

УДК: 551.501:004

## СПЕЦИАЛИЗИРОВАННЫЕ МАССИВЫ ДАННЫХ ДЛЯ КЛИМАТИЧЕСКИХ ИССЛЕДОВАНИЙ

**О. Н. Булыгина<sup>1</sup>, Н. Н. Коршунова<sup>2</sup>, В. Н. Разуваев<sup>3</sup>**

*ФГБУ «Всероссийский научно-исследовательский институт  
гидрометеорологической информации – Мировой центр данных»,*

*<sup>1</sup> bulygina@meteo.ru, <sup>2</sup> nnk@meteo.ru, <sup>3</sup> razuvaev@meteo.ru,*

### Введение

Эмпирико-статистический анализ данных наблюдений является одним из главных источников информации об изменении состояния климатической системы в прошлом и создает основу для оценки возможных изменений в будущем. Исследования изменений климата на основе имеющихся метеорологических данных российскими и зарубежными специалистами проводится уже в течение многих лет, и полученные результаты говорят о перспективности таких работ и в будущем.

Указанное направление представляется особенно важным в свете решений проведенной по инициативе ВМО Третьей Всемирной климатической конференции (Женева, 2009 г.), положившей начало созданию Рамочной основы для климатического обслуживания. В создаваемой схеме климатического обслуживания фактические данные о состоянии всех компонент климатической системы Земли играют первостепенную роль.

В России регулярные метеорологические наблюдения были начаты в конце XIX века, и к настоящему времени накоплен огромный объем данных, содержащий богатейшую информацию о климатических условиях на территории России за весь период наблюдений. В то же время существует ряд серьезных обстоятельств, препятствующих широкому развитию климатических исследований. В первую очередь, это связано с недостаточной доступностью исходных данных для анализа. Накопление данных исходных наблюдений в Государственном фонде данных Росгидромета (Госфонд Росгидромета) направлено на решение основополагающей задачи – надежное хранение данных. Принципы формирования массивов архивированных данных во многом ограничивают возможности их непосредственного использования

для исследовательских целей, что приводит к необходимости развития технологий формирования специализированных массивов данных, предназначенных для решения научно-прикладных задач в области исследования климата.

В последние годы в Росгидромете активизировались работы по созданию климатических баз данных, которые должны составить основу для проведения широкомасштабных работ по изучению климата России в соответствии с Климатической доктриной. Перечень создаваемых баз данных определен на основе требований Глобальной системы наблюдений за климатом и включает основные климатические параметры, такие как температура воздуха, атмосферные осадки, параметры свободной атмосферы и другие. Одним из главных требований к создаваемым массивам является их доступность. В рамках реализации концепции глобальной рамочной основы для климатического обслуживания (Принцип 6: содействовать бесплатному и открытому обмену данными наблюдений) должно быть обеспечено функционирование открытого полноценного доступа к специализированным климатическим массивам исторических данных через Интернет, осуществляться регулярное пополнение и расширение состава массивов. С этой целью было принято решение о размещении массивов на Web-сайтах организаций-участников данной работы для свободного использования. В настоящее время на Web-сайте ВНИИГМИ-МЦД уже размещены массивы данных по температуре воздуха, атмосферным осадкам, снежному покрову и ряду других климатических параметров.

### **1. Основные характеристики специализированных массивов для климатических исследований**

Под климатическими данными понимается набор геофизических элементов и их комплексов, необходимых для полного описания состояния климатической системы и ее изменений под действием различных факторов. Климатические данные представляются обычно в виде массивов климатических данных различной конфигурации в зависимости от поставленной задачи. Информационные базы климатических данных содержат кроме собственно данных средства управления данными, обобщенные климатические характеристики, дополнительные сведения о данных (метаданные).

Основные свойства специализированных массивов, предназначенных для климатических исследований:

*а) Доступность.*

Эффективное использование климатических баз данных возможно только при их доступности для широкого круга пользователей.

*б) Распределенность баз климатических данных.*

Специализированные массивы размещаются на сайтах организаций, ответственных за их создание и регулярное обновление.

*в) Качество данных.*

Важным свойством климатических баз данных является отсутствие в них грубых ошибок и пропусков. Кроме того, базы данных должны сопровождаться полным перечнем использованных процедур контроля качества.

*г) Наличие метаданных.*

Создание массивов метаданных является обязательным условием при подготовке климатических баз данных. Помимо метаданных, описывающих саму базу данных, для климатических исследований необходима информация, позволяющая устранить возможную неоднородность в рядах метеозлементов (сведения о переносе станции, информация о смене приборов и методик наблюдения, изменения в окружающем ландшафте и т.д.). Подготовка таких массивов данных приводит к необходимости решения сложных методических и технологических задач.

*д) Регулярное обновление баз данных.*

Создаваемые базы климатических данных должны регулярно обновляться. Периодичность пополнения баз данных определяется требованиями реализуемых на этих данных задач.

*е) Унификация баз данных.*

Создаваемые специализированные массивы имеют специфику, обусловленную конкретной исходной информацией. Однако должны соблюдаться общие требования и принципы организации массивов, обеспечивающие возможность применения их для климатических исследований.

Накопленный во ВНИИГМИ-МЦД опыт подготовки на современных технических носителях высококачественных массивов метеорологических данных позволил в 2009–2011 гг. в рамках Целевой научно-технической программы Росгидромета «Научные исследования и разработки в области гидрометеорологии

и мониторинга окружающей среды» подготовить и разместить на Web-сайте института для свободного бесплатного доступа специализированные массивы по основным метеорологическим параметрам различного временного разрешения (рис.). Созданные массивы данных для климатических исследований – плод многолетнего и кропотливого труда сотрудников отдела климатологии ВНИИГМИ-МЦД.



Рис. Специализированные информационные базы данных на сайте ВНИИГМИ-МЦД

## 2. Краткое описание специализированных массивов данных

Все специализированные массивы создавались по данным, содержащимся на технических носителях Госфонда Росгидромета, и по опубликованным источникам [1, 2, 3]. Перечень станций составлен на основании списка станций Росгидромета, включенных в Глобальную сеть наблюдений за климатом (утвержденного Руководителем Росгидромета 25 марта 2004 г.) и списка реперных метеорологических станций Росгидромета, подготовленного в Главной геофизической обсерватории им. А. И. Воейкова (исп. зав. ОМРЭИ ГГО В. И. Кондратюк). Массивы данных регулярно пополняются, обнаруженные ошибки исправляются. Информацию о внесенных исправлениях можно также найти на сайте в разделе «Обнаруженные и исправленные ошибки». Каждый массив имеет каталог станций с указанием географических координат, высоты метеоплощадки, начала наблюдений на станции, сведений о переносах метеоплощадки и др.

Подготовленные массивы подвергались различным методам контроля. Первый контроль массивы метеорологических данных проходят уже на этапе формирования их на технических носителях. Затем при организации архивного хранения массивы последовательно проходят несколько этапов контроля [4]. Осуществляемый контроль можно назвать технологическим. После осуществления технологического контроля имеющихся архивов анализировалось качество самих данных. В силу больших объемов массивов метеорологических данных и регулярного их пополнения текущей информацией визуальный контроль таких данных на наличие ошибочных значений отдельных метеозаписей практически невозможен из-за трудоемкости. В связи с этим использовались различные методы статистического контроля.

**1. Средняя месячная температура воздуха на станциях России** (авторы О. Н. Булыгина, В. Н. Разуваев, Л. Т. Трофименко, Н. В. Швець).

Массив содержит данные 518 станций от начала наблюдений на станции до 2010 года.

Массив имеет записи фиксированной длины 83 байта с символом конца строки. Каждая запись содержит 14 полей с фиксированным положением в записи и фиксированной длиной. Дополнительно все поля записи отделены друг от друга символом «пробел». Такое «двойное» форматирование записи позволяет

использовать любой из наиболее употребительных способов разделения полей («comma separate» или «fixed length fields») при вводе данных в программные средства хранения (СУБД), электронные табличные процессоры, статистические пакеты и средства графического представления информации.

Массив проконтролирован на экстремальные значения, а также статистическим методом, подробно описанным в [5]. Метод основан на анализе значений асимметрии и эксцесса, которые наиболее чувствительны к отдельным выбросам.

**2. Месячные суммы осадков на станциях России** (авторы *О. Н. Булыгина, В. Н. Разуваев, Н. Н. Кориунова, Н. В. Швець*).

Массив содержит данные 518 станций от начала наблюдений на станции до 2010 года.

В публикуемой версии массива ряды данных содержат информацию с 1966 года. Такой период обусловлен тем, что до 1966 года в наблюдениях за осадками, по крайней мере, трижды была нарушена однородность рядов: а) в 30-е годы прошлого столетия произвели массовый перенос станций на открытое место, репрезентативное для ряда элементов, но не для осадков; б) в начале 50-х годов массовое нарушение однородности вызвано сменой прибора, был установлен осадкомер с защитой Третьякова; в) с января 1966 года поправка на смачивание стала вводиться непосредственно на станции. После 1966 года никаких изменений в методиках измерений и обработки не происходило, т. е. не было массового нарушения однородности данных. Данный массив имеет структуру и формат записи, аналогичные массиву среднемесячной температуры. Массив проконтролирован на экстремальные значения, а также на совпадение месячных сумм осадков данного массива и расчетных сумм, которые получались из суточных сумм осадков.

**3. Среднемесячное давление воздуха на уровне станции** (авторы *О. Н. Булыгина, В. Н. Разуваев, Л. Т. Трофименко, Н. В. Швець*).

Массив содержит данные 518 станций от начала наблюдений на станции до 2010 года.

Массив имеет записи фиксированной длины 94 байта с символом конца строки. Каждая запись содержит 14 полей с фиксированным положением в записи и фиксированной длиной. Дополнительно, все поля записи отделены друг от друга символом «пробел».

**4. Суммарная за месяц продолжительность солнечного сияния на станциях России** (авторы Е. Г. Апасова, Л. К. Клещенко).

Массив содержит данные 399 метеорологических станций, т. к. гелиограф установлен не на всех станциях.

Наблюдения за солнечным сиянием производились с помощью гелиографов 5 моделей (Величко, Джордана, Берлин-Стекуц, Штаде-Беккер, Кемпбелла-Стокса), однако наибольшее распространение получила модель гелиографа Кемпбелла-Стокса, которая начала использоваться в конце XIX века, а со 2-й половины XX века гелиографы Кемпбелла-Стокса обыкновенной, полярной или универсальной модели установлены на всей сети станций, гелиографы других моделей не используются [6, 7]. Если в последней графе документа есть сведения о типе прибора, то ими нужно воспользоваться, а если сведений нет, то следует считать, что наблюдения производились по одной из моделей гелиографа Кемпбелла-Стокса. В документе «Высота гелиографа» помещены сведения об индексе ВМО и названии станции, о начале наблюдений, высоте гелиографа. Для заполярных станций (45 станций) указаны даты начала и окончания полярной ночи. Массив содержит записи фиксированной длины в коде ASCII с символом конца строки. Если наблюдения не производились, то значение продолжительности отсутствует. Если продолжительность равна нулю, значит, солнце в течение месяца не появлялось или же продолжительность его сияния составила менее получаса. Если не менее 27 дней зимнего месяца входят в период полярной ночи, то в графе продолжительность стоит константа 9999.

**5. Среднемесячное парциальное давление водяного пара на станциях России** (авторы О. Н. Булыгина, В. Н. Разуваев, Н. В. Швець, В. Н. Кузнецова).

Массив содержит данные 518 станций от начала наблюдений на станции до 2010 года. Данный массив имеет структуру и формат записи, аналогичные массиву среднемесячной температуры. Массив прошел такие же процедуры статистического контроля, что и массив средней месячной температуры воздуха.

**6. Суточная температура воздуха и количество осадков на станциях России и бывшего СССР** (авторы О. Н. Булыгина, В. Н. Разуваев, Т. М. Александрова).

Первоначальная версия массива создавалась на базе архива 223 станций на территории бывшего СССР, данные для которых



публиковались в «Метеорологическом ежемесячнике СССР, часть 1 “Ежедневные данные”». Эта версия была подготовлена в рамках международного сотрудничества (ВНИИГМИ-МЦД, Россия; CDIAC, США) и была опубликована в CDIAC на CD-ROM (США) как NDP-040 [8].

Данный массив содержит данные 600 станций. Перечень станций составлен на основании списка станций Росгидромета, включенных в Глобальную сеть наблюдений за климатом (утвержденного Руководителем Росгидромета 25 марта 2004 г.) и списка реперных метеорологических станций Росгидромета, подготовленного в Главной геофизической обсерватории им. А.И. Воейкова (исп. зав. ОМРЭИ ГГО В.И. Кондратюк). Данные по станциям, находящимся на территории независимых государств – бывших республик СССР, публикуются в соответствии с решениями Исполкома межгосударственного совета по метеорологии стран СНГ об утверждении списка станций для международного обмена (Алма-Аты, 1993 год). Для некоторых станций пополнение информации не осуществляется, так как:

- Станции закрыты (как на территории России, так на территории независимых государств, бывших республик СССР);
- Данные по станциям не представлены для подготовки «Метеорологического ежемесячника станций стран содружества независимых государств, часть 1 “Ежедневные данные”».

Массив состоит из 600 файлов данных в формате ASCII с именами вида IIII.dat, где IIII – синоптический индекс станции (индекс ВМО).

Записи в файлах данных упорядочены по возрастанию ключевых элементов:

- год;
- месяц;
- день.

С целью проинформировать пользователей о наличии ошибочных данных в архивных источниках, на основании которых осуществлялось формирование массива, в состав записи введен групповой признак качества для показателей температуры воздуха – TFLAG. Значения, которые может принимать этот признак, а также формат записи приведены на сайте. Над устранением этих ошибочных значений продолжается работа, которая, к сожалению, требует значительных временных затрат из-за необходимости



обращения к первичным материалам наблюдений (книжкам КМ-1 на метеостанциях).

**7. Срочные данные основных метеорологических параметров на станциях России** (авторы О. Н. Булыгина, В. М. Веселов, В. Н. Разуваев, Т. М. Александрова).

Массив создавался по данным, содержащимся на технических носителях Госфонда Росгидромета.

Массив содержит данные восьмисрочных наблюдений за основными метеорологическими параметрами с 1966 года. Наблюдения проводились в стандартные синоптические сроки с интервалом в 3 часа. Поскольку до 1993 года данные наблюдений фиксировались по московскому декретному времени, а с 1993 года – по Гринвичскому, в призначной части каждой записи приведены параметры, позволяющие определить время производства наблюдений по поясному зимнему декретному и Гринвичскому времени.

Состав метеорологических элементов, их физический смысл и описание формата записи приведены на сайте. Точность данных соответствует точности измерения метеорологических параметров, указанной в «Наставлениях метеорологическим станциям и постам», вып. 3, часть 1, 1985. Большинству элементов массива присвоены признаки качества, которые могут принимать следующие значения:

- 0 – значение элемента достоверно;
- 1 – значение элемента достоверно и восстановлено вручную;
- 2 – значение элемента достоверно и восстановлено автоматически;
- 3 – значение элемента сомнительно;
- 4 – значение элемента забраковано программами синтаксического и семантического контроля;
- 5 – значение элемента отсутствует, но наблюдения проводились;
- 6 – значение элемента забраковано на станции;
- 7 – значение элемента отсутствует, т. к. наблюдения не производились.

Все элементы проконтролированы на допустимые значения. Некоторые элементы имеют дополнительные характеристики, отражающие особенности производства наблюдения или кодирования этих элементов.

**8. Массив суточных данных о температуре почвы на глубинах до 320 см по метеорологическим станциям Российской Федерации (автор А. Б. Шерстюков).**

Массив содержит данные 431 метеорологической станции, т. к. наблюдения за температурой почвы на глубинах проводятся не на всех станциях. Наблюдения на станциях Российской Федерации начинались неодновременно. В связи с этим период наблюдений на станциях различен. Подробное описание массива приводится в [9].

**9. Характеристики снежного покрова на 600 метеорологических станциях на территории бывшего СССР (авторы О. Н. Булыгина, В. Н. Разуваев, Т. М. Александрова).**

Регулярные наблюдения за снежным покровом на метеостанциях в России проводятся с 1882 года. Ежедневные наблюдения за снежным покровом на метеорологических станциях включают измерения его высоты и определение степени покрытия снегом окрестности станции и характера залегания снежного покрова. Степень покрытия снегом окрестности станции и характер залегания снежного покрова определяются визуально в утренний срок наблюдений.

Массив создан по данным ежедневных наблюдений за высотой снежного покрова и степенью покрытия снегом окрестности станции.

Массив состоит из 600 файлов (метеостанций) данных в формате ASCII. Файлы данных содержат записи фиксированной длины 31 байт. Каждая запись содержит 9 полей с фиксированным положением в записи и фиксированной длиной. Дополнительно все поля записи отделены друг от друга символом «пробел».

Записи в файлах данных упорядочены по возрастанию ключевых элементов: год, месяц, день.

До 1959 года степень покрытия снегом окрестности станции кодировалась цифрами 0 (50 % и менее 50 % окрестности станции) и 1 (более 50 %). С августа 1959 г. по настоящее время степень покрытия окрестности станции снегом оценивается по 10-балльной шкале. Например, отсутствие снега – 0 баллов, 20 % окрестности станции покрыто снегом – 2 балла, 50 % окрестности станции покрыто снегом – 5 баллов и т. д. По высоте снежного покрова содержится дополнительная информация в виде дополнительного признака, который принимает следующие значения:

- 1 – отсутствие снега;
- 2 – снежный покров отсутствует на станции, однако в окрестностях станции снег есть;
- 3 – высота снега меньше 0,5 см.

При подготовке массива проведен контроль данных, результаты контроля отражены в признаках качества Q2 и Q3.

Q2 содержит информацию о сравнении последовательных значений высоты снежного покрова. Q2 равен 1 (значение сомнительное), если выполняется условие :

$$\text{abs}(D) > 30 \ \& \ \text{abs}(\Delta H_1) > 10 \ \& \ \text{abs}(\Delta H_2) > 10,$$

где  $D = \Delta H_2 - \Delta H_1$ ;  $\Delta H$  – разница между соседними значениями высоты снежного покрова.

Q3 отражает результаты анализа значений высоты снежного покрова при различных значениях среднесуточной и минимальной за сутки температуры воздуха. Q3 равен 1, если высота снега не нулевая, а среднесуточная температура выше 5 °С и минимальная за сутки температура выше 0 °С.

**10. Маршрутные снегомерные съемки** (авторы О. Н. Булыгина, В. Н. Разуваев, Т. М. Александрова).

Массив содержит характеристики снежного покрова по данным маршрутных снегомерных съемок. Во время снегомерных съемок определяются высота и плотность снежного покрова, толщина ледяной корки и слоя снега, насыщенного водой, а также степень покрытия снегом и ледяной коркой маршрута и состояние поверхности почвы под снегом. В поле длина маршрута составляет 2 000 или 1 000 м, в лесу – 500 м. Высота снежного покрова в лесу измеряется через 10 м, а в поле – через 20 м. Плотность измеряется на маршруте в 1000 и 500 м – через 100 м, на маршруте длиной 2 000 м – через 200 м. Снегомерные съемки начинают проводить с момента, когда снегом покрыта половина и более площади видимой окрестности 1 раз в 10 дней. Весной перед началом и в период таяния измерения проводят каждые 5 дней. На лесных маршрутах до 20 января измерения осуществляют 1 раз в месяц (20-го числа). Измерения плотности снега и запасов воды в снеге по данным маршрутных снегомерных съемок осуществляются в России с 1930 года.

Методика наблюдений за характеристиками снежного покрова неоднократно изменялась. После 1965 года нарушений однородности, вызванных изменением процедуры наблюдений, не было. В массиве приведены данные 517 метеорологических

станций России с 1966 года по настоящее время. Список станций формировался на основе списка ВМО (Publication N 9, Volume A, Observing Stations). Были отобраны станции, наблюдения на которых осуществлялись в течение всего периода начиная с 1966 года по настоящее время (отсутствие наблюдений не более двух лет).

Снегосъемки осуществляются отдельно для трех типов ландшафта: поле, лес и овраги. В оврагах измерения проводятся только по заданию Гидрометеорологической обсерватории.

Каждая запись в архиве содержит 19 характеристик. Содержится информация о степени покрытия окрестности станции и маршрута снегом, степени покрытия маршрута ледяной коркой, средней, наименьшей и наибольшей высоте снежного покрова на маршруте, средней плотности снега и средней толщине ледяной корки, толщине слоя снега, насыщенного водой, толщине слоя чистой воды, запасе воды в снеге, характере залегания снежного покрова и характере снежного покрова.

### **3. Организация доступа к специализированным массивам в сети Интернет**

На сервере ВНИИГМИ-МЦД доступ к массивам данных, выборка данных по интересующим пользователя станциям, их просмотр и копирование обеспечиваются специализированной технологией Аисори [10]. Авторы – канд. физ.-мат. наук В. М. Веселов и канд. техн. наук И. Р. Прибыльская.

Технология работает под управлением Web-сервера приложений Байконур [11]. Достоинством этого сервера приложений является то, что для разработки пользовательских приложений вместе с сервером Байконур поставляется библиотека HTML-компонентов, предназначенная для визуального построения приложений в среде Borland C++ Builder. Это позволяет разрабатывать приложения, доступные любому пользователю Интернет/Интранет с помощью браузера общего назначения, установленного на компьютере клиента. Конфигурация компьютера клиента и операционная система, установленная на нем, значения не имеют. Интерфейс технологии, обеспечивающий взаимодействие с клиентом, выполнен с использованием этой библиотеки. Пользователь формулирует запрос на выборку данных в режиме диалога. В процессе диалога ему показываются формы с окнами, которые он заполняет осуществляя выбор из предоставляемых ему вариантов.

## Заключение

Создание баз климатических данных органично связано с решением ряда проблем, решение которых позволит в полной мере провести комплексные исследования состояния климата на территории России. Необходимо осуществить сбор всей информации об изменении местоположения станций, изменениях в режиме наблюдений и смене приборов наблюдения, что позволит осуществить коррекцию данных с целью создания однородных рядов данных. Чрезвычайно важным представляется организация работ по исследованию климата с использованием вновь создаваемых массивов, привлечению новых участников, в том числе и зарубежных.

## СПИСОК ЛИТЕРАТУРЫ

1. *Климатический справочник СССР. Метеорологические данные за отдельные годы. Выпуски 1–34.* Л.: Гидрометеиздат, 1952–1957.
2. *Справочник по климату СССР. Метеорологические данные за отдельные годы. Выпуски 1–34.* Л.: Гидрометеиздат, 1966–1970.
3. *Метеорологический ежемесячник СССР. Выпуски 1–34, ч. 2.* 1961–1990.
4. *Веселов В.М.* Архивы Госфонда на ПЭВМ и технология их организации // Труды ВНИИГМИ-МЦД. 2002. Вып. 170. С. 16–30.
5. *Трофименко Л.Т., Шевченко Н.Н.* Анализ полей асимметрии и эксцесса как средство контроля качества архивных данных // Труды ВНИИГМИ-МЦД. 1985. Вып. 116. С. 67–77.
6. *Сохрина Р.Ф., Фан А.А.* О влиянии систематических ошибок гелиографов, методов наблюдений и обработки на нормы характеристик солнечного сияния // Труды ГГО. 1976. Вып. 335. С. 87–90.
7. *Быкова Л.С.* Об однородности рядов наблюдений за продолжительностью солнечного сияния // Труды ГГО. 1987. Вып. 515. С. 124–128.
8. *Razuvayev V.N., Apasova E.G., Martuganov R.A., Steurer P., Vose R.* Daily Temperature and Precipitation Data for 223 U.S.S.R. Stations. ORNL/CDIAC, Numerical data package – 040, Oak Ridge, Tennessee, USA: Oak Ridge National laboratory, 1993.
9. *Шерстюков А.Б.* Массив суточных данных о температуре почвогрунтов на глубинах до 320см по метеорологическим станциям Российской Федерации // Труды ВНИИГМИ-МЦД. 2012. Вып. 176. С. 233–256.
10. *Веселов В.М., Прибыльская И.Р., Проскурня В.И.* Система Аисори – стандартное средство работы с архивными гидрометеорологическими данными // Труды ВНИИГМИ-МЦД. 2000. Вып. 166. С. 8–25.
11. *Серверы приложений и средства для быстрой разработки Интернет-приложений.* Официальный сайт компании Epsilon Technologies. – <http://www.demo.ru>