

- на річкових ділянках Дунаю із завчасністю 10-30 діб // Наук. пр. УкрНДГМІ. – 2002. – Вип. 250. – С. 179-187.
3. Христюк Б.Ф. Прогнозування середніх, максимальних та мінімальних за декаду витрат води на Верхньому Дунаї // Наук. пр. УкрНДГМІ. – 2012. – Вип. 262. – С. 102-109.
  4. Дунай и его бассейн. Гидрологическая монография. – Л.: Гидрометеоздат, 1989. – Ч.1. – 396 с.
  5. Дунайська комісія / www.danubecommission.org.
  6. Розробка автоматизованої системи прогнозування рівнів води на судохідній частині р. Дунай. Звіт про НДР / УкрГМІ. – № д.р.0112U004678. – К., 2013. – 211 с.
  7. Лабор В.В. *Си Шарп*. Создание приложений для Windows. – Мн.: Харвест. 2003. – 384 с.
  8. Шилдт Г. Полный справочник по С#. – М.: Изд. дом «Вильямс», 2004. – 752 с.
  9. Руководство по гидрологическим прогнозам. Вып. 1. Долгосрочные прогнозы элементов водного режима рек и водохранилищ. – Л.: Гидрометеоздат, 1989. – 358 с.
  10. Руководство по гидрологическим прогнозам. Вып. 2. Краткосрочный прогноз расхода и уровня воды на реках. – Л.: Гидрометеоздат, 1989. – 246 с.
  11. Настанова з оперативної гідрології. Прогнози режиму вод суші. Гідрологічне забезпечення і обслуговування. – К.: Верлан, 2012. – 120 с.

Український науково-дослідний  
гідрометеорологічний інститут, Київ

УДК 551.501:681.3

**В.П. Евстигнеев, М.П. Евстигнеев, Н.И. Кульбида,  
В.А. Наумова, Н.И. Швень**

## **ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ СОЗДАНИИ БАЗЫ МЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ УКРАИНЫ**

Реализован первый этап создания национального архива метеорологических данных Украины на технических носителях с использованием современных информационных технологий. Разработана структура базы метео данных и осуществлен перенос данных из формата ЯОД в формат MS SQL Server. В процессе преобразования форматов выполнена корректировка ошибок данных исходного ЯОД архива.

**Ключевые слова:** метеорологические данные, ЯОД, MS SQL Server, контроль качества, климатическая однородность.

### **Введение**

В условиях современных тенденций изменения климата актуальным является вопрос связи гидрометеорологических параметров с глобальными климатическими процессами и, как

**Б.Ф. Христюк**

**Аналитико-экспертная система прогнозирования уровней воды судоходного Дуная «Истер»**

*На основе разработанных методик прогнозирования ежедневных, характерных декадных и месячных уровней воды судоходной части р. Дунай создано компьютерную программу «Аналитико-экспертная система прогнозирования уровней воды судоходного Дуная «Истер».*

**Ключевые слова:** уровни воды, прогностические зависимости, блок-схема, компьютерная программа.

**В.Ф. Khrystyuk**

**The analytical and expert system for the forecasting of the water levels of the navigable part of Danube «Ister»**

*On the basis of the developed methods of the forecasting of the daily and characteristic ten-day and monthly water levels of the navigable part of Danube the computer program «The analytical and expert system for the forecasting of the water levels of the navigable part of Danube «Ister» was created.*

**Keywords:** water levels, forecasting dependences, flow-chart, computer program.

ца (ICRC) 90 % стихийных бедствий связано с метеорологическими и климатическими факторами естественного происхождения [1]. Совершенствование методик прогноза опасных и стихийных явлений с целью снижения социально-экономических рисков от их возникновения и научные исследования в области климатологии требуют привлечения все большего объема информации о гидрометеорологических параметрах как на региональном уровне, так и в глобальном масштабе. По этой причине в мире большое внимание уделяется развитию систем комплексного анализа исторических рядов данных наблюдений за гидрометеорологическими элементами в отдельных регионах и современных массивов ретроспективных данных о состоянии климатической системы Земли как целого. Решение подобных задач связано с необходимостью обработки больших объемов данных с привлечением методов многомерного статистического анализа – задача, которая должна решаться с использованием современных вычислительных средств и программного обеспечения на основе оптимально структурированного архива гидрометеорологических данных.

Централизованное накопление гидрометеорологической информации в СССР началось в 1957 г. и было связано с организацией Фонда данных по гидрометеорологии и загрязнению природной среды (Постановление Совета Министров СССР № 1195 от 10.10.1957 г.) по предложению руководства Главного управления Гидрометеорологической службы. Уже тогда формировалось представление о ключевой роли вычислительной техники в архивации данных на технических носителях, их автоматизированной обработке и контроле качества [6]. Последовавшее развитие архивного фонда было сопряжено с развитием электронной вычислительной техники, а также с началом глобального мониторинга («спутниковая эра») состояния лито-, крио-, атмо- и гидросфер Земли. Процесс накопления данных осуществлялся централизованно, хотя и предполагал хранение части информации в территориальных Гидрометцентрах, в том числе в Украине [6]. Бурному накоплению гидрометеорологической информации также способствовал международный обмен между Мировыми центрами данных, один из которых располагался в бывшем СССР (ВНИИГМИ-МЦД, г. Обнинск, Россия). Помимо этих функций, ВНИИГМИ-МЦД осуществлял научно-методическое руководство автоматизированной обработкой информации, контролем качества, передачей и последующим ее хранением. Эта деятельность

включала разработку ряда специализированных наставлений и руководств, форматов хранения и систем управления архивной информацией. Хранение осуществлялось в формате ЯОД (Язык Описания Данных) [3, 4], а в качестве системы управления базой данных использовалась система АИСОРИ [5].

С распадом СССР (1991 г.) процесс накопления гидрометеорологических данных стал осуществляться, независимо друг от друга, Национальными гидрометслужбами стран постсоветского пространства, которые взяли на себя также и функцию научно-методического сопровождения всего цикла движения, преобразования и хранения гидрометеорологической информации. Взаимодействие с ВНИИГМИ-МЦД сводилось только к международному обмену. Однако именно с 90-х годов начался и продолжается по сей день стремительный рост качества и производительной мощности компьютерной техники, произошла смена парадигмы разработки программного обеспечения, баз данных и систем управления. Сложилась ситуация, когда созданные в советский период принципы архивации данных, форматы хранения, вся методическая основа этого вида деятельности и обеспечения потребителя довольно быстро устарели и потеряли свою эффективность, что было обусловлено следующим:

- быстрое развитие компьютерных технологий способствовало автоматизации цикла движения гидрометеорологических данных от первичной обработки данных наблюдений на гидрометеорологических станциях до полностью подготовленных к печати таблиц месячного, годового и многолетнего обобщения. В России автоматизированной системой обработки гидрометеорологических данных стал комплекс ПЕРСОНА [9]. В Украине с середины 90-х велись разработки нескольких программных приложений для обработки и обобщения данных гидрометеорологических наблюдений (например, программный комплекс для получения метеорологических ежемесячников [17]). Однако в настоящий момент наиболее динамичное развитие получил программный комплекс АССОКА (для обработки данных метеорологических, гидрологических и морских береговых наблюдений) [7];

- использование современных компьютерных технологий в архивных системах открывает новые возможности в повышении качества данных и их использовании;

- на основе накопленных данных продолжительных наблюдений (более 100 лет по отдельным станциям Украины) требуются более детальные

климатические исследования гидрометеорологического режима Украины и его изменения, в особенности, в условиях современных климатических тенденций [13, 15].

**Целью** работы является анализ текущего состояния архива метеорологических наблюдений на территории Украины и реализация первого этапа создания национального архива метеорологических данных Украины на технических носителях с использованием современных информационных технологий.

### Результаты и обсуждение

#### Объем архива данных метеорологических наблюдений Украины на технических носителях

Основным источником гидрометеорологических данных для украинской территории являются регулярные наблюдения, проводящиеся на сети станций и постов гидрометслужбы Украины. Материалы наблюдений хранятся в Отраслевом Государственном архиве (ОГА) Центральной геофизической обсерватории (г. Киев). Общее число станций на территории Украины, зарегистрированных в метеоархиве в разные периоды лет, составляет 242, из которых в настоящее время действующими являются 187. Для описания состава электронного архива на рис. 1а, б графически представлены объем информации как функция времени (рис. 1а) и распределение станций по длине архивных рядов данных (рис. 1б).

Следует отметить, что за последнее столетие количество пунктов наблюдений изменялось несколько раз. Первое существенное изменение их числа произошло в 1936 г. (см. рис. 1а), что было связано с расширением метеорологической сети по всей территории СССР. В эти же годы существенное изменение претерпела мето-

дика наблюдений – введен ночной срок и наблюдения стали круглосуточными (4-х срочные по местному времени). Вторая Мировая война внесла свои коррективы, уменьшив количество пунктов сети, однако, несмотря на это, общее количество действующих пунктов наблюдений все равно оставалось выше периода до 1936 г. В послевоенный период сеть постепенно расширялась, и очередной скачок развития сети пришелся на начало 60-х (см. рис. 1а). Тогда же (1966 г.) изменилась методика наблюдений, предусматривающая 8-ми срочные наблюдения по единому Московскому времени. В 80-х гг. был достигнут максимум как по количеству пунктов сети, так и по охвату территории. Период с 1985 г. и по настоящее время негативно сказался на состоянии гидрометеорологической сети бывшего СССР и, в последующем, стран постсоветского пространства, что выразилось в вынужденном сокращении количества пунктов наблюдений и отдельных видов работ.

В связи с указанными этапами развития гидрометеорологической сети распределение количества станций по длине архивных рядов данных оказалось сильно неравномерным (см. рис. 1б): большое количество станций имеют ряды данных длиной 50-55 лет (70 станций) и 70-75 лет (40 станций). Кроме того, в архиве имеются 11 станций, ряды данных которых имеют длину более 100 лет. Следует иметь в виду, что под длиной ряда понимается количество лет наблюдений, хранящихся в электронных архивах, причем эта длина может отличаться от реального периода функционирования станций. Ярким примером этого служит информация за 1976 год, представленная на рис. 1а. Как видно из рисунка, за данный год в электронном архиве отсутствуют

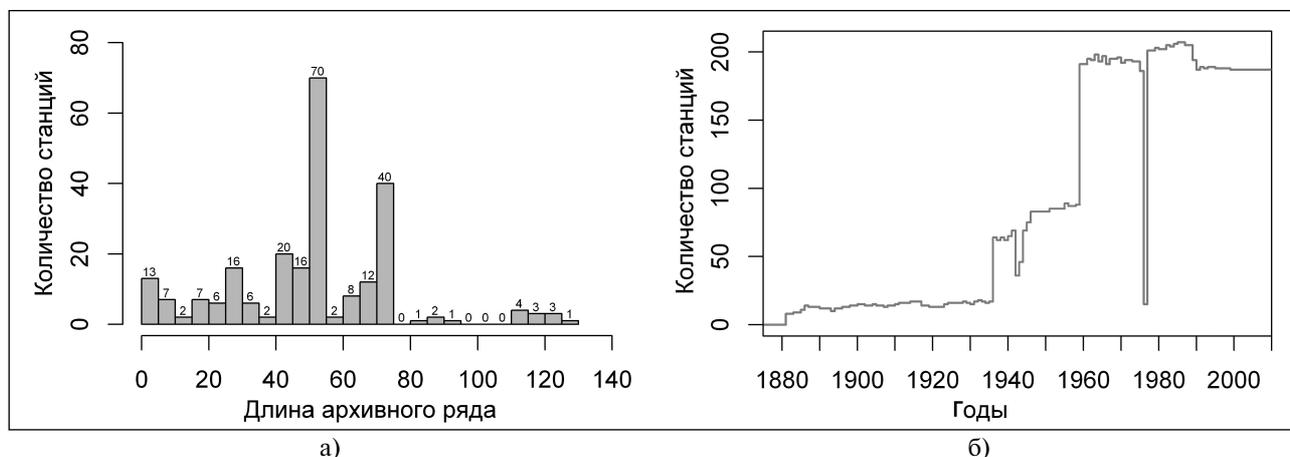


Рис. 1. Объем информации, содержащейся в метеорологическом архиве на технических носителях, как функция времени (а), и распределение числа станций в зависимости от длины архивного ряда (б)

данные по подавляющему большинству пунктов, что, вполне очевидно, не может быть связано с перерывом в наблюдениях. Другой иллюстрацией несоответствия указанных периодов является рис. 2, на котором представлены данные о полноте метеорологической информации для пунктов морской береговой сети Украины.

Как следует из рис. 2, информация по многим пунктам в электронном архиве отсутствует практически за половину исторического периода. По станции «Одесса-порт» она и вовсе отсутствует, несмотря на вековые наблюдения на этом пункте. Причины такого несоответствия периодов лет, имеющих в электронном архиве, и исторических массивов данных, по нашему мнению, кроются в следующем:

- исторические массивы данных на технических носителях начали формироваться с 60-х годов. На тот момент в силу технологических сложностей данного процесса занесение исторических данных до рассматриваемого периода было затруднительным. В таких условиях при-

оритет отдавался основным станциям оперативного обеспечения и международного обмена климатической информацией [6];

- отдельные пункты наблюдений (в особенности береговой зоны) за исторический период находились в подчинении различных ведомств, в том числе военных. В результате в период формирования архива на технических носителях у гидрометслужбы попросту не оказалось исходных данных метеорологических наблюдений.

Тем не менее, часть материалов на бумажных носителях (книжки наблюдений, месячные таблицы, ежегодные данные и т.д.) сохранилась, поэтому проблема восстановления архивных данных в электронном формате все еще остается актуальной и должна быть решена в рамках современных программных средств. Это, в свою очередь, требует приведения имеющихся электронных архивов к современным стандартам.

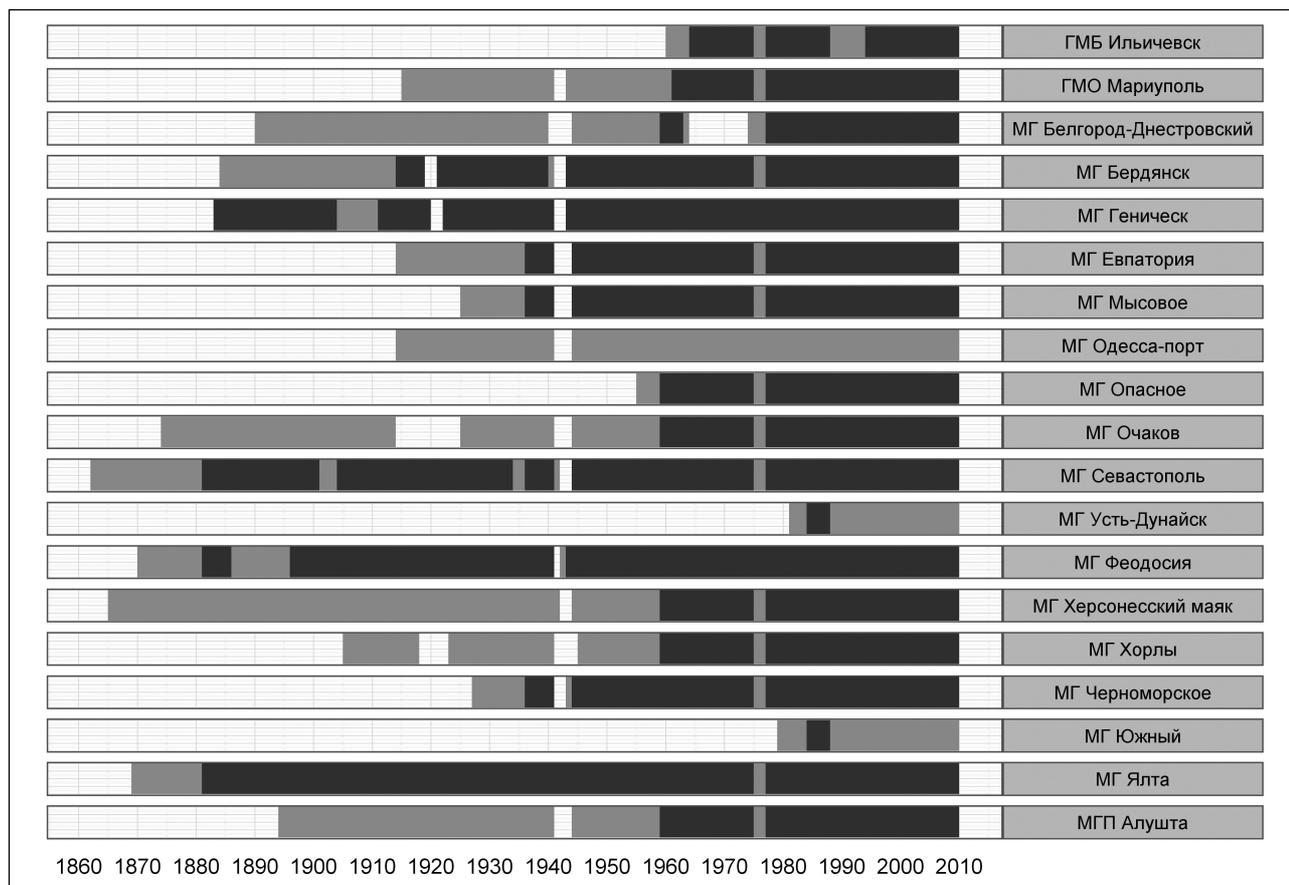


Рис. 2. Наличие информации в архиве на технических носителях и период функционирования пунктов морской береговой сети Украины (Темные полосы соответствуют периоду функционирования пунктов с учетом перерывов наблюдения; светлые полосы показывают объем информации, содержащийся в архиве на технических носителях)

Таблица 1

## Форматы хранения данных метеорологического архива

Данные	1881(1936)-1965	1966-1975	1976-1983	1984-1990	1991-2010
Срочные (только с 1936)	ТМ1-СРОКИ	ВОСХОД	ТММ1	ТММ1	ТМС
Суточные	ТМ1-СУТКИ	СУТКИ-76	—	ТМСС	ТМСС
Месячные	—	—	—	ТМСМ	ТМСМ

*Краткая характеристика текущего состояния архива данных метеорологических наблюдений Украины на технических носителях*

На протяжении всей истории существования гидрометслужбы стандарты хранения метеорологических данных на технических носителях неоднократно менялись. Это было обусловлено сменой методик наблюдения (1966 г.), изменениями самих технических носителей информации (перфокарты, магнитные ленты и т.д.) и способов ее обработки: на машиносчетных станциях до 1970-х годов, а в последующем – внедрение ЭВМ-ориентированных технологий [6]. В результате всех этих преобразований электронный метеоархив Украины представляет собой в значительной степени разнородный массив данных (табл. 1).

Исходные данные метеорологических наблюдений представлены на трех уровнях, а именно: срочного, суточного и месячного разрешения. Массивы срочных данных существуют в четырех различных базовых ЯОД-форматах в зависимости от периодов лет, в пределах которых существовала соответствующая схема архивирования данных наблюдений: ТМ1-СРОКИ, ВОСХОД, ТММ1, ТМС (см. табл. 1). При этом до 1976 г. данные наблюдений кодировались в текстовом формате, а после 1976 г. – в бинарном. Формат хранения и кодирования данных суточного и месячного обобщений в целом подобен базовым форматам срочных данных.

Физическое хранение осуществляется в виде отдельных файлов по схемам «один файл – одна станция за весь период лет» – для архивов ТМ1-СРОКИ, ТМ1-СУТКИ, ВОСХОД, СУТКИ-76 и «один файл – один месяц по всем станциям» – для архивов ТММ1, ТМС, ТМСС, ТМСМ. В отличие от суточных и месячных обобщений, данные срочных наблюдений присутствуют в метеоархиве, начиная с 1936 г., без глобальных перерывов, за исключением тех, которые указаны в табл. 2. Это, вообще говоря, позволяет восстанавливать суточные и месячные архивы путем перерасчета из срочных данных.

Таблица 2  
Глобальные пропуски в рядах данных метеорологического архива

Год	Месяцы	Отсутствуют данные
1978	10, 11	срочные и суточные
1983	6	срочные и суточные
1985	5	срочные
1993	9, 10	срочные

Краткий анализ текущего состояния метеоархива, а также проведенное нами детальное исследование качества имеющихся в метеоархиве архивных файлов и исходных базовых форматов позволяет выделить три основные сложности, неизбежно возникающие при использовании метеорологических данных в их нынешнем состоянии:

- разнородность форматов данных, существенно усложняющая произвольный доступ к данным;

- разнородность структуры и правил кодирования данных. Под этим подразумеваются различные правила кодирования некоторых гидрометеорологических элементов в разные периоды лет (например, группа атмосферных явлений, элементы влажности и пр.);

- ошибки формата и сбой записи в архивных файлах. Под этим подразумеваются нарушения внутренней структуры архивных файлов, наличие фрагментов «мусора» вместо данных наблюдений и прочие проблемы, возникшие, по всей видимости, в результате многократного перезаписывания архивных файлов с разных технических носителей (магнитные ленты, дискеты, жесткие диски).

Учитывая анализ текущего состояния метеоархива, проведенный выше, процесс создания архивной системы современного уровня в первом приближении может быть представлен в виде трех последовательных этапов:

- 1) создание структуры электронной базы данных и перевод имеющегося ЯОД-архива в формат базы данных;

- 2) заполнение имеющихся пропусков наблю-

дений и создание однородных рядов данных по всем станциям наблюдений;

3) создание современной системы управления базой данных.

В настоящей работе совместными усилиями специалистов ОГА и Севастопольской ГМО был реализован первый этап — создание единой структуры архивной системы на технических носителях и перевод метеорологической информации в формат базы данных.

#### *Реализация первого этапа создания современного метеоархива на технических носителях*

Принципиально важным для сохранения целостности архива и исключения потери данных при переборке в новый формат является корректировка ошибок формата и сбойных участков исходных архивных файлов. Наибольшее число проблемных ситуаций возникало с файлами архивов ВОСХОД (восстановлено 33 станции из 201), СУТКИ-76 (восстановлено 28 станций из 200), ТМ1-СУТКИ (восстановлено 30 станций из 220), ТМ1-СРОКИ (восстановлено 21 станция из 218), которые были созданы на основе перфокартотек. Как было отмечено в [2], со времени их создания и до ввода в ЭВМ перфокартотеки деградировали из-за старения основы (перфокарт) и хранения в неподходящих условиях. Для более поздних архивов ТМСС и ТММ 1 лишь отдельные файлы потребовали корректировки.

На основании имеющегося скорректированного набора архивных файлов была произведена разработка универсальной структуры базы данных метеорологических наблюдений, независимая от периода наблюдений и базового ЯОД-формата исходных данных. Для достижения этой цели была выполнена процедура согласования типа хранимых элементов и правил кодирования данных по всей вертикали имеющихся ЯОД-форматов, описанных выше и представленных в табл. 1. Данные каждого из массивов трех уровней (срочные, суточные, месячные) были распределены по таблицам соответствующих элементов наблюдений. Кроме того, структура предусматривает наличие таблицы паспортных данных пунктов наблюдений.

Разработанная универсальная структура метеоархива была реализована в современной базе данных MS SQL Server. Наличие унифицированной структуры позволило создать программу-загрузчик метеоданных из разнородных ЯОД-форматов (см. табл.1) в формат MS SQL Server и выполнить саму процедуру загрузки. В рамках этой программы также осуществляется

автоматическое пополнение метеобазы данными текущих наблюдений путем подключения к системе первичной обработки АССОКА (Метеорология) [8].

Полученный таким образом метеоархив в формате базы данных MS SQL Server инкапсулирует все функции современной клиент-серверной системы. В частности, реализует возможность произвольной выборки с любым числом входящих в запрос ограничений и наборов признаков, а также позволяет выполнять любые операции по обслуживанию самого архива. Однако использование электронной метеорологической базы данных возможно только при условии, если хранящиеся в ней данные обладают достаточным уровнем качества, в частности, содержат незначительное количество грубых ошибок [16, 18] и являются однородными. Только в этом случае можно заявлять о достоверности информации, получаемой на ее основе, например, при решении научно-прикладных задач. Ниже будет продемонстрировано решение задачи первичного контроля качества и однородности рядов данных в рамках созданной метеобазы.

#### *Первичный контроль качества данных архива метеорологических наблюдений*

Качество архивных данных зависит от погрешности измерительных приборов или их неисправности, ошибок метода измерений, интерполяции или осреднения во времени или пространстве. Стоит, однако, отметить, что на станциях и в отделах метеорологии региональных методических центров по гидрометеорологии при поступлении материалов текущих наблюдений осуществляется критический контроль качества данных [7], информация о результатах которого сохраняется в метеоархиве в виде признака качества при каждом значении метеорологической величины [6]. Кроме того, в архиве хранится информация о результатах автоматизированного контроля, проводившегося в вычислительном центре (ВНИИГМИ-МЦД) с появлением первых ЭВМ [6]. Тем не менее, методы контроля постоянно совершенствуются, становясь более эффективными. Не стоит забывать о том, что существенные ошибки могут быть внесены в процессе движения метеорологической информации на этапах ввода первичных данных, обработки, преобразования одних форматов хранения в другие (бумажные носители, перфокарты, магнитные ленты, диски, электронные форматы и т.д.). Таким образом, вопрос о качестве метеорологического архива актуален до сих пор.

Выделяют два типа ошибок, содержащихся

в архивах [16, 19]: единичные (изолированные в пространстве и/или во времени) и пакетные (систематически проявляющиеся в последовательности значений измеряемой величины). В настоящей работе мы сосредоточились на проведении контроля единичных ошибок, связанных с кратковременным действием внешнего фактора [16, 19]. Существует широкий спектр методов выявления и устранения этих типов ошибок, среди которых основное место занимают методы автоматизированного контроля [18]. В гидрометеорологической практике автоматизированный контроль подразумевает два метода: семантический и пространственный [18]. Семантические методы служат для обнаружения грубых ошибок, обусловленных влиянием случайных факторов, и применяются, как правило, к данным одной станции (внутристанционный контроль).

На основе созданной архивной системы нами проводится семантический контроль качества метеорологических данных, включающий как элементы логического контроля данных на непротиворечивость, так и статистические процедуры обнаружения грубых выбросов. При этом помимо известных тестов [16, 19] при контроле нами используются и современные статистические решения в этой области. Проиллюстрируем один из таких тестов на примере суточных данных о приземной температуре воздуха.

Понятие статистического выброса определяется диапазоном значений рассматриваемой метеорологической величины, попадание в который является практически невозможным событием. Границы этого диапазона выбираются эмпирически на основании знания статистических характеристик распределения величины [23].

Хорошо известно [11], что для температуры воздуха, как и для большинства метеорологических величин, характерна асимметрия распределения, знак которой (в смысле коэффициента асимметрии) меняется в зависимости от сезона. В связи с этим мы использовали метод, основанный на анализе специальных графиков «ящички с усами», которые служат для изображения одномерного распределения вероятностей [27]. Для построения такого типа графиков из эмпирического распределения величины определяют первый ( $Q_1$ ), второй ( $Q_2$  – медиана) и третий ( $Q_3$ ) квартили и межквартильное расстояние  $IQR = Q_3 - Q_1$  (рис. 3). Значения величины, попадающие внутрь интервала  $[c_1, c_2]$  (1), считаются относящимися к статистическому распределению; значения, выпадающие из этого интервала, считаются грубыми выбросами [27].

$$[c_1, c_2] = [Q_1 - 1,5IQR; Q_3 + 1,5IQR]. \quad (1)$$

Во многих работах было показано, что в случае распределений с выраженной асимметрией данный метод («ящик с усами») не обладает необходимой эффективностью, относя довольно большой процент значений к выбросам [20, 24]. Более надежным с этой точки зрения является метод приведенного «ящичка с усами» для асимметричных распределений [25]. Согласно этому методу по одномерной выборке  $\{x_1, \dots, x_n\}$  определяется робастный аналог коэффициента асимметрии распределения  $MC$  (*medcouple*) [21] по формуле:

$$MC = \underset{x_i \leq Q_2 \leq x_j}{med} \frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i}, \quad (2)$$

В зависимости от знака  $MC$ , а значит, от вида асимметрии, определяется допустимая область значений  $[c_1, c_2]$ :

$$[c_1, c_2] = \begin{cases} [Q_1 - 1,5e^{-4MC}IQR; Q_3 + 1,5e^{3MC}IQR], & MC > 0 \\ [Q_1 - 1,5e^{-3MC}IQR; Q_3 + 1,5e^{4MC}IQR], & MC < 0 \end{cases}. \quad (3)$$

На рис. 3б и 3в приведен пример построения обсуждаемых графиков для ноябрьских средне-суточных температур воздуха в Ялте. При сравнении рис. 3б и 3в отчетливо видно, что недоучет асимметрии распределения температур привел к переоценке отрицательных выбросов и полному отсутствию положительных. Эта разница примечательна и тем, что после коррекции допустимой области значений с использованием (3) (см. рис. 3в), положительных выбросов оказалось больше отрицательных.

Метод приведенного «ящичка с усами» был применен нами к архивным рядам суточных температурных характеристик приземного воздуха (средняя, максимальная, минимальная температуры). Задачей проверки было выявление суток с ошибочным значением температуры. Такая температура, либо будет совершенно невероятной, либо будет нехарактерной для выбранного дня в году или сезона для анализируемого пункта наблюдений. Таким образом, для поиска ошибки в исходную выборку должны войти температуры за данный календарный день, выбранные за все годы наблюдений на рассматриваемом пункте, вошедшие в архив. Тогда объем выборки будет ограничен длиной архивного ряда и для большого количества станций (около 60) составит не более 40 значений (см. рис. 1б). Для преодоления этой сложности рекомендуют [23, 26] дополни-

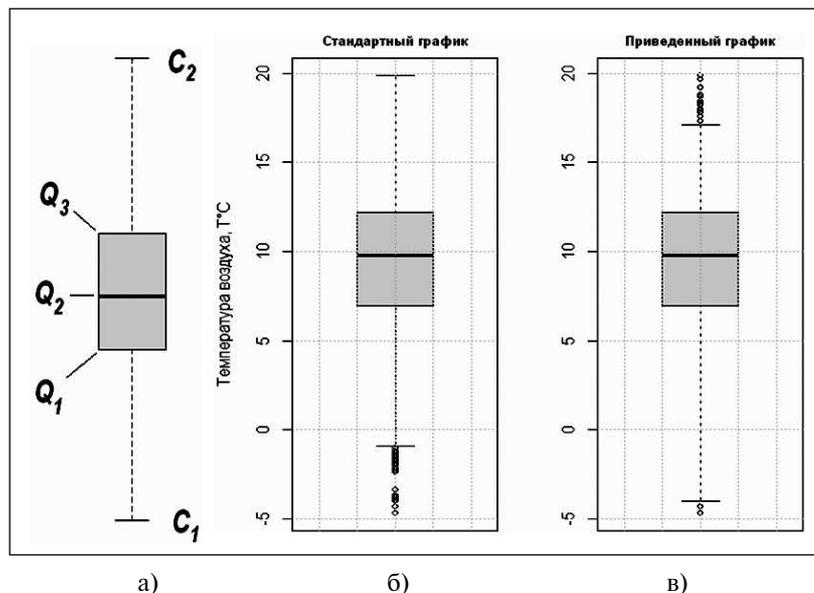


Рис. 3. График типа «ящик с усами» для ноябрьских температур воздуха в Ялте: а) схематичное представление графика «ящик с усами» (пояснение в тексте); б) стандартный график с использованием интервала (1) [27]; в) приведенный график с учетом асимметрии распределения (3) [25]

тельно включать в выборку данные по температурам за предыдущие и последующие сутки ( $\pm 3$  [26],  $\pm 4$  [22],  $\pm 5$  [23]). Предполагается, что искажение информации о проверяемом календарном дне несущественно, вследствие незначительного отличия среднесуточных значений температур смежных суток (и их изменчивости) [26]. Зато объем выборки, а, значит, надежность процедуры контроля, вырастает в разы. При этом в каждом тесте мы исключали контрольные сутки для того, чтобы минимизировать влияние потенциального выброса на результат расчетов.

Тестирование указанным выше методом проводили для пунктов с длиной архивного ряда, превышающей 5 лет (исключены 13 пунктов), причем значения, помеченные в метеобазе как забракованные или сомнительные, исключались из расчетов. Общее количество записей, выявленных тестом, составило 1,2 %, 1,5 % и 1,3 % от общего объема присутствующей в архиве информации о среднесуточных, максимальных и минимальных температурах соответственно. Стоит отметить, что этот тест является сугубо статистическим. Практически независимо от жесткости накладываемых условий существует вероятность, что грубый выброс, выявленный тестом, на самом деле окажется экстремальным значением температуры воздуха, ставшей таковой по вполне естественным причинам. В этой связи полученные значения температуры являются лишь

потенциальными выбросами и требуют дополнительной проверки специалистом. Тем не менее, полученные с использованием (3) температуры воздуха  $c_1$  и  $c_2$ , для каждой метеорологической станции определяют границы допустимых значений этого элемента и могут служить для оценки качества и наличия выбросов в текущих данных, поступающих с этих станций.

### Контроль однородности рядов данных наблюдений

Контроль качества данных метеорологической базы является относительно тривиальной задачей, которая может быть также решена в рамках любой системы обработки первичных данных наблюдений, таких как ПЕРСОНА или АССОКА, а также архивной системы АИСОРИ. Более сложной задачей, по нашим сведениям, не решаемой в рамках указанных систем, является контроль однородности рядов данных наблюдений.

Как известно [10], климатологически однородным считается ряд, межгодовая изменчивость элементов которого определяется естественной изменчивостью макропроцессов, формирующих погоду и климат данного района. Климатологическая однородность может быть нарушена такими факторами как: изменение местоположения станции, изменение растительных ландшафтов, застройка, изменение методики и условий наблюдений и прочее.

Для проверки рядов данных гидрометеорологических элементов на однородность в работе был использован метод (Wang et al.), практикующий пенализированный максимальный  $t$ - и  $F$ -тесты для поиска скачкообразных сдвигов в среднем значении ряда [28] и реализованный в виде рекурсивного алгоритма в библиотеке RhtestsV3 программной среды R Software (version 2.15.2) ([www.r-project.org](http://www.r-project.org)), подключаемой к метео-базе через интерфейс ODBC. Данный алгоритм является одним из наиболее эффективных в настоящее время для детектирования множественных скачкообразных сдвигов и был ранее успешно применен для выявления неоднородностей в многолетних рядах данных различных гидрометеорологических элементов.

Внедрение алгоритма (Wang et al.) позволяет проводить не только детальный анализ однородности рядов метеорологических элементов, но и устранение неоднородностей. С использованием созданной метеобазы данных, в сочетании с

указанной процедурой, мы впервые выполнили контроль климатологической однородности скорости ветра на морской береговой сети Украины (23 пункта) за весь доступный исторический ряд наблюдений. Для этого использовались ряды данных средней скорости ветра, суточного и месячного разрешений, предварительно пересчитанные на стандартную высоту 10 м, с учетом корректировки значений измеряемой флюгером скорости ветра при переходе от флюгера к анеморумбометру (в период 60-70-х годов 20 века) по методике Е.В. Майстрюковой [12]. В рядах данных суточного разрешения до середины 50-х годов было выявлено большое число сдвигов в средних значениях скорости ветра, которые не во всех случаях могли быть интерпретированы из-за отсутствия полной информации о ветроизмерительных приборах, высоте их установки в разные периоды, защищенности и др. Ряд наблюдений, начиная с середины 50-х годов, дает вполне объяснимый набор сдвигов, обусловленный в большинстве случаев такими факторами как перенос пункта наблюдения и изменение степени защищенности метеоплощадки. Информация о дате и величине выявленных сдвигов позволила выполнить процедуру приведения рядов скорости ветра к однородным, выполненную также средствами библиотеки RНtestsV3, сопряженной с метео-базой данных. Наличие однородных рядов открыло перспективу проведения полноценного исследования многолетней изменчивости характеристик ветра и анализа ветрового режима Азово-Черноморского региона Украины.

В настоящий момент работа по приведению рядов данных продолжается для тех метеорологических элементов (температура воздуха, осадки, давление и пр.), к которым эта процедура применима. При этом, для обеспечения контроля результатов трансформации исходных рядов на всех этапах, важной отличительной особенностью созданной метеорологической базы является возможность сохранения в ней как первичных данных наблюдений, так и приведенных к однородным для обеспечения контроля результатов трансформации исходных рядов на всех этапах.

### Выводы

В настоящей работе представлен итог реализации первого этапа создания национального архива метеорологических данных Украины с использованием современных информационных технологий, в ходе которого проведена корректировка ошибок исходного ЯОД-архива, создана структура базы данных и произведен перенос данных из формата ЯОД в формат MS SQL Server. Следующим этапом является создание специализированной системы управления базой

данных как основы метеорологического архива на технических носителях.

Работа выполнена при частичной поддержке гранта Президента Украины для поддержки научных исследований молодых ученых № GP/F49/110.

\* \*

1. Всемирная Метеорологическая Организация. Бюллетень. – Женева: Секретариат ВМО, 2006. – Т. 55, № 1. – С. 21-30.
2. *Быховец С.С.* История наблюдений за температурой почвы на сети метеорологических станций России // Криосфера Земли. – 2007. – Т. XI, № 1. – С. 7-20.
3. *Веселов В.М.* Язык описания гидрометеорологических данных // Тр. ВНИИГМИ-МЦД. – 1978. – Вып. 43. – С. 3-30.
4. *Веселов В.М.* Язык описания гидрометеорологических данных для IBM PC совместимых ПЭВМ // Тр. ВНИИГМИ-МЦД. – 1995. – Вып. 160. – С. 41-54.
5. *Веселов В.М.* Структура и информационные технологические баз и банков архивных данных Госфонда (1976-2004 гг.) // Тр. ВНИИГМИ-МЦД. – 2007. – Вып. 172. – С. 38-59.
6. *Григорьев В.И.* Автоматизированная обработка гидрометеорологической информации. – Л.: Гидрометеоиздат. – 1979. – 303 с.
7. *Евстигнеев В.П., Наумова В.А., Евстигнеев М.П.* Комплекс автоматизированных систем сбора, обработки, контроля, анализа и хранения данных гидрометеорологических наблюдений – АССОКА // Системы контроля окружающей среды. – Севастополь: МГИ НАН Украины. – 2009. – С. 157-162.
8. *Евстигнеев М.П., Евстигнеев В.П., Наумова В.А., Лантушенко О.О.* Свідчення про реєстрацію авторського права на твір №23207, дата реєстрації 01.04.2009.
9. *Зубрицкая Е.Н.* Опыт разработки программного обеспечения систем ПЕРСОН (МИС, МИП, БЕРЕГ) в среде Windows // Тр. ВНИИГМИ-МЦД. – 2010. – Вып. 174. – С. 5-6.
10. *Кобышева Н.В., Навролянский Г.Я.* Климатологическая обработка метеорологической информации. – Л.: Гидрометеоиздат, 1978. – 295 с.
11. *Костин С.И., Покровская Т.В.* Климатология. – Л.: Гидрометеоиздат. – 1953. – С. 269-276.
12. *Мастрюкова Е.В.* К методике расчета экстремальных скоростей ветра // Тр. ГГО. – 1987. – Вып. 515. – С. 36-41.
13. *Наумова В.А., Евстигнеев М.П., Евстигнеев В.П., Любарев Е.П.* Ветро-волновые условия Азово-Черноморского побережья Украины // Тр. УкрНИГМИ. – 2010. – Вып. 259. – С. 263-283.
14. *Николаев Е.А., Шаймарданов В.М.* Развитие архивной системы Росгидромета // Тр. ВНИИГМИ-МЦД. – 2010. – Вып. 174. – С. 3-10.
15. *Полонский А.Б.* Роль океана в изменениях климата. – К.: Наук. думка. – 2008. – 183 с.
16. *Пуголовкин В.В., Степаненко С.Р., Шаймарданов М.З.* О выявлении систематических ошибок в ме-

- теорологических архивах данных // Метеорология и гидрология. — 1996. — №1. — С. 58-67.
17. Звіт про НТР «Розроблення нових та удосконалення існуючих методів обробки й аналізу метеорологічної інформації, прогнозу атмосферних процесів та метеорологічних явищ». — 2007. — УкрНДГМІ. — С. 8-10.
  18. Рекомендации по анализу результатов пространственного контроля режимной метеорологической информации. — С. Пб.: Гидрометеиздат. — 1993. — 176 с.
  19. Степаненко С.Р., Воронцов А.А. Оценка качества массива суточных значений температуры воздуха // Тр. ВНИИГМИ-МЦД. — 2010. — Вып. 174. [http://www.meteo.ru/publish\\_tr/trudy174/st15.htm](http://www.meteo.ru/publish_tr/trudy174/st15.htm).
  20. Brys G., Hubert M., Rousseau P.J. A Robustification of Independent Component Analysis // J. Chemometrics. — 2005. — V.19. — P. 1-12.
  21. Brys G., Hubert M., Struyf A. A robust measure of skewness // Journal of Computational and Graphical Statistics. — 2004. — V. 13. — P. 996-1017.
  22. Domonkos P. Temporal accumulations of extreme daily mean temperature anomalies // Theor. Appl. Climatol. — 2001. — V. 68. — P. 17-32.
  23. Haylock et al. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006 // J. Geophys. Res. — 2008. — V. 113. — D20119.
  24. Hoaglin, D.C., Mosteller, F., Tukey, J.W. Understanding Robust and Exploratory Data Analysis. — New York: Wiley. — 1983. — P. 58-77.
  25. Hubert M., Vandervieren E. An adjusted boxplot for skewed distributions // Computational Statistics and Data Analysis. — 2008. — V. 52. — P. 5186-5201.
  26. Kalkstein L.S., Valimont K.M. An evaluation of summer discomfort in the United States using a relative climatological index // Bull. Amer. Meteor. Soc. — 1986. — V.67. — P. 842-848.
  27. Tukey J.W. Exploratory Data Analysis. — Massachusetts: Reading (Addison-Wesley). — 1977. — P. 39-49.
  28. Wang X. L., Wen Q. H., Wu Y. Penalized maximal test for detecting undocumented mean change in climate data series // J. Appl. Meteor. Climatol. — 2007. — V. 46. — № 6. — P. 916-931.

*Севастопольская гидрометеорологическая обсерватория  
Український гідрометцентр, Київ  
Центральная геофизическая обсерватория, Киев*

**Євстігнєєв В.П., Євстігнєєв М.П., Кульбіда М.І., Наумова В.А., Швень Н.І.**

**Використання сучасних інформаційних технологій у ході створення бази метеорологічних даних України**

*Реалізовано перший етап створення національного архіву метеорологічних даних України на технічних носіях з використанням сучасних інформаційних технологій. Розроблено структуру бази даних і здійснено перенесення даних із формату МОД у формат MS SQL Server. У процесі перетворення форматів виконано коригування помилок даних початкового архіву МОД.*

**Ключові слова:** метеорологічні дані, МОД, MS SQL Server, контроль якості, кліматична однорідність.

**Evstigneev V.P., Evstigneev M.P., Kulbida N.I., Naumova V.A., Shven N.I.**

**Application of the modern informational technologies to create meteorological database of Ukraine**

*The first stage of creation of national meteorological archive of Ukraine on technical medium has been accomplished using modern informational technologies. The structure of database has been developed and transfer of DDL data to MS SQL Server format has been done. Errors correction of initial data was carried out during formats transformation.*

**Keywords:** meteorological data, DDL, MS SQL Server, quality control, climatic homogeneity.