

В.М. Веселов

**Анализ пропусков
в метеорологических и гидрологических
архивах по территории России.
Нормализованные ЯОД- архивы**

<http://aisori.meteo.ru/aisori>

Апрель, 2007 г.

Источники формирования нормализованных ЯОД-архивов

Срочные наблюдения, архивы ручной подготовки

<i>TM0_SR</i>	<i>3-срочные данные</i>	<i>1878-1935</i>	<i>300</i>	<i>1.01</i>	<i>77</i>
<i>TM1_SR</i>	<i>4-срочные данные</i>	<i>1936-1965</i>	<i>3471</i>	<i>9.36</i>	<i>591</i>
<i>ВОСХОД</i>	<i>8-срочные данные.</i>	<i>1966-1976</i>	<i>3429</i>	<i>11.81</i>	<i>893</i>
<i>ВОСХОД-77</i>	<i>8-срочные данные</i>	<i>1977-1983</i>	<i>3628</i>	<i>9.36</i>	<i>671</i>

Суточные обобщения, архивы ручной подготовки

<i>TM1_SU</i>	<i>Суточные данные</i>	<i>1878-1965</i>	<i>3388</i>	<i>2.77</i>	<i>207</i>
<i>СУТ-66</i>	<i>Суточные данные</i>	<i>1966-1976</i>	<i>3423</i>	<i>1.31</i>	<i>128</i>
<i>СУТ-77</i>	<i>Суточные данные</i>	<i>1977-1983</i>	<i>3629</i>	<i>1.04</i>	<i>191</i>

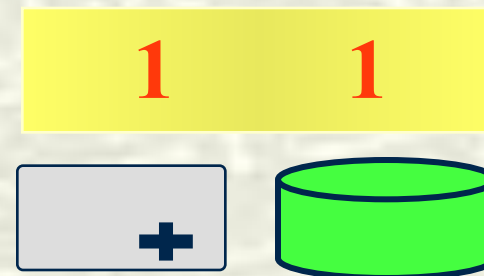
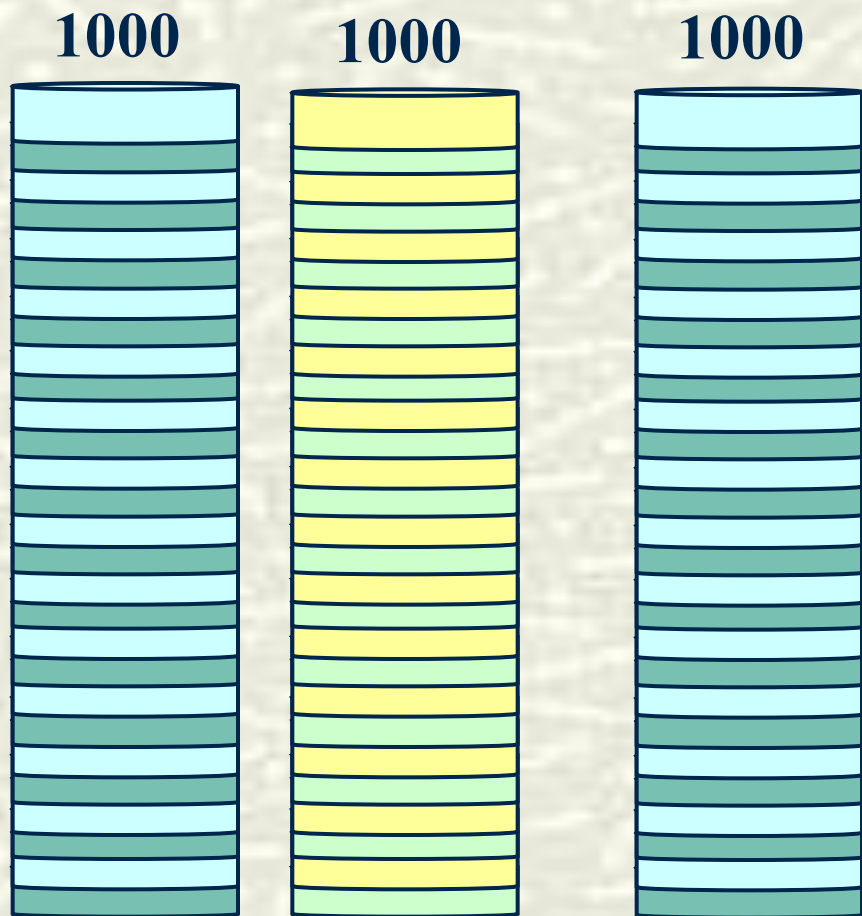
Данные автоматизированных систем первичной обработки

<i>TMM1</i>	<i>Наблюдения на станциях</i>	<i>1977-1984</i>	<i>3593</i>	<i>11.27</i>	<i>2155</i>
<i>TMC</i>	<i>Наблюдения на станциях</i>	<i>1984-2005</i>	<i>3649</i>	<i>18.64</i>	<i>955</i>
<i>TMCC</i>	<i>Суточные обобщения</i>	<i>1984-2005</i>	<i>3648</i>	<i>2.44</i>	<i>335</i>
<i>TMCM</i>	<i>Месячные обобщения</i>	<i>1984-2005</i>	<i>3648</i>	<i>1.30</i>	<i>333</i>

Всего:
08/13/2023

70.316536

Среда хранения и среда обработки архивов на примере архива ТМС–Текущая метеорология станций (1984 – 2007 гг.)



Хранение 1985-1993гг.
Обработка
08/13/2023

Хранение 1998 г.
Обработка

Хранение 2005 г.
Обработка

Основные недостатки базовых архивов

- наличие структурных дефектов, появившихся как во время хранения, так и вследствие дефектов технологий сбора;
- наличие дублей (полных и, еще хуже, частичных) как на уровне файлов, так и внутри файлов;
- наличие существенных пропусков в данных;
 - отсутствие сведений о полноте и пропусках в данных
- недостаточное качество данных,
- неудобный способ упорядочения по пространственно-временным признакам,
ориентированный на технологии сбора, а не на задачи обслуживания;
- необеспеченность эффективного доступа (за счет хранения на лентах и картриджах)

Дефекты структуры базовых архивов

- Потеря или появление лишних байтов, Инородные включения
- Нарушение структуры в смысле файловой системы ОС
- Ошибки значений счетчиков и длин
- Выход ключей за границы
- Нарушение постанционной структуры файлов
- Выход значений за естественные границы
- Ошибки формата элементов (буквы вместо цифр и т.п.)

Нормализованные архивы

- Нормализованные архивы получают из базовых архивов путем тщательного тестирования, исправления структурных дефектов, реструктуризации и упорядочения, удаления дублей, устранения пропусков и других процедур, повышающих их качество
- Для различных базовых архивов набор применяемых процедур может быть различным
- Нормализованные архивы сопровождаются подробным описанием выполненных процедур и общей оценкой полноты и качества данных
- Нормализованные архивы сохраняют содержание базового архива
- Нормализованные архивы используются как отправная точка при формировании всех других производных массивов и баз данных, а также при обслуживании потребителей
- Нормализованные архивы служат основой построения эффективной системы доступа к режимным данным

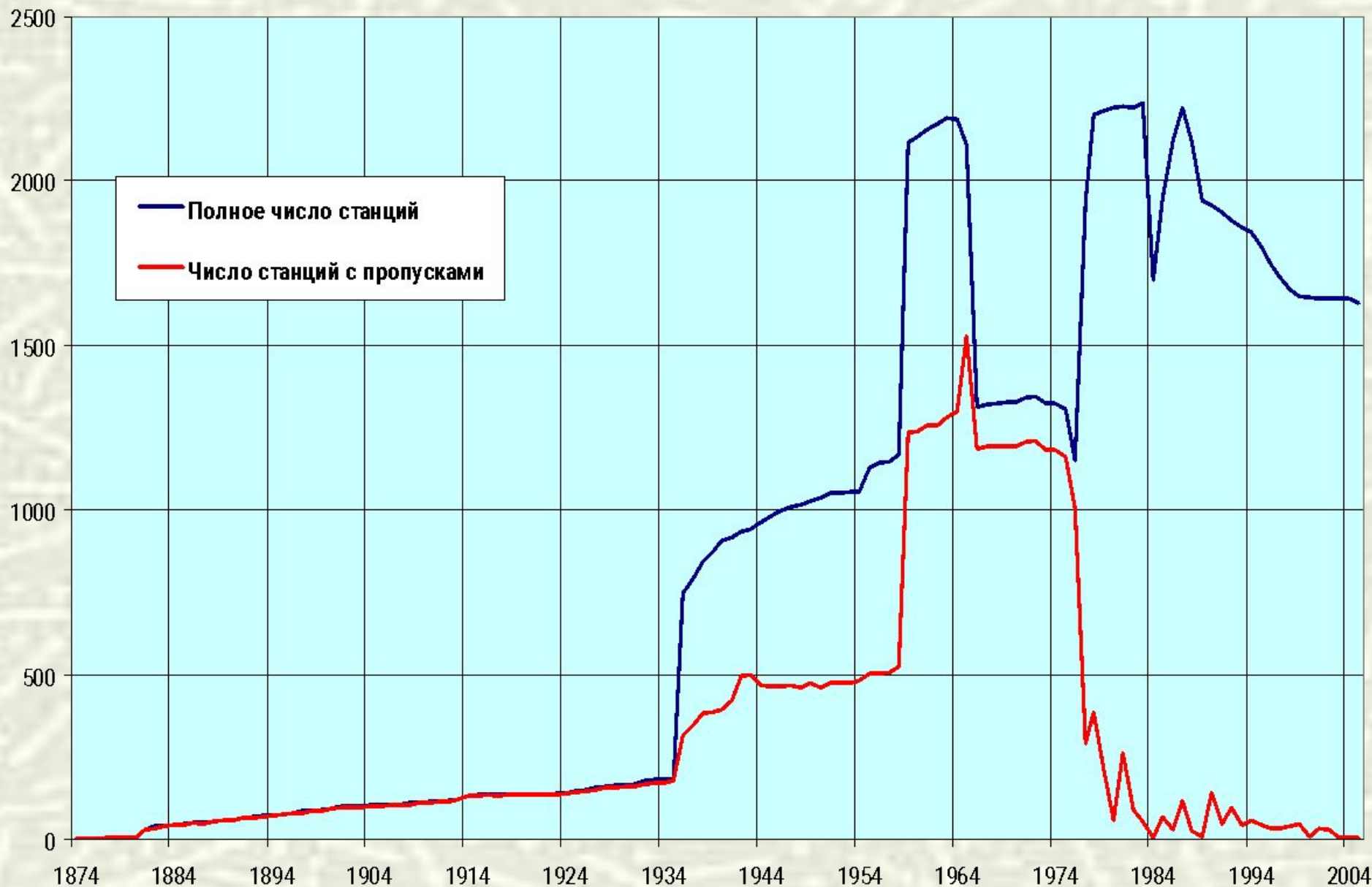
Нормализованный архив – это не просто производный массив, полученный применением определенных процедур к базовому архиву, а производный массив, наделенный статусом

Метеорологические станции России

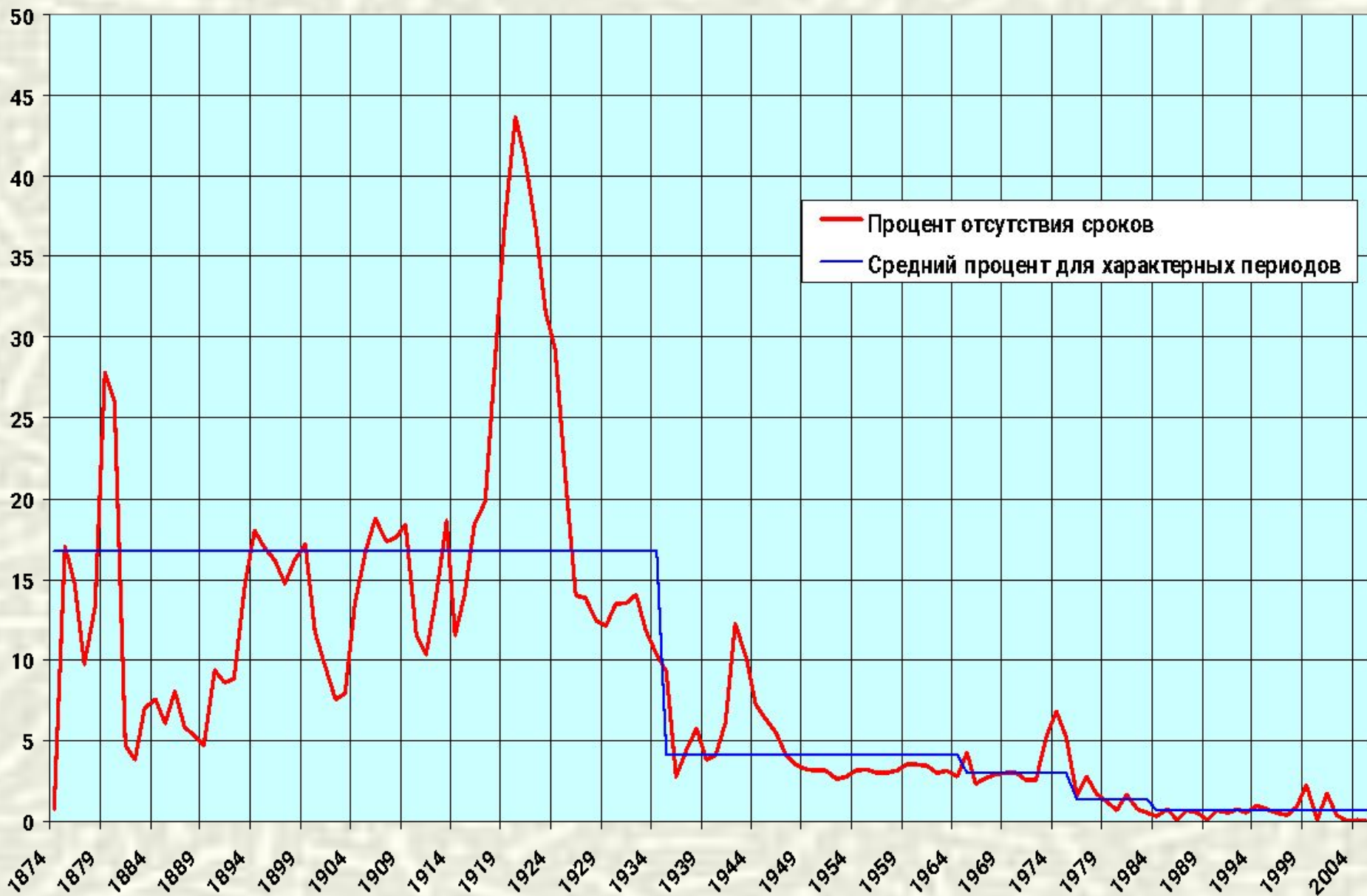


08/13/2023

Число метеостанций России в срочных архивах Госфонда по годам

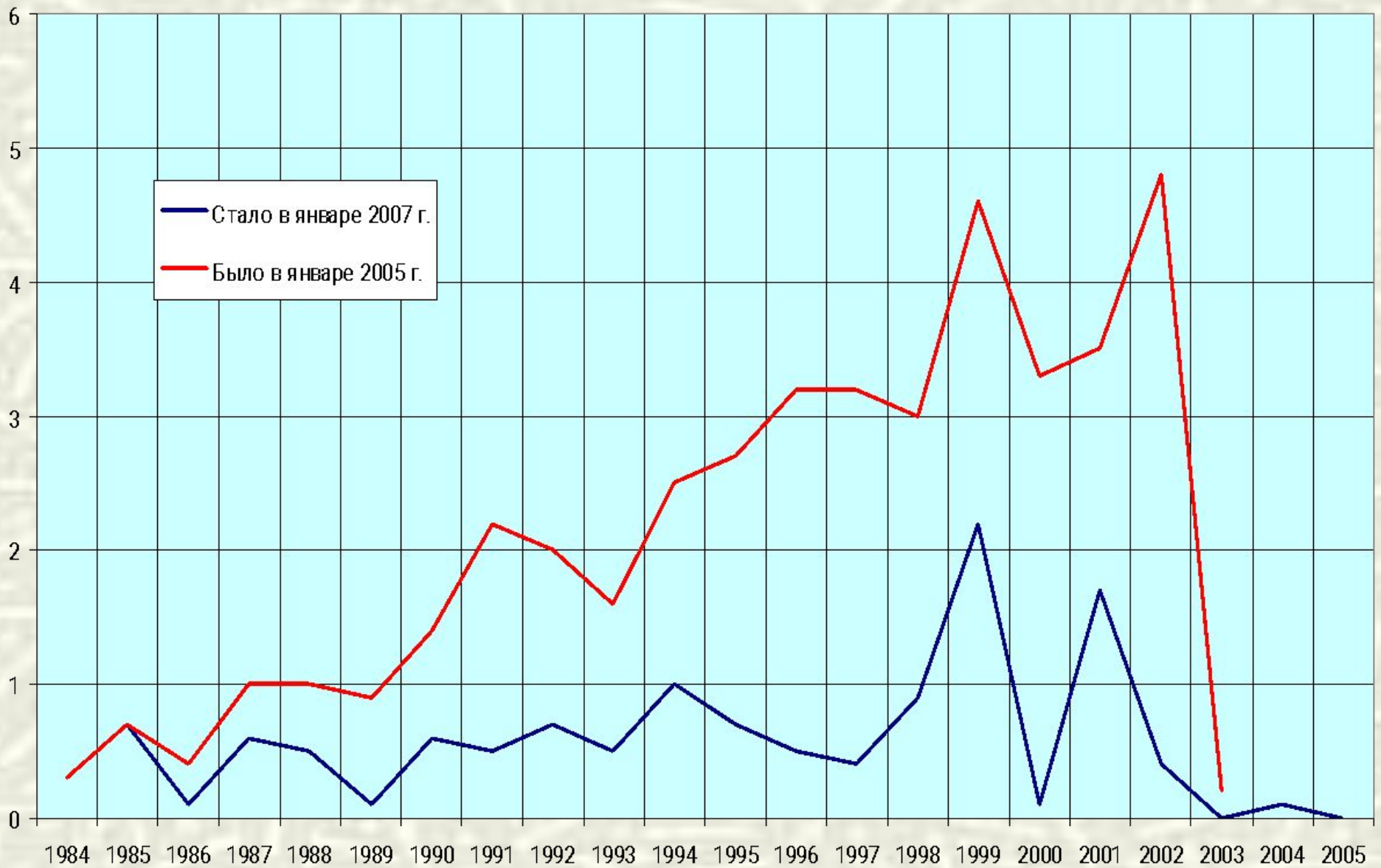


Пропуски в срочных данных для метеостанций России в массивах Госфонда



08/13/2023

Сравнение пропусков в срочных данных в 2005 и 2007 гг.



Полнота и отсутствие данных по станциям

19 - 9.3. Полнота по срокам (нет ни одного из осн. эл-тов)

№ станц	Название станции	Начало наблюдений		Окончание наблюдений		Число сроков		Полнота, %
		Год	Месяц	Год	Месяц	Расчетное	Отсутствуют	
5614040	Владимир, АМСГ	1986	1	2003	12	52592	3648	93
5453620	Калуга, Грабцево	1984	1	2003	12	58440	3	99
5593760	Москва, ВДНХ	1984	1	2003	12	58440	240	99
5463980	Рязань, смсоево, АС	1986	1	2003	12	52592	3648	93
5483210	Смоленск	1984	1	2003	12	58440	0	100
5693590	Тверь	1984	1	2003	12	58440	25	99
5423770	Тула	1984	1	1987	12	11688	0	100
5784000	Ярославль	1986	1	2003	12	52592	3168	93

9 : 1 Запрошено станций: 8 Найдено: 8

20 - 9.4. Пропуски сроков (нет ни одного из осн. эл-тов)

№ станц	Название станции	Год	Месяц	День	№ ср	Тип отс	Отсутствующие сроки или дни
5614040	Владимир, АМСГ	1990	9		240	Месяц	
		1992	4- 5		488	Месяцы	
		2002	1-12		2920	Год	
5453620	Калуга, Грабцево	2003	12	8	3	Сроки:	06 09 12
5593760	Москва, ВДНХ	1997	6		240	Месяц	
5463980	Рязань, смсоево, АС	1990	9		240	Месяц	
		1992	4- 5		488	Месяцы	
		2002	1-12		2920	Год	
5693590	Тверь	1990	11	18	1	Сроки:	21
		1990	11	19	3	Сроки:	00 03 06
		1998	12		16	Дни:	26 27
		1998	12	25	2	Сроки:	15 18
		1998	12	28	3	Сроки:	00 03 21
5784000	Ярославль	1992	5		248	Месяц	
		2002	1-12		2920	Год	

12 : 24 Запрошено станций: 8 Найдено: 6

Спасибо за внимание

08/13/2023