

Организация гидрометеорологической информации программного комплекса ГИС "Океан"

Иванов Борис Николаевич, кандидат физ.-мат. наук, доцент
Дальневосточный федеральный университет, (г. Владивосток)

Дается практическая реализация выделенного приложения базы данных гидрометеорологической информации. Отмечены ключевые свойства данных, положенных в основу файловой организации базы. Структурно отмечена организация доступа к базе ключевых компоненты программного комплекса, участвующих в обработке данных.

Ключевые слова: база данных, метеорологические наблюдения, гидрометеорологические данные.

1. Введение. Программный комплекс Геоинформационная система (ГИС) «Океан» — это автоматизированная динамическая модель состояния и прогноза погоды в атмосфере и океане. Задачи ГИС «Океан» — обеспечение народнохозяйственных организаций необходимыми сведениями о текущем и ожидаемом состоянии атмосферы и океана, предупреждение об опасных и особо опасных явлениях погоды [1]. Реализация программного комплекса выполнена на базе сети персональных ЭВМ в среде Microsoft Windows. Проведение расчётов возможно для любого района поверхности Земли.

Входной поток исходной гидрометеорологической информации для комплекса подразделяется на две составные части. Во-первых, данные фактических наблюдений за состоянием океана и атмосферы; во-вторых, фактические и прогностические поля данных, заданные в узлах регулярной сеточной области (далее сеточные поля), основных показателей гидрометеорологических элементов ведущих мировых центров прогноза погоды. Организация программного комплекса ГИС «Океан», во многом, предопределена наличием указанных потоков исходных данных. Как правило, важнейшие составляющие компоненты ГИС «Океан» ориентированы на обработку одного или другого потока данных [2]. Следует отметить, что решения многочисленных задач являются взаимозависимыми и требуют «погружения» их в объединяющую среду [3]. Проект является именно такой средой, которая помимо решения основных задач позволяет решать многочисленные вспомогательные или дополняющие задачи анализа и прогноза погоды [4, 5].

Гидрометеорологические данные упорядочены во времени и в пространстве (определяются на стандартных высотах). Это даёт возможность вести обработку данных в режиме формирования единого динамического составного поля, позволяющего наблюдать в режиме мультипликации изменение значений каждого из метеоэлементов в форматах полей изолиний, цифрового поля, градиентного заполнения полей изолиний и т. п.

2. Файловая структура базы данных исходной гидрометеорологической информации. Организация базы данных (БД или БД «Океана») явилась следствием её практического назначения. База данных обеспечивает:

- 1). Коллективный доступ к данным.
- 2). Долговременное хранение данных.

- 3). Выборку данных из общего потока информации.

- 4). Оперативную пересылку части базы данных по сети.

- 5). Простоту обновления данных.

- 6). Высокое быстродействие.

Обозначенные свойства базы данных являются взаимозависимыми. Из общего перечня свойств базы ключевым является возможность оперативной пересылки её частей по сети, с целью обработки на «другой стороне» указанных данных в рамках комплекса ГИС «Океана». При пересылке такая часть базы должна стать составной частью единой базы. Именно данное свойство базы стало определяющим при её реализации.

На рис. 1 представлена структурная схема доступа различных компонент ГИС «Океана» к распределённой базе данных. Суть организации базы данных — это множество файлов данных одного каталога. Каждый из файлов привязывается к сроку наблюдений и является независимой единицей, которую можно «безболезненно» удалить или привнести в БД. Файлы базы формируются автоматически по мере поступления данных. Для идентификации файлов в их имена включаются сроки наблюдения данных. Это позволяет привязать данные таких файлов ко времени, не открывая (не читая) сами файлы.

Описание данных каждого из файлов является его внутренней организацией. При чтении данных из такой БД её справочник формируется на основе файлов каталога, включенных во временной интервал запроса. При дополнении (или удалении) к каталогу новых файлов данных, будем наблюдать увеличение (уменьшение) справочника базы. В этом суть реализации пересылки частей базы данных по сети. Части базы выступают отдельные файлы, составляющие базу. Количество файлов в такой базе со временем может стать значительным. Чтобы выполнение операций с такой базой не зависело от числа файлов её составляющих, доступ к файлам базы выполняется по их именам-срокам наблюдения. Данное свойство делает все операции с базой одинаково быстрыми по времени вне зависимости от количества файлов, составляющих базу.

Реализация базы данных выполнена в формате самостоятельного приложения. Весь необходимый комплекс задач по ведению базы данных осуществляется в рамках основного окна приложения, представленного на рис. 1.

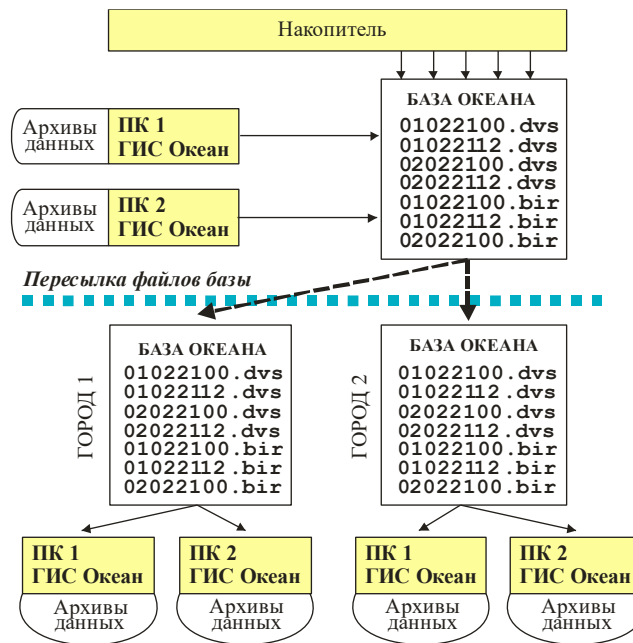


Рис. 1. Структурная схема подключения ГИС «Океана» к базе данных

3. Идентификация файлов фактических наблюдений базы данных. Фактические наблюдения за состоянием атмосферы подразделяются на два составляющих потока. Первый поток – это сводки приземных наблюдений. К ним относятся синоптические, судовые, морские и данные буйковых наблюдений. Такие данные передаются 8 раз в сутки и привязываются к срокам (часам) наблюдений: 00, 03, 06, 09, 21, 15, 18, 21. Объем данных по основным срокам (00 и 12) составляет порядка 5000 сводок по всей поверхности земли.

Второй поток – это данные аэрологических наблюдений. Сводки аэрологических наблюдений передаются 2 раза в сутки и привязываются к срокам (часам) наблюдений 00 и 12. Объем данных за каждый срок составляет порядка 500 сводок по всей поверхности земли.

Каждый из потоков данных упаковывается в свои файлы данных. Привязка файла к типу данных (приземные или аэрологические наблюдения) осуществляется по расширению файла. Срок наблюдения: день/месяц/год/срок составляет имя файла. Так например, в файле с именем 01022112.dvs упакованы данные типа «dvs» за 01 февраля 2021 за 12 часов. Тип данных и расширение файла – это принятая внутренняя кодировка.

4. Структура данных файлов фактических наблюдений базы данных. Практически, любую обработку данных сводок фактических наблюдений удобно вести, если они собраны в массив (вектор) данных [6]. Векторная структура данных фактических наблюдений распространяется и на их файловую организацию. В начале файла размещаются его сигнатура и справочник. Далее размещается вектор сводок наблюдений.

Пополнение базы данных из накопителя (рис. 1) выполняется равномерно через заданный интервал времени. Сводки за один срок могут пополняться в разные интервалы времен, возможны «опоздавшие» сводки, возможно и многократное повторение сво-

док из накопителя. Отметим, что прежде, чем разместить сводки данные в файле, выполняется их предварительная обработка. Например, сортировка данных по их индексам, проверка дублирования сводок и т. д. Индексы – это уникальные международные номера станций, формирующих сводки наблюдений. Поэтому при каждом пополнении данных необходим совместный анализ текущих файловых сводок и новых сводок из накопителя (рис. 1).

С этой целью перед пополнением файла новыми сводками из сводок данного файла формируется AVL-дерево [7, с. 536] по ключам поиска сводок. Ключ поиска состоит из индекса станции, идентификатора типа сводок (синоптические, судовые, морские и буйковые) и стандартного уровня наблюдений. Индекс станций является уникальным для станции. Идентификатор типа сводок включаются в ключ поиска с целью, чтобы дифференцировать сводки наблюдений в файле были по типам и уровням. Наличие уровня наблюдений является необходимым для аэрологических наблюдений.

Если обозначить через p число сводок за один срок наблюдения, то сложность формирования AVL-дерева [7, с. 536] составит $p(p \log p)$. Сложность поиска в таком дереве является логарифмической $p(\log p)$. Обход двоичного дерева слева-направо или наоборот позволяет получить сортированную по ключам поиска последовательность сводок за линейное время $p(p)$ [8, с. 217]. В таком сортированном формате сводки данных и размещаются в файле. Накладные расходы (временные и программные) обработки данных на AVL-дерево с лихвой окупаются при многократном обращении к накопителю за пополнением база данных.

5. Идентификация файлов данных сеточных полей базы данных. Данные сеточных полей определяются в узлах регулярной широтно-долготной сеточной области поверхности земли (сеточные поля или просто поля). Каждый центр прогноза погоды разбивает всю поверхность земли на районы или выделяет для себя специальные участки. Для каждого

из таких районов определяется буквенное обозначение и геометрические характеристики его сеточной области (тип сетки). На такие сеточные области на стандартных высотах выпускаются прогнозы погоды для основных типов данных метеоэлементов (давление, геопотенциал, температура, ветер, влажность, осадки и др.) [9]. Заблаговременность отдельных прогнозов может достигать 240 часов.

Каждое такое сеточное поле данных однозначно характеризуется следующими показателями: центром прогноза, типом данных, буквенным обозначением района, типом сетки поля, стандартным уровнем, временем исходных, заблаговременностью прогноза. Все упомянутые показатели составляют уникальный ключ сеточного поля данных.

Прогнозы вычисляются по текущим данным фактических наблюдений (приземных и аэрологических), относительно которых задаётся и заблаговременность прогнозов. Фактические наблюдения называются исходными данными.

Весь поток прогностических полей данных распределяется по файлам базы относительно сроков исходных наблюдений, по которым прогнозы были выпущены.

Имена таких файлов определяются по датам исходных фактических наблюдений, по которым составлялись прогнозы. Поэтому в одном файле размещаются сеточные поля прогностических данных с одними исходными по всем стандартным уровням. Объём такого файла может достигать 5 МБ и более.

Расширение файлов сеточных полей обеспечивает привязку данных к центрам прогнозов. Таким образом, для различных центров прогнозов погоды определяются свои файлы данных со своими расширениями. Так например, в файле с именем 01022112.brk размещаются сеточные поля, выпущенные по исходным данным фактических наблюдений за 01 февраля 2021 за 12 часов. Центр прогноза кодируется расширением «brk».

Данная файловая структура базы данных позволяет просто реализовать цикличность базы по времени.

6. Структура данных файлов сеточных полей базы данных. Локализация полей в файле данных выполняется по их ключам. Ключ поля данных составляют следующие его показатели (см. раздел 5): центр прогноза, тип данных, буквенное обозначение района, тип сетки поля, стандартный уровень, время исходных, заблаговременность прогноза.

Ключи полей данных файла составляют справочник файла. Справочник файла имеет распределённую блочную структуру и реализован связанным списком блоков справочника. Один блок справочника имеет конечное число элементов. В элементе справочника размещается ключ поля данных, смещение по файлу, где размещается само сеточное поле, и смещение по файлу для следующего блока справочника. Такая организация файла сеточных полей позволяет размещать в нём поля данных, не заботясь о переполнении справочника. На рис. 2} представлена структурная схема данных файла сеточных полей.

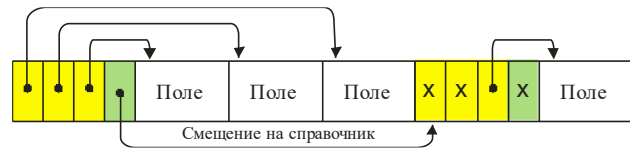


Рис. 2. Структурная схема данных файла сеточных полей

Пополнение базы данных из накопителя (рис. 1) выполняется равномерно через заданный интервал времени. Повторное чтение при каждом пополнении может приводить к дублированию полей. Не исключено и дублирование полей в самом накопителе. Поэтому при каждом пополнении данных необходим совместный анализ текущих полей файла и полей данных из накопителя (рис. 1).

С этой целью перед пополнением файла сеточных полей из элементов его справочника формируется AVL-дерево [7, с. 536]. По данному дереву за время $p(\log p)$, выполняется проверка наличия или отсутствия полей данных в файле, где p – число элементов в справочнике.

Новые поля данных, также как и новые блоки справочника, всегда добавляются в конец файла. Данная файловая структура базы позволяет просто реализовать цикличность базы по времени.

Компоненты ГИС «Океана» обработки сеточных полей базы данных вынесены в отдельный диалог, представленный на рис. 3. Число компонент таких более пятидесяти. При заказе на выполнение каждая из заказанных компонент формирует отдельную составляющую общего динамического поля, которое допускает просмотр режиме мультипликации (и пошаговым) по времени и в пространстве (по уровням) наблюдать изменения значений заказанных величин.

7. Локальный архив фактических наблюдений ГИС «Океана». Вся обработка данных фактических наблюдений выполняется в рамках локального архива фактических наблюдений ГИС «Океана».

Такой архив создаётся на каждом компьютере, где установлен ГИС «Океана» и ведётся обработка фактических наблюдений. Данный архив формируется чтением данных из общей базы данных (см. раздел 3 и 4). Цель такого «дублирования» фактических наблюдений заключается в том, что довольно много задач в рамках ГИС «Океана» ориентировано на применение фактических наблюдений. Локальный архив позволяет существенно повысить производительность работы как компьютера (не нужен постоянный сетевой доступ к общей базе данных), так специалистов.

Заказы на обработку данных локального архива выполняются в рамках его справочника. Сам справочник представляется в удобной матричной форме в диалоге (рис. 4) без каких-либо накладных расходов, связанных с сетевым доступом.

Матричная форма справочника позволяет просто выполнить заказ на обработку как отдельного срока наблюдений на одном уровне, так и произвольного интервала сроков наблюдений одновременно на нескольких уровнях.

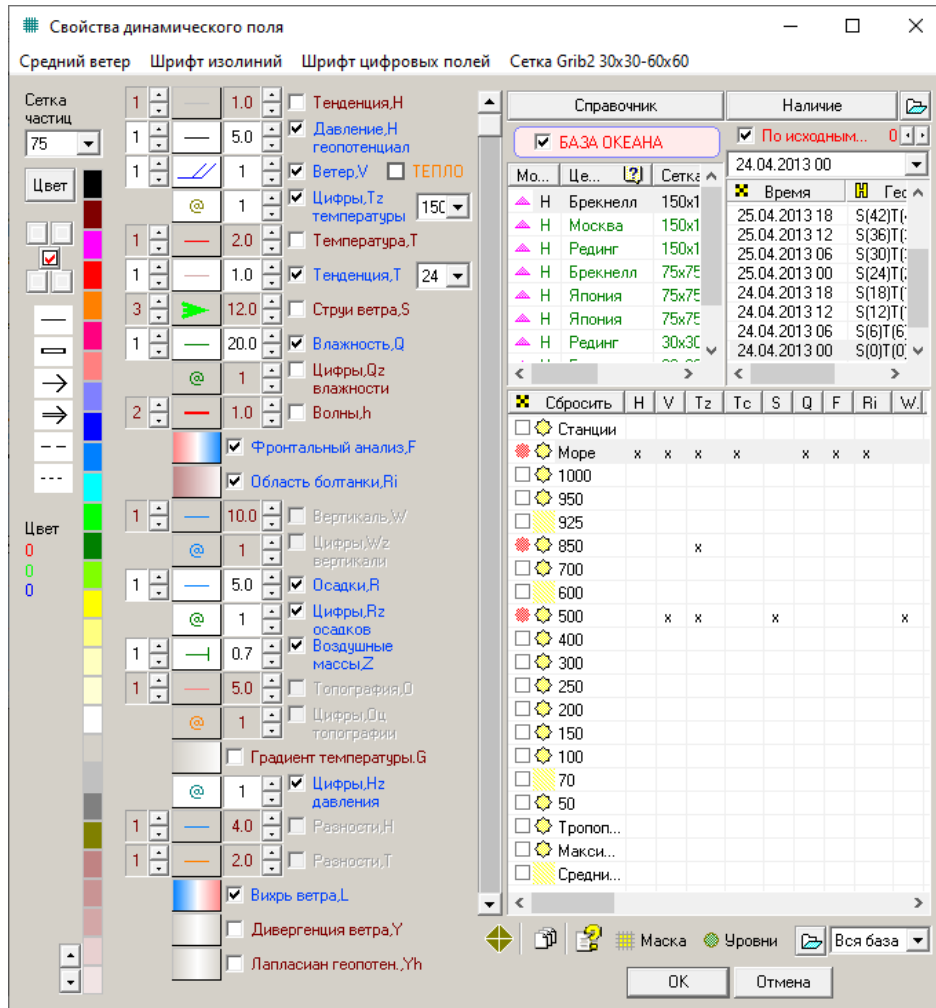


Рис. 2. Диалог обработки сеточных полей базы

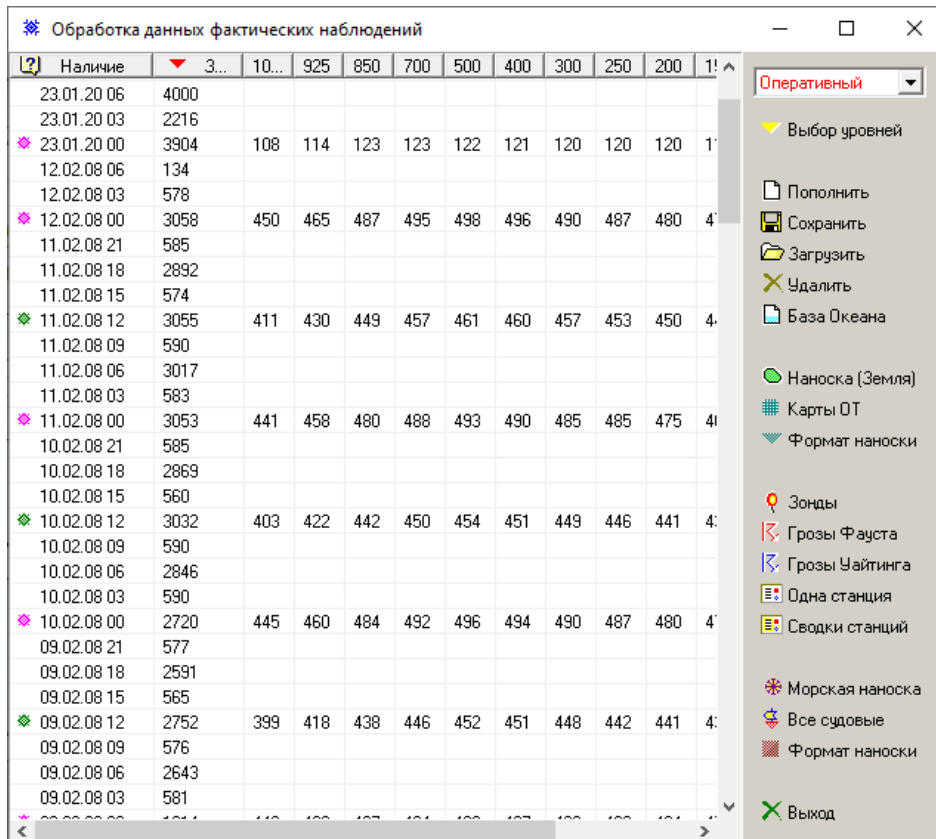


Рис. 4. Матричная форма справочника архива фактических наблюдений

Основные операции ведения архива: «Пополнить» и «Удалить». При «Пополнении» архива выполняется чтение данных из общей базы (см. раздел 3 и 4) и запись в локальный архив фактических наблюдений.

Реализация локального архива выполнена на двух файлах. Один файл является справочником архива. В другом файле хранятся данные сводок. Организация локального архива повторяет структуру данных файла сводок наблюдения (см. раздел 4). В локальном

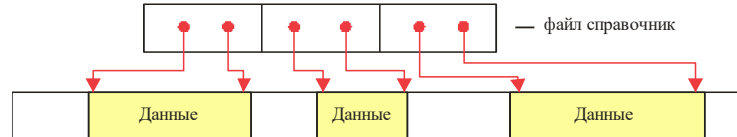


Рис. 5. Структурная схема локального архива сводок наблюдений

Файл сводок данных локального архива для больших интервалов времени (год более) может достигать значительных размеров. Для реализации используются длинные файлы. Объем такого файла ограничивается максимальным значением целого 64 разрядного числа. На рис. 5 представлена структурная схема локального архива сводок наблюдений.

8. Локальный архив сеточных полей ГИС «Океана». Локальный архив сеточных полей предназначен для сбора и хранения полей давления, геопотенциала, температуры, осадков и влажности на стандартных высотах. Основное назначение таких

архиве смешиваются данные приземных и аэрологических сводок наблюдений по всему временному интервалу.

Элемент файла справочника локального архива содержит ключ данных. Показателями ключа являются: время наблюдения, тип данных и уровень. Для указания на данные сводок наблюдения в элемент справочника включены два смещения по файлу данных локального архива. Одно смещение --- начало сводок данных, другое — конец сводок данных.

полей — это использование их в качестве статистического материала. Например, составляются декадные и месячные карты полей изолиний — это средние карты за указанный период.

Даже отдельные сеточные поля могут иметь значительный объем. Поэтому в локальном архиве сеточных полей сбор данных по времени для каждого типа поля одного уровня осуществляется в рамках двух файлов. Один файл является справочником данных, в другом — хранятся сами поля данных. Число файлов в локальном архиве сеточных полей в два раза больше числа типов полей (с учетом уровней), подлежащих статистическому сбору.

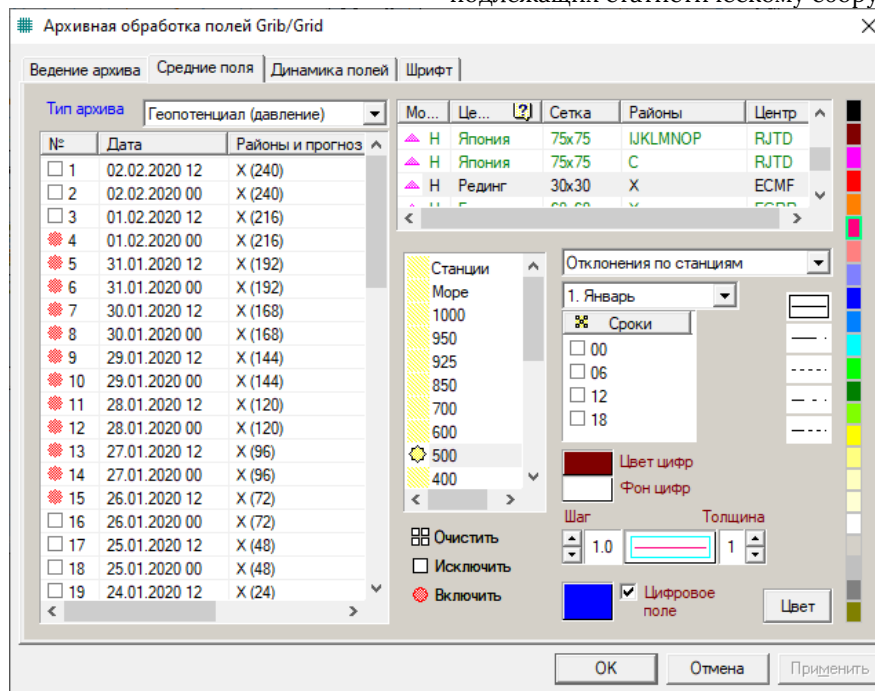


Рис. 6. Обработка данных локального архив сеточных полей

Запись справочника включает ключ поля и смещение по файлу данных. Смещение указывает начало поля данных. Структурная схема локального архива сеточных полей для одного типа поля повторяет структурную схему локального архива сводок наблюдения (рис. 5).

Ведение локального архив сеточных полей и обработка его данных осуществляется в рамках диалога, представленного на рис. 6. Диалог позволяет в

удобной форме показать состояние архива и выполнить заказ на обработку данных.

7. Заключение. Дается практическая реализация выделенной в отдельное приложение базы данных гидрометеорологических наблюдений. Отмечены ключевые свойства данных, положенных в основу файловой организации базы. Рассмотрена общая структура доступа и обработки данных базы. Структурно отмечена организация доступа к базе данных ключевых компоненты ГИС «Океана».

Литература:

1. Иванов Б.Н. Решение задачи расчёта оптимальных маршрутов судов в рамках геоинформационной системы “ОКЕАН” // Вычислительные методы и программирование. 2012. т 13, № 1. С. 226–234.
2. Иванов Б.Н. Геометрический подход к решению задачи построения траекторий циклонов и антициклонов // Вычислительные методы и программирование. 2014. т. 15. № 2. С. 370–382.
3. Иванов Б.Н. Сглаживающая адаптивная аппроксимация в задаче построения изолиний гидрометеорологических полей // Вычислительные методы и программирование. 2018. т. 19, № 4. С. 449–463.
4. Иванов Б.Н. Обзор практических методов сплайновой интерполяции в задаче построения изолиний гидрометеорологических полей. // Евразийское Научное Объединение. 2020. № 11-1 (69). С. 9–15.
5. Иванов Б.Н. Формирование вертикальных подвижных динамических разрезов синоптических процессов в задаче обеспечения безопасности полетов // Евразийское Научное Объединение. 2020. № 11-1 (69). С. 15–20.
6. Иванов Б.Н. Структуры вложенности поля изолиний в задаче градиентного заполнения // Вычислительные методы и программирование. 2006. т. 7, № 2. С. 30–40.
7. Кнут Д. Искусство программирования для ЭВМ. Т. 3. Сортировка и поиск. — М.: Мир, 1978. — 848 с.
8. Иванов Б.Н. Дискретная математика. Алгоритмы и программы. Полный курс. — М.: Физматлит, 2007 — 408 с.
9. Шмелькин Ю. Метеорологи 20 лет вместе с ГИС Метео // Электроэнергия. Передача и распределение. 2013. №4 (19). С. 50–53.