УДК 004.42

РАСПРЕДЕЛЕННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ХРАНЕНИЯ И ОБРАБОТКИ МЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ

Ботыгин И.А., Шерстнёв В.С., Шерстнёва А.И.

ФГБОУ ВПО «Национальный исследовательский Томский политехнический университет», Томск, e-mail: vss@tpu.ru, sherstneva@tpu.ru

В статье приведены результаты разработки программного обеспечения распределенного сетевого хранилища данных приземных метеорологических наблюдений. Данные представлены национальным вариантом международного кода для оперативной передачи данных о состоянии окружающей среды с метеорологических станций по всем регионам России. Они доступны для исследователей как в визуальной форме, так и в широко распространенных экспортных форматах. Спроектирована архитектура распределенного сетевого хранилища метеоданных, включающая следующие модули: модуль диспетчера (для мониторинга вычислительных узлов node, распределения метеоданных по узлам, приёма запросов от клиентов), клиентский модуль (для доступа к метеоданным сторонним исследователям), модуль терминалов (для внесения данных метеорологических наблюдений), модули обработки и хранения (узлы хранения распределенных метеоданных, включающие подмодуль обработки информации и подмодуль хранения данных). Приведены результаты практической апробации разработанного программного обеспечения сетевого хранилища данных приземных метеорологических наблюдений.

Ключевые слова: приземные метеорологические наблюдения, информационная технология, сетевое хранилище данных, распределенные системы, системы управления базами данных

DISTRIBUTED SOFTWARE OF METEOROLOGICAL DATA HOLDING AND PROCESSING

Botygin I.A., Sherstnev V.S., Sherstneva A.I.

National Research Tomsk Polytechnic University, Tomsk, e-mail: vss@tpu.ru, sherstneva@tpu.ru

In this paper we offer the results of the development of distributed network ground hydro-meteorological observations data warehouse software. In our research we use a data representation by the national analogue of international code for the transmission of environmental conditions data from Russian meteorological stations. Moreover, we also let work with data represented in different visual formats and in almost all of widely used export ones. The distributed network data warehouse architecture is designed. It includes the following modules: dispatcher' module (for computational node monitoring, meteorological data distribution to nodes, receiving of clients' requests), clients' module (for outside users' access to meteorological data), module of terminals (for loading of meteorological data), modules of processing and holding (nodes of distributed meteorological data holding with submodule of data processing and submodule of data holding). We adduce the results of practical approbation of the developed network ground hydro-meteorological observations data warehouse software.

Keywords: ground meteorological observations, information technology, network data warehouse, distributed systems, database control systems

В настоящее время основными источниками больших объемов данных являются, все же, данные научных и физических измерений. Например, для задач анализа атмосферных процессов, данных метеорологических наблюдений, системообразующих параметров природной среды, пространственно-временных наборов дендрохронологических данных, естественных импульсных электромагнитных характеристик Земли и т.д. используются петабайты данных, которые необходимо хранить, и для обработки которых необходимы значительные вычислительные ресурсы. Добавим, что из-за вариативности данных, вследствие изменения их состава и структуры при разработке новых информационных сервисов и услуг, возникают дополнительные проблемы в платформах сетевых хранилищ при занесении, хранении, управлении и анализе информации.

Одним из наиболее простых и эффективных решений указанной проблемы, по мнению авторов, является создание сетевого хранилища данных. По определению, сетевое хранилище подразумевает территориально распределенное хранение информации. Таким образом, сетевое хранилище данных можно представить распределенной вычислительной системой, предназначенной для хранения наборов данных, которые не только логически связаны между собой, но и физически (территориально) распределены по разным узлам компьютерной системы. В пользовательском смысле, основной задачей такой распределенной системы, как составного элемента централизованного Интернет-сервиса, является хранение данных и предоставление доступа к ним по мере необходимости со стороны неограниченного количества сетевых пользовательских сервисов.

В статье предлагается к обсуждению инфологическое, алгоритмическое и программное обеспечение распределённой си-

стемы хранения и обработки метеорологических данных, обладающей следующими основными преимуществами:

- позволяет получать данные из первоисточников в формате КН-01;
- обладает распределённой масштабируемой архитектурой;
- хранит метеоданные в реляционных СУБД.

На сегодня в России используется множество автоматических и полуавтоматических метеостанций, ежедневно передающих множество погодных сводок [1]. Соответственно, весь поступающий объем метеоданных требуется эффективно хранить и в последующем использовать для тех или иных задач, например для моделирования и прогнозирования погоды.

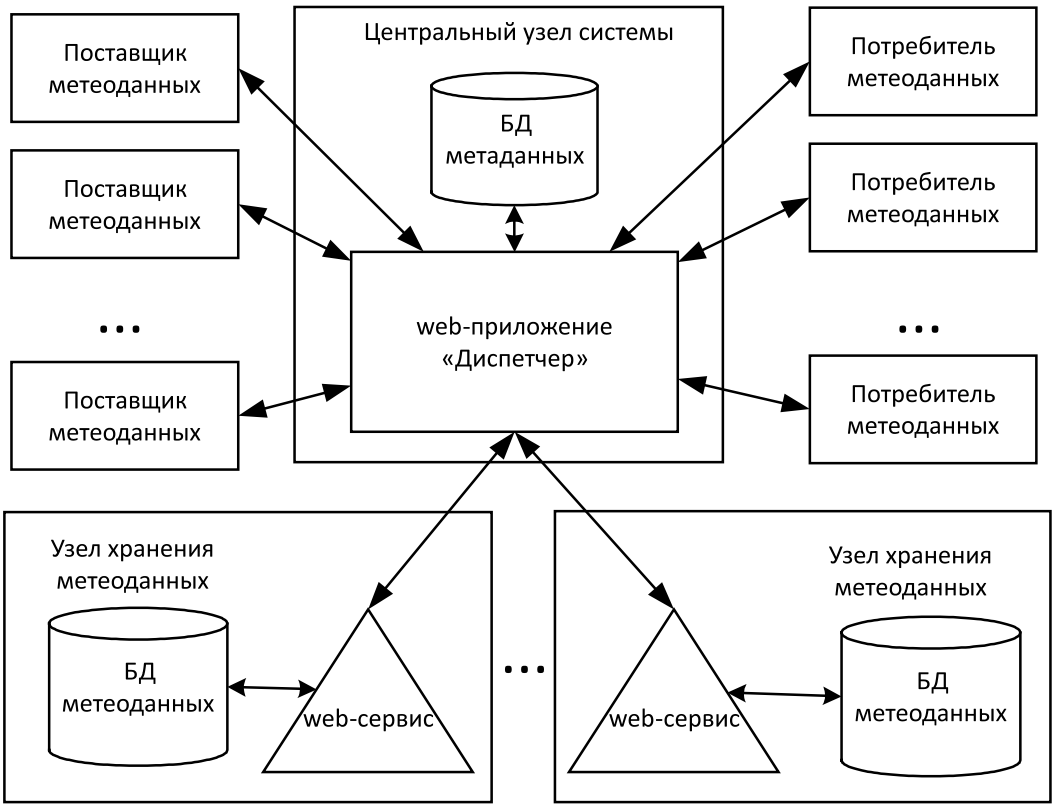

Рис. 1. Обобщённая архитектура распределённой системы хранения и обработки метеорологических данных

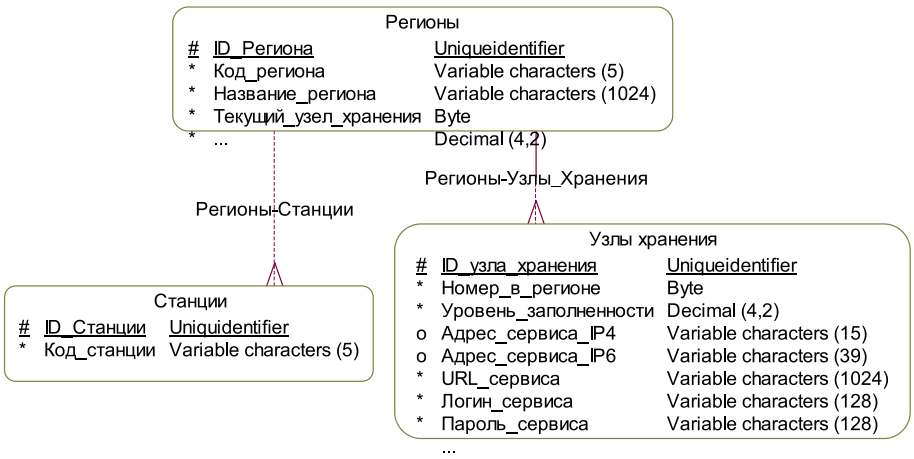

Рис. 2. Фрагмент концептуальной модели БД с метаинформацией системы

Хранение метеоданных погодных сводок возможно в оригинальном формате КН-01, но это нецелесообразно из-за потребности декодирования данных для пост-обработки. Особенно явно видится сложность использования формата КН-01 при хранении метеоданных с учётом больших объёмов поступающих метеоданных. Хранение и обработка больших объёмов слабо структурированных данных не рекомендуется в современных информационных системах. Это позволяет сделать вывод о целесообразности разработки средств для хранения и постобработки метеоданных, например средствами реляционных СУБД.

Код КН-01 представляет собой российский вариант международного кода FM 12 SYNOP [2, 3]. В нём имеется ряд отступлений от кода ВМО FM 12, допускаемых правилами всемирной метеорологической организации.

В классическом виде погодная сводка (сообщение с метеоданными) ПС представляет собой буквенно-числовую последовательность символов.

Погодная сводка российского формата КН-01 состоит из 4 разделов. Являясь подмножеством международного кода FM 12, российская версия не использует 2-й и 4-й. Каждый раздел делится на группы значений. Роль группы символов в начале каждого раздела предопределена – они являются маркерами, показывающими начало нового раздела погодной сводки.

В целом, обобщённая архитектура предлагаемой распределённой системы хранения и обработки метеорологических данных представлена на рис. 1.

На схеме обобщённой архитектуры распределённой информационной системы хранения и обработки метеорологических данных представлены:

- Центральный узел системы, состоящий из приложения «Диспетчер» и его базы метаданных.
- Узлы хранения, состоящие из веб-сервисов и баз данных с накопленными метеоданными.

Приложение «Диспетчер» предназначено для решения задачи получения погодного сообщения, выяснения принадлежности сводки некоторому географическому региону, нахождения узла хранения, соответствующего этому географическому региону, передачи сводки на нужный узел хранения.

Веб-сервисы в узлах хранения необходимы для получения запроса от приложения «Диспетчер», выполнения соответствующего запроса к СУБД узла хранения, возврата выбранных из СУБД метеоданных к приложению «Диспетчер».

В целом, база данных системы подразделяется на 2 части: БД метаданных и БД метеоданных. БД метаданных описывает систему в целом и хранит высокоуровневую информацию о системе, а БД метеоданных хранит непосредственно информацию из погодных сводок.

Модель БД с метаинформацией системы достаточно проста и представлена на рис. 2. База метаданных хранит служебную информацию о узлах хранения метеоданных: регион и адрес веб-сервиса узла хранения метеоданных по этому региону.

Модель позволяет хранить информацию о регионах и узлах хранения метеоданных для этих регионов. На основе этой информации приложение «Диспетчер» (зная источник метеоданных) выбирает узел хранения данных и пересылает погодную сводку на соответствующий web-сервис узла хранения. Соответствующий web-сервис узла хранения принимает погодную сводку и сохраняет её в своей БД.

База метеоданных в узлах хранения содержит погодные сводки в нормализованном виде, пригодном для дальнейшей постобработки. Для хранения метеоданных, расшифрованных из погодной сводки формата КН-01, предлагается использовать более сложную модель данных, фрагмент которой представлен на рис. 3.

Хотя в исходной погодной сводке формата КН-01 могут присутствовать секции с повторяющимися данными, структура БД приведена в 3-ю нормальную форму, что избавляет от хранения избыточной информации. При структурированном хранении метеоданных в реляционных СУБД данные могут быть легко использованы сторонними приложениями для дальнейшей постобработки. Структура метаданных системы позволяет контролировать наполненность каждого узла хранения и сигнализировать администратору системы об исчерпывании свободного пространства для хранения данных.

В случае критической наполненности текущего узла хранения и наличия пустых узлов хранения для метеоданных региона приложение «Диспетчер» автоматически меняет текущий заполненный узел хранения на следующий свободный.

В случае отсутствия новых пустых узлов администратор системы должен:

1) найти ресурсы для создания очередного узла хранения;

2) развернуть новый узел хранения;

3) внести информацию о нём в БД метаданных системы.

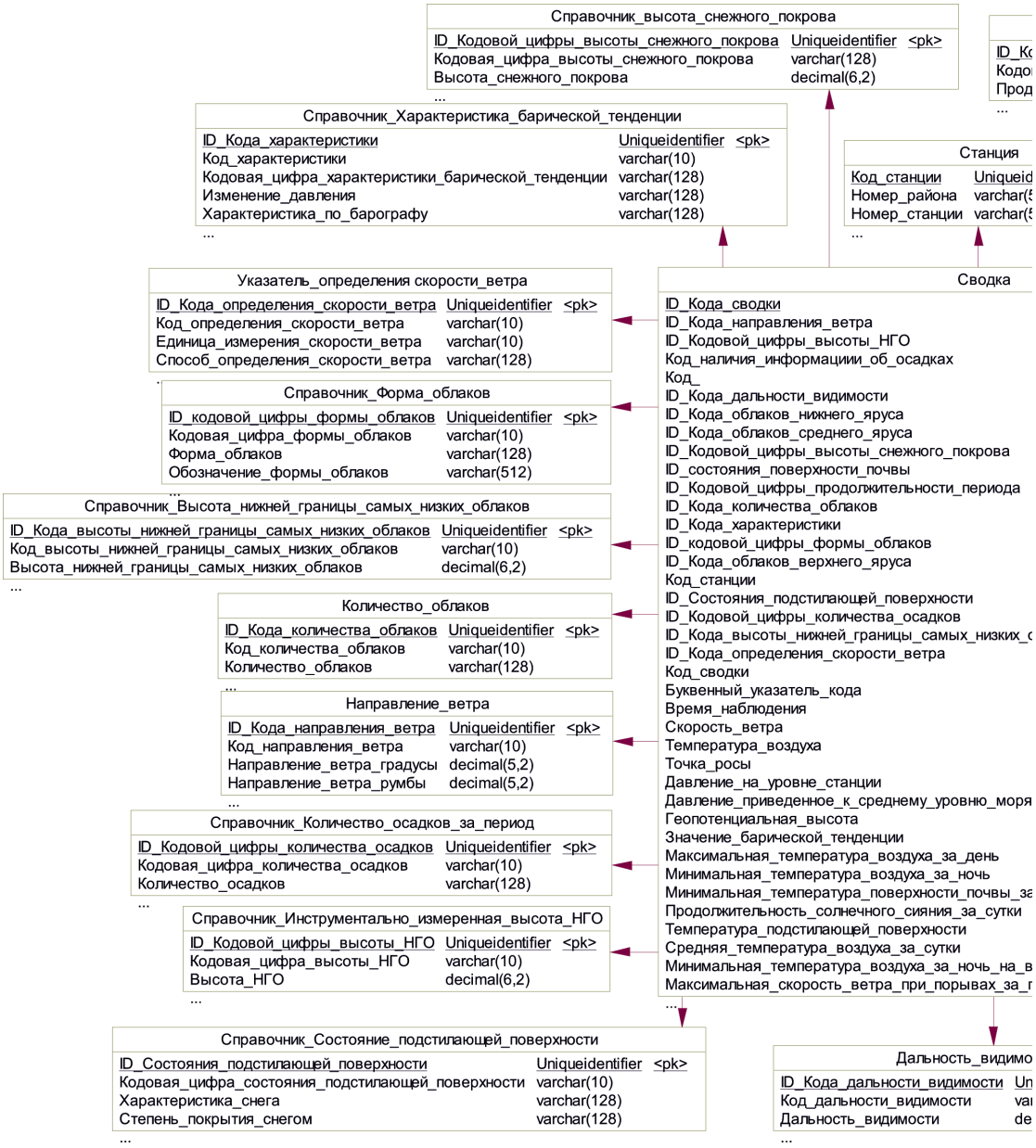

Рис. 3. Фрагмент физической модели БД для хранения метеоданных (начало)

В целом, для создания данной системы предлагается использовать кроссплатформенное программное обеспечение: программный код на языке Java [4], реляционную СУБД MySQL [5], стандарт ANSI-SQL92 [6].

Для сохранения метеоданных из погодной сводки в БД требуется произвести парсинг сводки формата КН-01, который может быть произведён средствами web-сервиса (на языке Java) или средствами бизнес-логики СУБД (на языке SQL).

Парсинг фрагментов текста (метеоданных погодных сводок) является распространённой задачей и может быть выполнен с помощью соответствующих регулярных выражений языка Java.

Так как структура погодных сводок является достаточно регулярной и предсказуемой, вариант парсинга с помощью бизнес-логики СУБД также доступен для реализации. Большинство современных СУБД обладают необходимым набором операторов для извлечения подстрок из текстовых выражений (иногда с помощью регулярных выражений). Таким образом, вся процедура парсинга погодной сводки может быть сведена к последовательному вызову хранимых процедур и функций. Каждая из соответствующих процедур

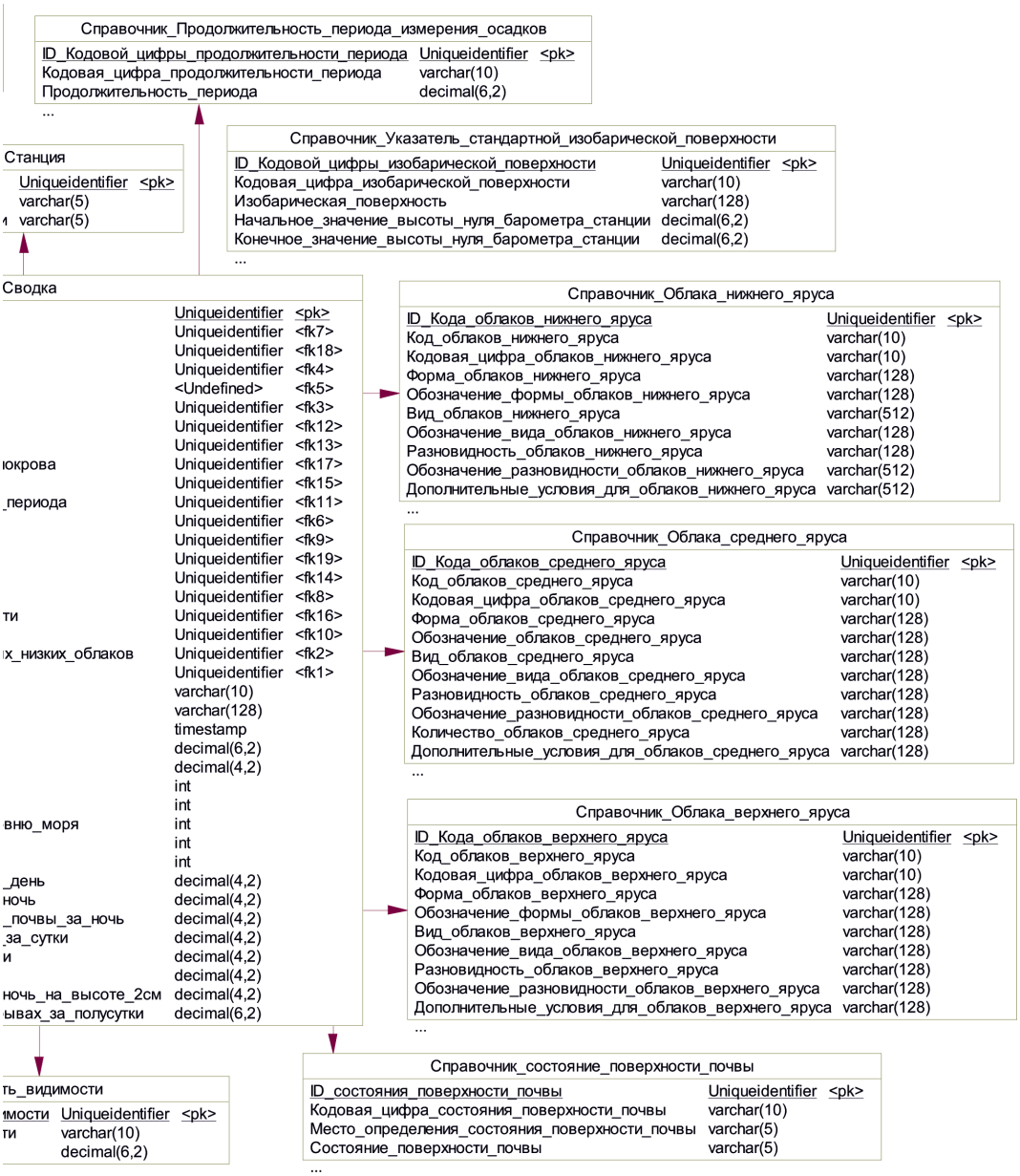

Рис. 3. Фрагмент физической модели БД для хранения метеоданных (окончание)

и функций СУБД должна выделять из метеоданных требуемую секцию и сохранять её в БД.

Благодаря декодированию погодных сводок формата КН-01 и сохранению метеоданных в реляционной СУБД есть возможность решать разнообразные задачи на основе этой создаваемой распределённой базы метеоданных.

Полученные результаты позволяют трансформировать наработки по управлению данными в формате SYNOP-FM12 и для оперативного сбора, систематизации, хранения и данных гидрометеорологических наблюдений, представленных в других буквенно-цифровых кодах (FM18 – BUOY, FM-32 – PILOT, … и т.д., вплоть до FM-75 – CLIMAT-TEMP). Кроме того, возможна адаптация системы и для наблюдений других категорий – AGRO, BUFR, RADR и др. Проведённые исследования позволили поставить задачи интеграции нескольких моделей и методов анализа данных в рамках единой распределенной информационной системы.

Примерами задач, которые могут быть решены на основе такой базы данных, могут быть задачи построения разнообразных тематических карт (температурных карт, карт давления, направления ветра и т.д.) [7,

8]. При наличии соответствующего программного конвертора, на основе собранной базы метеоданных возможно решать задачи моделирования с помощью пакетов моделирования WRF [9].

Список литературы

1. Городецкий О.А., Гуральник И.И., Ларин В.В. Метеорология, методы и технические средства наблюдений. – 2-е изд. – Л.: Гидрометеоиздат, 1991. – С. 8. – 336 с. – ISBN 5-268-00646-9.

2. Наставление по кодам. Международные коды. Том. I.1. Часть А. Буквенно-цифровые коды. Издание 2011 г. Секретариат ВМО. [Электронный ресурс]. – Режим доступа: http://www.wmo.int/pages/index_ru.html (01.05.2015).

3. Наставление по кодам. Региональные коды. Национальная практика кодирования. Том. II. Издание 2011 г. Секретариат ВМО. [Электронный ресурс]. – Режим доступа: http://www.wmo.int/pages/index_ru.html (01.05.2015).

4. Что такое Java? Oracle Corporation. [Электронный ресурс]. – Режим доступа: http://www.java.com/ru/download/whatis_java.jsp (01.05.2015).

5. The world's most popular open source database. Oracle Corporation. [Электронный ресурс]. – Режим доступа: http://www.mysql.com/ (01.05.2015).

6. SQL-92. Wikipedia. [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/SQL-92 (01.05.2015).

7. Институт радарной метеорологии. [Электронный ресурс]. – Режим доступа: http://www.iram.ru/iram/p26_expert_ru.php (01.05.2015).

8. Карты погоды. Интернет-журнал Meteoweb.ru. [Электронный ресурс]. – Режим доступа: http://meteoweb.ru/maps.php (01.05.2015).

9. The Weather Research and Forecasting (WRF) Model. [Электронный ресурс]. – Режим доступа: http://www.wrf-model.org/index.php (01.05.2015).

9. The Weather Research and Forecasting (WRF) Model. [Jelektronnyj resurs]. Rezhim dostupa: http://www.wrf-model.org/index.php (01.05.2015).