If I Had Not Believed It
I Would Not Have Seen It

T Cohn, D Koutsoyiannis, H Lins and A Montanari

H. E. Hurst

Courtesy S. Hurst, 2013

The Beginning: Hurst [1951]

AMERICAN SOCIETY OF CIVIL ENGINEERS

Founded November 5, 1852

TRANSACTIONS

Paper No. 2447

LONG-TERM STORAGE CAPACITY OF RESERVOIRS

BY H. E. HURST[1]

WITH DISCUSSION BY VEN TE CHOW, HENRI MILLERET, LOUIS M. LAUSHEY, AND H. E. HURST.

SYNOPSIS

A solution of the problem of determining the reservoir storage required on a given stream, to guarantee a given draft, is presented in this paper. For example, if a long-time record of annual total discharges from the stream is available, the storage required to yield the average flow, each year, is obtained by computing the cumulative sums of the departures of the annual totals from the mean annual total discharge. The range from the maximum to the minimum of these cumulative totals is taken as the required storage.

The results of the investigations given here were applied in computing the storage required in the Great Lakes of the Nile River Basin, the length of time chosen as a basis of the estimates of storage being 100 years. These reservoirs, in combination with the other projects described, would enable the Nile to be developed for irrigation to the fullest possible extent. Finally, a relation has been derived between the storage capacity on the main stream and the amounts required on the tributaries to produce its equivalent.

It is thought that the general theory may have other applications than the design of reservoirs for the storage of water.

1. INTRODUCTION

A number of projects has been investigated which will enable Egypt and the Sudan to develop irrigation from the Nile River to its fullest extent.[2,3] The broad plan involves storage of water from good years for use in bad ones, and necessitates reservoirs of sufficient capacity to meet the shortages that might occur during a century. Much research had been done on the capacity

NOTE.—Published in April, 1950, as *Proceedings-Separate No. 11*. Positions and titles given are those in effect when the paper or the discussion was received for publication.

[1] Scientific Consultant to the Ministry of Public Works, Egypt; formerly Director-Gen., Physical Dept., Cairo, Egypt.

[2] "The Nile Basin," Vol. VII, by H. E. Hurst, R. P. Black, and Y. M. Simaika, *Physical Department Paper No. 51*, Ministry of Public Works, Cairo Govt. Press, Cairo, Egypt, 1946.

[3] *Civil Engineering and Public Works Review*, September, 1948.

of such reservoirs and the results were used to determine the size of the reservoirs required in the Great Lakes of the Nile Basin for what was called "century storage."

Some preliminary results of investigations were released in 1938[4] which were amplified in 1946;[2] but in neither case was there an adequate statement of the theoretical investigation. The latter has been reserved for this paper.

Storage problems were treated first by the late Allen Hazen,[5] M. ASCE, in 1914, when he studied the discharge of thirteen American rivers, dealing with them mainly in a graphical manner, on probability paper. The longest record was one of 45 years, but longer sets of observations were made by combining records from several rivers. The objection to this procedure and the uncertainty in extending the normal Gaussian frequency distribution to the future were recognized by Mr. Hazen, but he thought it gave the best results then obtainable.

The work was extended by the late Charles E. Sudler,[6] M. ASCE, by similar graphical methods. Much of Mr. Sudler's work was based on artificial records in which information taken from a short record of a stream was extended to a period of 1,000 years by writing, say, 50 annual runoff values on cards and then shuffling and drawing a card from these 1,000 times. Again the actual data were necessarily confined to short periods since stream measurements in the United States had rarely been made for as long as 50 years. Thus, the results are mainly based on a few short-period records of natural river flow extended to long periods by random repetition. It is clear that the process can give very little information about what may happen over long periods and may be completely misleading since it excludes values higher or lower than those which have already occurred. This was recognized by some of those who contributed to the discussion on Mr. Sudler's paper. Moreover, the graphical methods employed do not lead to any concise and easily understandable presentation of results.

In the present investigation the methods of the theory of probability are applied, but the data are long-period records of natural events which are shown to have certain similarities, although they may be records of river discharge, rainfall, temperature, annual growth rings of trees, or annual deposits of clay in lakes. The average results can be expressed quite simply by two equations which apply to the storage of other substances as well as water. It is shown that, in regard to storage, records covering such short periods as 30 years or 40 years may be very misleading, but their usefulness can be extended by analyzing a large number of phenomena which have been recorded for periods as long as a century or more.

The problem of what storage is required on a stream to give a certain minimum discharge was first investigated in the case of Lake Albert (in the Belgian Congo and the Uganda Protectorate) to determine the size of an over-year storage reservoir which would be required in order to equalize the outflow

[4] "The Nile Basin," Vol. V, by H. E. Hurst and P. Phillips, Cairo Govt. Press, Cairo, Egypt, 1938, p. 86.

[5] "Storage to Be Provided in Impounding Reservoirs for Municipal Water Supply," by Allen Hazen, *Transactions*, ASCE, Vol. LXXVII, December, 1914, p. 1539.

[6] "Storage Required for the Regulation of Stream Flow," by Charles E. Sudler, *ibid.*, Vol. 91, 1927, p. 622.

Hurst 1951

Hurst's Major Contributions

- Science
  - Discovery of a rich correlation structure in large-scale hydrological systems
  - Application to a wide range of geophysical processes
- Mathematics/Statistics
  - Characterization and estimation of long-term dependence
  - Kolmogorov dynamics
- Scientific Process
  - Centrality of data (!!)
  - Acknowledgment of importance of stochastics
  - Consideration of phenomena in seemingly unrelated areas
  - Potential of seemingly mundane work
- Engineering

Theoretical Development

the range would be increased by four. For the present this difference is ignored, but it will be considered later.

In the case selected for analysis $2m$ coins are tossed N times, heads and tails or gains and losses are equal at the end of the trial, and n is written for $N m$.

The number of orders or arrangements in which n gains and n losses can occur is

$$ {}_{2n}C_n = \frac{(2n)!}{(n!)^2} \qquad (3)$$

in which ${}_{2n}C_n$ is the number of combinations of $2n$ different things taken n at a time. Among these orders there are ${}_{2n}C_{n+h}$ instances in which, at some point in the process of tossing, losses exceed gains by h or more.[8]

There are three cases:

Case	Condition	No. of arrangements
(a)	Losses never exceed gains	$\frac{{}_{2n}C_n}{n+1}$
(b)	Gains never exceed losses	$\frac{{}_{2n}C_n}{n+1}$
(c)	Losses never exceed gains and gains never exceed losses	$(n-1)\frac{{}_{2n}C_n}{n+1}$

In case (a) the range is the maximum value of gains minus losses; in case (b) the range is the maximum value of losses minus gains; and in case (c) the range is the sum of the maximum of gains minus losses, and the maximum of losses minus gains.

Summing all values of the range to obtain a mean value, and writing G for number of gains and L for number of losses,

$$\text{Mean range} = \frac{\text{maximum }(G-L) + \text{maximum }(L-G)}{{}_{2n}C_n}$$

$$= \frac{2\text{ maximum }(G-L)}{{}_{2n}C_n} \qquad (4)$$

To find the mean range see Table 4. There are ${}_{2n}C_{n+1}$ arrangements in which, at some point, gains exceed losses by one or more; and there are ${}_{2n}C_{n+2}$

TABLE 4.—COMPUTATIONS TO DETERMINE THE MEAN RANGE

Gains minus losses	No. of arrangements	Products
1	${}_{2n}C_{n+1} - {}_{2n}C_{n+2}$	$1\,({}_{2n}C_{n+1} - {}_{2n}C_{n+2})$
2	${}_{2n}C_{n+2} - {}_{2n}C_{n+3}$	$2\,({}_{2n}C_{n+2} - {}_{2n}C_{n+3})$
3	${}_{2n}C_{n+3} - {}_{2n}C_{n+4}$	$3\,({}_{2n}C_{n+3} - {}_{2n}C_{n+4})$
....	....	....
....	....	....
$n-1$	${}_{2n}C_{2n-1} - {}_{2n}C_{2n}$	$(n-1)\,({}_{2n}C_{2n-1} - {}_{2n}C_{2n})$
n	${}_{2n}C_{2n}$	$n\,({}_{2n}C_{2n})$
Sum	${}_{2n}C_{n+1}$	${}_{2n}C_{n+1} + {}_{2n}C_{n+2} + \cdots + {}_{2n}C_{2n}$

[8] "Choice and Chance," by W. A. Whitworth, Cambridge, England, 4th Ed., 1886, Chapter on Priority.

in which gains exceed losses by two or more. That is, there are ${}_{2n}C_{n+1} - {}_{2n}C_{n+2}$ arrangements in which gains exceed losses by one.

Referring to Table 4, $2^{2n} = 1 + {}_{2n}C_1 + {}_{2n}C_2 + \cdots + {}_{2n}C_{n-1} + {}_{2n}C_n + {}_{2n}C_{n+1} + \cdots + {}_{2n}C_{2n}$. Hence, the sum of the products is equal to $\frac{1}{2}(2^{2n} - {}_{2n}C_n)$. The arrangements in Table 4 include all except those in which gains never exceed losses, whose number is $\frac{{}_{2n}C_n}{n+1}$ (or ${}_{2n}C_n - {}_{2n}C_{n+1}$), and in which the maximum of gains minus losses is 0. Hence, the mean maximum value of gains minus losses is $\frac{2^{2n-1} - \frac{1}{2}\,{}_{2n}C_n}{{}_{2n}C_n}$ and this is also the mean maximum of losses minus gains. Therefore, the mean range is

$$R = \frac{2^{2n}}{{}_{2n}C_n} - 1 \qquad (5)$$

Since n is large, Eq. 5 can be simplified by the approximation for $n!$ of James Stirling which is as follows:

$$n! = \sqrt{2n\pi}\left(\frac{n}{e}\right)^n \qquad (6)$$

Hence,

$${}_{2n}C_n = \frac{(2n)!}{n!\,n!} = \frac{\sqrt{4n\pi}\,(2n/e)^{2n}}{[\sqrt{2n\pi}\,(n/e)^n]^2} = \frac{2^{2n}}{\sqrt{n\pi}} \qquad (7)$$

and

$$R = \sqrt{n\pi} - 1 = \sqrt{N m \pi} \qquad (8)$$

since the product $N m$ is large.

Thus the average range of the accumulated sums (number of heads minus number of tails), when $2m$ coins are tossed N times ($N m$ being large), increases as the square root of the number of tosses, exactly like the accumulated error on a line of leveling.

The standard deviation of the binomial distribution produced by tossing $2m$ coins is

$$\sigma_r = \sqrt{2m \times \tfrac{1}{2} \times \tfrac{1}{2}} = \sqrt{\frac{m}{2}} \qquad (9)$$

This is the standard deviation of the number of heads (r) or tails ($2m - r$). The standard deviation σ_d of the number of heads minus the number of tails is twice σ_r, or $\sqrt{2m}$. Substituting for $\sqrt{2m}$ in Eq. 8,

$$R = \sigma_d \sqrt{\tfrac{1}{2} N \pi} = 1.25\,\sigma_d \sqrt{N} \qquad (10)$$

In Eq. 10, R is the range of the continued sum of $2mN$ tosses of a single coin, and is a little larger than the range obtained from N tosses of a set of $2m$ coins. The average difference d between the two computations for range can be found as follows:

Toss a coin a large number of times and plot the continued sums of heads minus tails, marking a given number of them off in sets of $2m$. The set of

Hurst's Fundamental Finding:

$$\frac{R}{\sigma} = (N/2)^{0.72} = 0.61N^{0.72}$$

$$N^{0.72}$$

Data

TABLE 7.—(*Continued*)

Phenomena	No. of cases	N years	$\frac{R}{\sigma}$	Log N	Log $\frac{R}{\sigma}$	K
(*e*) Group of 90 Cases						
Thickness of annual layers of mud; Tamiskaming, Ont., Canada, and Moen, in the Sogne District, Norway	44	50	10.9	1.70	1.02	0.73
	22	100	21.3	2.00	1.31	0.77
	11	200	42.7	2.30	1.58	0.79
	4	300	82.5	2.48	1.90	0.87
	3	400	126	2.60	2.09	0.91
	4	550	115	2.74	2.00	0.82
	2	1,100	181	3.04	2.19	0.80
Mean of 90 cases	....	....	....	1.98	1.30	0.77
(*f*) Group of 114 Cases						
Thickness of annual layers of mud, Lake Saki in the Crimea	40	50	9.7	1.70	0.98	0.70
	40	100	15.3	2.00	1.17	0.69
	20	200	25.0	2.30	1.39	0.70
	8	500	47.9	2.70	1.66	0.69
	4	1,000	84.0	3.00	1.91	0.71
	2	2,000	179.0	3.30	2.24	0.75
Mean of 114 cases	....	....	....	2.06	1.22	0.69
(*g*) Group of 25 Cases						
Sunspot numbers and wheat prices	12	64	12.4	1.77	1.06	0.72
	6	124	22.1	2.09	1.34	0.75
	7	237	16.9	2.36	1.22	0.60
Mean of 25 cases	....	....	....	2.01	1.17	0.69
(*h*) Group of 259 Cases						
Weight mean of 259 cases[c]	....	97	17.8	....	....	....

[a] Rainfall stations with one value of R; includes temperature at one station. [b] Rainfall stations with two groups of values of R; includes temperature and one pressure. [c] N ranges from 81 to 120.

tion of log N and can be expressed as

$$\log \frac{R}{\sigma} = K \log \frac{N}{2} \quad \text{(12)}$$

In each main group of observations a line drawn through the point $\log (R/\sigma) = 0$, $\log N = 0.30$ $(N = 2)$, and the center of gravity of all the observations is a good fit. The slopes of these lines (K) vary from 0.69 to 0.80.

Whether there is any theoretical significance in the fact that K is approximately $\frac{3}{4}$ is not known. The value of K has been calculated for each of the 690 individual values of R/σ and the frequency curve shown in Fig. 5 has been drawn. The data for this curve are as follows:

Description	K
Number of values	690
Mean value	0.729
Standard deviation	0.092
Range of individual values	0.46 to 0.96

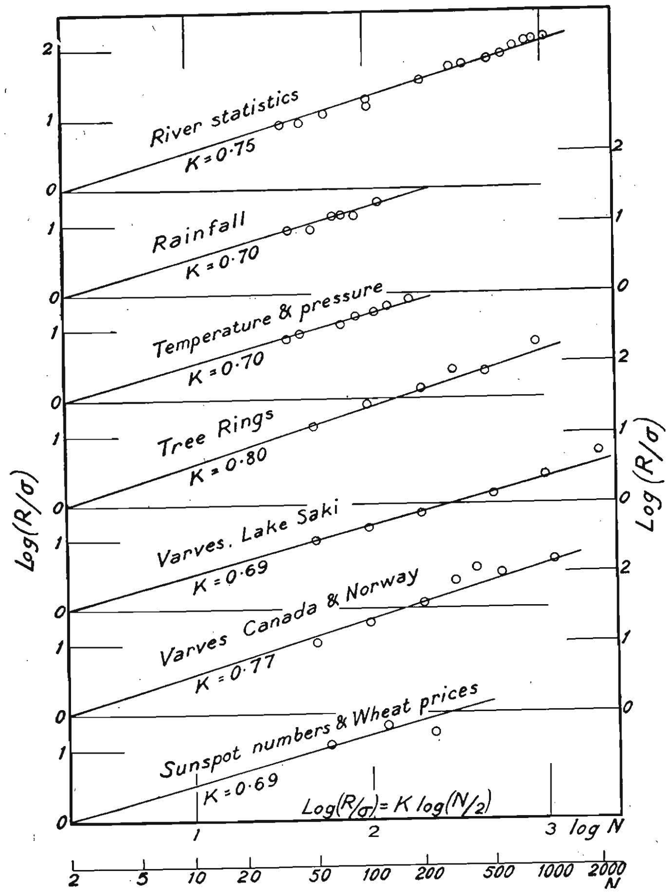


FIG. 4.—RELATION BETWEEN RANGE OF SUMMATION CURVE R, THE STANDARD DEVIATION σ, AND THE YEARS OF RECORD N

Fig. 5 shows that there is a slight skewness in the distribution of K, the mode tending to be slightly greater than the mean. However, it will be seen that the normal curve which has been fitted is a close approximation to the observed distribution. Referring to Table 7, a summary of K-values can be compiled as in Table 8.

Diverse Processes and Data [Hurst, 1951]

- Runoff Statistics
- Rainfall
- Temperature and Pressure
- Tree Rings
- Varves, Lake Saki
- Varves, Canada and Norway
- Sunspot Numbers and Wheat Prices

Scientific Process:

- Why Hurst is notable in Hydrology
  - Described a new, universal and fundamental characteristic of natural, and occasionally human-dominated, systems

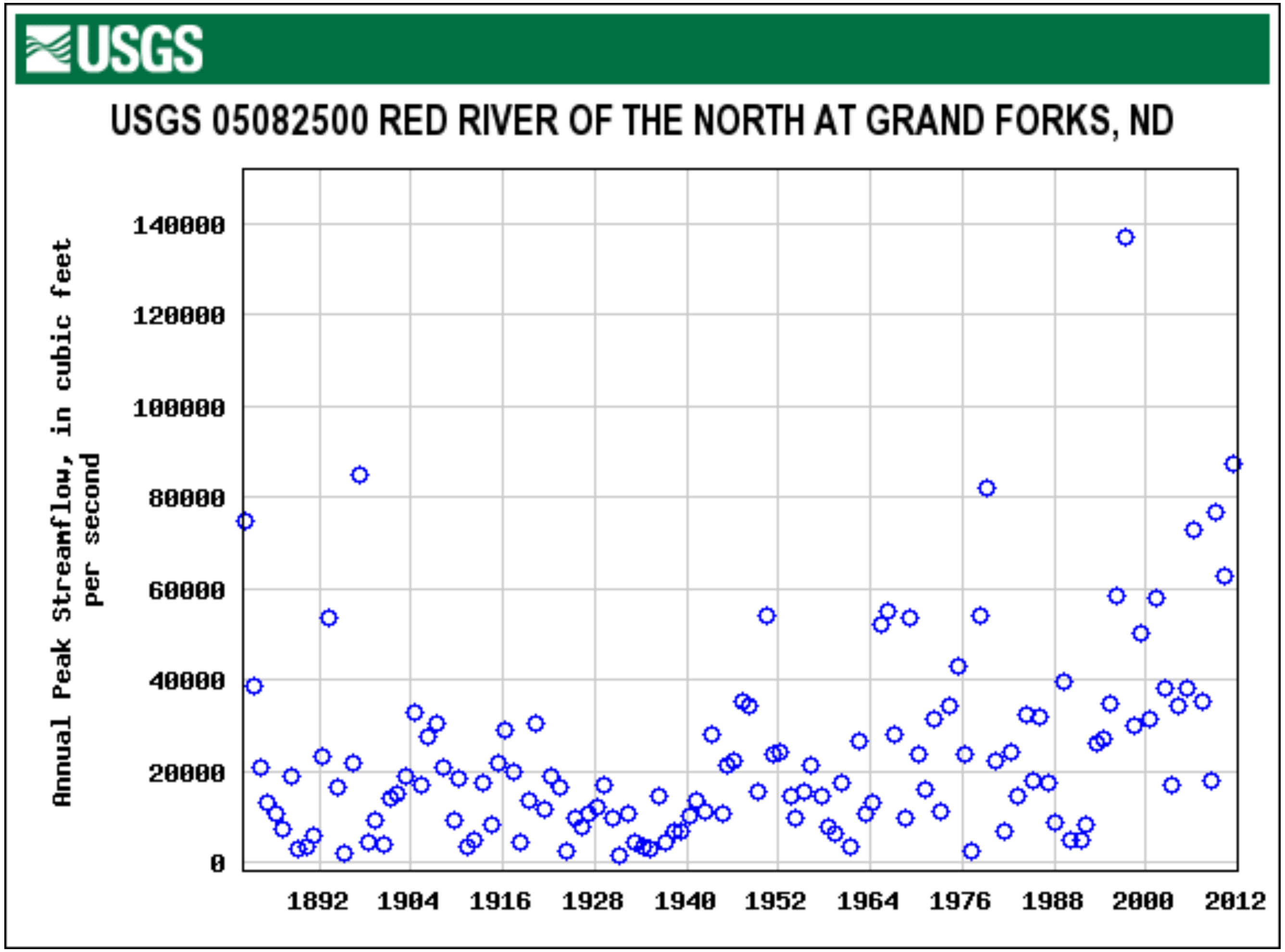
Scientific Process:

- Why Hurst is notable in Hydrology
  - Described a new, universal and fundamental aspect of natural systems
- Why Hurst is notable in history of math/ science
  - Calculus / Classical Physics
  - Riemannian Geometry / General Relativity
  - Hilbert Spaces / Quantum Mechanics
  - Kolmogorov Dynamics / Hurst Phenomenon

Role of B. B. Mandelbrot

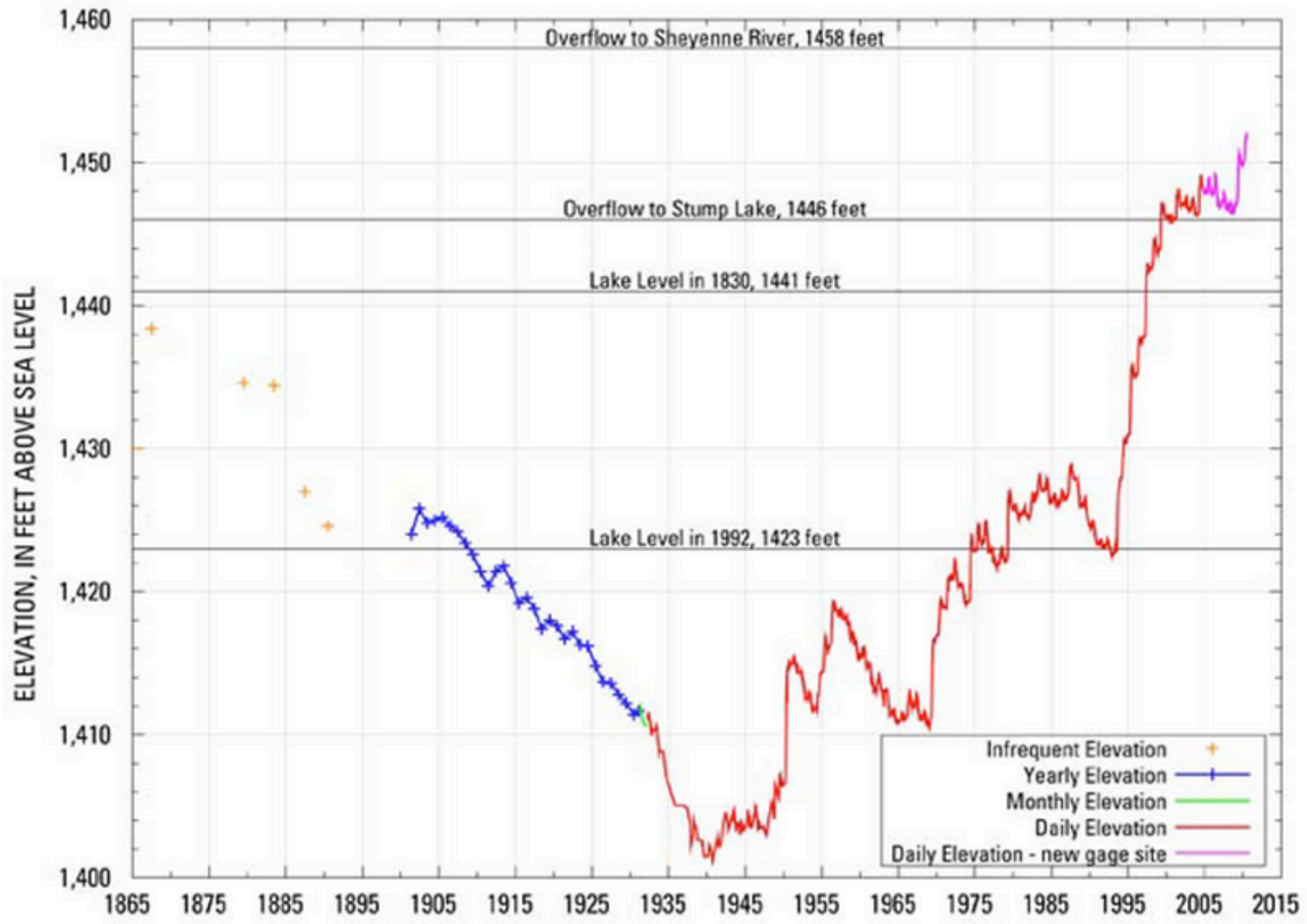
- Promoted Hurst's main findings
- Introduced catchy terminology: fractals, Joseph effect, Noah effect
- Developed spectacular graphics
- Emphasized static versus stochastic
- Emphasized micro versus macro perspective
- Did not promote Kolmogorov, although he acknowledged Kolmogorov in his papers

Where Can We See H-K Dynamics?



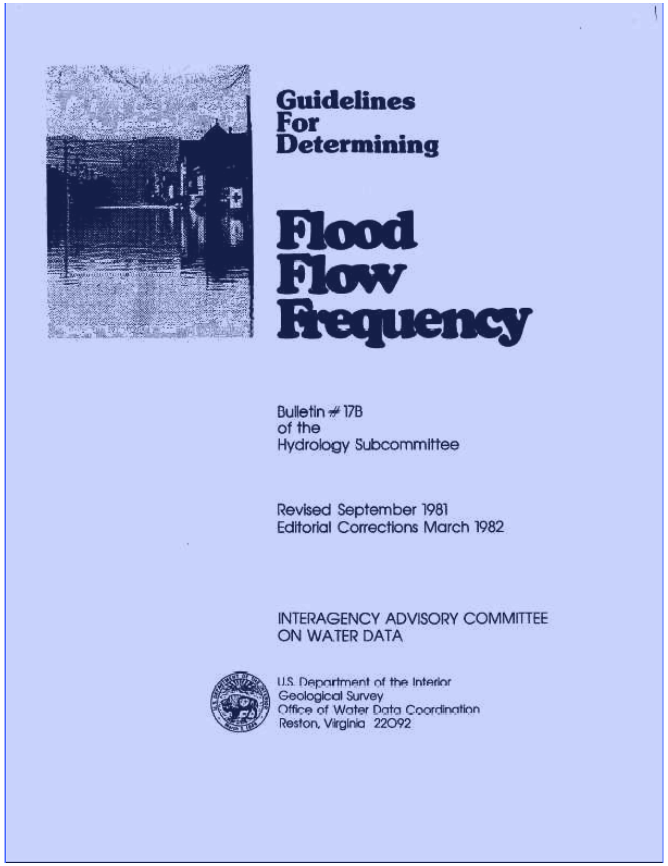
At A Longer Time Scale

Water Surface Elevations, Devils Lake, ND (USGS)



Is Hurst Universally Embraced?

Flood Guidelines?


Guidelines For Determining

Flood Flow Frequency

Bulletin # 17B
of the
Hydrology Subcommittee

Revised September 1981
Editorial Corrections March 1982

INTERAGENCY ADVISORY COMMITTEE
ON WATER DATA

U.S. Department of the Interior
Geological Survey
Office of Water Data Coordination
Reston, Virginia 22092

Bulletin 17B [1981]

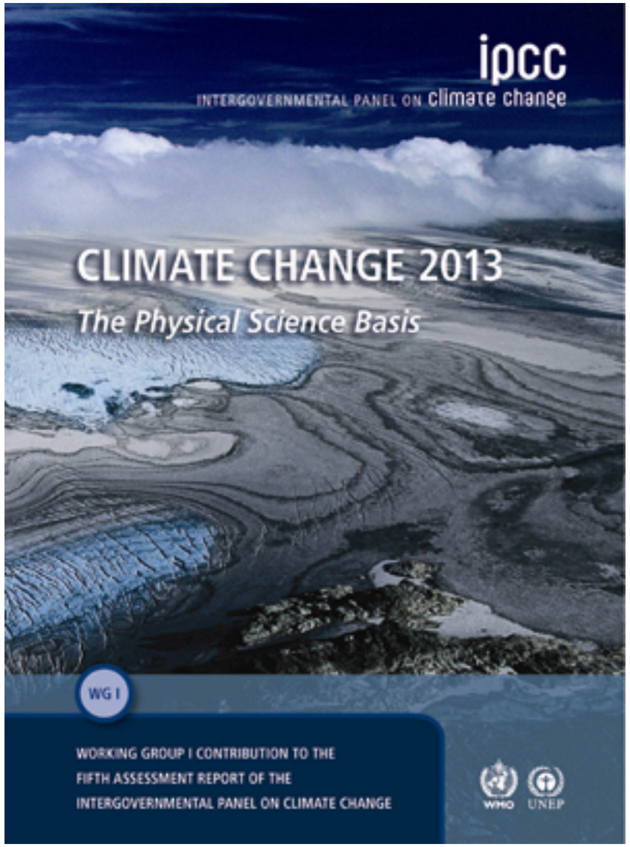
A. <u>Climatic Trends</u>

There is much speculation about climatic changes. Available evidence indicates that major changes occur in time scales involving thousands of years. In hydrologic analysis it is conventional to assume flood flows are not affected by climatic trends or cycles. Climatic time invariance was assumed when developing this guide.

B. <u>Randomness of Events</u>

In general, an array of annual maximum peak flow rates may be considered a sample of random and independent events. Even when statistical tests of the serial correlation coefficients indicate a significant deviation from this assumption, the annual peak data may define an unbiased estimation of future flood activity if other assumptions are attained.

Does IPCC Acknowledge Hurst?

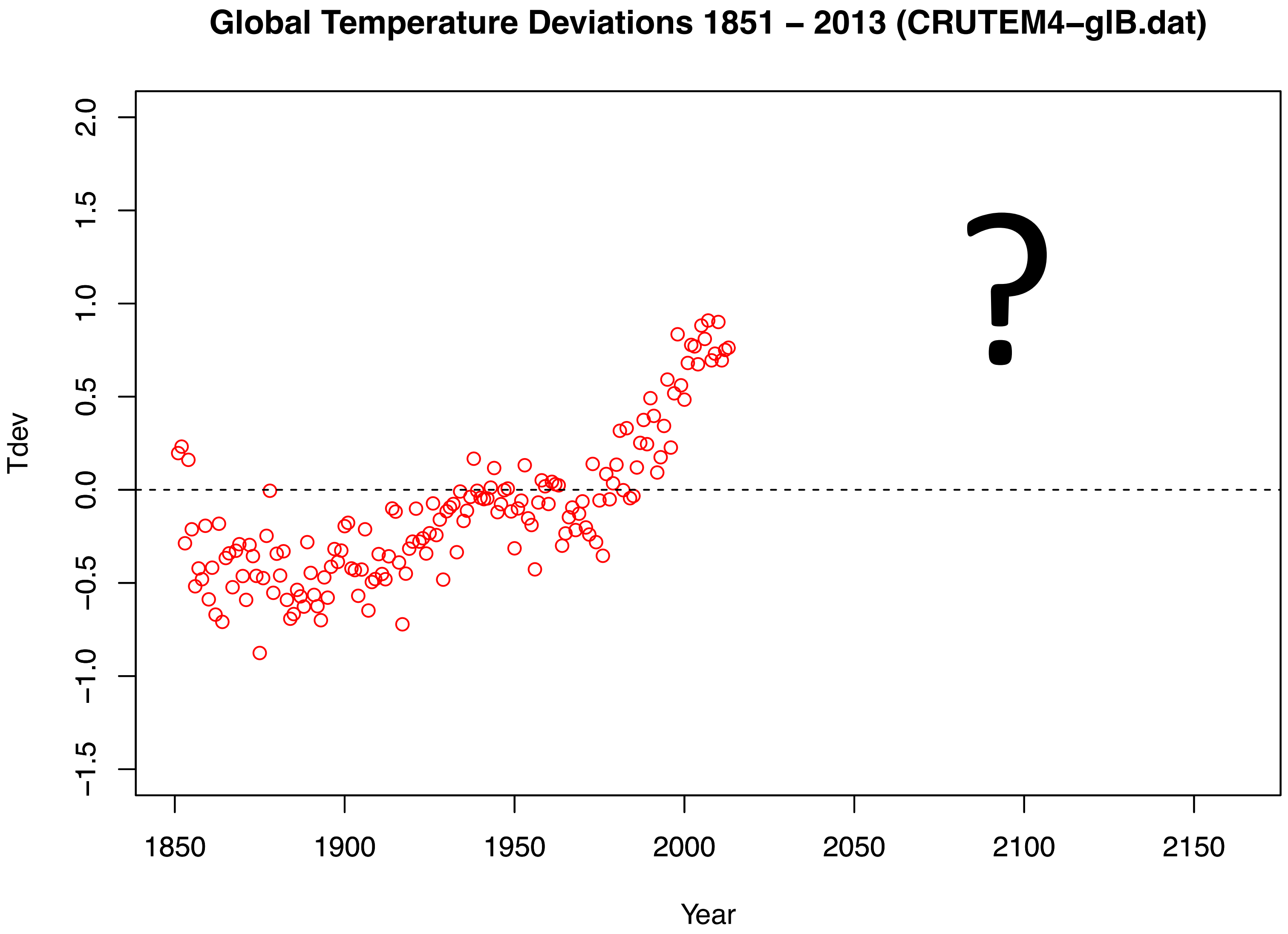


- "No attempt has been undertaken to further describe and interpret the observed changes in terms of multidecadal oscillatory (or low frequency) variations, (long-term) persistence and/or secular trends (e.g., as in Cohn and Lins, 2005; Koutsoyiannis and Montanari, 2007; …)"
- "Trends that appear significant when tested against an AR(1) model may not be significant when tested against a process which supports this 'long-range dependence' (Franzke,2010)."
- However the SPM does not mention LTP, although is speaks about the internal climate variability, e.g.:
  "Internal variability will continue to be a major influence on climate, particularly in the near-term and at the regional scale."

Still, Hurst-Kolmogorov Dynamics Permeates Science

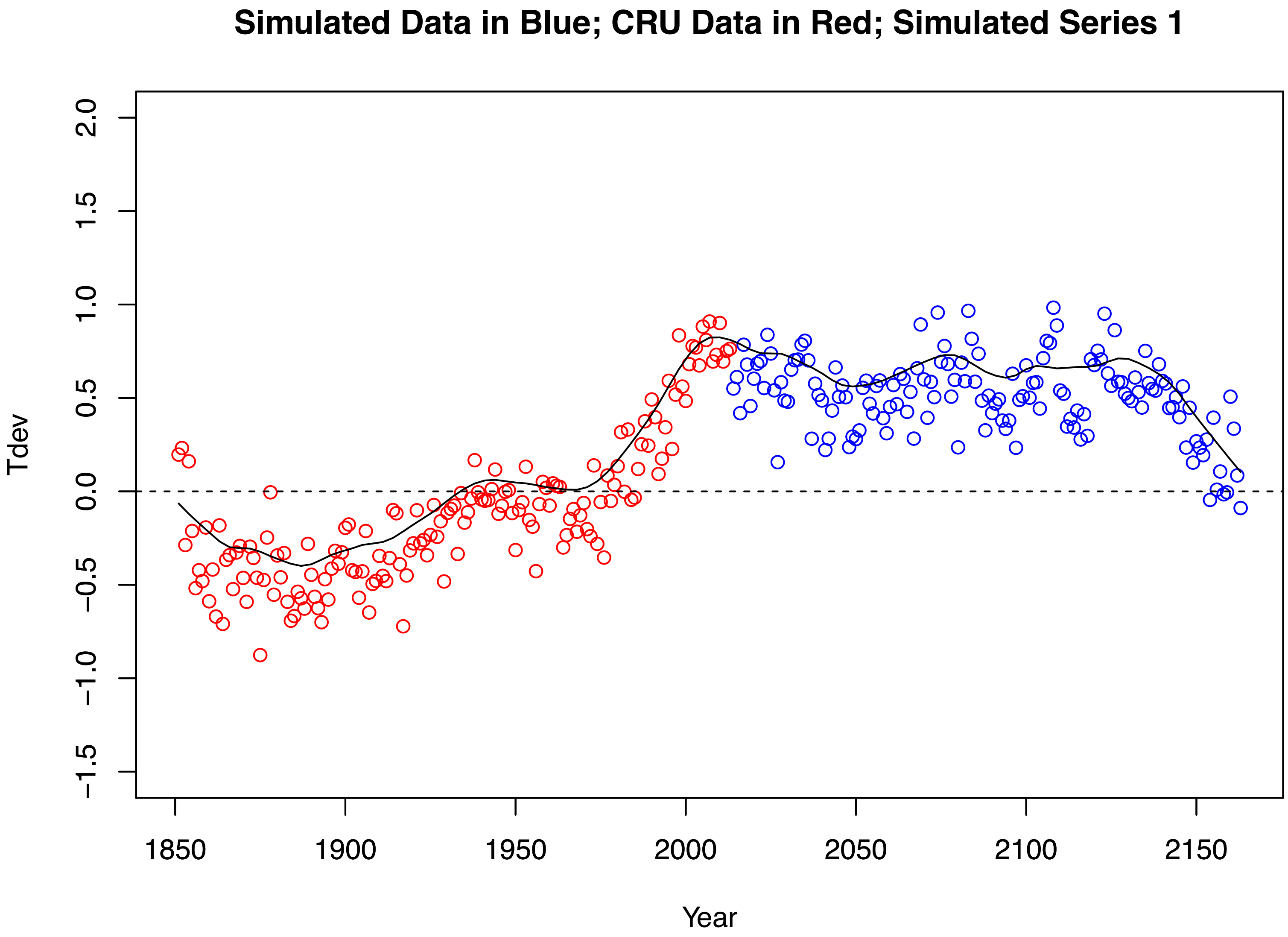
- **Data**: Value of long records
- **Science**: Realistic view of inverse problems
- **Forecasting**: Recognition that predicting the long-term behavior of complex systems is fundamentally challenging.

Predicting the Future

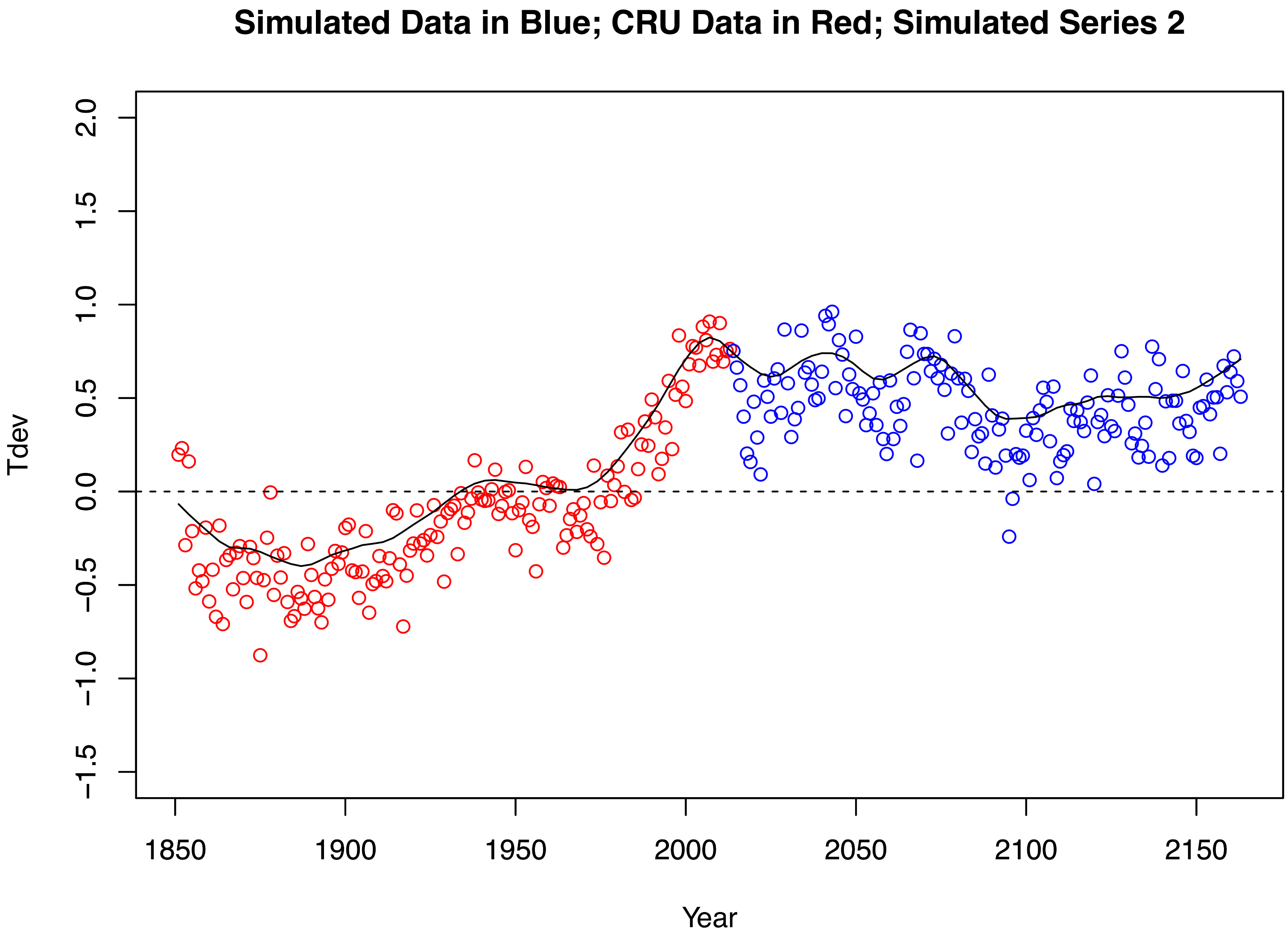


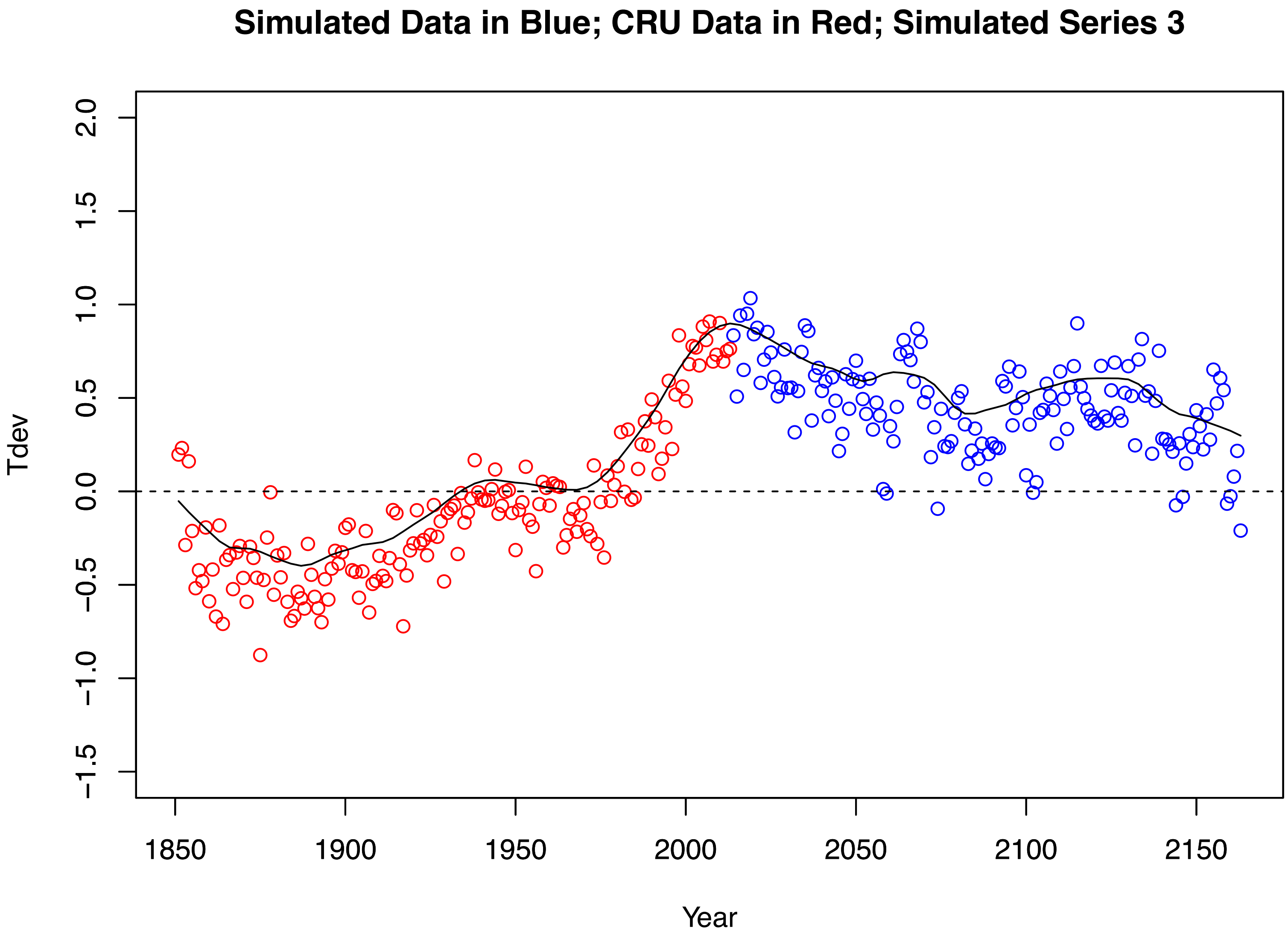
Extending the Temperature Record to 2163

- Stationary FARIMA Model
  - Short and Long-term persistence [Farima(1,1,0)]
  - Fitted to 1851-2012 CRU annual global temperature anomaly data
  - Preserves correlations with historical record
  - Trend-free
- Innovations generated using random number generator with seed=i for simulation i

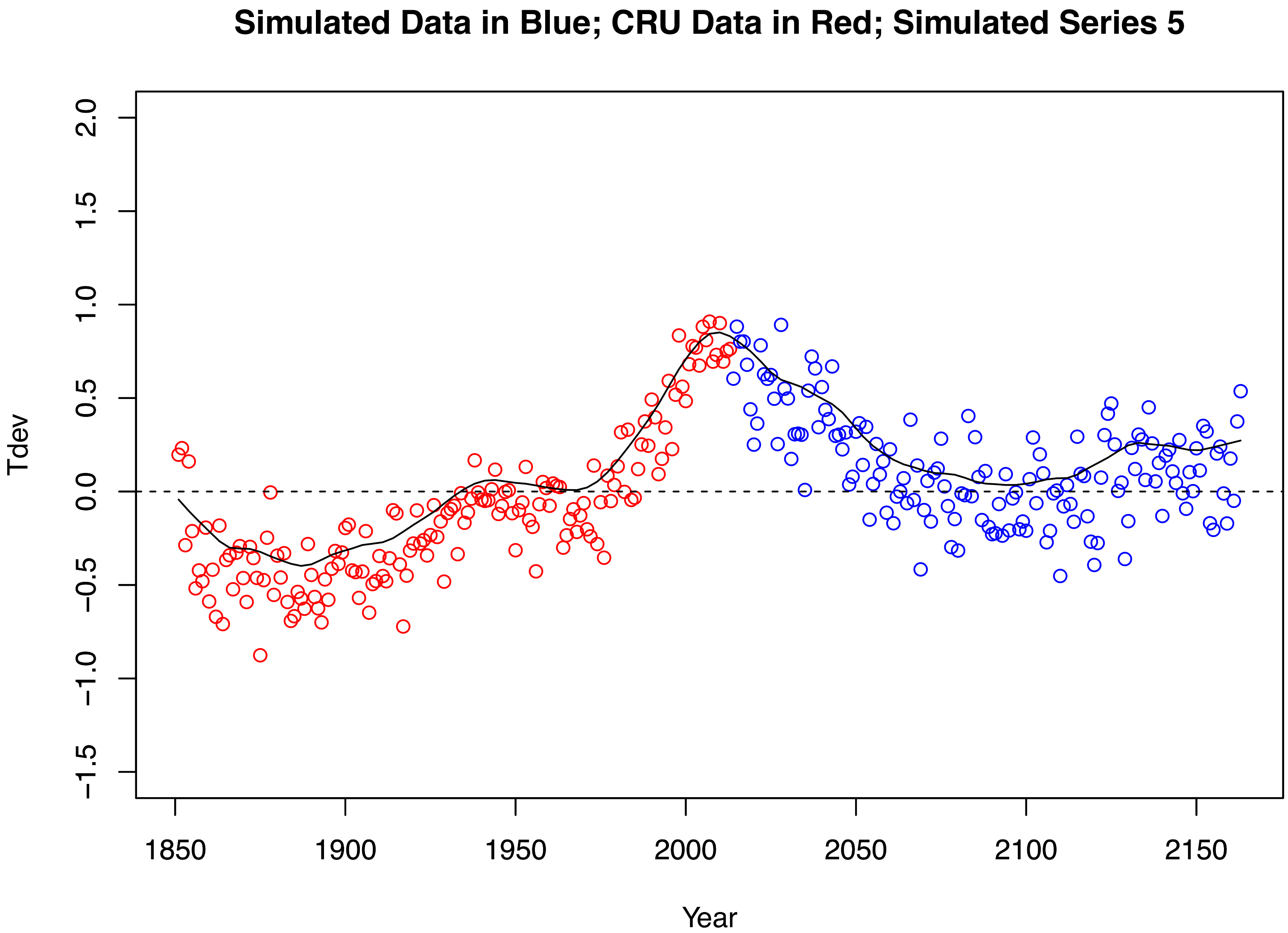


Simulated Data in Blue; CRU Data in Red; Simulated Series 1
Tdev
Year
1850
1900
1950
2000
2050
2100
2150
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

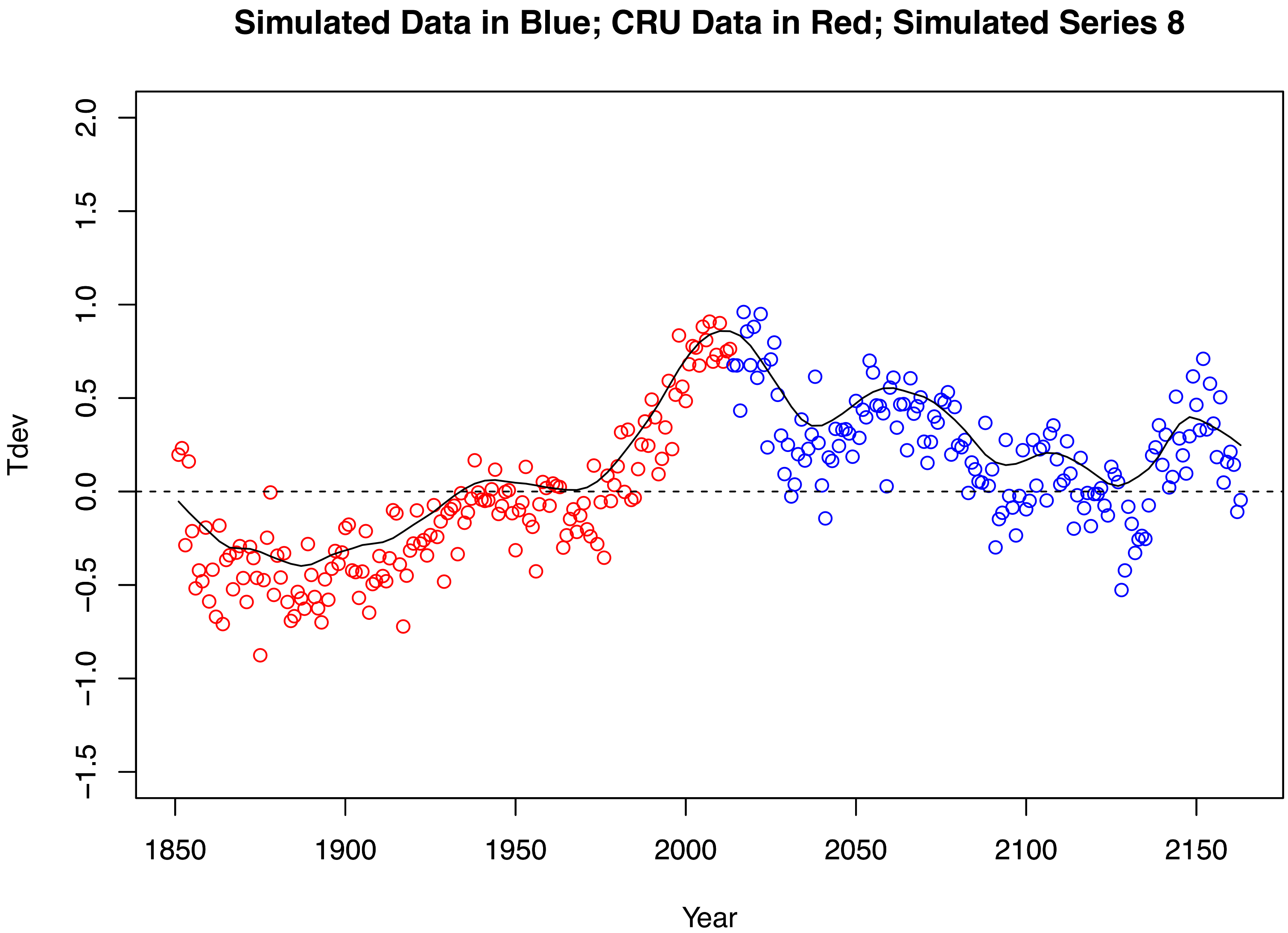

Simulated Data in Blue; CRU Data in Red; Simulated Series 2
Tdev
Year
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1850
1900
1950
2000
2050
2100
2150

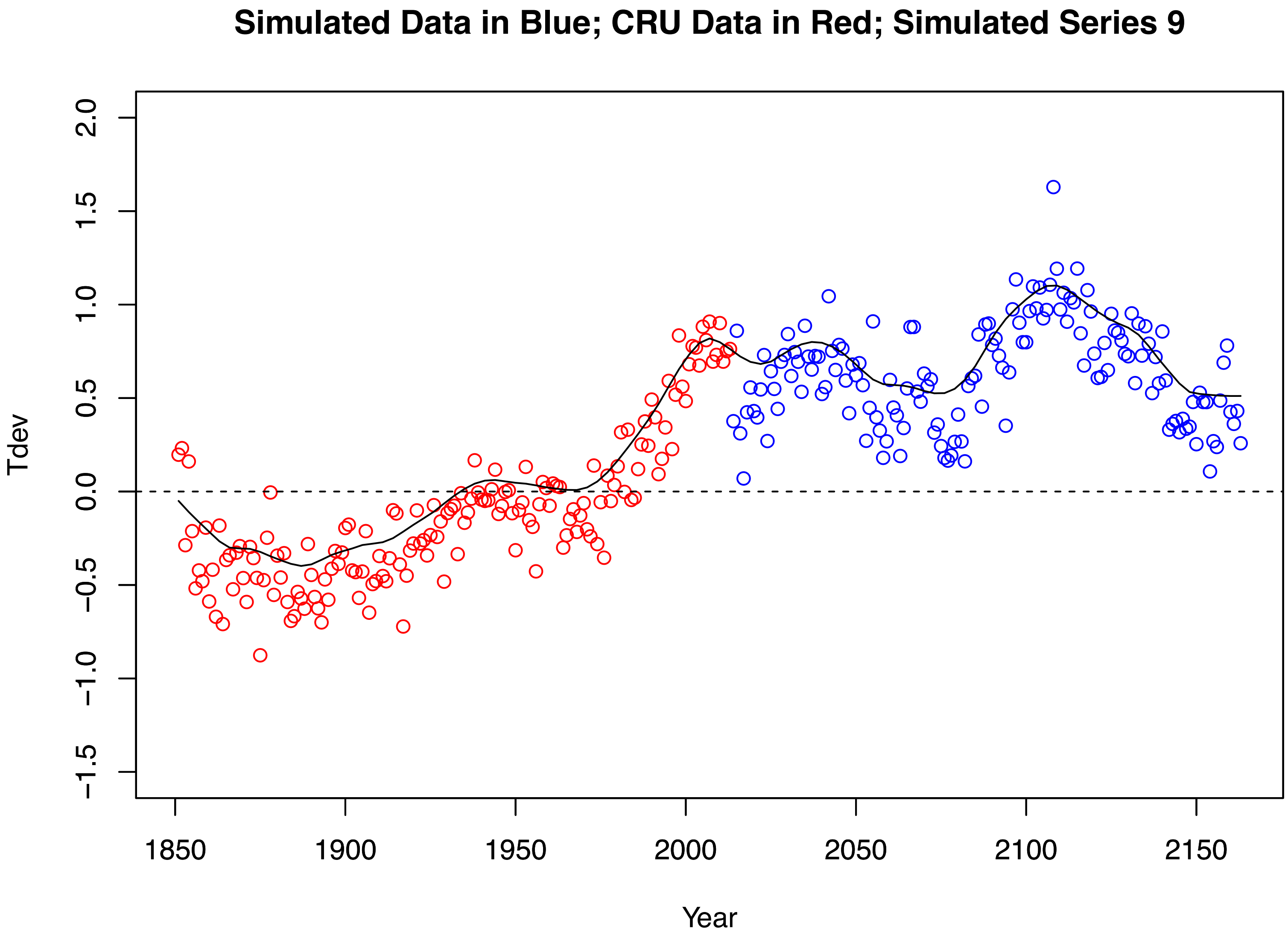


Simulated Data in Blue; CRU Data in Red; Simulated Series 3
Tdev
Year
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1850
1900
1950
2000
2050
2100
2150

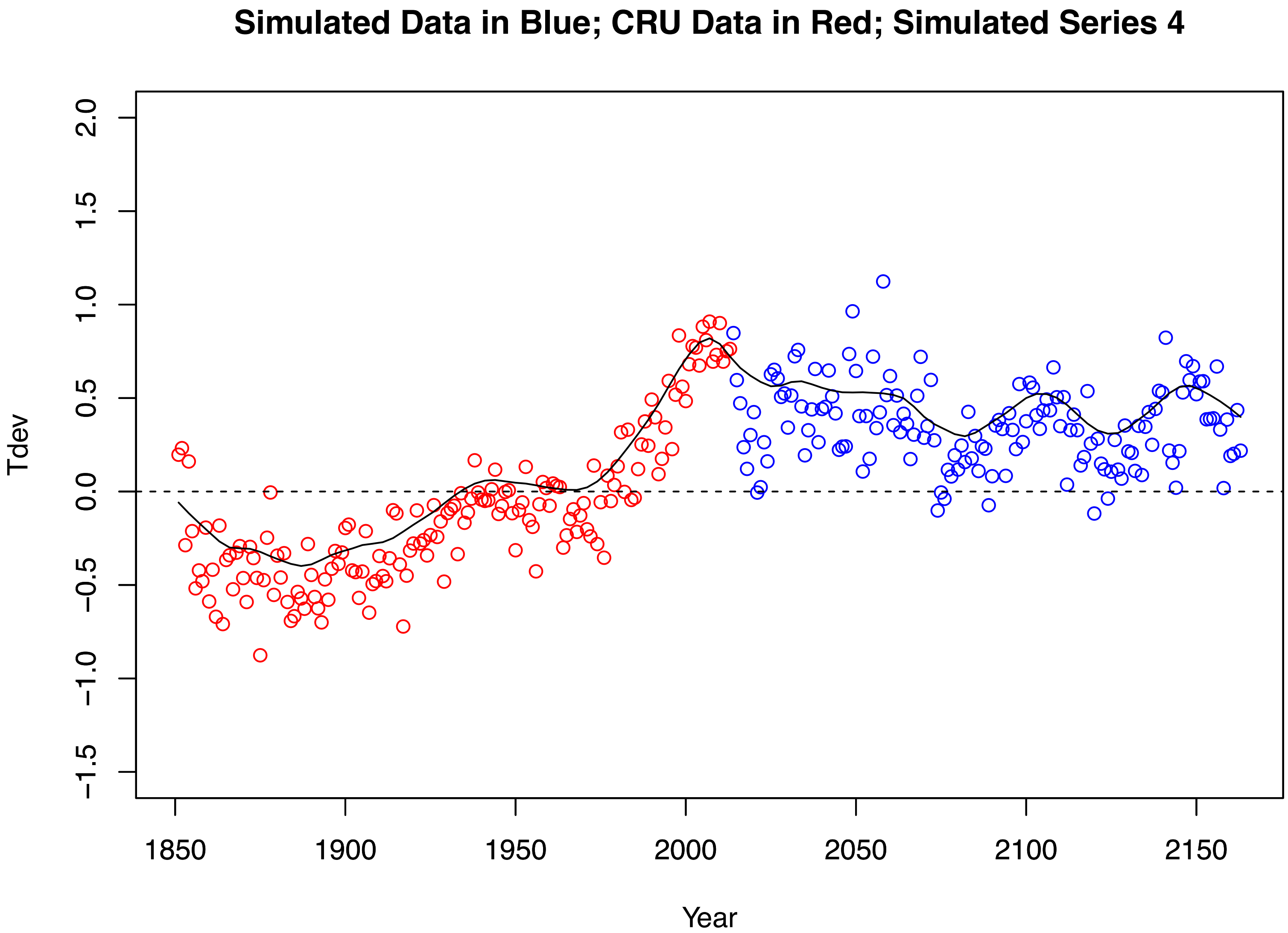


Simulated Data in Blue; CRU Data in Red; Simulated Series 5
Tdev
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1850
1900
1950
2000
2050
2100
2150
Year

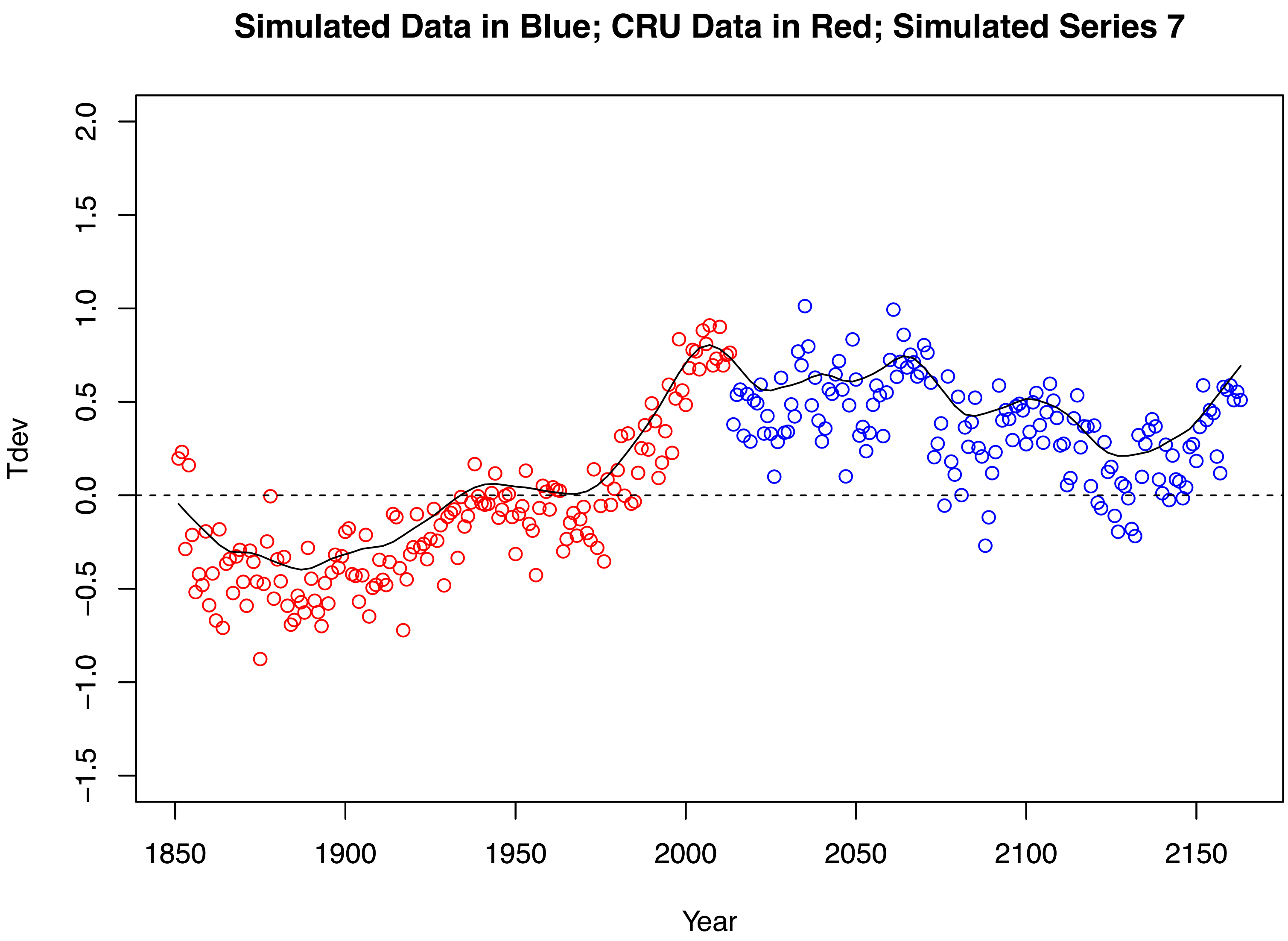

Simulated Data in Blue; CRU Data in Red; Simulated Series 8
Tdev
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1850
1900
1950
2000
2050
2100
2150
Year


Simulated Data in Blue; CRU Data in Red; Simulated Series 9
Tdev
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1850
1900
1950
2000
2050
2100
2150
Year

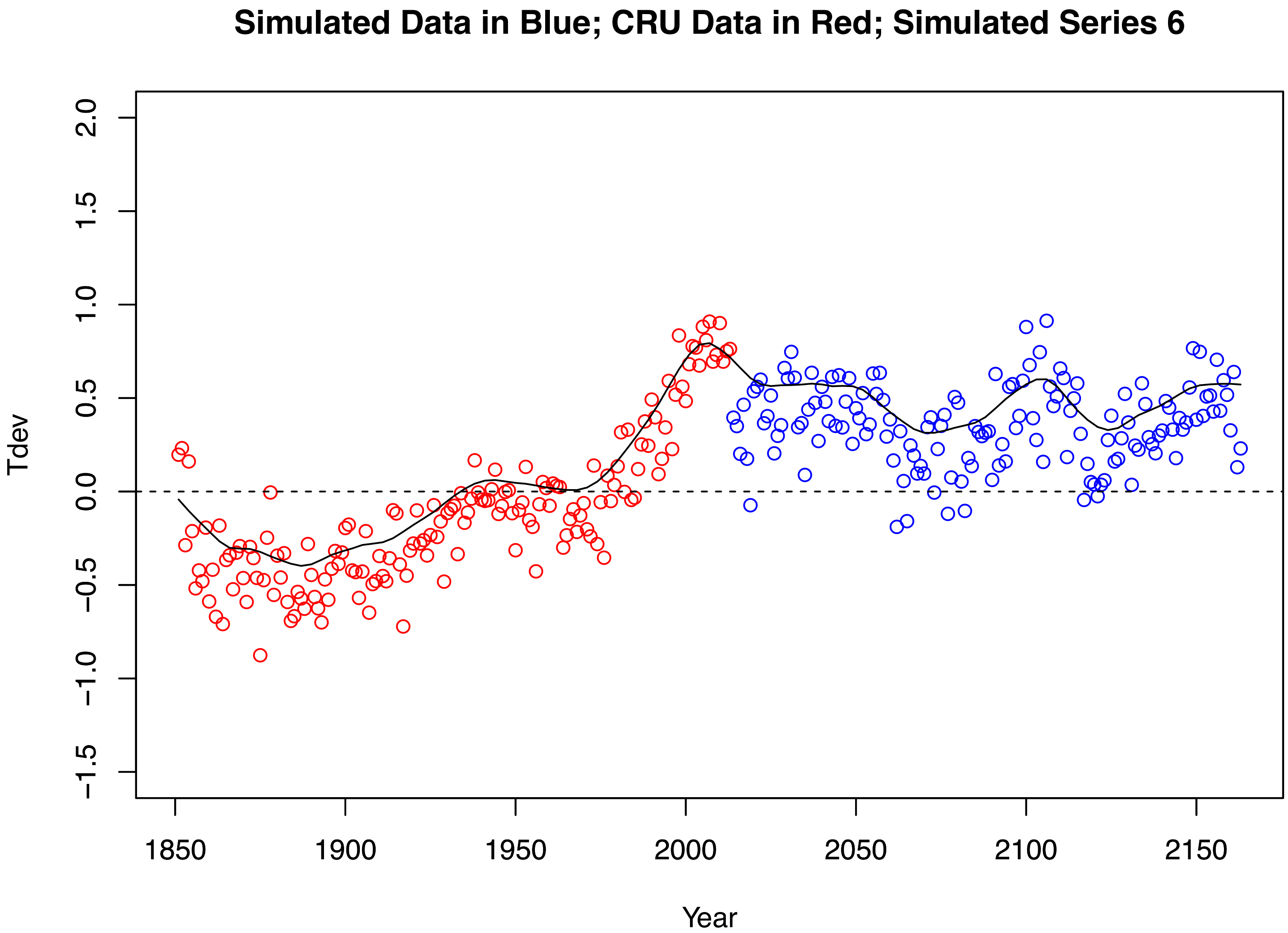
Thank you!



Simulated Data in Blue; CRU Data in Red; Simulated Series 4
Tdev
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
1850
1900
1950
2000
2050
2100
2150
Year



Simulated Data in Blue; CRU Data in Red; Simulated Series 7
Tdev
Year
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1850
1900
1950
2000
2050
2100
2150


Simulated Data in Blue; CRU Data in Red; Simulated Series 6
Tdev
Year
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
1850
1900
1950
2000
2050
2100
2150