УДК 004.658.2

ОРГАНИЗАЦИЯ ПЕРВИЧНЫХ АРХИВОВ МЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ, МАТЕМАТИЧЕСКИЙ АППАРАТ ЕГО ОБРАБОТКИ И ВИЗУАЛИЗАЦИИ РЕЗУЛЬТАТОВ

Я. В. Петросян

Военный авиационный инженерный университет (г. Воронеж)

Поступила в редакцию 1.03.2009 г.

Аннотация. В работе исследуется вопрос построения системы извлечения информации из первичных архивов гидрометеорологических данных, методические подходы к построению математического аппарата обработки пространственно-временных данных, его технической реализации и представления полученных результатов.

Ключевые слова: геоинформационная система, база данных, гидрометеорология, пространственно-временные данные, архив гидрометеорологической информации, трехмерное представление данных, визуализация.

Abstract. In the article the question of the system for extracting information from initial hydrometeorological data archives building, technical ways for mathematical mechanism of space-time data processing building, its technical realization and presentation of results are developed.

Keywords: geographic information system (GIS), database, hydrometeorology, space-time data, archive of hydrometeorological information, three-dimensional data representation, visualization.

ВВЕДЕНИЕ

Геоинформационные системы (ГИС) в последнее время получили огромное развитие благодаря тому, что до 90% информации имеет пространственную привязку [1]. Огромный интерес представляют собой исследовательские задачи в области гидрометеорологии, на примере которых можно рассматривать технологию проектирования ГИС. Огромные объёмы данных, собираемых по всей планете, представляют большой интерес для учёных и энтузиастов в области метеорологии. Однако не всегда есть средства для проверки гипотез о климате и текущей погоде, как в плане получения исходных данных, так и в плане программных средств их обработки.

Очевидно, что при решении исследовательских задач основную роль в ГИС играет пакет анализа. По мере развития ГИС сложился перечень функций, используемых при анализе, однако далеко не все ГИС снабжены возможностями специализированного анализа, например, ориентированного на вопросы гидрометеорологии или экологии. Связано это с тем, что чёткой схемы проведения таких работ не существует и организации, занимающиеся ими, предпочитают проводить анализ по собственным методикам и правилам. Выход заключается в том, чтобы создать систему обработки пространственных данных, которая позволяла бы создавать новые инструменты анализа на базе уже имеющихся.

Для построения такой системы необходимо решить следующие задачи: получение данных, представление данных в удобном для вычисления и управления виде, представление результатов.

1. НАКОПЛЕНИЕ И ПОДГОТОВКА ДАННЫХ

Метеорологические данные могут быть получены по сети Transmet. К этой сети подключены гражданские и военные аэропорты, а также другие потребители первичной метеорологической информации. По ней поступают данные наблюдений с метеорологических и аэрологических станций в виде телеграмм в метеорологических кодах, а также карты погоды и снимки облачных полей с метеорологических искусственных спутников земли (МИСЗ). В среднем ежедневно полный набор данных по планете занимает объём от 10 до 15 МБт текстовой информации в метеорологических кодах.

© Петросян Я.В., 2009

Таким образом, данные за год по всей планете могут быть размещены на одном DVD диске. Если доступа к этим данным нет, то они могут быть получены по сети Интернет с использованием программы Digital Atmosphere или просто в виде таблиц.

Накопление метеорологического архива позволит решать задачи исследовательского характера. Однако первичная метеорологическая информация имеет весьма большой объём, кодированный вид и потому не пригодна для анализа человеком. Поэтому, целесообразно создание программного обеспечения (ПО) для работы с архивом на ПК.

При формировании архива данных можно сформулировать следующие требования:

- он должен быть полным, то есть получаемая информация не должна усекаться (ведь заранее неизвестно, какая информация понадобится);
- при помещении первичной информации в архив не должно происходить ошибок раскодирования информации из телеграмм (нужно оговориться, что при передаче по каналам связи возможны ошибки);
- так как архив ежедневно пополняется, то необходимо чтобы его можно было разместить на нескольких носителях (рано или поздно одного носителя окажется мало);
- включение новой информации в архив требовало минимума затрат рабочего времени.

Ввиду перечисленных требований использование реляционных баз данных (БД) для этой задачи неприемлемо по следующим причинам:

- таблицы для хранения раскодированной информации имеют много полей, которые остаются пустыми ввиду существования редко наблюдаемых явлений погоды и редко измеряемых параметров. В связи с этим размеры файлов с таблицами неоправданно увеличиваются;
- дополнение БД новой информацией требует много времени и специальных программных средств;
- появление новых метеорологических кодов приведёт к серьёзному изменению структуры БД, что может оказаться неприемлемым.

В конечном итоге для проверки каких-либо гипотез придётся по БД формировать выборку. Процесс этот не ежедневный и не срочный, поэтому предлагается создание специального ПО для формирования выборок непосредственно по архиву телеграмм. Это позволит избавиться от недостатков, связанных с хранением информации в формате реляционных баз данных.

Рассмотрим, как сформировать архив, чтобы с ним было удобно работать. Необходимо, чтобы информация за каждый день была разделена по отдельным файлам. Имя файла должно нести информацию о дате наблюдения, время и место наблюдения заложено в самих телеграммах. Формат имени файла: ДДММГГГГ.txt, где: ДД - дата; ММ - месяц; ГГГГ - год наблюдения.

В одном файле содержится вся информация, которая приходила в этот день в любых метеорологических кодах. Таким образом, архив формируется из совокупности несвязанных файлов с гидрометеорологическими телеграммами.

При такой организации архива задача поиска информации сводится к трём этапам:

- поиск телеграмм;
- раскодирование полученных телеграмм и выделение значений требуемых метеовеличин;
- представление полученных данных в удобном для дальнейшей обработки виде.

Заранее нельзя охватить и предусмотреть все возможности, для которых будет использоваться этот архив, поэтому программа для его обработки должна представлять собой отдельный модуль, компоненту или библиотеку, на основе которой можно конструировать различные программы. В целом идея заключается в том, чтобы иметь некоторого «посредника» между архивом данных и программой пользователя, который берёт на себя всю рутинную работу по поиску необходимой информации и представлению её в удобном для обработки виде. Такие промежуточные программные единицы обычно называют двигателями баз данных (ДБД). Для выполнения этих функций создана библиотека MeteoArh.dll, содержащая функции для поиска и раскодирования запрашиваемой информации, а также интерфейс пользователя для формирования запросов.

Общая схема работы ДБД представлена на рис.1.

Запрос к архиву представляется в виде файла инициализации, где каждый раздел содержит параметры отдельного запроса. Запрос на выборку телеграмм представляет собой раздел в файле, который используется ДБД для организации своей работы. Далее этот файл будем

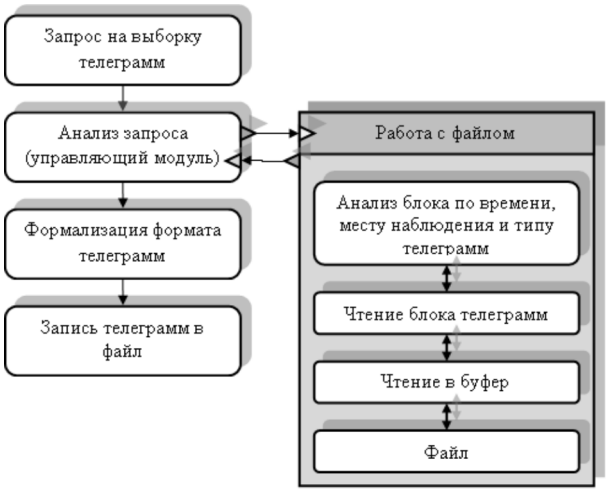

Рис. 1. Структурная схема алгоритма выполнения запроса на выборку телеграмм из архива

называть файлом проекта. Использование файла проекта позволяет свести к минимуму количество параметров при вызове функций в ДБД и организовать пакетную обработку.

При анализе запроса из файла проекта выбираются невыполненные запросы, определяются их параметры и определяются файлы в архиве, которые необходимо обработать. Для работы с файлом создан специальный модуль, в задачи которого входит отработка запроса на одном файле. Результатом работы модуля является набор телеграмм, удовлетворяющих запросу. Данный модуль вызывается необходимое количество раз.

Чтение данных из файла с телеграммами производится порциями, которые помещаются в буфер. Буфер представляет собой участок оперативной памяти произвольной длинны, куда помещаются фрагменты информации из файла, последовательно друг за другом, до тех пор, пока файл не закончится.

Такой алгоритм чтения данных из файла вызван тем, что файлы с гидрометеорологическими наблюдениями достаточно большие 8-10 мБт и при чтении его целиком в оперативную память начинает задействоваться файл подкачки и операции обработки данных занимают достаточно много времени, иногда несколько минут. Использование предложенной схемы позволяет ускорить процесс до нескольких секунд. Однако и при этом подходе возникает вопрос с выбором оптимального размера буфера.

С одной стороны с небольшим буфером (несколько килобайт) приходится чаще обращаться к жёсткому диску, что занимает дополнительное время, но с небольшими порциями информации легче оперировать в оперативной памяти. С увеличением размера буфера до тысяч килобайт работа с жёстким диском идёт быстрее, но он (буфер) должен целиком помещаться в оперативной памяти. Очевидно, существует оптимальный размер буфера относительно времени обработки, причём для ПК с разной производительностью центрального процессора и объёмом оперативной памяти размер буфера может быть разным.

Для проверки этой гипотезы и возможности определения оптимального размера буфера была написана процедура тестирования времени работы системы по выполнению запросов, которая размещена в ДБД и доступна для проведения тестов на ПК различной производительности.

Был проведён эксперимент на ПК с центральным процессором Intel Celeron M 1.5 гHz и объёмом оперативной памяти 512 мБт. Процедура тестирования выводит график зависимости времени обработки одного мегабайта информации от размера буфера (рис. 2), определяет оптимальный размер и использует его при последующей работе ДБД. Для указанного ПК оптимальный размер оказался 11000 байт.

Для подтверждения результатов эксперимента был проведён контрольный отбор всех телеграмм с номерами 27xxx (Московский регион) на объёме архивной информации в один год. Использовался оптимальный размер буфера, заниженный и завышенный размер.

Эксперимент показал, что подбор размера буфера приводит к увеличению производительности в несколько минут при работе с архивом в один год. При больших объёмах архива это будет значительная экономия времени, особенно это актуально для построения серверов.

После выполнения запроса на отбор телеграмм, результаты его выполнения записываются в файл проекта. Это ссылки на файлы с отобранными телеграммами и указание на то, что запрос выполнен.

Далее полученные файлы телеграмм можно раскодировать и получить итоговые таблицы для статистического анализа.

Декодеры (программные компоненты, предназначенные для представления информации метеорологических телеграмм в приемлемом для обработки виде) написаны для

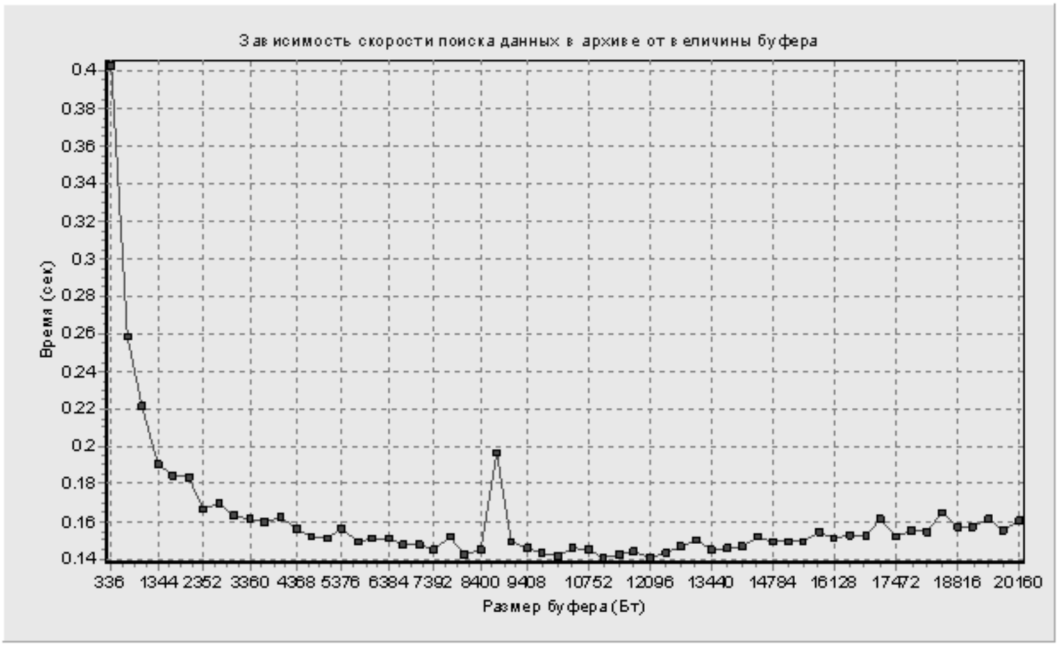

Рис. 2. Зависимость скорости чтения одного мегабайта информации от размера буфера

Таблица 1

Зависимость времени выполнения запроса на поиск телеграмм по станциям Московского региона на объёме архива в один год

Размер буфера (байт)	Затраченное время (мин.)
11000	13.2
1024	21.6
300000	16.7

каждого типа телеграммы в виде классов. Каждый класс имеет стандартную архитектуру, в которой содержится свойство для записи телеграммы и множество функций для получения соответствующих величин. Каждая функция возвращает либо значение соответствующей величины из телеграммы, либо сообщение о причинах его отсутствия. Созданные классы позволяют легко конструировать программы для работы с архивом.

На основе этих классов построена процедура для раскодирования телеграмм, которые отбираются на первом этапе работы с архивом. В качестве параметра ей передаётся имя файла проекта, в котором должна быть секция с параметрами запроса. Результатом является текстовый файл с таблицей метеорологических величин, которые были запрошены пользователем в формате CSV и по умолчанию открывается в приложении MS Excel.

Для формирования запроса на раскодирование также созданы диалоги, содержащие структуру метеорологических кодов и позволяющие сформировать текст запроса. После выполнения запроса на раскодирование, в файле проекта сохраняются ссылки на созданные файлы таблиц и запись о выполнении запроса.

Для работы с библиотекой MeteoArh.dll создана программа MeteoExtractor, обеспечивающая интерфейс доступа к процедурам библиотеки. Главная форма программы представлена на рис. 3.

Для просмотра таблиц выборок может использоваться текстовый редактор или MS Excel. Это зависит от настроек программы, которые пользователь может редактировать.

Во время работы программы пользователь информируется о процессах, которые проводятся программой. Программа не имеет ограничений на объём информации в архиве и количеству отбираемых телеграмм. Размер файла выборки может достигать 2 гБт, но традиционные средства обработки, электронные таблицы, ма-

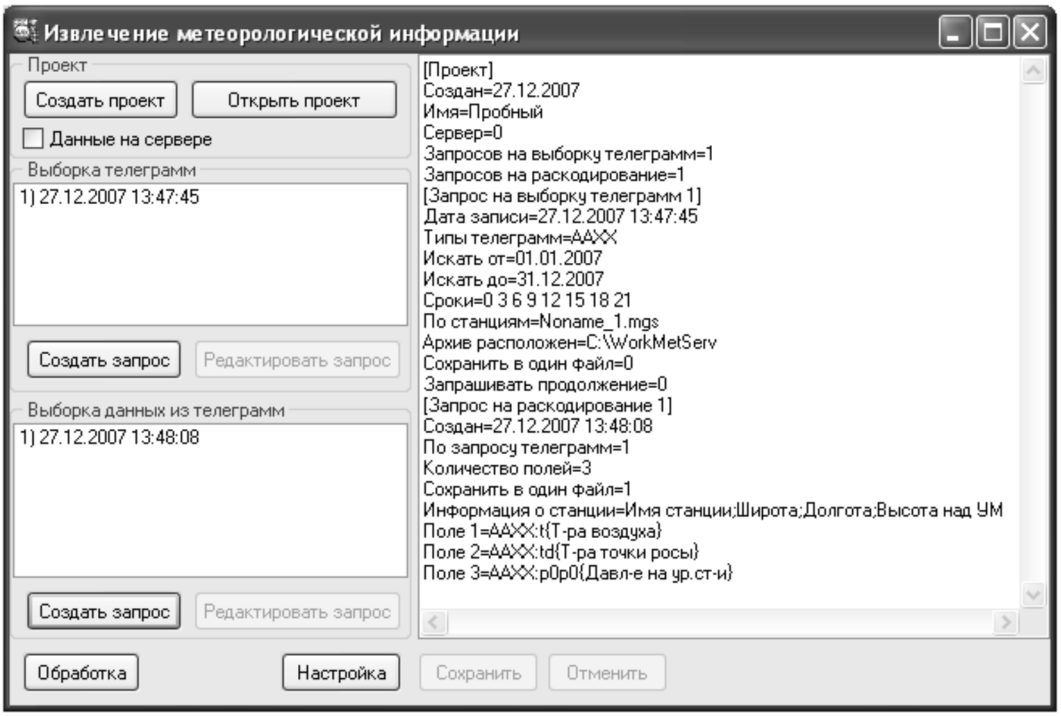

Рис. 3. Главная форма программы интегрированной работы с архивом

тематические и статистические пакеты, не всегда позволяют обрабатывать столько данных, кроме того не всегда существуют удобные инструменты для решения исследовательских задач.

2. МЕТОДИКА ОБРАБОТКИ ДАННЫХ

Для построения системы обработки данных, допускающей собственную достройку, необходимо определить структуру организации пространственных данных, удобную для управления, а также базовый (минимальный) набор операций над этими структурами, позволяющий создавать более сложные алгоритмы их обработки.

Рассмотрим структуру организации исходных данных. Пространственно распределённые данные представляют собой измерения некоторых количественных или качественных параметров в пространственно-временном поле планеты, то есть множество значений, где каждому значению соответствуют координаты пространства-времени.

Пространственно-временные координаты (**параметры ключа**) могут быть представлены значениями года, месяца, даты, часа, минуты, широты, долготы, высоты над уровнем моря или земли в любых сочетаниях. Множество значений одного параметра в совокупности с пространственно-временными координатами для каждого значения будем называть **геополем**. **Ключом геополя** будем называть множество координат, а **значением геополя** – множество значений. В дальнейшем геополе рассматривается как самостоятельная информационная сущность, над которой можно проводить операции. Операция над геополем означает операции над его значениями с учётом значения ключа. Ключ геополя будем обозначать $\tilde{A}$ элемент ключа $\tilde{a}$, значение геополя A, элемент значения a. Тогда геополе, которое состоит из множества значений с присоединённым к нему ключом, обозначим $A_{\tilde{A}}$, а отдельное значение $a_{\tilde{a}}$.

Для ключей, как для множества точек в пространстве-времени, определяются действия традиционные для множеств (объединение $\tilde{A} \cup \tilde{B}$, пересечение $\tilde{A} \cap \tilde{B}$, проверка принадлежности элемента множества А множеству В $\tilde{a} \in \tilde{B}$).

Определим значение простых математических действий {+ - * /} над геополями:

$$C_{\tilde{C}=\tilde{A}\cap\tilde{B}} = \left(A_{\tilde{A}}\{+-*/\}B_{\tilde{B}}\right)\Big|_{\tilde{c}=\tilde{a}=\tilde{b}}.$$

Данная запись означает, что геополе C получается в результате проведения действий с геополями A и B таким образом, что сначала для геополя C рассчитывается ключ $\tilde{C} = \tilde{A} \cup \tilde{B}$, а

затем элемент геополя c рассчитывается как $c=a\{+ - * /\}b$ при условии равенства ключевых элементов $\tilde{c} = \tilde{a} = \tilde{b}$. Очевидно, что производить действия над геополями можно, если у них имеются ключи одинаковой структуры. К примеру, сумма двух геополей $A+B$ приводит к вычислению суммы элементов этих полей с равными значениями ключа. Если геополя A и B будут располагаться в непересекающихся географических регионах планеты или иметь разные структуры ключей, то в результате сложения получим пустое множество. Таким образом, описание пространства-времени вводится в базовые операции над полями.

При проведении статистического анализа возникает задача рассматривать геополе как набор подмножеств. Например, если есть данные о температуре воздуха над Европой за пять лет, то можно разделить эти данные на подмножества за каждый год и рассматривать статистические характеристики этих подмножеств, анализируя временные изменения температуры над всей территорией. С другой стороны, можно разделить это геополе на подмножества территориально (подмножества для каждой наблюдательной станции) и рассматривать территориальные изменения температуры. Задача заключается в том, чтобы ввести минимальное количество новых действий, обеспечив широту и универсальность создаваемых алгоритмов.

Введём новую функцию $Reduction(\tilde{A}, \Theta)$ (сокращение), где Θ - сокращённый перечень параметров ключа $\tilde{A}$. Работа функции заключается в том, чтобы сократить параметры ключа $\tilde{A}$ до Θ с удалением повторяющихся элементов.

Для примера рассмотрим случай, когда есть ключ $\tilde{A}$ только с временными данными (год, месяц, дата). После применения функции сокращения с параметрами «Год, Месяц» получим ключ $\tilde{B}$, в котором будет отсутствовать параметр «Дата» и за счёт сокращения степени свободы ключа сократится количество элементов ключа $\tilde{B}$, так как повторяющиеся элементы не допускаются. Для наглядного представления примера приведена таблица 2. Теперь несколько элементов ключа $\tilde{A}$, которые можно рассматривать как отдельные подмножества, принадлежат одному элементу ключа $\tilde{B}$. Например одному значению $\tilde{b} = (2007,1)$ соответствует 31 значение в ключе $\tilde{A}$ - это все дни января 2007 года. В таблице 2 элементы $\tilde{B}$ и подмножества $\tilde{A}$ представлены соответствующими штриховками.

Значения для геополя по ключу $\tilde{B}$ можно рассчитать, используя какие либо функции, применимые к множеству данных, например расчёт средних значений, поиск экстремальных значений или каких-либо статистических оценок. Каждое значение $b_{\tilde{b}}$ будет рассчитываться по подмножеству $A_{\tilde{A}}$, элементы которого удовлетворяют условию $\tilde{a} \in \tilde{b}$.

К примеру, мы рассчитываем среднее значение температуры с ключом $\tilde{A}$ в таблице 2. Если мы сократим исходный ключ до года, то получим среднегодовые температуры, если до месяца, то среднемесячные и сможем оценить годовой ход температуры, если до года и месяца, то средние температуры за каждый месяц каж-

Таблица 2

Пример сокращения параметров ключа

Ключ $\tilde{A}$			Сокращение	Ключ $\tilde{B}$	
Год	Месяц	Дата		Год	Месяц
2007	1	1		2007	1
2007	1	2		2007	2
2007	1	...		2007	...
2007	1	31		2007	12
2007	2	1	→Reduction($\tilde{A}$,(**Год, Месяц**))→		
2007	2	2			
2007	...	...			
2007	12	1			
2007	...	...			
2007	12	31			

дого года. Если вообще сократить ключ, то получим среднее значение по всему геополю. Таким образом, сокращая ключ до нужной детализации, мы определяем параметры пространства-времени, в узлы которого необходимо произвести пересчёт исходных геополей.

Описав рассмотренные действия над геополями на алгоритмическом языке можно построить систему программирования, в которой описать статистические функции и методы анализа в рамках принятой методологии.

Для технической реализации описанной методологии обработки пространственной информации используется среда объектно-ориентированного программирования Delphi 2005 for .Net, которая допускает перегрузку операторов для новых структур данных [2].

Геополя описываются как объекты, которые использую файловые потоки для хранения данных. Ключ и величина геополя размещаются в различных файлах, так чтобы на один ключ могли ссылаться несколько величин, при этом в оперативной памяти хранятся только текущие данные, с которыми производятся операции. Так как файлы целиком никогда не загружаются, то файл подкачки не используется для выполнения операций над большими объёмами информации и вычисления происходят сравнительно быстро.

Объекты геополей используются для хранения, поиска и производства простых математических вычислений над данными и ключами в пространственно-временном поле планеты. Этот уровень представляет собой самый низкий уровень системы программирования, на его основе создаётся описание статистических функций, фильтров и представления данных. Описание статистических методов ведётся на основе первых двух уровней и представляет собой описание действий над геополями, где результатом решения являются также геополя, содержащие поля коэффициентов искомых статистических зависимостей.

3. ПРЕДСТАВЛЕНИЕ ДАННЫХ

Результирующие геополя необходимо представлять в удобном для восприятия виде. При этом не малую роль играет интуиция и воображение человека, особенно когда речь идёт о новых знаниях. В этом может помочь использование технологий виртуальной реальности.

В этом случае актуальной является следующая задача:

- визуализация земной поверхности с учётом рельефа и точное ориентирование в пространстве;
- визуализация по данным в узлах регулярной сетки точек полей данных;
- визуализация динамики изменения процессов.

Для построения графических трёхмерных моделей используется графический сервер OpenGL [3], который поставляется с операционными системами, начиная с версии Windows-98 и выше. Тем не менее, для более корректной работы программы рекомендуется использовать операционные системы, начиная с версии Windows-2000 и выше. Этот сервер позволяет отрисовывать на мониторе высококачественные графические образы, учитывая их пространственное положение, текстуры, прозрачность, а также моделировать атмосферные явления, такие как дымка или туман.

Использование OpenGL позволяет получать картинку настолько чёткую и детальную, насколько это позволит графическая аппаратная часть ПК. То есть программа может работать и на бытовой вычислительной технике и на мощных графических станциях, но качество картинки и скорость обработки будет разной. При разработке программ ориентация идёт на ПК общего назначения, у которых тактовая частота процессора не менее 2 гГц, оперативная память не менее 512 мБ, видеокарта может не иметь графического акселератора.

Исходными данными являются геополя с данными в узлах регулярной сетки точек. На рис. 4, 5 представлены примеры изображений.

Пространственное представление данных пока носит экспериментальный характер. Основная её задача состоит в том, чтобы сделать как можно более наглядным и реалистичным представление информации в пространстве планеты и динамику её изменений, при этом возникают некоторые проблемы. Так, например, при отображении изобарических поверхностей абсолютной и относительной топографии над полушариями они выглядят плоскими. Связано это с тем, что толщина атмосферы по отношению к радиусу Земли очень мала. Поэтому возникает дилемма: сохранить чёткий масштаб сцены или представить сцену так, чтобы поверхности имели объёмный вид, хорошо

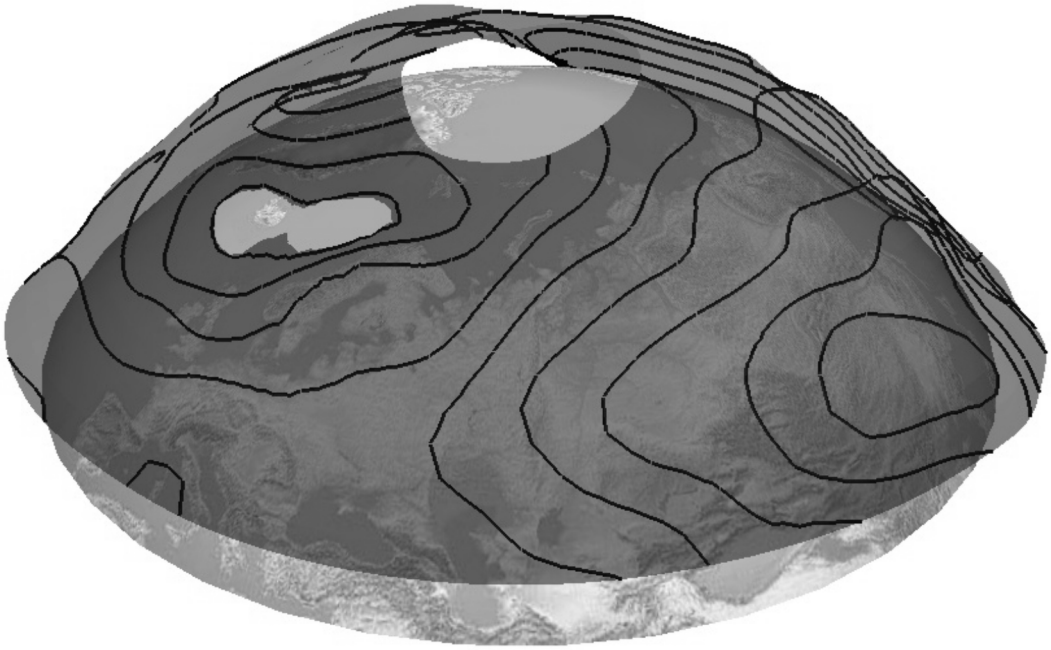

Рис. 4. Пространственное представление поля АТ-1000 (умножение в 5000 раз) над широтами 40-80° Северного полушария. Земная поверхность отображается как сфера с текстурой.

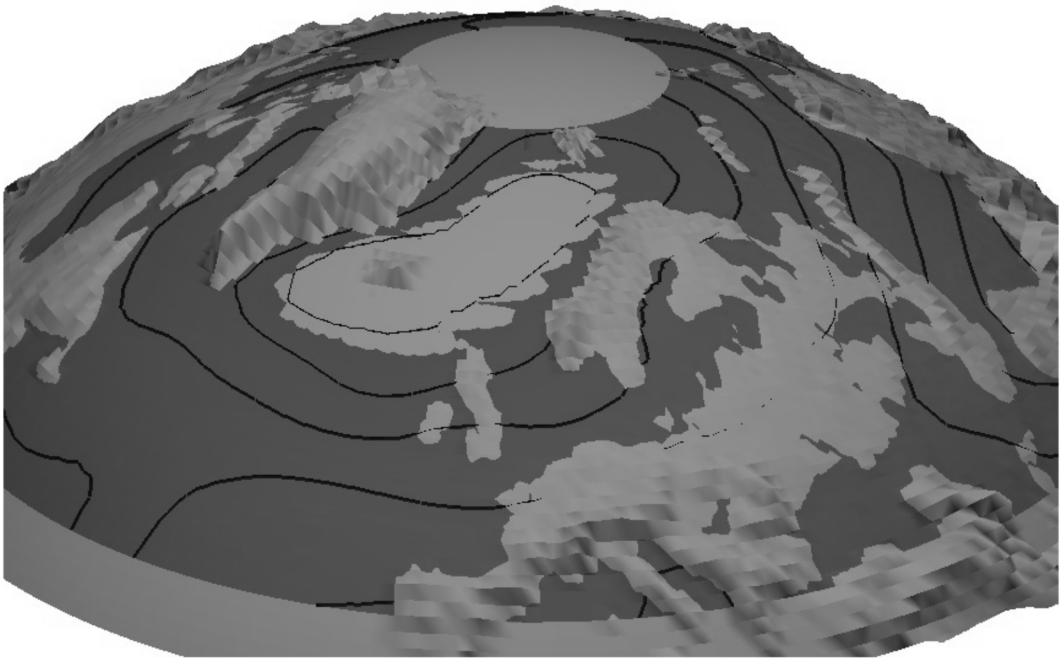

Рис. 5. Пространственное представление АТ-1000 над широтами 40-80° Северного полушария (умножение в 100 раз высоты поля и рельефа Земли). Земная поверхность отображается с использованием карты высот.

понятный человеку. Пока проблема решается путём умножения высоты поверхности, или увеличения её дисперсии. При этом пользователь должен понимать, какие преобразования с поверхностями происходят для демонстрации её объёмного вида. Та же ситуация происходит и с отображением рельефа земной поверхности.

При отображении метеорологической информации на более мелких районах земной поверхности планируется отрисовывать сцены в строгом соответствии с масштабами.

Данная технология представления позволяет быстро создавать объёмные сцены. Самый простой и дешёвый способ – это формирование двух одинаковых снимков сцены с разных точек зрения (левая и правая). Каждый кадр переводится в градации серого и затем записывается в красную и зелёную составляющую цвета RGB. При просмотре такого изображения через очки с красным и зелёным светофильтром ясно чувствуется объём сцены.

ЗАКЛЮЧЕНИЕ

Использование подобных методик представления, обработки и отображения информации позволит создать полноценную систему, пригодную для проведения исследовательской работы с пространственно-временными данными. При этом геополя выступают в качестве универсального формата различных типов данных, которые пригодны для управления и отображения, а также для удобного программирования вычислений. Эти методы позволяю описывать статистические функции и методы анализа, применительно к пространственно-временным данным, что позволит говорить о методологии обработки такого рода данных вообще. Выражение этой методологии должно найти в создании специализированного языка программирования, разработка которого ведётся под рабочим названием AlgoMet.

СПИСОК ЛИТЕРАТУРЫ

1. *Геоинформатика:* в 2 кн. Кн. 1: учебник для студ. высш. учеб. заведений / Е. Г. Капралов, А. В. Кошкарев, В.С. Тикунов и др.; под ред. В. С. Тикунова. – 2-е изд., перераб. и доп. – М.: Издательский центр «Академия», 2008. – 384 с. [16] с. цв. ил.: ил.

2. *Кэнту М.* Delphi 2005. Для профессионалов. – СПб.: Питер, 2006. – 907с.: ил.

3. *Краснов М.В.* OpenGL. Графика в проектах Delphi. – СПб.: БХВ-Петербург, 2004. – 352 с.: ил.

Петросян Яков Владимирович – кандидат географических наук, преподаватель кафедры теоретической гидрометеорологии военного авиационного инженерного университета (г. Воронеж).

Petrosyan Yakov Vladimirovich - candidate of geographical science, lecturer of hydro-meteorological department of Military Aviation Engineering University (Voronezh).