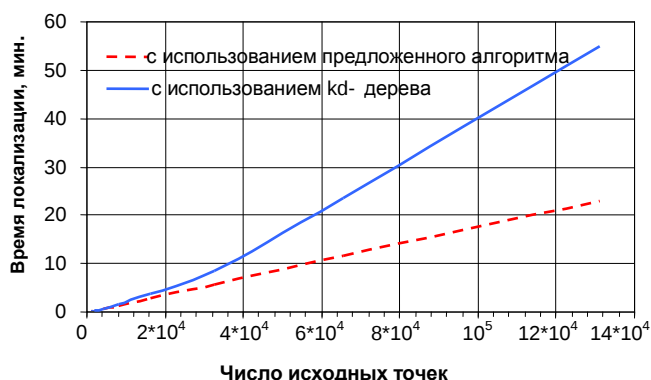


Рис. 4 Зависимость времени поиска ближайших точек от числа исходных точек

в среднем на 12% по сравнению со временем работы алгоритма kd-дерева. При $N > 65 \cdot 10^4$ время локализации с использованием этого алгоритма по сравнению со временем локализации с использованием алгоритма kd-дерева уменьшилось в 2 раза.

Описываются алгоритмы анализа геополей, основанные на методах вариограммного анализа данных и кригинга. Предложен подход к автоматизированному подбору параметров вариограммы. Его суть заключается в использовании при тестировании результатов восстановления геополя кросс-валидации — эмпирического оценивания обобщающей способности алгоритмов, обучаемых по прецедентам. Показано, что если число всех возможных значений параметров вариограммы невелико, то решение задачи поиска значений входных параметров может быть найдено простым перебором за приемлемое время. В противном случае такой перебор потребует больших временных затрат. Для их снижения предложены различные варианты сокращения затрат, в том числе распараллеливание процесса подбора параметров.

Приведены исследования комплексного подхода к восстановлению пространственно-временных геополей, использующего понятие анизотропии или псевдоевклидова пространства Минковского. Показано, что использование пространства Минковского дает сравнимую ошибку при восстановлении геополя с ошибками в случае использования анизотропии, однако в отличие от последнего не требует расчета радиусов влияния для каждого набора данных.

Проведены численные эксперименты для изучения точности восстановления геополя предложенным многоэтапным способом, методом простого кригинга и детерминистическим методом обратных взвешенных расстояний (IDW). В экспериментах использовался генератор случайных чисел, с помощью которого формировались различные наборы нерегулярных точечных сетей с равномерным распределением в интервале -5.12 до 5.12 для каждой из координат x и y . Для каждой точки (x, y) вычислялось значение геополя f в рамках сферической и эллипсоидальной моделей и модели Розенброка. По каждому из этих наборов восстанавливался геополь — регулярная сеть размером 50×50 ячеек. Для сравнения результатов в каждом узле восстановленной регулярной сети рассчитывалась относительная среднеквадратическая ошибка.

Результаты исследования эффективности предложенного алгоритма приведены на рис. 4. Видно, что этот алгоритм позволил существенно снизить временные затраты на локализацию исходного набора данных. Так, даже для сравнительно небольшого количества исходных точек ($N < 8 \cdot 10^3$) использование этого алгоритма позволило сократить время локализации в

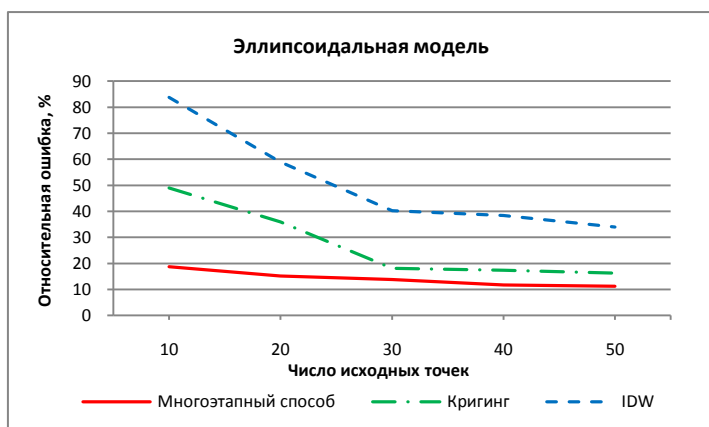


Рис.5 Зависимость ошибки восстановления геополья от числа исходных точек

На рис. 5 в качестве примера для случая эллипсоидальной модели показана зависимость ошибки восстановления геополья, выполненного методом IDW, методом кригинга и многоэтапным способом. Видно, что ошибки восстановления геополья предложенным многоэтапным способом ниже ошибок других методов. При уменьшении количества исходных точек, использованных в расчетах, ошибка восстановления

многоэтапным способом растет не так быстро, как при использовании других методов. Поэтому предложенный способ может быть эффективен даже в тех случаях, когда число исходных точек невелико.

В четвертой главе рассматриваются программный комплекс для анализа геопольей и созданные на его основе и апробированные при решении практических задач информационные технологии.

Предложена структура программного комплекса для анализа геопольей (рис. 6).

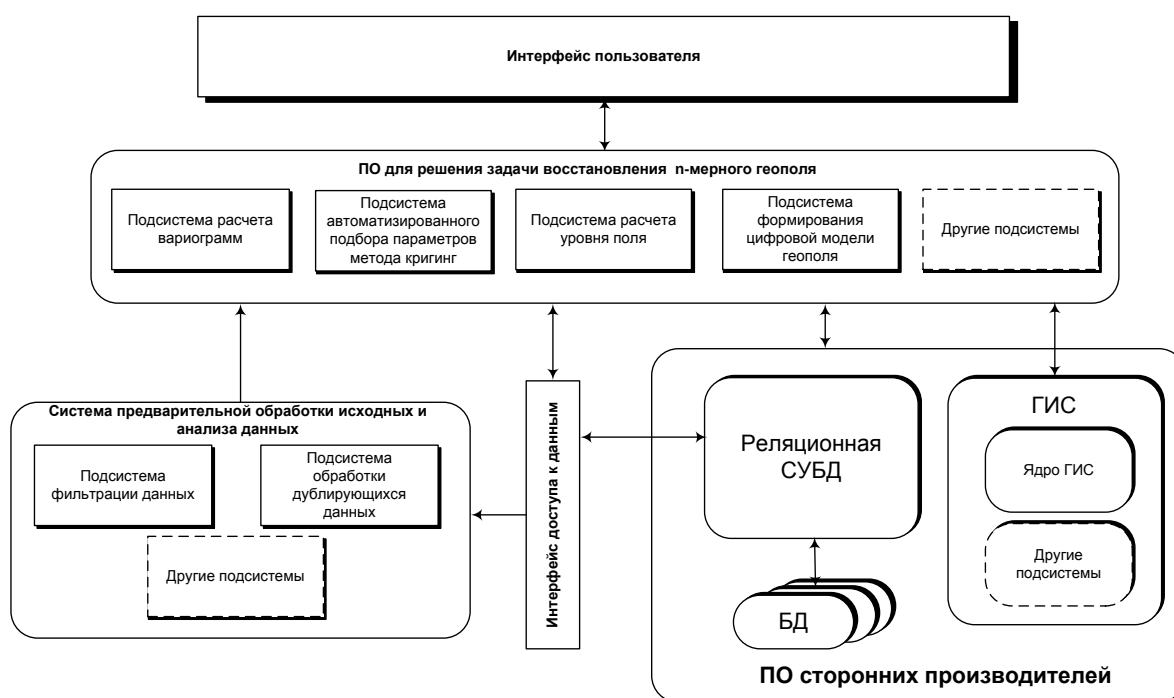


Рис. 6. Схема структуры программного комплекса для анализа геопольей

В основе комплекса лежит как оригинальное ПО (авторская разработка), непосредственно нацеленное на решение прикладных задач анализа геопольей, так и ПО сторонних производителей. Пространственный анализ данных осуще-

ствляется ПО универсальной векторной ГИС MapInfo Professional. Функции хранения и оперирования данными реализованы через модуль *«Интерфейс доступа к данным»* на основе современной реляционной СУБД MS SQL 2008. ПО для решения задач восстановления *n*-мерного геополья включает функциональные модули, реализующие базовые операции восстановления геопольей. Система предварительной обработки и анализа исходных данных является связующим звеном между модулями ПО для восстановления *n*-мерных геопольей, СУБД и ГИС. Интерфейс пользователя, реализованный на языках высокого уровня, расширяет интерфейс пользователя ГИС с помощью дополнительных меню, панелей инструментов и окон.

Большинство модулей программного комплекса реализованы ГИС-независимыми и оформлены в виде динамических библиотек (DLL). Это позволяет создавать информационные технологии в случае использования универсальных векторных ГИС, отличных от ГИС MapInfo Professional.

Комплекс программ может включать в себя дополнительные модули и подсистемы (показаны на рис. 6 пунктиром), которые могут потребоваться при реализации информационной технологии для решения той или иной конкретной прикладной задачи.

Рассмотрены особенности создания информационной технологии для решения задачи восстановления недостающих значений потенциала (напряжения), приходящих от контрольно-измерительных пунктов (КИП) установок катодной защиты (УКЗ) на газопроводах (нефтепроводах), используемых для дальнейшего построения диаграмм потенциалов «труба-земля». Значения потенциала снимаются каждым КИП автоматически через каждые 15 минут, передаются по корпоративной сети и хранятся в БД КГСУ «Магистраль-Восток». Информационная технология создана путем встраивания комплекса программ (за исключением блока *«ПО сторонних производителей»*) в КГСУ «Магистраль-Восток». Рассмотрим основные этапы использования этой технологии.

Этап 1. Из БД КГСУ «Магистраль-Восток» в программный комплекс для анализа геопольей загружаются данные КИП всех УКЗ газопровода или нефтепровода. Технология апробировалась на данных по газопроводу «Мыльджинское ГKM – Вертикос». Всего у данного газопровода 8 КИП, для каждого из них в БД хранится архив значений потенциала «труба-земля». Пространственные координаты для КИП в явном виде отсутствуют, но каждому из них соответствует собственный пикет — километровая отметка, указывающая как далеко от начала газопровода установлен КИП. Газопровод «Мыльджинское ГKM – Вертикос» отрисован на электронной карте в виде полилинии. Для получения пространственных географических координат каждого КИП в соответствии с их километровой отметкой вычислялось положение КИП на трубе и затем уже определялись пространственные координаты. Определение координат пикетов и последующее преобразование из географических координат (долготы-широты) в прямоугольную систему координат производилось средствами геоинформационной подсистемы КГСУ «Магистраль-Восток». Время *t* измерения потенциала «труба-земля» в БД хранится в формате UTC, поэтому преобразо-

вания временной координаты не потребовалось. Таким образом, результатом этапа 1 является для каждого КИП набор данных, каждый элемент которого состоит из вектора пространственных координат $\vec{s}(s_1, s_2)$, t – времени измерения потенциала «труба-земля» и f – значения потенциала «труба-земля». В итоге данные от всех КИП описаны как пространственно-временное геополе.

Этап 2. Каждое значение потенциала имеет атрибут достоверности. Этот атрибут может быть получен непосредственно с устройств телеметрии КИП, либо изменен в КГСУ «Магистраль-Восток». Он может принимать 3 значения: «Достоверно», «Не достоверно», «Достоверность не определена». Для восстановления недостающих значений потенциала можно использовать только достоверные архивные значения потенциалов. Недостоверные значения, а также значения потенциала, достоверность которых не определена, были исключены на этапе фильтрации данных. Исследование объемов данных в БД КГСУ «Магистраль-Восток» в части данных от КИП показало, что за несколько лет работы в системе накоплен огромный архив исходных данных (значений потенциалов). Такой объем данных является избыточным и ведет к существенному увеличению времени расчета восстанавливаемых значений потенциала. Поэтому из всего архива данных оставили только те значения потенциала, которые были измерены в те сутки, когда имелись подлежащие восстановлению недостающие значения потенциала. Возможный диапазон изменений потенциала невелик и составляет интервал от $-2,5$ В до 0 В, поэтому все значения, которые находятся вне указанного интервала, также исключались с помощью подсистемы фильтрации данных. Дублирующиеся значения отсутствуют, т.к. фильтруются ПО КГСУ «Магистраль-Восток» на стадии внесения данных в систему.

Этап 3. По отфильтрованным данным строилась экспериментальная вариограмма. Эмпирически было установлено, что наиболее точно описывает экспериментальную вариограмму сферическая модель теоретической вариограммы.

Этап 4. С учетом результатов этапа 3 проводилось восстановление недостающих значений потенциала.

Для определения пределов применимости предложенных алгоритмов восстановления пропущенных значений потенциала с помощью разработанной информационной технологии был проведен ряд численных экспериментов. Для этого из БД КГСУ «Магистраль-Восток» случайным образом было выбрано 10 дней, в которые производились измерения значений потенциала. В первом эксперименте был исключен из расчета только потенциал, полученный от КИП в $t = 12$ часов и для него в последующем восстанавливалось значение (результаты ошибки восстановления приведены на рис. 7, группа 1). Во втором эксперименте были исключены значения потенциала, снимаемые в 15 минут и 45 минут каждого часа, а также было исключено и впоследствии восстановлено значение потенциала на 12 часов (результаты приведены на рис. 7, группа 2). В третьем эксперименте исключено значение потенциала в точке с $t = 12$ часов (для нее в

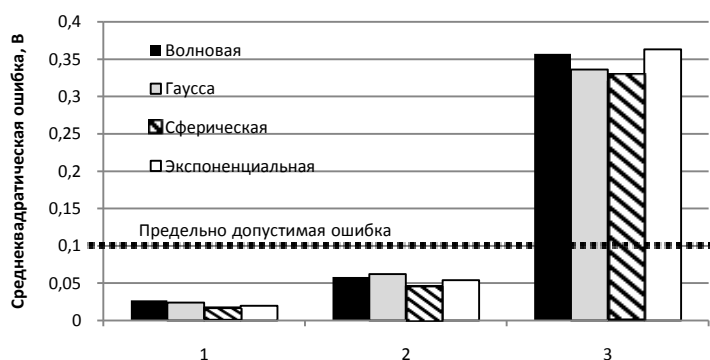


Рис. 7. Среднеквадратичная ошибка восстановления значения потенциала в зависимости от частоты измерения потенциала.

дальнейшем ведется восстановление значения потенциала), а также исключены значения потенциала, снимаемые в 15, 30 и 45 минут каждого часа (ошибка восстановления приведена на рис. 7, группа 3). Как видно из рис. 7 при использовании сферической модели вариограммы ошибка минимальна для всех наборов данных. Для восстановления значения потенциала со среднеквадратической ошибкой

менее 0,1 В (предельная ошибка при измерении потенциала применяемая в нефтегазовой отрасли) достаточно значений потенциала, измеренных каждые полчаса.

Рассмотрены особенности построения информационной технологии для решения задачи получения недостающих значений буферных давлений газовых скважин. ПО этой технологии включено в состав информационно-аналитической подсистемы АРМ геолога КГСУ «Магистраль-Восток». Проведена апробация созданной информационной технологии при анализе данных со скважин ОАО «Томскгазпром»

В заключении приведены основные результаты и выводы диссертационной работы.

В приложение вынесены акты о внедрении полученных результатов.

Основные результаты и выводы по работе

В ходе выполнения диссертационной работы были получены следующие основные научные и практические результаты.

1. Проведен анализ проблемы автоматизации задач анализа n-мерных геополей, позволивший сделать вывод об актуальности тематики по созданию моделей, алгоритмов и программных средств анализа геополей на основе методов геостатистики.

2. Предложена концепция создания технологии (включает модели, алгоритмы, программные средства и методики их применения) анализа n-мерных геополей. Разработана обобщенная структура программного комплекса для проведения такого анализа.

3. Предложен многоэтапный способ восстановления n-мерных геополей по точечным данным, включающий предварительную обработку и вариограммный анализ исходных данных. Результаты численных экспериментов позволили сделать вывод о том, что этот способ позволяет повысить точность восстановления геополей по сравнению с детерминистическими методами и методом кригинга.

4. Разработан адаптивный алгоритм поиска ближайших точек к восстанавливаемой точке, отличающийся от известных тем, что позволяет исключить появление неоцененных зон.

5. Разработан алгоритм локализации исходных точечных данных для восстановления регулярной сети позволяющий существенно сократить время локализации исходных данных при восстановлении геополей.

6. Предложен комплексный подход к восстановлению пространственно-временных геополей, основанный на использовании псевдоевклидова пространства Минковского и понятия анизотропии.

7. Разработана структура комплекса программ и на основе предложенных алгоритмов созданы программные средства для анализа n-мерных геополей.

8. Проведена апробация разработанных алгоритмических и программных средств путем создания информационных технологий решения двух практически важных задач восстановления пространственно-временных геополей в нефтегазовой отрасли. Результаты исследования этих технологий и внедрения их в производство подтвердили эффективность разработанных алгоритмов и программ для анализа n-мерных геополей, включая пространственно-временные геополя.

Основные публикации по теме диссертации

Статьи, опубликованные в изданиях из перечня ВАК:

1. Копнов М.В., Ковин Р.В. Восстановление двумерных геополей: проблемы и решения // Известия Томского политехнического университета. Томск: Изд-во ТПУ, 2007. Т.311. №5. С. 72-75.

Статьи, тезисы докладов на международных и всероссийских конференциях:

2. Ковин Р.В., Кудинов А.В., Копнов М.В., Мирошниченко Е.А., Шерстнев В.С. Корпоративная геоинформационная система для управления производством ОАО «Востокгазпром» // Материалы 3-й Межрегиональной научно-практической конференции «Газораспределительные системы. АГНКС. АГЗС». Томск: Изд-во МДМЦ, 2003. С. 112–119.

3. Копнов М.В., Гутов Д.В., Марков Н.Г. Разработка методов, алгоритмов и программных средств восстановления трехмерных геополей // Сборник материалов Всероссийского конкурса инновационных проектов аспирантов и студентов по приоритетному направлению развития науки и техники «Информационно-телекоммуникационные системы» / Под. ред. А.О. Сергеева. М.: ГНИИ ИТТ «Информика», 2005. С. 114.

4. Копнов М.В., Ковин Р.В. Использование алгоритмов геометрического поиска при восстановлении геополей // Материалы X Юбилейной Международной научно-практической конференции студентов, аспирантов и молодых ученых «Современные техника и технологии». Томск: Изд-во Томского политехнического университета, 2004. С. 158–160.

5. Копнов М.В., Ковин Р.В. Минимизация времени восстановления геополей при решении геологических задач // Материалы научно-практической

конференции «Проблемы и перспективы развития минерально-сырьевого комплекса и производительных сил Томской области». Новосибирск: СНИИГ-ГиМС, 2004. С. 265–267.

6. Копнов М.В., Ковин Р.В., Марков Н.Г. Автоматизация процесса построения прогнозных карт и оценки их качества // Сборник трудов VII Всероссийской научно-практической конференции-конкурса «Технологии Microsoft в теории и практике программирования». Томск: Изд-во Томского политехнического университета, 2010. С. 57–58.

7. Копнов М.В., Кудинов А.В., Марков Н.Г. Использование пространства Минковского при восстановлении геополей // Сборник трудов VI Всероссийской научно-практической конференции-конкурса «Технологии Microsoft в теории и практике программирования». Томск: Изд-во Томского политехнического университета, 2009. С. 97–99.

8. Копнов М.В., Марков Н.Г. Способ повышения вычислительной эффективности алгоритмов восстановления геополей // Труды XI Международной научно-практической конференции студентов, аспирантов и молодых ученых «Современные техника и технологии». Томск: Изд-во Томского политехнического университета, 2005. С. 165–166.

9. Копнов М.В., Марков Н.Г. Локализация исходных данных при восстановлении геополей // Тезисы докладов конференции-конкурса «Технологии Microsoft в теории и практике программирования». Новосибирск: ЗАО РИЦ «Прайс-курьер», 2005. С. 123–124.

10. Копнов М.В., Марков Н.Г. Линеаризация вложенных циклов при Восстановлении многомерных геополей // Тезисы докладов конференции-конкурса «Технологии Microsoft в теории и практике программирования». Новосибирск: ЗАО РИЦ «Прайс-курьер», 2006. С. 188–189.

11. Копнов М.В., Ковин Р.В., Марков Н.Г. Особенности поиска значений параметров при автоматизированном построении прогнозных карт // Сборник трудов XVI Международной научно-практической конференции студентов, аспирантов и молодых ученых «Современные техника и технологии». Томск: Изд-во Томского политехнического университета, 2010. С. 342–343.

12. 90. R.V. Kavin, A.V. Kudinov, N.G. Markov, E.A. Miroshnichenko, P.M. Ostrast, V.S. Sherstnyov, N.A. Shestakov Designing Production Management Systems of Oil-And-Gas Holding Companies // Proceedings of 8th Korea-Russia International Symposium on Since and Technology KORUS. 2004. vol.1. pp.91-95.