

ПОДГОТОВКА ДАННЫХ СТАЦИОННЫХ НАБЛЮДЕНИЙ ДЛЯ ПРОВЕДЕНИЯ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ ПО МЕЗОМЕТЕОРОЛОГИЧЕСКИМ МОДЕЛЯМ

* Исследование выполнено при финансовой поддержке Иркутского государственного университета в форме гранта за счёт средств темы 091-15-236 (в соответствии с приказом № 701 от 13.11.2015 г. "О результатах конкурса грантов ИГУ для аспирантов и молодых сотрудников").

METHODS OF PREPARING THE DATA OF STATION OBSERVATIONS FOR NUMERICAL EXPERIMENTS ON THE MESOSCALE MODES

P. Naidenov

Annotation

The results of development of special software that allows automated using of hydrometeorological information archive TMS-84 for verification of mesometeorological models. Proprietary software used by the author to verify the simulation results mesometeorological trials of Lake Baikal using non-hydrostatic numerical model WRF.

Keywords: verification of the non-hydrostatic models, archive TMS-84, an array of data, automated using methods.

Найденов Павел Андреевич
Аспирант, Иркутский
государственный университет

Аннотация

Представлены результаты разработки специального программного обеспечения, позволяющего автоматизировано использовать гидрометеорологическую информацию архива ТМС-84 для верификации мезомасштабных моделей атмосферы. Собственное программное обеспечение используется автором для верификации результатов моделирования мезометеорологических процессов над озером Байкал с использованием численной негидростатической модели WRF.

Ключевые слова:

Данные для верификации мезометеорологических моделей, архив ТМС-84, массив данных, методика автоматизированного использования.

Введение

В настоящее время использование сетевых метеорологических данных является неотъемлемой частью большинства исследовательских работ в области верификации моделей мезометеорологических процессов. Например, в наших исследованиях планируется оценить качество детализации мезометеорологических полей над озером Байкал, полученных в результате динамической интерполяции с помощью математической модели атмосферы WRF, разработанной Национальным Центром Атмосферных Исследований США [1]. Планируется также провести такую оценку путем сравнения с данными наблюдений на метеорологических станциях и постах, расположенных по периметру озера Байкал.

Эта задача требует использования данных из электронного архива таблиц метеорологических наблюдений ТМС-84 Иркутского УГМС и Забайкальского УГМС. Данные архивы насчитывают несколько десятков тысяч текстовых файлов, каждый из которых содержит основные метеорологические параметры (температура, давление, ветер, облачность, характеристики влажности и др.) для

определенной станции по восьми срокам наблюдения за один календарный месяц. Вся информация проходит многоступенчатый числовой контроль, прежде чем попасть в ТМС-84.

Таблицы архивов ТМС-84 существуют только в виде текстовых файлов, так как программы ЭВМ на сети метеорологических станций генерируют их именно в таком формате. Работая с текстовыми файлами, невозможно оперативно производить различные выборки необходимой информации. Попытка сделать подобные манипуляции вручную при таком значительном объеме архива неизбежно приведет к нерациональной трате времени и накоплению ошибок.

Цель работы

Цель данной работы заключается в том, чтобы создать специальное программное обеспечение, позволяющее оперативно переносить данные архивов ТМС-84, принадлежащих территориальным УГМС России, в электронные таблицы Microsoft Excel. При этом для удобства использования вместо множества отдельных текстовых

файлов целесообразно было создать 8 электронных таблиц. Эти таблицы используются для верификации модели WRF. Преимущество электронных таблиц по отношению к текстовым файлам заключается в возможности оперативно производить любые выборки нужной информации. Для решения этой задачи используется программное обеспечение Microsoft Visual Basic и Microsoft Excel. Кроме того, используются образцы таблиц архива ТМС-84 Иркутского УГМС.

В соответствии с целью были поставлены следующие задачи:

1. Разработать необходимые алгоритмы и написать программу распознавания текстовых файлов архивов ТМС-84 и перевода метеорологической информации в электронные таблицы Microsoft Excel.
2. Оценить точность и корректность разработанной методики на примере тестовых образцов.

Краткое описание модели WRF

Модель WRF негидростатическая, основана на системе уравнений гидротермодинамики, записанных в системе координат. В модели применяется параметризация процессов подсеточного масштаба: конвекции в кучевых облаках, крупномасштабной конденсации, микрофизических процессов, параметризация радиации, планетарного пограничного слоя, процессов на подстилающей поверхности [1].

В WRF используются картографические проекции земной поверхности: стереографическая полярная, ламбертовская конформная и меркаторовская. Необходимость использования картографических проекций возникает в связи с тем, что земную поверхность невозможно спроектировать на плоскость без искажения [1].

В качестве начальных полей метеозлементов в моде-

ли используются данные реанализа (NCEP/NCAR Reanalysis). Метеорологические поля Реанализа записаны в коде GRIB, применяемом для хранения гидрометеорологической информации и читаемом большинством зарубежных атмосферных моделей, в том числе и WRF. Существуют также программы для перехода от кода GRIB к обычному текстовому формату [1].

Для того чтобы оценить качество модельных данных результаты моделирования должны сопоставляться с данными станционных наблюдений.

Проблемы алгоритмизации чтения текстовых файлов архива ТМС-84

Для успешного выполнения поставленных задач необходимо было разработать подходящий алгоритм автоматизированного чтения текстовых файлов и одновременного извлечения необходимой информации в память компьютера.

При этом основные сложности были связаны со следующим:

1. Большой объем информации был рассредоточен более чем по 16000 отдельных текстовых файлов.
2. Каждый из этих файлов следовало проверять на кодировку, так как некоторые из них кодировались по таблице DOS (IBM CP 866), а другие по таблице WIN1251.
3. Некоторые из файлов были испорчены вследствие технического сбоя, поэтому их нужно было исключить.

На рис. 1 представлен фрагмент типичного текстового файла из архива ТМС-84. Из этого фрагмента видно, что таблицы текстовых файлов содержат основную метеорологическую информацию (температура воздуха, характеристики влажности, облачность, ветер и др.).

станция Белозиминская										N станции 5360051										Год 1999										Месяц 5										Гринвич. время (поясное) 15										,18										(22										,01)										стр. 2									
число	Температура, град.			парц. давл. вод. п гпа	отн. вл. , проц	Дефицит насыщ гпа	Атм. давление, гпа		Бар. тенд		Види м-сть, шифр	О б л а ч н о с т ь										погода, шифр		Ветер		осадки мм																																																															
	воздуха	пов. почвы	точки росы				на ур. станции	на ур. моря	вид	вел. гпа		кол-во баллов	Сi	формы	NS	высо-	ни-	же	ст.	w	ww	напр град	ск, м/с																																																																		
1	-0.9	-6	-8.7	3.17	55	2.55	919.5	1011.5	2	1.6	10	0	5	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0																																																													
2	2.4	-2	2.4	7.26	100	0.00	926.2	1017.6	8	0.3	10	10	8	8	0	0	2	*600	6	65	0	0	0	0	6	65	0	0																																																													
3	4.9	-2	1.6	6.87	79	1.79	931.0	1022.0	1	1.0	10	0	5	1	0	0	0		2	2	0	0	0	2	2	0	0																																																														
4	1.9	-3	-3.5	4.72	67	2.28	930.2	1022.2	3	1.2	6	0	5	0	0	0	0		2	2	0	0	0	2	2	0	0																																																														
5	1.9	-2	-7.9	3.38	48	3.62	928.1	1020.0	2	1.0	0	0	0	0	0	0	0		0	2	0	0	0	0	2	2	0	0																																																													
6	1.9	-0	-7.9	3.38	48	3.62	927.2	1019.0	1	0.6	10	0	3	0	0	0	0		2	2	0	0	0	2	2	0	0																																																														
7	3.5	-0	-5.0	4.20	54	3.65	923.8	1014.6	1	0.2	6	0	3	0	0	0	0		2	2	0	0	0	2	2	0	0																																																														
8	5.9	3	-4.6	4.34	47	4.94	921.2	1011.0	2	1.3	6	3	5	0	0	2	0	*800	2	2	0	0	0	2	2	0	0																																																														
9	5.5	3	2.6	7.34	81	1.69	927.9	1018.4	1	0.5	10	5	5	0	0	2	0	*800	2	2	0	0	0	2	2	0	0																																																														
10	-0.6	-2	-0.6	5.85	100	0.00	929.5	1022.2	1	0.3	10	10	8	8	2	0	3	*500	8	86	0	0	0	8	86	0	0																																																														
11	-2.7	-4	-3.0	4.88	97	0.13	932.9	1026.8	5	0.0	6	0	5	0	0	0	0		2	2	0	0	0	2	2	0	0																																																														
12	1.5	-0	-5.1	4.18	61	2.63	923.4	1015.3	4	0.0	8	6	0	1	0	2	0	*800	2	2	0	0	0	2	2	0	0																																																														
13	4.4	3	-0.7	5.80	69	2.56	925.1	1015.7	3	3.7	10	6	0	1	0	2	0	*800	2	2	360	2	0	2	2	360	2																																																														
14	4.2	2	-5.6	4.02	49	4.22	928.2	1019.3	4	0.0	3	0	1	0	0	0	0		0	2	0	0	0	0	2	2	0	0																																																													
15	7.9	7	4.5	8.4	79	2.2	925.4	1014.8	3	0.8	10	6	1	0	0	2	0	*800	2	2	0	0	0	2	2	0	0																																																														
16	3.2	-0	-2.2	5.20	68	2.48	932.1	1023.9	2	2.1	5	0	1	0	0	0	0		0	2	0	0	0	0	2	2	0	0																																																													

Рисунок 1. Фрагмент текстового файла из архива ТМС-84.

Однако, поскольку это текстовый файл, у пользователя возникают трудности при выборке информации. Например, если необходимо извлечь из представленной таблицы информацию об атмосферном давлении на уровне моря за период с 1 по 5 число, понадобится 5 раз провести операции копирования и вставки.

Обоснование выбора типа электронных таблиц

Перед началом работы нужно было выбрать программное обеспечение для создания будущей базы данных. Выбор был между MicrosoftAccess и MicrosoftExcel. В случае затруднений в выборе между программами ExcelиAccess нужно ответить на важный вопрос: данные должны быть организованы в виде реляционной структуры или в виде плоской структуры? Дело в том, что программа Access предназначена для работы с реляционными данными, а приложение Excel лучше подходит для работы с данными в виде плоской структуры [2]. В реляционной структуре данные логически разделены на части, каждая из которых помещена в отдельную таблицу. В отличие от реляционной структуры данных плоская структура представляет собой таблицу, поля которой не связаны с другими данными.

Данные архива ТМС-84 было решено привести к плоской структуре для использования в программеExcel. Однако для этого потребовалось в начале каждой строки с метеорологической информацией помещать соответствующие данные о метеостанции, дату и срок наблюдения.

Электронная таблица Excel, фрагмент которой представлен на рис. 2, позволяет оперативно производить выборку информации, перенесенной из архива текстовых файлов ТМС-84. Оперативность достигается за счет возможности применения фильтров отдельно для каждого столбца таблицы.

Обоснование разделения архива на несколько электронных таблиц

Прежде всего был оценен объем будущей электронной таблицы. Оказалось, что в случае помещения всей накопленной информации в одну электронную таблицу, последняя содержала бы более 4 000 000 строк. Такая таблица занимала бы более 1 гигабайта дискового пространства и ее невозможно было бы открыть с помощью программы MicrosoftExcel, так как даже ее современная версия позволяет создавать и читать файлы с количеством строк не более 1 000 000.

Поэтому было решено создать 8 электронных таблиц, каждая из которых должна была содержать информацию по всем станциям за весь период наблюдений, но только за один срок наблюдения, например, 00, 03 и так далее. Такой подход обеспечивает приемлемое количество строк в каждой из восьми электронных таблиц (около 500000) и в то же время позволяет достаточно оперативно работать с существующими данными и добавлять вновь поступающие.

Способ добавления исходной информации в начало каждой строки с метеорологическими данными электронной таблицы

Все текстовые файлы архива ТМС-84 содержат информацию о названии станции, ее номере и времени проведения наблюдений. Однако, локализация этой информации меняется от одного файла к другому. Следовательно, в алгоритме невозможно указать конкретный номер строки с которой эту информацию можно считать. Поэтому был разработан алгоритм поиска нужной строки при помощи определенного шаблона которому удовлетворяет только нужная строка с исходной информацией.

Таким образом, при чтении каждого файла, программа сначала находит строку с исходной информацией, заносит ее в память компьютера и затем, уже при чтении и

Станция	Номер	Год	Месяц	Срок (поясное)	Срок (Принвич)	Число	Темп воздуха	Темп почвы	Точка росы	Парци давление	Отн влажность	Деф. Насыщения	Облачность						
													Бальность общая	Нижний ярус	Ci Cc Cs	Ac As	Cu Cb	St Sc	Ns Ft-rb
Альг джер	5339830	1999	март	15	22	1	-6.1	-12	-17.4	1.57	40	2.31	10	0	5	0	0	0	0
Альг джер	5339830	1999	март	15	22	2	-8.8	-10	-12.6	2.32	74	0.83	10	0	8	4	0	0	0
Альг джер	5339830	1999	март	15	22	3	-13.0	-14	-17.3	1.58	69	0.68	10	0	8	4	0	0	0
Альг джер	5339830	1999	март	15	22	4	-12.9	-16	-17.2	1.59	71	0.68	10	0	8	4	0	0	0

Рисунок 2. Фрагмент электронной таблицы Excelс данными из архива ТМС-84.

помещении метеорологической информации в электронные таблицы, добавляет ее в начало каждой строки. Затем при переходе к следующему файлу, исходная информация изменяется.

Метод определения кодировки читаемого текстового файла

Различия в кодировке текстовых файлов архива ТМС-84 представляли определенную трудность при разработке алгоритма переноса информации из текстовых файлов в электронные таблицы. Каждый из этих файлов следовало проверять на кодировку, так как некоторые из них кодировались по таблице DOS (IBMCP 866), а другие по таблице WIN1251. Потому был разработан простой и надежный способ проверки кодировки текста. Блок-схема этого алгоритма [3] представлена на рис. 3.

Представленная блок-схема иллюстрирует работу отдельного участка разработанного автором программного обеспечения. На данном этапе программы выполняется предварительное чтение текстового файла из архива ТМС-84. При этом используется кодировка DOS (IBMCP 866). Далее программа сравнивает результат предварительного чтения контрольной строки файла с эталонным шаблоном, заложенным в нее заранее. В случае совпадения результата чтения с образцом, программа записывает в память компьютера информацию о том, что при чтении текущего текстового файла нужно использовать

кодировку DOS. В случае несовпадения результата чтения с образцом, программа изменяет используемую кодировку на WIN1251

Метод исключения испорченных текстовых файлов

Некоторые файлы архива ТМС-84 были испорчены. Всего в архиве ТМС-84 Иркутского УГМС оказалось три таких файла. Их пришлось исключать из формируемых электронных таблиц. Наиболее простым и надежным методом представлялось использование функции определения кодировки, поскольку испорченные файлы не поддавались чтению даже при использовании не только описанных выше, но и всех доступных автору кодировок.

В программе было задано следующее условие: если контрольная строка не соответствует шаблону при использовании кодировок DOS и WIN1251, то текущий файл следует считать испорченным и не переносить из него информацию в электронные таблицы.

Это позволило исключить испорченные файлы из итоговых электронных таблиц. Из-за утери части информации ряды метеорологических данных претерпели разрыв. Однако это не повлияет на оценку качества детализации мезометеорологических полей над озером Байкал, поскольку утерянные файлы относятся к метеорологическим станциям, находящимся в западной части Иркутской области.

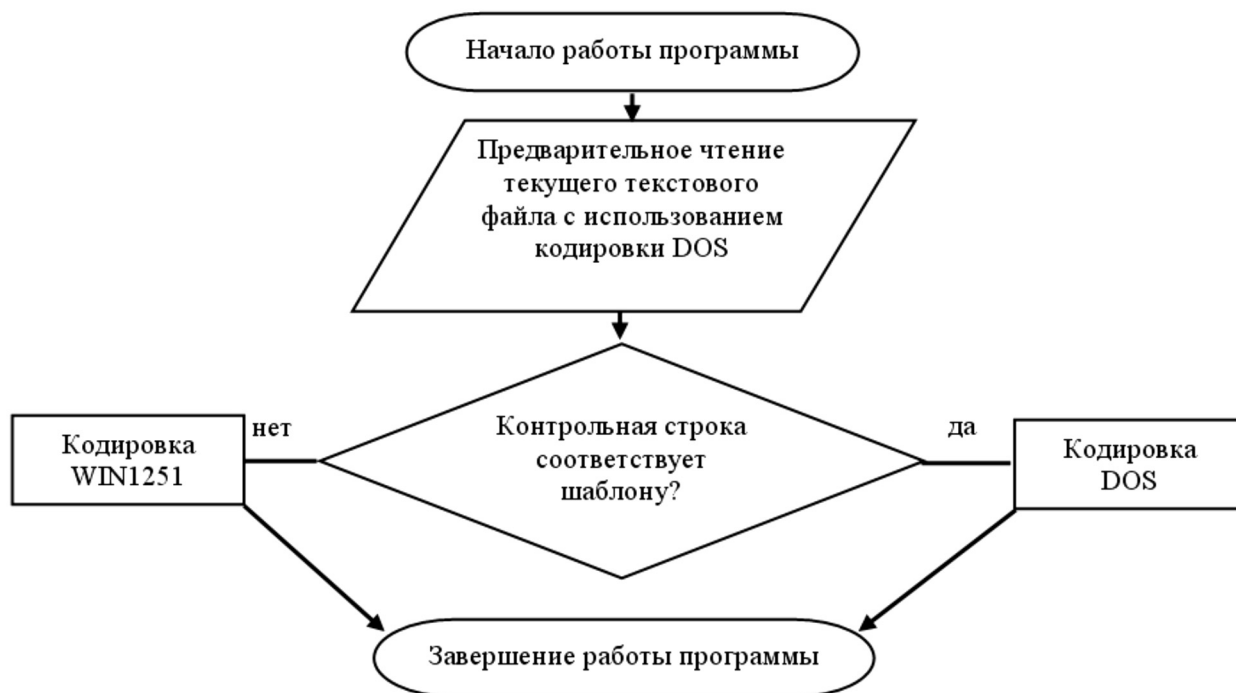


Рисунок 3. Блок-схема алгоритма определения кодировки текущего текстового файла.



Рисунок 4. Блок-схема алгоритма разработанного программного обеспечения.

Общий алгоритм чтения текстовых файлов архива ТМС-84 и создания на их основе электронных таблиц Excel

В итоге был разработан алгоритм, представленный на рис. 4. На основе данного алгоритма была написана программа на языке VisualBasic 2010с использованием платформы NET.Framework 4.5 [4].

На представленной блок-схеме проиллюстрированы основные этапы работы разработанного автором программного обеспечения, позволяющего оперативно переносить метеорологическую информацию из текстовых файлов архива ТМС-84 в электронные таблицы Excel.

В начале работы программы происходит получение списка всех текстовых файлов, содержащихся в директории архива ТМС-84. Затем для каждого текстового файла производится проверка его кодировки.

После этого производится перенос информации в электронные таблицы. При этом в начало каждой строки помещается соответствующие данные о метеостанции, дате и сроке наблюдения. При достижении последней строки текущего текстового файла, программа приступает к работе со следующим.

При достижении конца списка файлов директории архива ТМС-84, программа завершается.

Результаты работы

1. Данные архива ТМС-84 переведены в электронные таблицы с помощью разработанного автором программного обеспечения
2. Полученные электронные таблицы можно оперативно использовать в метеорологических исследованиях,

в том числе, для оценки качества модельных мезометеорологических полей для территории Иркутской области

3. Разработанное программное обеспечение будет внедряется в работу Иркутского гидрометцентра
4. Представленная методика может использоваться для архивов метеорологической информации из других регионов при условии сходной входной информации.

ЛИТЕРАТУРА

1. WRFmodel [Электронный ресурс] // The weather research and forecasting model URL: <http://www.wrf-model.org> (дата обращения: 10.10.2013).
2. Выбор между Accessи Excel [Электронный ресурс] // База знаний Microsoft. URL: <http://office.microsoft.com/ru-ru/training/RZ010253998.aspx?section=2> (дата обращения: 08.10.2013).
3. Блок-схема [Электронный ресурс] // Википедия URL: <http://ru.wikipedia.org/wiki/Блок-схема> (дата обращения: 10.10.2013).
4. Массивы в Visual Basic [Электронный ресурс] // База знаний MSDN. URL: <http://msdn.microsoft.com/ru-ru/library/wak0wfyf.aspx> (дата обращения: 06.10.2013).

© П.А. Найденов, (pavelnaidenov@geogr.isu.ru), Журнал «Современная наука: актуальные проблемы теории и практики»,



КАЗАНСКАЯ
ЯРМАРКА

23-я международная специализированная выставка

НЕФТЬ ⚡ ГАЗ ⚡ НЕФТЕХИМИЯ



16+

7-9
СЕНТЯБРЯ
Казань, 2016

Проектирование и строительство объектов для нефтяной, нефтехимической и газовой промышленности

При поддержке:
Правительства Республики Татарстан и Президента Республики Татарстан



ТАТАРСТАНСКИЙ
НЕФТЕГАЗОХИМИЧЕСКИЙ
ФОРУМ

деловая программа выставки - активная площадка, содействующая развитию науки и бизнеса

www.oilexpo.ru

РЕКЛАМА





Добыча нефти и газа




420059, г. Казань, Оренбургский тракт, 8
ОАО «Казанская ярмарка»
тел./факс: +7 (843) 570-51-14, 570-51-11
e-mail: d2@expokazan.ru, d3@expokazan.ru
www.expokazan.ru