

Environmental Sciences Division

**UNITED STATES HISTORICAL CLIMATOLOGY NETWORK
DAILY TEMPERATURE AND PRECIPITATION DATA**

Contributed by

P. Y. Hughes

E. H. Mason

T. R. Karl

W. A. Brower

National Oceanic and Atmospheric Administration
National Climatic Data Center
Asheville, North Carolina

Prepared by D. P. Kaiser*

Carbon Dioxide Information Analysis Center

Environmental Sciences Division

Publication No. 3778

Date Published: February 1992

*Energy, Environment, and Resources Center

University of Tennessee

Knoxville, Tennessee

Prepared for the

Carbon Dioxide Research Program

Environmental Sciences Division

Office of Health and Environmental Research

Budget Activity Number KP 05 00 00 0

Prepared by the

OAK RIDGE NATIONAL LABORATORY

Oak Ridge, Tennessee 37831-6335

managed by

MARTIN MARIETTA ENERGY SYSTEMS, INC.

for the

U.S. DEPARTMENT OF ENERGY

under contract DE-AC05-84OR21400

MASTER

EP

TABLE OF CONTENTS

	<u>Page</u>
LIST OF FIGURES	v
LIST OF TABLES	vii
ABSTRACT	ix
PART 1: INFORMATION ABOUT THE DATA PACKAGE	1
1. NAME OF THE NUMERIC DATA PACKAGE	3
2. CONTRIBUTORS	3
3. KEYWORDS	3
4. BACKGROUND INFORMATION	3
5. SOURCE AND SCOPE OF THE DATA	4
6. APPLICATIONS OF THE DATA	10
7. LIMITATIONS AND RESTRICTIONS	11
8. REFERENCES	12
9. QUALITY ASSURANCE OF THE HCN/D DATA BASE	14
10. HOW TO OBTAIN THE DATA PACKAGE	18
PART 2: INFORMATION ABOUT THE MAGNETIC TAPE	19
11. CONTENTS OF THE MAGNETIC TAPE	21
12. DESCRIPTIVE FILE ON THE MAGNETIC TAPE	23
13. LISTINGS OF THE FORTRAN IV DATA RETRIEVAL PROGRAMS	37
14. LISTINGS OF THE SAS* INPUT/OUTPUT RETRIEVAL PROGRAMS	40

*SAS is a registered trademark of SAS Institute, Inc., Cary, North Carolina 27511-8000.

15. VERIFICATION OF DATA TRANSPORT	44
APPENDICES	55
APPENDIX A: DATA QUALITY ASSESSMENTS FOR THE HCN/D STATIONS	A-1
APPENDIX B: REPRINTS OF PERTINENT LITERATURE	B-1
Relationship between Decreased Temperature Range and Precipitation Trends in the United States and Canada, 1941-80. T. R. Karl, G. Kukla, and J. Gavin. 1986.	B-3
Climatic Temperature Normals. N. B. Guttman and M. S. Plantico. 1987.	B-13
On an Additive Model of Daily Temperature Climates. N. B. Guttman and M. S. Plantico. 1989	B-21
Statistical Descriptors of Climate. N. B. Guttman. 1989	B-25
Detecting Climate Variations and Change: New Challenges for Operational Observing and Data Management Systems. T. R. Karl, R. G. Quayle, and P. Ya. Groisman. 1992.	B-31

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Spatial distribution of the 1219 monthly temperature and precipitation reporting stations in the U.S. Historical Climatology Network (HCN), with larger locator symbols depicting the locations of the 138 stations selected for the HCN/D	5

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Station inventory of the 138-station HCN/D data base	6
2. HCN/D stations requiring calendar day shifting of maximum temperature observations for portions of their periods of record	16
3. Characteristics of numeric variables from the HCN/D station inventory file	45
4. Characteristics of numeric variables from the HCN/D station history file	46
5. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 1-10 (File 10 on the magnetic tape)	49
6. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 11-20 (File 11 on the magnetic tape)	50
7. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 21-30 (File 12 on the magnetic tape)	51
8. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 31-40 (File 13 on the magnetic tape)	52
9. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 41-48 (File 14 on the magnetic tape)	53

ABSTRACT

HUGHES, P. Y., E. H. MASON, T. R. KARL, and W. A. BROWER. 1992.
United States Historical Climatology Network Daily Temperature and
Precipitation Data. ORNL/CDIAC-50, NDP-042. Carbon Dioxide Information
Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
140 pp.

This document describes a data base containing daily observations of maximum and minimum temperature and precipitation amounts from 138 U.S. stations. These stations are a specially chosen subset of the 1219-station U.S. Historical Climatology Network (HCN), compiled by the National Climatic Data Center (Asheville, North Carolina) and described in Karl et. al (1990). The daily data network (herein referred to as the HCN/D) consists of stations considered to be the best of those from the HCN, selected to provide reasonably homogeneous spatial coverage of the contiguous U.S. after considering the temporal homogeneity of each station's observing times, instrument types/positions, and surroundings. The data for each station extend through 1987, and most station records are complete for at least 80 years. The daily resolution of these data lends maximum flexibility for studies attempting to detect and monitor long-term climatic changes on a regional scale. Studies using daily data may be able to detect changes in regional climate that would not be apparent from analysis of the more commonly used monthly temperature and precipitation data. Such studies may include analyses of trends in maximum/minimum temperatures, temperature extremes, daily temperature range, precipitation "event size" frequency, and the magnitude and duration of wet and dry periods. Other applications of the data include planning and risk assessment in areas such as agriculture, natural resource exploration, and construction.

These data are available free of charge as a numeric data package (NDP) from the Carbon Dioxide Information Analysis Center. The NDP consists of this document and a magnetic tape containing the daily data, the HCN/D station inventory and station histories, a detailed description of all data files, and computer software routines for reading the data. This document describes how the stations in the HCN/D were selected, defines limitations and restrictions of the data, describes the format and contents of the magnetic tape, and provides reprints of literature pertinent to the collection and application of daily climate data.

PART 1

INFORMATION ABOUT THE DATA PACKAGE

1. NAME OF THE NUMERIC DATA PACKAGE

United States Historical Climatology Network Daily Temperature and Precipitation Data

2. CONTRIBUTORS

P. Y. Hughes, E. H. Mason, T. R. Karl, and W. A. Brower
National Oceanic and Atmospheric Administration
National Climatic Data Center
Asheville, North Carolina 28801

3. KEYWORDS

Climate change; daily data; historical data; precipitation; maximum temperature; minimum temperature; temperature extremes.

4. BACKGROUND INFORMATION

Over the past few decades, numerous global, hemispheric, and regional meteorological data bases have been assembled for use in studying the nature and variability of the earth's climate. This work has been largely inspired by growing international concern over potential impacts of increasing atmospheric concentrations of greenhouse gases. While the parameters important in the study of climate change are myriad, those that seem to have received the most attention are near-surface air temperature (herein referred to as temperature) and precipitation. There are many reasons for this, including (1) the spatial and temporal variability of these parameters affects agriculture, water supply, and energy production/consumption; (2) instrumental records of these variables are relatively long, beginning in the 1800s in many regions of the northern hemisphere; and (3) temperature, in particular, is widely considered to be the variable best suited for detecting or "finger printing" an enhanced greenhouse effect (NASA 1991).

The suitability of modern historical temperature and precipitation data for climate change studies depends on their reliability and accuracy. Most records of significant length, regardless of source, are likely to contain biases or inhomogeneities resulting from changes in the environment or operation of individual observing sites (e.g., urbanization, station moves, and instrument and time of observation changes). The process of identifying and removing these nonclimatic effects is complex and tedious, and has been undertaken on large scales in such studies as Jones et al. (1986) (temperature only), Vinnikov et al. (1990), and Quinlan et al. (1987). The work of Quinlan et al. (1987) involved the compilation of a data base containing monthly temperature and precipitation data from a network of 1219 U.S. stations known as the Historical Climatology Network (HCN). The compilation was performed at the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA) in Asheville, North Carolina, and sponsored by the Carbon Dioxide Research Program of the U.S. Department of Energy. The project arose from the need for an accurate, unbiased, and

modern historical climate record suitable for detecting and monitoring secular changes in regional climate in the contiguous United States. The quality of the HCN data was assured with the use of outlier and areal edits, and the data were corrected for time of observation differences, instrument changes, instrument moves, station relocations, and urbanization effects (Karl et al. 1986a; Karl and Williams 1987). Some of the stations in the HCN are first-order weather stations, but the majority were selected from approximately 5000 U.S. cooperative weather stations. The HCN and the resulting monthly data base (extending through 1984) were documented by Quinlan et al. (1987) and made available free of charge through the Carbon Dioxide Information Analysis Center (CDIAC). These data were later modified to remove biases introduced by urbanization effects (Karl et al. 1988) and were extended through 1987 (Karl et al. 1990).

The data presented in this package are daily observations of maximum/minimum temperature and precipitation amount from a carefully selected 138-station subset of the HCN (herein referred to as the HCN/D to denote the daily resolution of the data). The temperature and precipitation records from these stations are considered to be the most reliable, internally consistent, and unbiased records from the HCN. Data from these stations were subjected to extensive manual and automated quality assurance (QA) checks (Sect. 9). The resulting high-quality daily data base offers a wide variety of applications (Sect. 6).

5. SOURCE AND SCOPE OF THE DATA

The HCN/D data base contains station histories, daily maximum and minimum temperatures, and daily precipitation amounts that were compiled by NCDC after being extracted from digital and nondigital data sets archived at NCDC. These data sets come from a variety of sources, including climatological publications, universities, federal agencies, individuals, and data archives. Figure 1 shows the distribution of the 1219 HCN stations and the locations of the 138 stations selected for the HCN/D. The station inventory for the HCN/D is given in Table 1.

The HCN/D stations were selected from the HCN to provide a reasonably homogeneous spatial distribution of stations within the contiguous United States after consideration of the following.

1. The degree to which each station maintained a constant observation time for maximum and minimum temperatures, excursions from a station's predominant observing time of no more than four years being desired.
2. At least 95% of a station's pre-1951 data should be contained in NCDC digital daily archives.
3. A station's potential for heat island bias over time should be low.
4. Quality assessments based upon the decile ranking assigned by Karl et al. (1990) to the stations' *monthly* maximum/minimum temperature data for the quality characteristics numbered 5-8 in Appendix A. A station should rank within the lower 9 deciles for each characteristic.

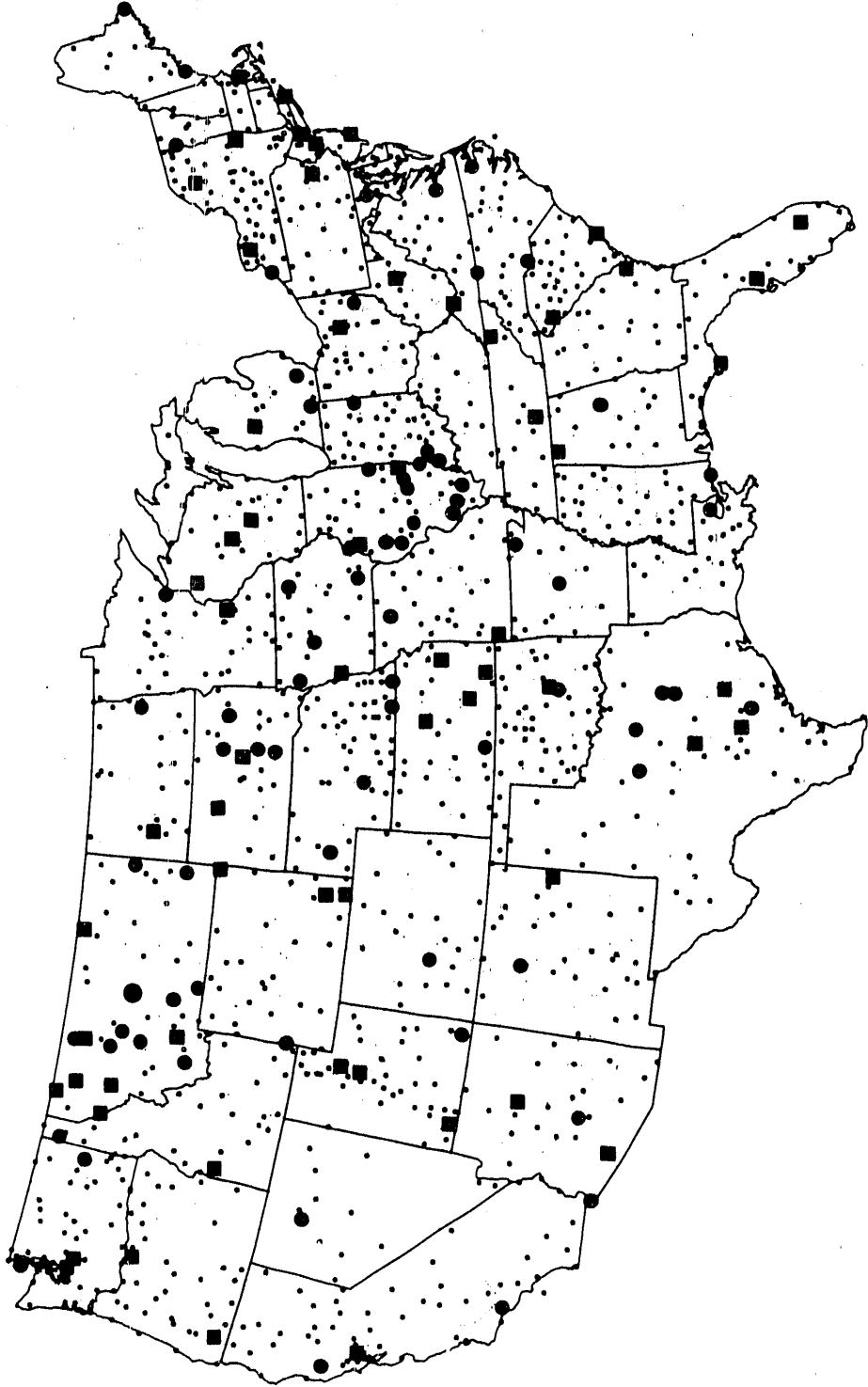


Fig. 1. Spatial distribution of the 1219 monthly temperature and precipitation reporting stations in the U.S. Historical Climatology Network (HCN), with larger locator symbols depicting the locations of the 138 stations selected for the HCN/D. HCN/D stations marked with squares are superior with regard to temporal homogeneity of operations and completeness of data.

Table 1. Station inventory of the 138-station HCN/D data base

State code	Sta. No.	State	Station name ^a	Beginning MO/YR ^b	TOBS ^c	Latitude ^d	Longitude ^e
01	5749	AL	Muscle Shoals FAA AP	12/1940	MD	34 45	85 37
01	8024	AL	Tallapoosa	09/1893	PM	33 26	86 05
02	0080	AZ	Ajo	01/1915	AM	32 22	112 52
02	3160	AZ	Fort Valley	01/1909	AM	35 16	111 44
02	5467	AZ	Mesa Exp. Farm	03/1896	AM	33 25	111 52
02	9652	AZ	Yuma Citrus	09/1920	AM	32 37	114 39
03	1596	AR	Conway	01/1897	PM	35 06	92 25
03	5820	AR	Pocahontas 1	04/1894	AM/PM	36 16	90 58
04	6074	CA	Napa State Hospital	01/1893	PM	38 17	122 16
04	6399	CA	Ojai	01/1906	PM	34 27	119 14
04	9122	CA	Ukiah	01/1893	PM	39 09	123 12
05	3662	CO	Gunnison	07/1893	PM	38 32	106 56
08	0211	FL	Apalachicola WSO AP	10/1903	PM/MD	29 44	84 59
08	0611	FL	Belle Glade Exp. Station	05/1924	AM	26 39	80 38
08	7851	FL	Saint Leo	03/1895	PM	28 20	82 16
09	7847	GA	Savannah WSO AP	01/1874	MD	32 08	81 12
10	1380	ID	Caldwell	10/1904	PM	43 40	116 41
10	7386	ID	Pricst River Exp. Station	02/1898	PM	48 21	116 50
11	0072	IL	Aledo 5NNE	01/1901	PM	41 14	90 44
11	1280	IL	Carlinville 4E	01/1893	PM	39 17	89 52
11	1436	IL	Charleston 2WNW	02/1896	PM	39 29	88 10
11	2483	IL	Duquoin 4SE	05/1898	PM/MD	37 59	89 12
11	3717	IL	Griggsville	01/1893	PM	39 43	90 44
11	3879	IL	Harrisburg 5N	03/1898	PM	37 45	88 33
11	4198	IL	Hoopeston 1NE	06/1902	PM	40 28	87 40
11	5768	IL	Monmouth	02/1893	PM	40 55	90 38
11	6558	IL	Palestine	01/1893	PM	39 00	87 37
11	6610	IL	Paris Waterworks	04/1893	AM/PM	39 38	87 42
11	7551	IL	Rushville	01/1893	PM	40 07	90 33
11	8147	IL	Sparta	01/1893	PM	38 08	89 42
11	9354	IL	Windsor	01/1904	PM	39 26	88 36
12	0676	IN	Berne	01/1910	PM/MD	40 40	84 57
12	7125	IN	Princeton 1W	01/1899	PM/MD	38 21	87 35
12	9253	IN	Washington	08/1896	AM/PM	38 40	87 11
13	2789	IA	Fairfield	01/1893	PM/MD	41 02	91 57
13	4735	IA	Le Mars	03/1896	PM	42 48	96 10
13	4894	IA	Logan	03/1893	PM	41 38	95 48
13	5952	IA	New Hampton	03/1897	PM	43 03	92 19
13	7161	IA	Rockwell City	09/1894	PM	42 24	94 37

Table 1 (continued)

State code	Sta. No.	State	Station name ^a	Beginning MO/YR ^b	TOBS ^c	Latitude ^d	Longitude ^e
14	2401	KS	El Dorado	01/1893	PM	37 49	96 50
14	3954	KS	Independence	01/1893	PM/MD	37 15	95 42
14	5173	KS	Medicine Lodge	05/1895	PM	37 17	98 35
14	5363	KS	Minneapolis	01/1892	PM	39 08	97 42
14	6128	KS	Ottawa	05/1895	PM	38 37	95 17
16	2151	LA	Covington 4NNW	05/1893	AM/PM	30 32	90 07
17	2426	ME	Eastport	04/1873	PM/MD	44 55	67 00
17	6905	ME	Portland WSMO AP	04/1874	MD	43 39	70 19
18	9750	MD	Woodstock	01/1893	PM	39 20	76 52
19	0736	MA	Blue Hill	04/1896	PM/MD	42 13	71 07
20	0230	MI	Ann Arbor Univ. of MI	10/1891	PM	42 18	83 43
20	0779	MI	Big Rapids Waterworks	04/1896	PM/MD	43 42	85 29
20	1675	MI	Coldwater State School	01/1898	PM/MD	41 57	85 00
21	1630	MN	Cloquet	04/1911	PM	46 42	92 31
21	5435	MN	Minneapolis WSFO AP	01/1891	MD	44 53	93 13
22	0792	MS	Biloxi Hiller Park	06/1893	PM	30 24	88 57
23	5976	MO	Neosho	01/1893	PM	36 52	94 22
23	8444	MO	Trenton	09/1895	PM	40 05	93 38
24	0364	MT	Augusta	06/1896	PM	47 29	112 23
24	0780	MT	Big Timber	04/1894	PM	45 50	109 57
24	1552	MT	Cascade 5S	04/1904	PM	47 13	111 43
24	2173	MT	Cutbank FAA AP	12/1903	PM/MD	48 36	112 22
24	2409	MT	Dillon WMCE	01/1895	PM	45 12	112 38
24	2689	MT	Ekalaka	01/1897	PM	45 43	104 32
24	3089	MT	Forks 4NNE	01/1915	PM	48 47	107 28
24	3139	MT	Fortine 1N	03/1906	PM	48 47	114 54
24	3984	MT	Haugan 3E	02/1912	PM	47 23	115 21
24	4055	MT	Helena WSO	01/1893	PM	46 36	112 00
24	4558	MT	Kalispell WSO AP	06/1896	MD	48 18	114 16
24	5761	MT	Moccasin Exp. Station	04/1909	AM	47 03	109 57
24	6157	MT	Norris Madison PH	01/1907	PM	45 29	111 38
24	6918	MT	Red Lodge	04/1894	AM/PM	45 11	109 15
24	7286	MT	Saint Ignatius	02/1896	PM	45 19	114 06
24	7382	MT	Savage	09/1905	PM	47 27	104 21
24	8501	MT	Valier	08/1911	PM	48 19	112 15
25	1145	NE	Bridgeport	05/1897	PM	41 40	103 06
25	2820	NE	Fairbury 2SSE	08/1895	AM	40 07	97 10
25	3365	NE	Gothenburg	09/1894	PM	40 56	100 10
25	6570	NE	Pawnee City	11/1903	PM	40 06	96 09

Table 1 (continued)

State code	Sta. No.	State	Station name ^a	Beginning MO/YR ^b	TOBS ^c	Latitude ^d	Longitude ^e
26	9171	NV	Winnemucca WSO AP	07/1877	MD	40 54	117 48
28	0325	NJ	Atlantic City State Marina	01/1874	MD	39 23	74 26
28	6055	NJ	New Brunswick	01/1893	AM/PM	40 28	74 26
29	4369	NM	Jemez Springs	10/1910	PM	35 46	106 41
29	7867	NM	San Jon	06/1907	PM	35 07	103 20
30	0042	NY	Albany WSFO AP	01/1874	MD	42 45	73 48
30	0443	NY	Batavia	01/1932	PM/MD	42 59	78 11
30	0889	NY	Bridgehampton	08/1930	PM	40 57	72 18
30	3033	NY	Fredonia	01/1914	PM/MD	42 27	79 18
30	5801	NY	New York Central Park	01/1876	MD	40 47	73 58
30	8944	NY	Wanakena Ranger School	06/1910	PM	44 09	74 54
31	2635	NC	Edenton	02/1896	PM	36 03	76 37
31	5771	NC	Monroe 4SE	02/1896	PM	34 58	80 30
31	5890	NC	Mount Airy	01/1893	PM	36 31	80 37
32	2188	ND	Dickinson Exp. Station	01/1896	AM/PM	46 53	102 48
32	5660	ND	Mayville	03/1896	PM	47 30	97 19
33	1152	OH	Cadiz	09/1903	PM	40 16	81 00
33	9312	OH	Wooster Exp. Station	06/1893	PM/MD	40 47	81 56
34	4235	OK	Holdenville	01/1901	AM/PM	35 05	96 24
34	6638	OK	Okemah	04/1912	PM	35 26	96 18
35	2440	OR	Dufur	06/1910	PM	45 27	121 08
35	3445	OR	Grants Pass	01/1893	PM	42 25	123 20
36	6689	PA	Palmerton	05/1917	MD	40 48	75 37
38	0165	SC	Anderson	06/1901	PM	34 32	82 40
38	1549	SC	Charleston	01/1871	MD	32 47	79 56
39	0043	SD	Academy 2NE	09/1898	PM	43 30	99 04
39	1739	SD	Clark	01/1893	PM	44 53	97 44
39	2429	SD	Dupree	03/1922	PM	45 03	101 36
39	2927	SD	Faulkton 1NW	01/1893	PM	45 02	99 08
39	3217	SD	Gann Valley 2NW	01/1893	PM	44 03	99 02
39	3832	SD	Highmore 1W	01/1893	PM	44 31	99 28
40	7884	TN	Rogersville 1NE	02/1896	PM	36 25	82 59
40	9155	TN	Tullahoma	01/1896	PM	35 21	86 12
41	0120	TX	Albany	11/1901	AM/PM	32 44	99 17
41	1048	TX	Brenham	01/1897	AM	30 09	96 24
41	2019	TX	Corsicana	01/1893	AM	32 05	96 28
41	3873	TX	Hallettsville 2N	01/1893	PM	29 28	96 57

Table 1 (continued)

State code	Sta. No.	State	Station name ^a	Beginning MO/YR ^b	TOBS ^c	Latitude ^d	Longitude ^e
41	5018	TX	Lampasas	01/1897	AM	31 03	98 11
41	5429	TX	Luling	01/1897	AM	29 40	97 39
41	5869	TX	Mexia	09/1904	AM	31 41	96 29
41	9532	TX	Weatherford	10/1896	AM	32 46	97 49
42	0788	UT	Bluff	06/1911	PM	37 17	109 33
42	7909	UT	Snake Creek PH	12/1913	PM	40 33	111 30
42	8119	UT	Spanish Fork PH	07/1909	PM	40 05	111 36
42	9717	UT	Zion National Park	01/1904	PM	37 13	112 59
43	1081	VT	Burlington	01/1875	PM/MD	44 28	73 09
44	4101	VA	Hopewell	10/1916	PM	37 18	77 18
45	0945	WA	Buckley 1NE	01/1913	PM	47 10	122 00
45	6096	WA	Olga 2SE	07/1891	PM	48 37	122 48
45	6803	WA	Puyallup Exp. Station	01/1914	PM	47 12	122 20
45	7938	WA	Spokane WSO AP	02/1881	MD	47 38	117 32
46	1220	WV	Buckhannon 2W	10/1891	PM	39 00	80 16
46	3353	WV	Gary	02/1920	AM	37 22	81 33
47	3405	WI	Hancock Exp. Farm	11/1902	PM	44 07	89 32
47	5120	WI	Marshfield Exp. Farm	01/1913	PM	44 39	90 08
47	8027	WI	Spooner Exp. Farm	08/1894	PM	45 49	91 53
48	0915	WY	Border 3N	01/1902	PM	42 15	111 02
48	1675	WY	Cheyenne WSFO	11/1872	MD	41 09	104 49
48	1730	WY	Chugwater	11/1900	PM	41 45	104 49
48	1905	WY	Colony	01/1915	PM	44 56	104 12

^aAP = Airport
 Exp. = Experiment
 FAA = Federal Aviation Administration
 PH = Power House
 WMCE = Western Montana College of Education
 WSFO = Weather Service Forecast Office
 WSMO = Weather Service Meteorological Observatory
 WSO = Weather Service Office

^bBeginning date (month/year) of the daily maximum/minimum temperature data provided in the HCN/D data set. Precipitation data may begin on a different date.

^cTime of observation; represents the predominant category during which temperature readings are historically observed at the site: morning (AM), evening (PM), or midnight (MD). Combinations of these codes indicate sites at which the excursion from a constant TOBS exceeded the 4-year limit imposed by the selection criteria given in Sect. 5.

^dDegrees and minutes north.

^eDegrees and minutes west.

The first of these criteria was given the greatest weight, the importance of maintaining a consistent schedule for observing daily maximum and minimum temperature having been illustrated by several studies, such as Mitchell (1958), Baker (1975), and Schaal and Dale (1977). These studies examined the effects of changing observation time on the daily mean temperature, customarily determined for U.S. stations by adding the maximum and minimum temperature observed over a prescribed 24-hour observing day and dividing by 2. The systematic biases introduced to the daily means by varying observing times can have far-reaching effects, as the daily mean temperatures form the basis of monthly and annual mean temperature, and also monthly, seasonal, and annual heating degree days (HDD), cooling degree days (CDD), and growing degree days (GDD). Twenty-four (~17%) of the 138 HCN/D stations do not strictly meet the four-year limit for excursions from the predominant observing time. Fifteen stations experienced a change in observing time either from PM to midnight (MD) or MD to PM, and 9 stations changed either from AM to PM or PM to AM. These stations are identified in Table 1 by their combination of time of observation codes (herein referred to as "TOBS" codes). For example, Pocahontas 1, Arkansas has a TOBS code of "PM/AM" to denote a change in observing time *from PM to AM* at some point in its period of record. Information on the exact local standard time (LST) of maximum/minimum temperature observations at each station is contained in a station history file for the HCN/D which is described in Sect. 12. Further discussion of the effects of varying observing times on the homogeneity of climate records is contained in Sect. 7.

6. APPLICATIONS OF THE DATA

The temperature and precipitation records from the HCN/D are among the most accurate, internally consistent, and unbiased daily climate records available for the contiguous United States. These records represent many of the best stations from the HCN, and because of their daily resolution offer maximum flexibility for studies attempting to detect and monitor long-term climatic changes on a regional scale.

Particular applications of these data may include analyses of long-term trends in maximum/minimum temperatures, temperature extremes, daily temperature range, and precipitation amount. For example, Karl et al. (1984) used daily data from 130 mostly nonurban U.S. and Canadian stations to identify a statistically significant decrease in diurnal temperature range over the period 1941-1980. This decrease was shown to be linked to changes in precipitation (and probably humidity) by Karl et al. (1986b). Daily temperature data also allow analysis of the frequency of various threshold temperatures and the magnitude and frequency of temperature extremes, information that is critical in providing risk assessments in areas such as agriculture, natural resource exploration, and construction.

The HCN/D data base also provides opportunities for studying the variability of maximum/minimum temperature and precipitation amount on submonthly time scales. Guttman and Plantico (1987) analyzed daily maximum/minimum temperature data from 74 stations in the eastern half of the U.S. and found the published normals* and the 30-year

*Normals are defined in WMO (1984) as "period averages computed for a uniform and relatively long period comprising at least three consecutive 10-year periods." The published 1951-1980 daily normals of maximum and minimum temperature were prepared by interpolating between mean monthly values over the same period.

averages of daily maximum/minimum temperature at these stations to be different over the period 1951-1980. These differences were highlighted by "singularities" (warm/cold events), which often occurred in certain regions at a particular time of year (e.g., the often alluded to January thaw). Given the high quality and long-term nature of the HCN/D records, they would seem well suited for further investigations of this sort, perhaps those comparing temperature normals and averages over periods on the order of 100 years. As pointed out in Guttman and Plantico (1987), energy users/planners and farmers could benefit if the daily temperature climate could be more accurately described through the use of computations other than the published normals.

The HCN/D records of daily precipitation amount allow analyses of "event size" frequency and the magnitude and duration of wet and dry periods. They also could be used to precisely examine historical trends in the timing of seasonal changes in precipitation pattern, frequency, and accumulation. It seems likely that studies using daily data, such as those mentioned in the preceding, may be able to detect changes in regional climate that would not be apparent from analysis of the more commonly used monthly temperature and precipitation data.

7. LIMITATIONS AND RESTRICTIONS

The stations comprising the HCN/D are among the best stations from the HCN. They have been carefully selected, emphasizing the consistency of elements such as observing time, instrument type and location, and station surroundings. Nevertheless, no station is completely free of changes which are likely to effect its instrumental record, therefore it is recommended that users make full use of the information contained in the station histories when performing analyses with these data. While attempting to minimize the degree of nonclimatic effects at stations selected, the data have not been adjusted for station relocations, heat island effects, instrument changes, or time of observation biases. The nature of inhomogeneities arising from such factors depends on a station's climatic regime. Owing to the complexity of these problems, a significant amount of work remains to be done before daily climate data can be economically tested and adjusted.

As pointed out in Sect. 5, the criterion deemed most important in the station selection process was the degree to which a station maintained a constant observing time, i.e., a fixed observing "day", for maximum and minimum temperatures. At first-order National Weather Service (NWS) stations (several of which are included in the HCN/D), the 24-hour observing day ends at or near local midnight. Monthly and annual mean temperatures derived using the mean of the daily maximum and minimum temperatures from such stations have been shown by Baker (1975) and Mitchell (1958) to correspond closely with those computed using the stations' hourly observations. While this evidence lends clear support to the practice of ending the observing day at midnight, cooperative observers (comprising most of the HCN/D stations) generally do not take readings at this hour. Most end their observing day in the late afternoon or early evening, with a smaller but increasing number choosing a time between 0700 and 0800 LST. Combining daily temperature (or precipitation) data from stations which use different observing days complicates data compilation and quality control and also distorts areal patterns; however, of more fundamental importance in studies of climatic trends is the homogeneity of observing practices at individual stations. As noted in Sect. 5, despite systematic station selection efforts, twenty-four stations included in the HCN/D departed significantly from a

single, fixed observation time for maximum/minimum temperatures over their period of record (Table 1). The user is referred to the work of Mitchell (1958), Baker (1975), and Schaal and Dale (1977) for detailed illustrations of how such changes are likely to bias calculations involving maximum/minimum temperature data. Two main conclusions common to all three studies are (1) mean temperature calculations using 24-hour maximum/minimum temperatures from PM observations are biased high with respect to midnight observations, while those from AM observations are biased low, and (2) the magnitude of these biases is dependent upon time of year and a station's climatic regime.

Another factor of which users should be aware pertains to thermometers used at the HCN/D stations. In 1984, the NWS introduced a new Maximum/Minimum Temperature System (MMTS) at cooperative network observing stations. From 1983 to 1987, the percentage of HCN stations having an MMTS rose annually from 2% to 11, 24, 41, and 45%, respectively. As of 1987, 73 (~53%) of the HCN/D stations had MMTS installations. Concerns have arisen about the calibration of this system as compared to that of the earlier thermometric system. The new system is thermistor-based with a "beehive like" instrument shelter, whereas the older systems consisted of liquid-in-glass thermometers, mounted inside a Cotton Region Shelter (Stevenson Screen). Quayle et al. (1991) looked into performance differences of the two systems and found that the new system produces maximum temperatures ~0.3°C lower and minimum temperatures ~0.4°C higher than the old system. Unfortunately, because large samples of side-by-side overlapping measurements are not available, site-specific corrections cannot yet be derived and only large-scale temperature changes can be corrected. Furthermore, daily biases, which are likely to be dependent on synoptic conditions, are unlikely to be the same from day to day. Thus, to this time there has been no attempt to adjust the daily temperature data from the HCN/D for these instrument-induced biases.

Finally, it must be acknowledged that the spatial distribution of the stations making up the HCN/D (Fig. 1) is not completely uniform. Coverage is relatively dense in the central U.S. (13 stations in Illinois alone), and also in Montana (17 stations). In contrast, the extreme western U.S. is rather data sparse. Users should also note the lack of data from 5 states; Connecticut, Delaware, Kentucky, New Hampshire, and Rhode Island.

8. REFERENCES

- Baker, D. G. 1975. Effect of observation time on mean temperature estimation. *J. Appl. Meteor.* 14:471-76.
- Guttman, N. B., and M. S. Plantico. 1987. Climatic temperature normals. *J. Clim. Appl. Meteor.* 26:1428-35.
- Jones, P. D., S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly, and T. M. L. Wigley. 1986. Northern Hemisphere surface air temperature variations 1851-1984. *J. Clim. Appl. Meteor.* 25:161-79.
- Karl, T. R., G. Kukla, and J. Gavin. 1984. Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980. *J. Clim. Appl. Meteor.* 23:1489-1504.

- Karl, T. R., G. Kukla, and J. Gavin. 1986a. Relationship between decreased temperature range and precipitation trends in the United States and Canada, 1941-80. *J. Clim. Appl. Meteor.* 25:1878-86.
- Karl, T. R., C. N. Williams, Jr., P. J. Young, and W. M. Wendland. 1986b. A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. *J. Clim. Appl. Meteor.* 25:145-60.
- Karl, T. R., and C. N. Williams, Jr. 1987. An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Clim. Appl. Meteor.* 26:1744-63.
- Karl, T. R., H. F. Diaz, and G. Kukla. 1988. Urbanization: Its detection and effect in the United States climate record. *J. Clim.* 1:1099-1123.
- Karl, T. R., C. N. Williams, Jr., and F. T. Quinlan. 1990. United States Historical Climatology Network (HCN) serial temperature and precipitation data. ORNL/CDIAC-30, NDP-019/R1. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.
- Mitchell, J. M., Jr. 1958. Effects of changing observation time on mean temperature. *Bull. Amer. Meteor. Soc.* 39:83-89.
- National Aeronautics and Space Administration. 1991. NASA SAFISY-GEDEX Atmospheric Temperature Workshop Report, Columbia, Maryland, July 9-11, 1991 (in preparation).
- Quayle, R. G., D. R. Easterling, T. R. Karl, and P. J. Hughes. 1991. Effects of recent thermometer changes in the cooperative station network. *Bull. Amer. Meteor. Soc.* 72:1718-23.
- Quinlan, F. T., T. R. Karl, and C. N. Williams, Jr. 1987. United States Historical Climatology Network (HCN) serial temperature and precipitation data. NDP-019. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, Oak Ridge, Tenn.
- Schaal, L. A. and R. F. Dale. 1977. Time of observation temperature bias and "climatic change". *J. Appl. Meteor.* 16:215-22.
- Vinnikov, K. Ya., P. Ya. Groisman, and K. M. Lugina. 1990. Empirical data on contemporary global climate changes (temperature and precipitation). *J. Clim.* 3:662-67.
- World Meteorological Organization. 1984. Technical regulations, Vol. I. WMO-NO. 49. Geneva, Switzerland.

9. QUALITY ASSURANCE OF THE HCN/D DATA BASE

An important part of the numeric data packaging process at CDIAC is the quality assurance (QA) of data before distribution. Data received at CDIAC are rarely in perfect condition for immediate distribution, regardless of their source. To guarantee data of the highest possible quality, CDIAC conducts extensive QA reviews. Reviews involve examining the data for completeness, reasonableness, and accuracy. Although they have common objectives, these reviews are tailored to each data set, often requiring extensive programming efforts. Although time-consuming, the QA process is an important component in the value-added concept of ensuring accurate, usable data for researchers.

NCDC conducted extensive manual and automated QA assessments of the HCN/D data. Although the data sent by NCDC was in excellent condition, CDIAC still conducted QA checks on the data and found some minor discrepancies. The following summarizes the QA work performed by NCDC and CDIAC, respectively.

NCDC QA Checks and Adjustments

1. Monthly mean values of maximum and minimum temperature, computed from the HCN/D data, were compared to their respective unadjusted monthly means from the HCN. All conflicts were investigated and resolved, with verification based on manuscript or published sources.*
2. Checks were performed to ensure that no monthly mean values of maximum and minimum temperature calculated from a station's daily data were above (below) the monthly state extremes of maximum (minimum) temperature.
3. Any daily precipitation total exceeding 5 in. was verified against manuscript or published sources.
4. Checks were implemented to ensure that maximum temperatures were never less than minimum temperatures on the day of occurrence, the preceding day, and the following day. Conversely, checks were performed to ensure that minimum temperatures were never greater than maximum temperatures on the day of occurrence, the preceding day, and the following day.
5. Temperature data from stations that took readings during the morning over some period have been checked for any date shifting resulting from observers assigning readings to the calendar day of occurrence (the previous day in the case of maximum temperature) rather than the observation day. Such readings were switched back to the day of observance as part of the manual QA checks on the HCN/D data. Of the 14 stations in the HCN/D that now take only morning readings, the records of 10 stations were found to include instances in which the maximum temperature was

*The primary digital source of HCN values is the NCDC Summary of the Month (SOM, TD-3220); whereas the primary digital source of HCN/D data is the NCDC Summary of the Day (SOD, TD-3200). Differences between the data contained in these major archives arise from differences in the key entry procedures and validation techniques of these data bases.

entered on the calendar day of occurrence. Similar "shifting" occurred at 12 other stations during the brief periods for which these sites took morning readings as well. These identifiable periods of record are detailed in Table 2.

CDIAC QA Checks and Modifications

1. Because each record in an HCN/D file contains 31 daily data elements (to allow for 31 days in a month), elements pertaining to nonexistent dates were checked for missing data indicators with blank flag spaces (the prescribed conventions). Records for which this was not the case were modified accordingly. (Records occasionally contained artifacts such as a monthly mean maximum or minimum temperature occupying element 31 for months with less than 31 days.)
2. Several types of data source, measurement, and quality flags were found in the data that were not included in the documentation provided by NCDC. A list of these flags was submitted to NCDC. Several of these proved to be meaningless, resulting from data entry errors. Others were real data flags whose meanings were then related to CDIAC for documentation in this package.
3. The logical record length of the HCN/D files was shortened from 402 to 270 characters. This was accomplished by deleting unnecessary information and compressing the width of each daily data field.
4. All data records were checked to ensure that the number of days in the month (specified in each record) was correct for the year and month of each record. Eighteen February records from 18 different stations required correction; some specifying 29 days in non-leap years, others specifying only 28 days in leap years.

Table 2. HCN/D stations requiring calendar day shifting of maximum temperature observations for portions of their periods of record

State code	Sta. No.	State	Station name	Period of calendar day shifting ^a	Period length ^b (years/months)
02	0080	AZ	Ajo	03/01/21-12/31/23	2.10
				01/01/26-02/28/26	0.2
				01/01/35-12/31/41	6.2
02	3160	AZ	Fort Valley	01/01/37-12/31/40	4.0
02	5467	AZ	Mesa Exp. Farm	01/01/34-12/31/44	11.0
12	9253	IN	Washington	01/01/03-08/31/10	7.8
16	2151	LA	Covington 4 NNW	08/01/03-06/30/09	5.11
				01/01/14-08/31/18	4.8
24	7382	MT	Savage	09/01/09-09/30/11	2.1
25	1145	NE	Bridgeport	04/01/42-09/30/42	0.6
25	2820	NE	Fairbury 2 SSE	04/01/05-07/31/48	43.4
32	2188	ND	Dickinson Exp. Station	06/01/16-08/14/18	2.3
				04/01/19-06/30/48	29.3
32	5660	ND	Mayville	11/01/06-02/28/07	0.4
33	9312	OH	Wooster Exp. Station	01/10/38-12/31/41	4.0
34	4235	OK	Holdenville	09/01/03-11/30/12	9.3
				04/01/13-08/31/18	5.5
31	6638	OK	Okemah	04/01/14-11/30/18	4.8
36	6689	PA	Palmerton	05/01/17-06/30/17	0.2
				03/01/18-04/30/18	0.2
41	0120	TX	Albany	07/01/17-12/31/46	29.6
41	1048	TX	Brenham	06/01/03-10/31/18	15.5
				04/01/22-07/31/47	25.4

Table 2 (continued)

State code	Sta. No.	State	Station name	Period of calendar day shifting ^a	Period length ^b (years/months)
41	2019	TX	Corsicana	06/01/03-10/31/18	15.5
				05/01/19-10/31/19	0.6
				04/01/20-10/31/20	0.7
				04/01/21-06/30/47	26.3
41	5018	TX	Lampasas	06/01/03-12/31/07	4.7
				02/01/08-03/31/08	0.2
				05/01/08-10/17/18	10.6
				04/01/19-11/30/19	0.8
				06/01/21-09/30/21	0.4
04/01/22-06/30/47	25.3				
41	5429	TX	Luling	06/01/03-08/31/18	15.3
				05/01/21-09/30/21	0.5
				04/01/22-07/31/47	25.4
41	5869	TX	Mexia	09/01/04-08/31/18	14.0
				04/01/21-04/30/21	0.1
				04/01/22-06/30/47	25.3
41	9532	TX	Weatherford	06/01/03-09/30/18	15.4
				10/01/19-10/31/19	0.1
				01/01/23-12/31/23	1.0
				04/01/24-07/31/47	23.4
46	3353	WV	Gary	02/01/20-12/31/25	5.11

^aBeginning and ending dates of the period in month/day/year format.

^bTotal length of period in years and months, e.g., 5.11 indicates 5 years and 11 months. Periods involving a fraction of a month have their lengths rounded upward.

10. HOW TO OBTAIN THE PACKAGE

The HCN/D data files are available on request on a 9-track magnetic tape from CDIAC. This document and the magnetic tape are available free of charge. Requests for the magnetic tape should include any specific instructions for transmitting the data (e.g., 1600 or 6250 BPI density, labeled or non-labeled, ASCII or EBCDIC characters, and block size or record length constraints) required by the user to access the data. Requests not accompanied by specific instructions will be filled on 9-track, 6250 BPI, standard-labeled tapes with characters written in EBCDIC and files formatted as shown in Sect. 11. Because of the size of these data files, it is not possible to fill data requests for the entire data base described in NDP042 on floppy diskettes. It may be possible for users interested in receiving data only for select stations to receive data on IBM-formatted floppy diskettes. This will depend on the number of stations desired and the resulting file sizes. Requests should be addressed to:

Carbon Dioxide Information Analysis Center
Oak Ridge National Laboratory
Post Office Box 2008
Oak Ridge, Tennessee 37831-6335, U.S.A.

Telephone: (615) 574-0390
 FTS 624-0390

Fax: (615) 574-2232
 FTS 624-2232

Electronic Mail: BITNET eMail: CDP@ORNLSTC
 INTERNET: CDP@STC10.CTD.ORNL.GOV
 OMNET: CDIAC

PART 2
INFORMATION ABOUT THE MAGNETIC TAPE

11. CONTENTS OF THE MAGNETIC TAPE

The following is a list of files distributed on magnetic tape by CDIAC along with this documentation.

File number ^a and description	File size (kB)	Logical records	Record format	Block size	Record length
1. General descriptive information file	41.81	630	FB ^b	6400	128
2. FORTRAN IV retrieval code to read and print the HCN/D station inventory (File 8)	2.67	33	FB	8000	80
3. FORTRAN IV retrieval code to read and print the HCN/D station histories (File 9)	1.05	13	FB	8000	80
4. FORTRAN IV retrieval code to read and print any of the HCN/D data files (Files 10-14)	3.08	38	FB	8000	80
5. SAS ^c retrieval code to read and print the HCN/D station inventory (File 8)	1.62	20	FB	8000	80
6. SAS retrieval code to read and print the HCN/D station histories (File 9)	2.10	26	FB	8000	80
7. SAS retrieval code to read and print any of the HCN/D data files (Files 10-14)	3.48	43	FB	8000	80
8. HCN/D station inventory	10.07	138	FB	7200	72
9. HCN/D station histories	399.28	1840	FB	5400	216
10. HCN/D temperature and precipitation data from states with codes 1-10 (AL-ID; 18 stations)	14,696.60	54,231	FB	5400	270

File number and description	File size (kB)	Logical records	Record format	Block size	Record length
11. HCN/D temperature and precipitation data from states with codes 11-20 (IL-MI; 34 stations)	30,518.67	112,615	FB	5400	270
12. HCN/D temperature and precipitation data from states with codes 21-30 (MN-NY; 37 stations)	31,054.16	114,591	FB	5400	270
13. HCN/D temperature and precipitation data from states with codes 31-40 (NC-TN; 22 stations)	18,999.27	70,108	FB	5400	270
14. HCN/D temperature and precipitation data from states with codes 41-48 (TX-WY; 27 stations)	22,414.68	82,711	FB	5400	270
Total size and records	118,148.54	437,037			

^aFile numbers are for 6250 BPI density tapes. File numbers would differ for 1600 BPI density tapes.

^bFB = fixed block.

^cSAS is a registered trademark of SAS Institute, Inc., Cary, North Carolina 27511-8000.

12. DESCRIPTIVE FILE ON THE MAGNETIC TAPE

The following is a listing of the first file on the magnetic tape distributed by CDIAC. This file is intended to provide the details (i.e., variable descriptions, formats, units, etc.) of each file associated with this numeric data package.

TITLE OF THE DATA SET

United States Historical Climatology Network Daily Temperature and Precipitation Data

DATA CONTRIBUTORS

P. Y. Hughes, E. H. Mason, T. R. Karl, and W. A. Brower
National Climatic Data Center
National Oceanic and Atmospheric Administration
Asheville, North Carolina 28801

SOURCE AND SCOPE OF THE DATA

The 138 station HCN/D data base contains station histories, daily maximum and minimum temperatures, and daily precipitation amounts that were compiled by NCDC after being extracted from digital and non-digital data sets archived at NCDC. These data sets come from a variety of sources, including climatological publications, universities, federal agencies, individuals, and data archives. The HCN/D stations were selected from the HCN so as to provide a reasonably homogeneous spatial distribution of stations within the contiguous United States after consideration of the following.

1. The degree to which each station maintained a constant observation time for maximum and minimum temperatures. This criterion was given the greatest weight, excursions from a station's predominant observing time of no more than 4 years being desired. (For 24 (~17%) of the 138 stations this condition is not met; 15 of the 24 stations changed their observing time either from PM to midnight (MD) or MD to PM, and 9 stations changed either from AM to PM or PM to AM.)
2. At least 95% of a station's pre-1951 data should be contained in NCDC digital daily archives.
3. A station's potential for heat island bias over time should be low.
4. Quality assessments based upon the decile ranking assigned by Karl et al. (1990) to the stations' *monthly* maximum/minimum temperature data for the quality characteristics numbered 5-8 in Appendix A. A station should rank within the lower 9 deciles for each characteristic.

DATA FORMAT

The information on this tape is arranged in fourteen files, containing the following:

- this descriptive file
- a FORTRAN IV Input/Output (I/O) routine for the HCN/D station inventory file
- a FORTRAN IV I/O routine for the HCN/D station history file
- a FORTRAN IV I/O routine for the HCN/D data files
- a SAS* I/O routine for the HCN/D station inventory file
- a SAS I/O routine for the HCN/D station history file
- a SAS I/O routine for the HCN/D data files
- the HCN/D station inventory file
- the HCN/D station history file
- five files containing HCN/D temperature and precipitation data

The format and contents of each data file are described in the following.

STATION INVENTORY FILE FOR THE HCN/D DATA SET

The station inventory file for the HCN/D data set (File 8 on the magnetic tape) is sorted by two-digit state code and four-digit Cooperative Network Index, with one record per station containing state code, Cooperative Network Index, state abbreviation, station name, beginning month and year of data, time of observation, latitude, and longitude.

The file may be read using the following FORTRAN format:

```

      INTEGER BYEAR,LATDEG,LATMIN,LONDEG,LONMIN
      CHARACTER*2 STCODE,STATE,BMON
      CHARACTER*4 CNI
      CHARACTER*23 STNAME
      CHARACTER*5 TOBS
      READ(5,100,END=99)STCODE,CNI,STATE,STNAME,BMON,BYEAR,TOBS,
+   LATDEG,LATMIN,LONDEG,LONMIN
100  FORMAT(1X,A2,2X,A4,2X,A2,2X,A23,2X,A2,2X,I4,2X,A5,
+   2X,I2,2X,I2,2X,I3,2X,I2)

```

or by using the SAS format:

```

DATA INVENT;
LENGTH STNAME $ 23;
INFILE IN;
INPUT STCODE $ 2-3 CNI $ 6-9 STATE $ 12-13 STNAME 16-38 BMON $ 41-42
      BYEAR $ 45-48 TOBS $ 51-55 LATDEG 58-59 LATMIN 62-63
      LONDEG 66-68 LONMIN 71-72;

```

*SAS is a registered trademark of SAS Institute, Inc., Cary, North Carolina 27511-8000.

Stated in tabular form, the contents of the station inventory file include the following.

Variable	Variable type	Variable width	Starting column	Ending column
STCODE	Character	2	2	3
CNI	Character	4	6	9
STATE	Character	2	12	13
STNAME	Character	23	16	38
BMON	Character	2	41	42
BYEAR	Numeric	4	45	48
TOBS	Character	5	51	55
LATDEG	Numeric	2	58	59
LATMIN	Numeric	2	62	63
LONDEG	Numeric	3	66	68
LONMIN	Numeric	2	71	72

where

- STCODE is the two-digit state code (01–48), defined as character to allow for preserving leading zeros upon output;
- CNI is the four-digit Cooperative Network Index, defined as character above to allow for preserving leading zeros upon output;
- STATE is the two-letter state abbreviation;
- STNAME is the station name;
- BMON is the beginning month of the daily maximum/minimum temperature record for a station;
- BYEAR is the beginning year of the daily maximum/minimum temperature record for a station. Precipitation data may begin in a different year;
- TOBS is the time of observation; the predominant time at which temperature readings are historically taken at the site: morning (AM), evening (PM), or midnight (MD). Combinations of these codes indicate sites at which the excursion from a constant TOBS exceeded the 4-year limit imposed by the selection criteria;
- LATDEG is the degrees (north) portion of the station's latitude;
- LATMIN is the minutes portion of the station's latitude;

LONDEG is the degrees (west) portion of the station's longitude; and

LONMIN is the minutes portion of the station's longitude.

STATION HISTORY FILE

The station history file (File 9 on the tape) provides valuable information concerning each station in the HCN/D. This file documents station moves and instrument changes, lists station observers and observation times, and identifies suspect fields.

The file may be read using the following FORTRAN format:

```
DIMENSION DATA(54)
READ(5,100,END=99) (DATA(I),I=1,54)
100 FORMAT(54A4)
```

or using the SAS format:

```
DATA HISTORY (DROP=X);
RETAIN STANUM STATE DIVISION CURRNAME COUNTY XREF;
INFILE IN MISCOVER LS=216;
INPUT @45 X $1. @;
IF X NE '' THEN DO;
INPUT STANUM 1-6 STATE $ 8-9 DIVISION 11-12 CURRNAME $ 14-43
COUNTY $ 45-60 XREF $ 62-86 ;
END;
ELSE
INPUT STANUM2 1-6 MOBEG 8-9 DAYBEG 11-12 YRBEG 14-17 MOEND 19-20
DAYEND 22-23 YREND 25-28 SUSPLAT 30 SUSPLONG 31 SUSPLOC 32
SUSPELEV 33 SUSPPO 34 SUSPNAME 35 SUSPQUAL 36 SUSPINST 37
SUSPTIME 38 SUSPHTS 39 SUSPPUBS 40 SUSPBEG 41 SUSPEND 42
SUSPOBS 43 SUSPOTHR 44 LATNORTH $ 46-51 LONGWEST $ 53-59
DISTANCE 61-63 DIRECT $ 65-67 ELEV 69-73 DISTPO 75-77
DIRECTPO $ 79-81 NAME $ 83-110 QUALIF $ 112-121 ADDINST 123
COTTON 124 DBULB 125 EVAPSTA 126 FISHPORT 127 HYGRO 128
MINTHERM 129 MAXTHERM 130 NORIVGAG 131 RAINGAGE 132
SHELTER 133 RECRIVER 134 RECRRAIN 135 SNOW 136 STORAGE 137
STDRAIN 138 STDSELT 139 THERMOGR 140 DIGTHERM 141
TIPBUCK 142 OTHEVAP 143 MAXMIN 144 TIMEOBS $ 146-149 PCPHT $
151-152 PCTHT $ 154-155 BULLETW 157 COMBBUL 158 CLIMDATA 159
RIVSTAGE 160 HYDROBUL 161 PRECDATA 162 SNOWBULL 163
NOTPUB 164 CWB 165 MONTHREV 166 STATEPUB 167 LCD 168 BQ 169
SGPD 170 WWR 171 MYB 172 OBSNAME $ 174-213 NUMOBS 215-216;
```

Stated in tabular form, the contents of the station history file include the following.

Variable	Variable type	Variable width	Starting column	Ending column
X	Alphanumeric	1	45	45
STANUM	Numeric	6	1	6
STATE	Character	2	8	9
DIVISION	Numeric	2	11	12
CURRNAME	Alphanumeric	30	14	43
COUNTY	Alphanumeric	16	45	60
XREF	Alphanumeric	25	62	86
STANUM2	Numeric	6	1	6
MOBEG	Numeric	2	8	9
DAYBEG	Numeric	2	11	12
YRBEG	Numeric	4	14	17
MOEND	Numeric	2	19	20
DAYEND	Numeric	2	22	23
YREND	Numeric	4	25	28
SUSPLAT	Numeric	1	30	30
SUSPLONG	Numeric	1	31	31
SUSPLOC	Numeric	1	32	32
SUSPELEV	Numeric	1	33	33
SUSPPO	Numeric	1	34	34
SUSPNAME	Numeric	1	35	35
SUSPQUAL	Numeric	1	36	36
SUSPINST	Numeric	1	37	37
SUSPTIME	Numeric	1	38	38
SUSPHTS	Numeric	1	39	39
SUSPPUBS	Numeric	1	40	40
SUSPBEG	Numeric	1	41	41
SUSPEND	Numeric	1	42	42
SUSPOBS	Numeric	1	43	43
SUSPOTHR	Numeric	1	44	44
LATNORTH	Alphanumeric	6	46	51
LONGWEST	Alphanumeric	7	53	59
DISTANCE	Numeric	3	61	63
DIRECT	Alphanumeric	3	65	67
ELEV	Numeric	5	69	73
DISTPO	Numeric	3	75	77
DIRECTPO	Alphanumeric	3	79	81
NAME	Character	28	83	110
QUALIF	Alphanumeric	10	112	121
ADDINST	Numeric	1	123	123
COTTON	Numeric	1	124	124

Variable	Variable type	Variable width	Starting column	Ending column
DBULB	Numeric	1	125	125
EVAPSTA	Numeric	1	126	126
FISHPORT	Numeric	1	127	127
HYGRO	Numeric	1	128	128
MINTHERM	Numeric	1	129	129
MAXTHERM	Numeric	1	130	130
NORIVGAG	Numeric	1	131	131
RAINGAGE	Numeric	1	132	132
SHELTER	Numeric	1	133	133
RECRIVER	Numeric	1	134	134
RECRAIN	Numeric	1	135	135
SNOW	Numeric	1	136	136
STORAGE	Numeric	1	137	137
STDRAIN	Numeric	1	138	138
STDSHELT	Numeric	1	139	139
THERMOGR	Numeric	1	140	140
DIGTHERM	Numeric	1	141	141
TIPBUCK	Numeric	1	142	142
OTHEVAP	Numeric	1	143	143
MAXMIN	Numeric	1	144	144
TIMEOBS	Alphanumeric	4	146	149
PCPHT	Alphanumeric	2	151	152
PCTHT	Alphanumeric	2	154	155
BULLETW	Numeric	1	157	157
COMBBUL	Numeric	1	158	158
CLIMDATA	Numeric	1	159	159
RIVSTAGE	Numeric	1	160	160
HYDROBUL	Numeric	1	161	161
PRECDATA	Numeric	1	162	162
SNOWBULL	Numeric	1	163	163
NOTPUB	Numeric	1	164	164
CWB	Numeric	1	165	165
MONTHREV	Numeric	1	166	166
STATEPUB	Numeric	1	167	167
LCD	Numeric	1	168	168
BQ	Numeric	1	169	169
SGPD	Numeric	1	170	170
WWR	Numeric	1	171	171
MYB	Numeric	1	172	172
OBSNAME	Alphanumeric	40	174	213
NUMOBS	Numeric	2	215	216

where

- X** is a dummy variable used in the above SAS program to differentiate header records from data records;
- STANUM** is the station identification number, composed of the two-digit state code followed by the four-digit Cooperative Network Index;
- STATE** is the two-letter state abbreviation;
- DIVISION** is the station division number;
- CURRNAME** is the most current station name;
- COUNTY** is the county in which the station is currently located;
- XREF** is a station cross-reference, representing the cooperative network index of the station or the county name that the current station moved to or from;
- STANUM2** is the station identification number, composed of the two-digit state code followed by the four-digit Cooperative Network Index;
- MOBEG** is the month the data record started (missing values are represented by 99);
- DAYBEG** is the day the data record started (missing values are represented by 99);
- YRBEG** is the year the data record started;
- MOEND** is the month the data record ended (missing values are represented by 99);
- DAYEND** is the day the data record ended (missing values are represented by 99);
and
- YREND** is the year the data record ended (missing values are represented by 9999).

The next 15 variables represent suspect fields in the station history file. The values for these variables will be either 0 or 1. Values of 1 represent fields flagged as suspect by the pre-key editor.

- | | |
|-------------|----------------------|
| 1. SUSPLAT | Latitude |
| 2. SUSPLONG | Longitude |
| 3. SUSPLOC | Previous location |
| 4. SUSPELEV | Elevation |
| 5. SUSPPO | Post office location |
| 6. SUSPNAME | Station name |
| 7. SUSPQUAL | Qualifier |
| 8. SUSPINST | Instruments |
| 9. SUSPTIME | Observation time |

10. SUSPHTS	Instrument heights
11. SUSPPUBS	Publications
12. SUSPBEG	Beginning date
13. SUSPEND	Ending date
14. SUSPOBS	Observer
15. SUSPOTHR	Other observers

LATNORTH is the current station latitude expressed in degrees and minutes north;

LONGWEST is the current station longitude expressed in degrees and minutes west;

DISTANCE is the distance, in tenths of miles, from the previous station location (e.g., 015 = 1.5 miles), with unknown distances represented by 999;

DIRECT is the direction (16 point) of a station move from the previous location. The location of the temperature instrument defines the official station location. Values may be blank, character, or numeric. Unknown direction is represented by 999. Some examples of DISTANCE and DIRECT combinations are:

999 999 = first record of new station or distance and direction unknown;
 015 NW = station moved 1.5 miles NW from previous location;
 000 000 = no change in station (or instrument) location;
 000 ESE = moved <0.1 mile east-southeast (ESE) from previous location;
 000 999 = moved <0.1 mile, direction unknown;
 902 ESE = temperature instrument moved 0.2 miles ESE and precipitation instrument either did not move or was moved to a location different than that of the temperature instrument;
 800 000 = precipitation instrument moved <0.1 mile, but the temperature instrument did not move; and
 999 NW = distance unknown, direction NW;

ELEV is the ground elevation at the station, expressed in whole feet above or below mean sea level;

DISTPO is the distance, in tenths of miles, from the nearest post office (e.g. 015 = 1.5 miles), with unknown distances represented by 999;

DIRECTPO is the direction on a 16-point compass from the nearest post office. Values may be either blank, character, or numeric. Unknown directions are represented by 999. Some examples of DISTPO and DIRECTPO combinations are:

999 999 = distance and direction unknown;
 015 NW = 1.5 miles NW of post office;
 000 NW = <0.1 mile NW from post office;
 000 999 = <0.1 mile from post office, direction unknown; and

000 000 = at the post office.

NAME is the full station name; and

QUALIF is a qualifier or description that is added to the proper name of the station (e.g., Charleston 2WNW).

The next 22 variables represent the following instruments and classifications. If an instrument was used at a particular station or if a particular classification is appropriate for that station, the variable will have a value of 1; if it was not used, the variable will have a value of 0.

1. ADDINST	Additional instrument (wind, pressure, etc.)
2. COTTON	Cotton region shelter (official, CRS)
3. DBULB	Dry bulb thermometer
4. EVAPSTA	Class "A" evaporation station
5. FISHPORT	Fisher-Porter gage
6. HYGRO	Hygrothermograph
7. MINTHERM	Minimum thermometer
8. MAXTHERM	Maximum thermometer
9. NORIVGAG	Nonrecording river gage
10. RAINGAGE	Nonstandard rain gage
11. SHELTER	Nonstandard shelter
12. RECRIVER	Recording river gage
13. RECRAIN	Recording rain gage
14. SNOW	Snow density gage
15. STORAGE	Storage gage
16. STDRAIN	Standard rain gage (SRG)
17. STDSHELT	Standard shelter (official)
18. THERMOGR	Thermograph
19. DIGTHERM	Digital thermometer
20. TIPBUCK	Tipping bucket gage
21. OTHEVAP	Other than class "A" evaporation station
22. MAXMIN	Max/min temperature system

TIMEOBS are the observation times (2 characters each) for precipitation and temperature, respectively, if both times are known. Values may be either numeric (rounded to the nearest whole hour), character, or alphanumeric. Codes which relate to one or both of the times may also be present. Possible values and their meanings include the following:

0719	= precipitation amount read at 0700 LST (local standard time), temperatures read at 1900 LST;
SRSS	= precipitation amount read at sunrise, temperatures read at sunset;
SS99	= precipitation amount read at sunset, time of temperature observations either unknown or no temperature data was available for that period of the record;

06HR	= station observed 6 hours per day (not to be confused with a 6-hourly synoptic observing schedule). How these observations were used to produce daily precipitation amount and maximum/minimum temperatures is unclear;
9079	= ambiguous form; station records only gave one observation time (0700 LST), but it is unknown if this time applies to both precipitation and temperature;
TRID	= Tri-daily temperature observations (TAVG = [7AM + 2PM + (2 x 9PM)]/4), but time of observation for precipitation amount is unknown; and
RSSS	= Precipitation amounts read on a rotating schedule (SR during crop season, i.e., April/May–October/November, but SS otherwise), temperatures read at sunset;
PCPHT	is the height of the precipitation instrument above ground level. Values may be numeric or character, with numeric values expressed to the nearest whole foot; and
PCTHT	is the height of the temperature instrument above ground level. Values may be numeric or character, with numeric values expressed to the nearest whole foot. Potential values for both PCPHT and PCTHT include the following:
01-97	= actual height;
98	= ≥ 98 feet;
99	= missing; and
RF	= roof, actual height above ground level unknown.

The next 16 variables represent the following forms of publications. If the data from a particular station appeared in a publication, the variable will have a value of 1; if not, the variable will have a value of 0. The variables and their corresponding forms of publications are as follows:

1. BULLETW	Bulletin W
2. COMBBUL	<i>Combined Bulletin</i>
3. CLIMDATA	<i>Climatological Data</i>
4. RIVSTAGE	<i>Daily River Stages</i>
5. HYDROBUL	<i>Hydrologic Bulletin</i>
6. PRECDATA	published as hourly precipitation data
7. SNOWBULL	<i>Snow Bulletin</i>
8. NOTPUB	not published
9. CWB	Report to the chief of the U.S. Weather Bureau
10. MONTHREV	<i>Monthly Weather Review</i>
11. STATEPUB	published in state publications
12. LCD	<i>Local Climatological Data</i>
13. BQ	<i>Bulletin Q</i> , 1870–1903.
14. SGPD	<i>Storage Gage Precipitation Data, Western United States</i>
15. WWR	<i>Weekly Weather Review</i>
16. MYB	<i>U.S. Meteorological Yearbook</i>

OBSNAME is the observer's name (may include more than one name per record);

NUMOBS is the number of observers participating during the time of record for an agency.

HCN/D DATA FILES

The HCN/D data files (Files 10-14 on the magnetic tape) contain daily maximum and minimum temperatures ($^{\circ}$ F), precipitation amounts (hundredths of inches), and data flags from the 138 HCN/D stations. The files are sorted by six-digit station number (the two-digit state code followed by the four-digit Cooperative Network Index), year, and month, with one record per month containing station number, data type, year, month, number of days in the month, and 31 daily data values with their respective flags. The data are divided among the five files according to state code (see Sect. 11. for an exact breakdown).

The files may be read using the following FORTRAN format:

```

      INTEGER YEAR,MON,DAYS,VALUE(31)
      CHARACTER*1 SF(31),DMF(31),DQF(31)
      CHARACTER*2 STCODE
      CHARACTER*4 DATTYP
      CHARACTER*6 STAID
      NREC=0
      1 CONTINUE
      READ(5,100,END=99) STAID,DATTYP,YEAR,MON,DAYS,
      + (SF(I),VALUE(I),DMF(I),DQF(I),I=1,31)
      100 FORMAT(1X,A6,1X,A4,1X,I4,I2,1X,I2,31(1X,A1,I4,2A1))

```

or by using the SAS format:

```

DATA HCND;
  ARRAY DAY {31} $ DAY1-DAY31;
  INFILE IN LRECL=270;
  INPUT @2 STAID $ 2-7 DATTYP $ 9-12 YEAR 14-17 MON 18-19
        DAYS 21-22 @23 (DAY1-DAY31) ($CHAR8.);

```

(The respective flag and data values contained in each of the 31 elements of the array DAY in the SAS format may be extracted using the SAS code contained in File 7 on the magnetic tape and also printed in Sect. 14.)

Stated in tabular form (using variable names from the FORTRAN format), the contents of an HCN/D data file include the following.

Variable	Variable type	Variable width	Starting column	Ending column
STAID	Character	6	2	7
DATTYP	Character	4	9	12
YEAR	Numeric	4	14	17
MON	Numeric	2	18	19
DAYS	Numeric	2	21	22
SF(1)	Alphanumeric	1	24	24
VALUE(1)	Numeric	4	25	28
DMF(1)	Alphanumeric	1	29	29
DQF(1)	Alphanumeric	1	30	30
SF(2-31)	Alphanumeric	1	*	*
VALUE(2-31)	Numeric	4	*	*
DMF(2-31)	Alphanumeric	1	*	*
DQF(2-31)	Alphanumeric	1	*	*

*May be obtained using: $COL(N) = COL(1) + (N * 8) - 8$, where $COL(N)$ is the starting/ending column for $SF(N)$, $VALUE(N)$, $DMF(N)$, or $DQF(N)$; $COL(1)$ is the starting/ending column for $SF(1)$, $VALUE(1)$, $DMF(1)$, or $DQF(1)$; and N is the day of the month (2-31).

where

- STAID is the station identification number, composed of the two-digit state code followed by the four-digit Cooperative Network Index (defined as character to preserve leading zeros upon output);
- DATTYP is the data type (TMAX = maximum temperature, TMIN = minimum temperature, and PRCP = precipitation amount). Some stations do not always have records for all three data types in a given month;
- YEAR is the year of the data;
- MON is the month of the data;
- DAYS is the number of days in the month;
- SF(1-31) are the source flags for the daily data values;
- VALUE(1-31) are daily data values, with temperatures in whole degrees Fahrenheit and precipitation amounts in hundredths of inches;

DMF(1-31) are the data measurement flags for the daily data values; and

DQF(1-31) are the data quality flags for the daily data values.

Flag codes for the HCN/D data

SF is a code indicating the source of the daily data value. The codes and their meanings are as follows:

- 0 = NCDC Tape Deck 3200, *Summary of the Day* Element Digital File;
- 3 = Manuscript-Original Records, NCDC;
- 4 = *Climatological Data (CD)* (monthly NCDC publication);
- 5 = *Climate Record Book*; as described within: *History of Climatological Records Books*, U.S. Department of Commerce, Weather Bureau, U.S. Government Printing Office (1960);
- Blank = manually estimated (see DQF flag) or missing data value.

DMF is the data measurement flag, which describes how the daily value was measured. The codes and their meanings are as follows:

- A = amount of accumulated precipitation since last measurement;
- B = amount of accumulated precipitation since last measurement (includes estimated values);
- E = manual or automated estimated value (see DQF flag for the particular estimation procedure);
- I = value determined by spatial interpolation using data from surrounding HCN stations;
- S = data value is included in a subsequent value;
- T = Trace of precipitation (data value should equal 0 for a trace); and
- Blank = valid original data (no flag needed) or missing data value.

DQF is the data quality flag. In January 1982, NCDC instituted a greatly enhanced computer algorithm for automated validation of digital data archives. The system checks the internal consistency of a station's data and compares each station's observations to prescribed climatological limits and observations from surrounding stations. Numeric DQF codes apply only to NCDC's digital data, i.e., where the source flag (SF) is equal to "0" for a particular value. Alphabetic codes describe the particular manual or automated NCDC procedure employed to correct or estimate a data value. The codes and their meanings are as follows:

- 0 = valid data;
- 1 = valid data (Pre-1982 quality control methods were employed, with only a gross check of the magnitude of the value.);

- 3 = invalid data—no edited data value available;
- 4 = validity unknown—automated quality control procedures have not been applied;
- A = substituted temperature from time of observation for TMAX or TMIN;
- B = time-shifted value;
- F = adjusted TMAX or TMIN by a multiple of $\pm 10^\circ$;
- L = switched TMAX and TMIN;
- M = switched temperature from time of observation with TMAX or TMIN;
- N = substituted the mean of values taken from the three nearest cooperative weather stations;
- O = snow and precipitation columns were switched in station's report;
- R = precipitation amount was not reported, "0" has been inserted;
- S = manually edited value (derived using one of the procedures described by data quality flags A–R);
- T = data value failed internal consistency check; and
- Blank = valid data value with source flag other than "0" or missing data value.

REFERENCE

Karl, T. R., C. N. Williams, Jr., and F. T. Quinlan. 1990. United States Historical Climatology Network (HCN) serial temperature and precipitation data. ORNL/CDIAC-30, NDP-019/R1. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

13. LISTINGS OF THE FORTRAN IV DATA RETRIEVAL PROGRAMS

The following is a listing of the FORTRAN IV data retrieval program provided on magnetic tape (File 2 on the tape) by CDIAC for reading and printing the station inventory file for the HCN/D data set (File 8 on the tape). The job control language (JCL) statements (preceded by // or /*) shown in the following are not provided in the file on the magnetic tape. The JCL statements required will vary for each operating system. The JCL statements in the following are provided to illustrate the statements required by an individual at ORNL who has requested these data on a 9-track, 6250 BPI, standard-labeled tape with characters written in EBCDIC and who is attempting to read the tape on an IBM mainframe (e.g., IBM 3090).

```
//UID=INV JOB(12345),'USER ADDRESS'
//OUT OUTPUT DEFAULT=YES,JESDS=ALL,DEST=LOCAL
// EXEC FORTVCLG
//FORT.SYSIN DD *
C FORTRAN RETRIEVAL CODE FOR READING AND PRINTING THE HCN/D
C STATION INVENTORY FILE (FILE 8 ON THE MAGNETIC TAPE)
  INTEGER BYEAR,LATDEG,LATMIN,LONDEG,LONMIN
  CHARACTER*2 STCODE,STATE,BMON
  CHARACTER*4 CNI
  CHARACTER*23 STNAME
  CHARACTER*5 TOBS
C INITIALIZE A RECORD COUNTER
  NREC=0
  1 CONTINUE
  READ(5,100,END=99) STCODE,CNI,STATE,STNAME,BMON,BYEAR,
  + TOBS,LATDEG,LATMIN,LONDEG,LONMIN
  100 FORMAT(1X,A2,2X,A4,2X,A2,2X,A23,2X,A2,2X,I4,2X,A5,
  + 2X,I2,2X,I2,2X,I3,2X,I2)
C
C WRITE OUT A HEADER ON EACH PAGE OF THE LISTING.
  IF(MOD(NREC,58).EQ.0) WRITE(6,200)
  200 FORMAT('1',' ST',' CNI ',' ST',' STATION NAME',13X,'MO YEAR',
  + ' TOBS ',' DEG MIN DEG MIN',/)
C
  WRITE(6,100) STCODE,CNI,STATE,STNAME,BMON,BYEAR,TOBS,
  + LATDEG,LATMIN,LONDEG,LONMIN
  NREC=NREC+1
  GO TO 1
  99 CONTINUE
  STOP
  END
/*
//GO.FT05F001 DD UNIT=TAPE62,VOL=SER=TAPENUMB,DISP=(,PASS),
// DSN=MFF.NDP042.INVENTORY.DATA,LABEL=(8,SL,EXPDT=98000)
//GO.FT06F001 DD *
//
```

The following is a listing of the FORTRAN IV data retrieval program provided on magnetic tape (File 3 on the tape) by CDIAC for reading and printing the HCN/D station history file (File 9 on the tape). The JCL statements (preceded by // or /*) shown in the following are not provided in the file on the magnetic tape. The JCL statements required will vary for each operating system. The JCL statements in the following are provided to illustrate the statements required by an individual at ORNL who has requested these data on a 9-track, 6250 BPI, standard-labeled tape with characters written in EBCDIC and who is attempting to read the tape on an IBM mainframe (e.g., IBM 3090).

```
//UIDHIS JOB (12345),'USER ADDRESS'
//OUT OUTPUT DEFAULT=YES,JESDS=ALL,DEST=LOCAL
// EXEC FORTVCLG
//FORT.SYSIN DD *
C
C FORTRAN RETRIEVAL CODE TO READ AND PRINT THE HCN STATION
C HISTORY FILE (FILE 9 ON THE MAGNETIC TAPE)
C
  DIMENSION DATA(54)
1  CONTINUE
  READ(5,100,END=99) (DATA(I),I=1,54)
100 FORMAT(54A4)
  WRITE(6,101) (DATA(I),I=1,54)
101 FORMAT(1X,32A4,/,1X,22A4)
  GO TO 1
99  CONTINUE
  STOP
  END
/*
//GO.FT05F001 DD UNIT=TAPE62,VOL=SER=TAPEVOL1,DISP=(,PASS),
// DSN=MFF.NDP042.STATNHIS.DATA,LABEL=(9,SL)
//GO.FT06F001 DD *
//
```

The following is a listing of the FORTRAN IV data retrieval program provided on magnetic tape (File 4 on the tape) by CDIAC for reading and printing any of the HCN/D data files (Files 10-14 on the tape). The JCL statements (preceded by // or /*) shown in the following are not provided in the file on the magnetic tape. The JCL statements required will vary for each operating system. The JCL statements in the following are provided to illustrate the statements required by an individual at ORNL who has requested these data on a 9-track, 6250 BPI, standard-labeled tape with characters written in EBCDIC and who is attempting to read the tape on an IBM mainframe (e.g., IBM 3090).

```

//UIDHCND JOB(12345),'USER ADDRESS'
//OUT OUTPUT DEFAULT=YES,JESDS=ALL,DEST=LOCAL
// EXEC FORTVCLG
//FORT.SYSIN DD *
C *** FORTRAN RETRIEVAL CODE TO READ AND PRINT ANY OF THE
C *** HCN/D DATA FILES (FILES 10-14 ON THE MAGNETIC TAPE)
      INTEGER YEAR,MON,DAYS,VALUE(31)
      CHARACTER*1 SF(31),DMF(31),DQF(31)
      CHARACTER*2 STCODE
      CHARACTER*4 DATTYP
      CHARACTER*6 STAID
      NREC=0
      1 CONTINUE
      READ(5,100,END=99) STAID,DATTYP,YEAR,MON,DAYS,
      + (SF(I),VALUE(I),DMF(I),DQF(I),I=1,31)
      100 FORMAT(1X,A6,1X,A4,1X,I4,I2,1X,I2,31(1X,A1,I4,2A1))
C *** STRIP OFF STATE CODE FROM STATION ID
      STCODE=STAID(1:2)
C
C *** PRINT A HEADER AT THE TOP OF EACH OUTPUT PAGE
      IF(MOD(NREC,19).EQ.0) WRITE(6,201)
      201 FORMAT('1', STAID TYPE YR/MO 1/14/27|2/15/28|3/16/29|',
      + '4/17/30|5/18/31| 6/19 | 7/20 | 8/21 | 9/22 |',
      + ' 10/23 | 11/24 | 12/25 | 13/26',/,
      + 23X,12('S MQ|'),'S MQ',/,23X,12(' |'))
      WRITE(6,200) STAID,DATTYP,YEAR,MON,DAYS,
      + (SF(I),VALUE(I),DMF(I),DQF(I),I=1,31)
      200 FORMAT(1X,A6,1X,A4,1X,I4,I2,1X,I2,13(1X,A1,I4,2A1),/,
      + 22X,13(1X,A1,I4,2A1),/,22X,5(1X,A1,I4,2A1))
      NREC=NREC+1
      GO TO 1
      99 CONTINUE
      STOP
      END
/*
//GO.FT05F001 DD UNIT=TAPE62,VOL=SER=TAPENUMB,DISP=(,PASS),
// DSN=MFF.NDP042.HCNDALID.DATA,LABEL=(10,SL,EXPDT=98000)
//GO.FT06F001 DD *
//

```

14. LISTINGS OF THE SAS INPUT/OUTPUT RETRIEVAL PROGRAMS

The following is a listing of the SAS data retrieval program provided on magnetic tape (File 5 on the tape) by CDIAC for reading and printing the station inventory file for the HCN/D (File 8 on the tape). The JCL statements (preceded by // or /*) shown in the following are not provided in the file on the magnetic tape. The JCL statements required will vary for each operating system. The JCL statements shown in the following are provided to illustrate the statements required by an individual at ORNL who has requested these data on a 9-track, 6250 BPI, standard-labeled tape with characters written in EBCDIC and who is attempting to read the tape on an IBM mainframe (e.g., IBM 3090).

```
//UIDINV JOB (12345),'USER ADDRESS'
//OUT OUTPUT DEFAULT=YES,JESDS=ALL,DEST=LOCAL
// EXEC SAS,SASRGN=4096K,WORK=1600
//IN DD UNIT=TAPE62,VOL=SER=TAPENUMB,DISP=(,PASS),
// DSN=MFF.NDP042.INVENTORY.DATA,LABEL=(8,SL,EXPDT=98000)
//GO.FT06F001 DD *
//SYSIN DD *
```

* SAS RETRIEVAL CODE TO READ AND PRINT THE STATION INVENTORY
FILE FOR THE HCN/D DATA SET (FILE 8 ON THE MAGNETIC TAPE);

```
DATA INVENT;
LENGTH STNAME $ 23;
INFILE IN;
INPUT STCODE $ 2-3 CNI $ 6-9 STATE $ 12-13 STNAME 16-38 BMON $ 41-42
BYEAR $ 45-48 TOBS $ 51-55 LATDEG 58-59 LATMIN 62-63
LONDEG 66-68 LONMIN 71-72;
```

```
FILE PRINT NOTITLE HEADER=NEWPAGE;
PUT _INFILE_;
RETURN;
```

```
NEWPAGE:
PUT @2 'ST CNI ST STATION NAME' @41 'MO YEAR TOBS DEG MIN'
@66 'DEG MIN';
RETURN;
```

```
RUN;
//
```

The following is a listing of the SAS data retrieval program provided on magnetic tape (File 6 on the tape) by CDIAC for reading and printing the station history file for the HCN/D (File 9 on the tape). The JCL statements (preceded by // or /*) shown in the following are not provided in the file on the magnetic tape. The JCL statements required will vary for each operating system. The JCL statements shown in the following are provided to illustrate the statements required by an individual at ORNL who has requested these data on a 9-track, 6250 BPI, standard-labeled tape with characters written in EBCDIC and who is attempting to read the tape on an IBM mainframe (e.g., IBM 3090).

```
//UIDHIS JOB (12345),'USER ADDRESS'
//OUT OUTPUT DEFAULT=YES,JESDS=ALL,DEST=LOCAL
//STEP1 EXEC SAS,SASRGN=3500K,WORK=500
//IN DD UNIT=TAPE62,VOL=SER=TAPEVOL1,DISP=(,PASS),
// DSN=MFF.NDP042.STATNHIS.DATA,LABEL=(9,SL)
//FT06F001 DD SYSOUT=A
//SYSIN DD *
DATA HISTORY (DROP=X);
  RETAIN STANUM STATE DIVISION CURRNAME COUNTY XREF;
  INFILE IN MISSEVER LS=216;
  INPUT @45 X $1. @;
  IF X NE '' THEN DO;
  INPUT STANUM 1-6 STATE $ 8-9 DIVISION 11-12 CURRNAME $ 14-43
    COUNTY $ 45-60 XREF $ 62-86 ;
  END;
ELSE
INPUT STANUM2 1-6 MOBEG 8-9 DAYBEG 11-12 YRBEG 14-17 MOEND 19-20
  DAYEND 22-23 YREND 25-28 SUSPLAT 30 SUSPLONG 31 SUSPLOC 32
  SUSPELEV 33 SUSPPO 34 SUSPNAME 35 SUSPQUAL 36 SUSPINST 37
  SUSPTIME 38 SUSPHTS 39 SUSPPUBS 40 SUSPBEG 41 SUSPEND 42
  SUSPOBS 43 SUSPOTHR 44 LATNORTH $ 46-51 LONGWEST $ 53-59
  DISTANCE 61-63 DIRECT $ 65-67 ELEV 69-73 DISTPO 75-77
  DIRECTPO $ 79-81 NAME $ 83-110 QUALIF $ 112-121 ADDINST 123
  COTTON 124 DBULB 125 EVAPSTA 126 FISHPORT 127 HYGRO 128
  MINTHERM 129 MAXTHERM 130 NORIVGAG 131 RAINGAGE 132
  SHELTER 133 RECRIVER 134 RECRAIN 135 SNOW 136 STORAGE 137
  STDRAIN 138 STDSELT 139 THERMOGR 140 DIGTHERM 141
  TIPBUCK 142 OTHEVAP 143 MAXMIN 144 TIMEOBS $ 146-149
  PCPHT $ 151-152 PCTHT $ 154-155 BULLETW 157 COMBBUL 158
  CLIMDATA 159 RIVSTAGE 160 HYDROBUL 161 PRECDATA 162
  SNOWBULL 163 NOTPUB 164 CWB 165 MONTHREV 166 STATEPUB 167
  LCD 168 BQ 169 SGPD 170 WWR 171 MYB 172 OBSNAME $ 174-213
  NUMOBS 215-216;
PROC PRINT;
RUN;
```

The following is a listing of the SAS data retrieval program provided on magnetic tape (File 7 on the tape) by CDIAC for reading and printing any of the HCN/D data files (Files 10-14 on the tape). The JCL statements (preceded by // or /*) shown in the following are not provided in the file on the magnetic tape. The JCL statements required will vary for each operating system. The JCL statements shown in the following are provided to illustrate the statements required by an individual at ORNL who has requested these data on a 9-track, 6250 BPI, standard-labeled tape with characters written in EBCDIC and who is attempting to read the tape on an IBM mainframe (e.g., IBM 3090).

```
//UIDINV JOB (12345),'USER ADDRESS'
//OUT OUTPUT DEFAULT=YES,JESDS=ALL,DEST=LOCAL
// EXEC SAS,SASRGN=4096K,WORK=1600
//IN DD UNIT=TAPE62,VOL=SER=TAPENUMB,DISP=(,PASS),
// DSN=MFF.NDP042.HCNDALID.DATA,LABEL=(10,SL,EXPDT=98000)
//GO.FT06F001 DD *
//SYSIN DD *
```

* SAS RETRIEVAL CODE TO READ AND PRINT ANY OF THE HCN/D FILES
(FILES 10-14 ON THE MAGNETIC TAPE);

DATA HCND;

```
ARRAY DAY {31} $ DAY1-DAY31;
ARRAY CHSF {31} $ 1 CHSF1-CHSF31;
ARRAY DMF {31} $ DMF1-DMF31;
ARRAY DQF {31} $ DQF1-DQF31;
ARRAY CHVAL {31} $ 4 CHVAL1-CHVAL31;
ARRAY VAL {31} VAL1-VAL31;
```

```
INFILE IN LRECL=270;
INPUT @2 STAID $ 2-7 DATTYP $ 9-12 YEAR 14-17 MON 18-19 DAYS 21-22
      @23 (DAY1-DAY31) ($CHAR8.);
```

*STATEMENTS WITHIN THE LOOP ARE INCLUDED FOR USE IF ONE
WISHES TO USE THIS PROGRAM AS A TEMPLATE FOR A PROGRAM
WHICH DOES MORE THAN SIMPLY PRINT THE FILE;

```
DO I=1 TO 31;
  CHSF{I}=SUBSTR(DAY{I},2,1);
  CHVAL{I}=SUBSTR(DAY{I},3,4);
  DMF{I}=SUBSTR(DAY{I},7,1);
  DQF{I}=SUBSTR(DAY{I},8,1);
  * VAL{I} WILL CONTAIN INTEGER VALUES;
  VAL{I}=INPUT(CHVAL{I},4.);
END;
```

```
FILE PRINT LRECL=126 NOTITLE HEADER=NEWPAGE;
PUT STAID 2-7 DATTYP 9-12 YEAR 14-17 MON 18-19 DAYS 21-22
    @23 (DAY1-DAY13) ($CHAR8.) / @23 (DAY14-DAY26) ($CHAR8.) /
    @23 (DAY27-DAY31) ($CHAR8.);
```

RETURN;

NEWPAGE:

PUT @2 'STAI D TYPE YR/MO 1/14/27|2/15/28|3/16/29|4/17/30|5/18/31|'

@64 ' 6/19 | 7/20 | 8/21 | 9/22 | 10/23 | 11/24 | 12/25 |'

@120 ' 13/26' / @24 12*'S MQ|' @120 'S MQ' /

@24 12*' |';

RETURN;

RUN;

//

15. VERIFICATION OF DATA TRANSPORT

The HCN/D station inventory, station history, and data files may be read using the FORTRAN or SAS I/O routines provided. Users should verify that the files have been correctly transported to their systems; to do so, they should generate some or all of the statistics presented in Tables 3-9. These statistics were generated using SAS (PROC MEANS) but can be duplicated in other statistical packages or languages. If the statistics generated by the user differ from those presented here, the files may have been corrupted in transport.

These statistics are presented only as a tool to ensure proper reading of the data sets. They are not to be construed as either a summary of the climatic data or an indicator of trends in the climatic data.

Table 3. Characteristics of numeric variables from the HCN/D station inventory file

Variable	Number of observations	Minimum value	Maximum value	Mean
STCODE	138	1.000000	48.000000	25.072464
STANUM	138	42.000000	9750.000000	4333.789855
STMO	138	1.000000	12.000000	4.036232
STYEAR	138	1871.000000	1940.000000	1898.804348
LATDEG	138	26.000000	48.000000	39.717391
LATMIN	138	0.000000	59.000000	28.050725
LONDEG	138	67.000000	123.000000	95.739130
LONMIN	138	0.000000	59.000000	30.231884

The following is a listing of the SAS program used to generate the statistics in the table.

```
*PROC MEANS FOR VERIFICATION OF TRANSPORT;
DATA INVENT;
INFILE 'FILE8';
INPUT STCODE 2-3 STANUM 6-9 STMO 41-42 STYEAR 45-48
      LATDEG 58-59 LATMIN 62-63 LONDEG 66-68 LONMIN 71-72;
PROC MEANS N MEAN MIN MAX MAXDEC=6;
RUN;
```

Table 4. Characteristics of numeric variables from the HCN/D station history file

Variable	Number of observations	Minimum value	Maximum value	Mean
STANUM	1840	15749.000000	481905.000000	249798.441304
DIVISION	1840	0.000000	10.000000	4.832065
STANUM2	1702	15749.000000	481905.000000	249371.955934
MOBEG	1702	1.000000	99.000000	6.626910
DAYBEG	1702	1.000000	99.000000	12.010576
YRBEG	1702	1738.000000	1987.000000	1935.353702
MOEND	1702	1.000000	99.000000	14.851351
DAYEND	1702	1.000000	99.000000	30.632785
YREND	1702	1759.000000	9999.000000	2592.659224
SUSPLAT	1702	0.000000	1.000000	0.022914
SUSPLONG	1702	0.000000	1.000000	0.023502
SUSPLOC	1702	0.000000	1.000000	0.005288
SUSPELEV	1702	0.000000	1.000000	0.060517
SUSPPO	1702	0.000000	1.000000	0.022914
SUSPNAME	1702	0.000000	0.000000	0.000000
SUSPQUAL	1702	0.000000	0.000000	0.000000
SUSPINST	1702	0.000000	1.000000	0.005875
SUSPTIME	1702	0.000000	1.000000	0.014101
SUSPHTS	1702	0.000000	1.000000	0.001175
SUSPPUBS	1702	0.000000	0.000000	0.000000
SUSPBEG	1702	0.000000	1.000000	0.001763
SUSPEND	1702	0.000000	1.000000	0.001175
SUSPOBS	1702	0.000000	1.000000	0.000588
SUSPOTHR	1702	0.000000	0.000000	0.000000
DISTANCE	1702	0.000000	999.000000	357.295535
ELEV	1702	8.000000	7760.000000	1649.367215
DISTPO	1702	0.000000	999.000000	273.493537
ADDINST	1702	0.000000	1.000000	0.205053
COTTON	1702	0.000000	1.000000	0.843126
DBULB	1702	0.000000	1.000000	0.045241
EVAPSTA	1702	0.000000	1.000000	0.055229
FISHPORT	1702	0.000000	1.000000	0.048766
HYGRO	1702	0.000000	1.000000	0.051704
MINTHERM	1702	0.000000	1.000000	0.903055
MAXTHERM	1702	0.000000	1.000000	0.903055
NORIVGAG	1702	0.000000	1.000000	0.028202
RAINGAGE	1702	0.000000	1.000000	0.002938
SHELTER	1702	0.000000	1.000000	0.015864
RECRIVER	1702	0.000000	1.000000	0.019389
RECRAIN	1702	0.000000	1.000000	0.121034

Table 4 (continued)

Variable	Number of observations	Minimum value	Maximum value	Mean
SNOW	1702	0.000000	1.000000	0.009401
STORAGE	1702	0.000000	1.000000	0.005288
STDRAIN	1702	0.000000	1.000000	0.978261
STDSHELT	1702	0.000000	0.000000	0.000000
THERMOGR	1702	0.000000	1.000000	0.016451
DIGTHERM	1702	0.000000	0.000000	0.000000
TIPBUCK	1702	0.000000	1.000000	0.051704
OTHEVAP	1702	0.000000	1.000000	0.000588
MAXMIN	1702	0.000000	1.000000	0.053467
BULLETW	1702	0.000000	1.000000	0.240893
COMBBUL	1702	0.000000	1.000000	0.002938
CLIMDATA	1702	0.000000	1.000000	0.886604
RIVSTAGE	1702	0.000000	1.000000	0.004113
HYDROBUL	1702	0.000000	1.000000	0.024677
PRECDATA	1702	0.000000	1.000000	0.115159
SNOWBULL	1702	0.000000	0.000000	0.000000
NOTPUB	1702	0.000000	0.000000	0.000000
CWB	1702	0.000000	1.000000	0.001763
MONTHREV	1702	0.000000	1.000000	0.013514
STATEPUB	1702	0.000000	1.000000	0.013514
LCD	1702	0.000000	1.000000	0.004700
BQ	1702	0.000000	1.000000	0.000588
SGPD	1702	0.000000	0.000000	0.000000
WWR	1702	0.000000	1.000000	0.001175
MYB	1702	0.000000	0.000000	0.000000
NUMOBS	1702	0.000000	9.000000	0.099882

The following is a listing of the SAS program used to generate the statistics in Table 4.

```

DATA HISTORY (DROP=X);
  RETAIN STANUM DIVISION;
  INFILE 'FILE9' MISSEVER LS=216;
  INPUT @45 X $1. @;
  IF X NE '' THEN DO;
    INPUT STANUM 1-6 DIVISION 11-12;
  END;
ELSE
  INPUT STANUM2 1-6 MOBEG 8-9 DAYBEG 11-12 YRBEG 14-17 MOEND 19-20
    DAYEND 22-23 YREND 25-28 SUSPLAT 30 SUSPLONG 31 SUSPLOC 32
    SUSPELEV 33 SUSPPO 34 SUSPNAME 35 SUSPQUAL 36 SUSPINST 37
    SUSPTIME 38 SUSPHTS 39 SUSPPUBS 40 SUSPBEG 41 SUSPEND 42 SUSPOBS 43
    SUSPOTHR 44 DISTANCE 61-63 ELEV 69-73 DISTPO 75-77
    ADDINST 123 COTTON 124 DBULB 125 EVAPSTA 126
    FISHPORT 127 HYGRO 128 MINTHERM 129 MAXTHERM 130 NORIVGAG 131
    RAINGAGE 132 SHELTER 133 RECRIVER 134 RECRAIN 135 SNOW 136
    STORAGE 137 STDRAIN 138 STDSHELT 139 THERMOGR 140 DIGTHERM 141
    TIPBUCK 142 OTHEVAP 143 MAXMIN 144 BULLETW 157 COMBBUL 158
    CLIMDATA 159 RIVSTAGE 160 HYDROBUL 161 PRECDATA 162 SNOWBULL 163
    NOTPUB 164 CWB 165 MONTHREV 166 STATEPUB 167 LCD 168 BQ 169
    SGPD 170 WWR 171 MYB 172 NUMOBS 215-216;
PROC MEANS N MEAN MIN MAX MAXDEC=6 FW=13;
RUN;

```

Table 5. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 1-10 (File 10 on the magnetic tape)

Variable	Number of observations	Minimum value	Maximum value	Mean
STAIID	54,231	15749.000000	107386.000000	54359.347366
YEAR	54,231	1871.000000	1987.000000	1943.926149
MON	54,231	1.000000	12.000000	6.504693
DAYS	54,231	28.000000	31.000000	30.437093
VALUE(1)	54,231	-999.000000	700.000000	38.285667
VALUE(2)	54,231	-999.000000	790.000000	39.013332
VALUE(3)	54,231	-999.000000	646.000000	39.492652
VALUE(4)	54,231	-999.000000	696.000000	38.911471
VALUE(5)	54,231	-999.000000	878.000000	38.971898
VALUE(6)	54,231	-999.000000	749.000000	39.549427
VALUE(7)	54,231	-999.000000	1090.000000	39.380207
VALUE(8)	54,231	-999.000000	1108.000000	39.279766
VALUE(9)	54,231	-999.000000	875.000000	39.331692
VALUE(10)	54,231	-999.000000	677.000000	39.334495
VALUE(11)	54,231	-999.000000	806.000000	39.732976
VALUE(12)	54,231	-999.000000	505.000000	39.399384
VALUE(13)	54,231	-999.000000	917.000000	39.729306
VALUE(14)	54,231	-999.000000	829.000000	39.233188
VALUE(15)	54,231	-999.000000	1312.000000	39.346333
VALUE(16)	54,231	-999.000000	902.000000	39.218934
VALUE(17)	54,231	-999.000000	903.000000	39.765927
VALUE(18)	54,231	-999.000000	753.000000	40.048422
VALUE(19)	54,231	-999.000000	812.000000	39.495916
VALUE(20)	54,231	-999.000000	710.000000	39.672918
VALUE(21)	54,231	-999.000000	644.000000	39.767790
VALUE(22)	54,231	-999.000000	763.000000	39.248529
VALUE(23)	54,231	-999.000000	565.000000	39.298409
VALUE(24)	54,231	-999.000000	917.000000	39.452933
VALUE(25)	54,231	-999.000000	1173.000000	39.222400
VALUE(26)	54,231	-999.000000	815.000000	39.444414
VALUE(27)	54,231	-999.000000	1012.000000	39.612989
VALUE(28)	54,231	-999.000000	725.000000	39.671774
VALUE(29)	54,231	-999.000000	838.000000	-25.767900
VALUE(30)	54,231	-999.000000	781.000000	-46.286681
VALUE(31)	54,231	-999.000000	731.000000	-395.087478

Table 6. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 11-20 (File 11 on the magnetic tape)

Variable	Number of observations	Minimum value	Maximum value	Mean
STAID	112,615	110072.000000	201675.000000	140776.755077
YEAR	112,615	1872.000000	1987.000000	1940.774355
MON	112,615	1.000000	12.000000	6.501567
DAYS	112,615	28.000000	31.000000	30.436461
VALUE(1)	112,615	-999.000000	503.000000	32.722248
VALUE(2)	112,615	-999.000000	820.000000	33.153692
VALUE(3)	112,615	-999.000000	665.000000	33.028673
VALUE(4)	112,615	-999.000000	720.000000	33.208738
VALUE(5)	112,615	-999.000000	1040.000000	33.416188
VALUE(6)	112,615	-999.000000	1050.000000	33.396084
VALUE(7)	112,615	-999.000000	577.000000	33.271829
VALUE(8)	112,615	-999.000000	655.000000	33.124974
VALUE(9)	112,615	-999.000000	600.000000	33.194051
VALUE(10)	112,615	-999.000000	548.000000	33.525596
VALUE(11)	112,615	-999.000000	764.000000	33.329414
VALUE(12)	112,615	-999.000000	576.000000	33.494623
VALUE(13)	112,615	-999.000000	662.000000	34.023647
VALUE(14)	112,615	-999.000000	605.000000	33.798615
VALUE(15)	112,615	-999.000000	675.000000	33.969933
VALUE(16)	112,615	-999.000000	810.000000	33.811766
VALUE(17)	112,615	-999.000000	690.000000	34.157963
VALUE(18)	112,615	-999.000000	625.000000	34.635599
VALUE(19)	112,615	-999.000000	807.000000	34.385624
VALUE(20)	112,615	-999.000000	735.000000	34.486116
VALUE(21)	112,615	-999.000000	660.000000	34.304622
VALUE(22)	112,615	-999.000000	1147.000000	34.187231
VALUE(23)	112,615	-999.000000	606.000000	34.111015
VALUE(24)	112,615	-999.000000	477.000000	33.940665
VALUE(25)	112,615	-999.000000	900.000000	34.248706
VALUE(26)	112,615	-999.000000	829.000000	34.238308
VALUE(27)	112,615	-999.000000	769.000000	33.862718
VALUE(28)	112,615	-999.000000	1020.000000	33.922977
VALUE(29)	112,615	-999.000000	876.000000	-31.187462
VALUE(30)	112,615	-999.000000	900.000000	-50.928562
VALUE(31)	112,615	-999.000000	503.000000	-398.121582

Table 7. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 21-30 (File 12 on the magnetic tape)

Variable	Number of observations	Minimum value	Maximum value	Mean
STAIID	114,591	211630.000000	308944.000000	257520.169158
YEAR	114,591	1874.000000	1987.000000	1943.245813
MON	114,591	1.000000	12.000000	6.503015
DAYS	114,591	28.000000	31.000000	30.436762
VALUE(1)	114,591	-999.000000	800.000000	28.693807
VALUE(2)	114,591	-999.000000	504.000000	29.157761
VALUE(3)	114,591	-999.000000	695.000000	29.347811
VALUE(4)	114,591	-999.000000	820.000000	29.499184
VALUE(5)	114,591	-999.000000	752.000000	29.452880
VALUE(6)	114,591	-999.000000	615.000000	29.837012
VALUE(7)	114,591	-999.000000	666.000000	29.802410
VALUE(8)	114,591	-999.000000	740.000000	29.719498
VALUE(9)	114,591	-999.000000	816.000000	29.912742
VALUE(10)	114,591	-999.000000	718.000000	30.437364
VALUE(11)	114,591	-999.000000	746.000000	29.881308
VALUE(12)	114,591	-999.000000	635.000000	30.480858
VALUE(13)	114,591	-999.000000	856.000000	30.535158
VALUE(14)	114,591	-999.000000	733.000000	30.425208
VALUE(15)	114,591	-999.000000	1060.000000	30.592909
VALUE(16)	114,591	-999.000000	850.000000	30.536438
VALUE(17)	114,591	-999.000000	445.000000	30.519718
VALUE(18)	114,591	-999.000000	897.000000	30.588231
VALUE(19)	114,591	-999.000000	563.000000	30.538280
VALUE(20)	114,591	-999.000000	588.000000	30.386854
VALUE(21)	114,591	-999.000000	812.000000	30.616139
VALUE(22)	114,591	-999.000000	688.000000	30.517589
VALUE(23)	114,591	-999.000000	915.000000	30.321517
VALUE(24)	114,591	-999.000000	700.000000	30.282500
VALUE(25)	114,591	-999.000000	782.000000	30.377569
VALUE(26)	114,591	-999.000000	592.000000	30.045754
VALUE(27)	114,591	-999.000000	635.000000	30.362891
VALUE(28)	114,591	-999.000000	766.000000	30.387910
VALUE(29)	114,591	-999.000000	675.000000	-34.112408
VALUE(30)	114,591	-999.000000	1050.000000	-54.690892
VALUE(31)	114,591	-999.000000	670.000000	-400.799347

Table 8. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 31-40 (File 13 on the magnetic tape)

Variable	Number of observations	Minimum value	Maximum value	Mean
STAID	70,108	312635.000000	409155.000000	361853.155603
YEAR	70,108	1871.000000	1987.000000	1942.219190
MON	70,108	1.000000	12.000000	6.503908
DAYS	70,108	28.000000	31.000000	30.437340
VALUE(1)	70,108	-999.000000	665.000000	32.355965
VALUE(2)	70,108	-999.000000	590.000000	32.953329
VALUE(3)	70,108	-999.000000	825.000000	32.841202
VALUE(4)	70,108	-999.000000	606.000000	32.696996
VALUE(5)	70,108	-999.000000	937.000000	33.180864
VALUE(6)	70,108	-999.000000	1033.000000	32.775375
VALUE(7)	70,108	-999.000000	600.000000	33.110216
VALUE(8)	70,108	-999.000000	647.000000	33.046186
VALUE(9)	70,108	-999.000000	526.000000	32.555586
VALUE(10)	70,108	-999.000000	490.000000	32.622796
VALUE(11)	70,108	-999.000000	1038.000000	32.816512
VALUE(12)	70,108	-999.000000	625.000000	33.491927
VALUE(13)	70,108	-999.000000	945.000000	33.457009
VALUE(14)	70,108	-999.000000	1005.000000	33.654503
VALUE(15)	70,108	-999.000000	568.000000	33.748189
VALUE(16)	70,108	-999.000000	525.000000	34.054402
VALUE(17)	70,108	-999.000000	600.000000	34.148257
VALUE(18)	70,108	-999.000000	679.000000	34.060863
VALUE(19)	70,108	-999.000000	616.000000	34.418269
VALUE(20)	70,108	-999.000000	636.000000	34.199535
VALUE(21)	70,108	-999.000000	662.000000	33.785403
VALUE(22)	70,108	-999.000000	626.000000	33.874651
VALUE(23)	70,108	-999.000000	676.000000	34.126705
VALUE(24)	70,108	-999.000000	505.000000	33.947666
VALUE(25)	70,108	-999.000000	1165.000000	33.785745
VALUE(26)	70,108	-999.000000	640.000000	33.107163
VALUE(27)	70,108	-999.000000	723.000000	33.460746
VALUE(28)	70,108	-999.000000	662.000000	33.427255
VALUE(29)	70,108	-999.000000	550.000000	-31.141567
VALUE(30)	70,108	-999.000000	663.000000	-51.168626
VALUE(31)	70,108	-999.000000	534.000000	-398.560806

Table 9. Characteristics of numeric variables from the HCN/D data file containing data from states with codes 41–48 (File 14 on the magnetic tape)

Variable	Number of observations	Minimum value	Maximum value	Mean
STAIID	82,711	410120.000000	481905.000000	444440.983932
YEAR	82,711	1871.000000	1987.000000	1943.885710
MON	82,711	1.000000	12.000000	6.498531
DAYS	82,711	28.000000	31.000000	30.436725
VALUE(1)	82,711	-999.000000	1130.000000	33.533774
VALUE(2)	82,711	-999.000000	757.000000	33.843213
VALUE(3)	82,711	-999.000000	1002.000000	34.277061
VALUE(4)	82,711	-999.000000	2905.000000	34.082335
VALUE(5)	82,711	-999.000000	661.000000	34.287785
VALUE(6)	82,711	-999.000000	618.000000	34.224154
VALUE(7)	82,711	-999.000000	808.000000	34.046814
VALUE(8)	82,711	-999.000000	537.000000	34.029742
VALUE(9)	82,711	-999.000000	674.000000	34.522119
VALUE(10)	82,711	-999.000000	996.000000	34.167982
VALUE(11)	82,711	-999.000000	616.000000	34.458379
VALUE(12)	82,711	-999.000000	685.000000	34.644497
VALUE(13)	82,711	-999.000000	728.000000	34.515433
VALUE(14)	82,711	-999.000000	599.000000	34.826311
VALUE(15)	82,711	-999.000000	556.000000	34.661665
VALUE(16)	82,711	-999.000000	810.000000	34.397263
VALUE(17)	82,711	-999.000000	651.000000	34.595290
VALUE(18)	82,711	-999.000000	612.000000	34.551208
VALUE(19)	82,711	-999.000000	748.000000	34.547436
VALUE(20)	82,711	-999.000000	490.000000	34.852015
VALUE(21)	82,711	-999.000000	705.000000	34.675339
VALUE(22)	82,711	-999.000000	850.000000	34.689715
VALUE(23)	82,711	-999.000000	708.000000	35.341164
VALUE(24)	82,711	-999.000000	732.000000	35.269892
VALUE(25)	82,711	-999.000000	675.000000	34.855581
VALUE(26)	82,711	-999.000000	1250.000000	34.768495
VALUE(27)	82,711	-999.000000	590.000000	34.745076
VALUE(28)	82,711	-999.000000	646.000000	34.595096
VALUE(29)	82,711	-999.000000	608.000000	-30.251224
VALUE(30)	82,711	-999.000000	650.000000	-51.318797
VALUE(31)	82,711	-999.000000	640.000000	-397.616097

The following is a listing of the SAS program used to generate the statistics in Tables 5--9.

```
DATA HCND;  
ARRAY VAL {31} VAL1-VAL31;  
INFILE 'FILE10' LRECL=270;  
INPUT @2 STAID 2-7 YEAR 14-17 MON 18-19 DAYS 21-22  
      @25 (VAL1-VAL31) (4.,+4 );  
PROC MEANS N MEAN MIN MAX MAXDEC=6 FW=13;  
RUN;
```

APPENDICES

APPENDIX A

DATA QUALITY ASSESSMENTS FOR THE HCN/D STATIONS

The process of selecting stations for the HCN/D included examining data quality assessments derived for the monthly HCN data by Karl et al. (1990), which are either paraphrased or excerpted in the following to briefly outline the nature of the quality assessment scheme.

A relative rating system was established for eight distinct station characteristics. These include:

1. The serial completion date.
2. What percentage of original data was available.
3. What percentage of adjusted data was estimated.
4. The beginning date of the original data.
5. The average confidence factor over the period of adjusted data (CONF).
6. The number of potential discontinuities during the period of adjusted data (#M).
7. For the *adjusted* data, the consistency of the station on an annual basis with its nearest 20 neighbors over the past 40 years (CONS40).
8. For the *adjusted* data, the consistency of the station on an annual basis with its nearest 20 neighbors for the full period of record (CONSFULL).

". . . In order to judge the quality of each station relative to the network, each category was assigned a numeric value between 0 and 9. These are related to the decile ranking (R) of each category; 0 implies the lower 10 percentile, 1 implies 10–20 percentile, . . . 9 implies 90–100 percentile. For the eight characteristics of data quality (listed above), the lowest deciles are assigned to:

1. Longest records.
2. Greatest data available.
3. Stations with fewest estimated adjusted data.
4. Longest records.
5. Lowest confidence factor (narrowest intervals).
6. Fewest discontinuities.
7. Most consistent, lowest difference.
8. Most consistent, lowest difference. . . ."

REFERENCE

Karl, T. R., C. N. Williams, Jr., and F. T. Quinlan. 1990. United States Historical Climatology Network (HCN) serial temperature and precipitation data. ORNL/CDIAC-30, NDP-019/R1. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

APPENDIX B

REPRINTS OF PERTINENT LITERATURE

Relationship between Decreased Temperature Range and Precipitation Trends in the United States and Canada, 1941-80

THOMAS R. KARL

National Climatic Data Center, NESDIS, NOAA, Asheville, NC 28801

GEORGE KUKLA AND JOYCE GAVIN

Lamont-Doherty Geological Observatory, Columbia University, Palisades, NY 10964

(Manuscript received 18 February 1986, in final form 14 June 1986)

ABSTRACT

Previous work has shown significant decreases of the diurnal temperature range (1941-80) across a network of 130 stations in the United States and Canada. In the present study, changes in monthly total precipitation at these same stations were related to the decrease in temperature range using various Monte Carlo tests. These tests indicate that factors other than those related to precipitation contributed to the decrease of daily temperature range. Further study of the mechanisms responsible for the decreased temperature range is warranted, based on these results. The decreased range may be one of the few pieces of evidence available in North America that is consistent with potential impacts of increased greenhouse gases and/or anthropogenic aerosols.

1. Background

Using a network of 130 nonurban (population $\leq 50\,000$) stations in the United States and Canada, Karl et al. (1984) investigated the local and field significance of trends in the mean monthly maximum, minimum and diurnal temperature range over the 1941-80 period. The term "local" is used in reference to specific stations, whereas "field" is used to refer to all stations in the network. The most striking result of that study was the high field significance of decreasing diurnal temperature range during all months of the year. Several factors were suggested as possible mechanisms responsible for such a result; these included various greenhouse mechanisms, namely, changes in cloudiness, aerosol loading, atmospheric water vapor, and/or carbon dioxide or other anthropogenic trace gases. The purpose of the present investigation is to determine to what degree the decreasing trends of diurnal temperature range can be explained by changes of atmospheric water vapor as reflected by changes in total precipitation.

Precipitation is the only climatic variable indirectly related to humidity whose records are available at the cooperative stations used in the analysis of the diurnal temperature range. Although it is not a perfect surrogate by any means, we argue that total monthly precipitation is a reasonable surrogate for cloudiness and humidity at these rural stations, where cloud cover and humidity measurements are unavailable. We cite the work of Baker (1969), Lowry and Glahn (1969), Younkin et al. (1965), and Spar (1953), who demon-

strate both statistically and physically the important relationship between integrated water vapor (or precipitable water) and cloudiness and precipitation. In fact, Lowry and Glahn (1969) indicate that over 85% of the variance of precipitable water can be explained by surface dew point, cloud cover and precipitation.

Aside from the practical importance of changes in temperature range, our interest in pursuing the cause of the observed decreasing trends of the diurnal temperature range stems from its potential association with increasing concentration of atmospheric CO_2 and other greenhouse gases. To determine how much the trends of the diurnal temperature range were influenced by precipitation trends, we designed a strategy based on the following.

1) Increased precipitation is likely to be associated with increased humidity and with a decrease in the diurnal temperature range and vice versa. Stations that have this negative correlation between precipitation and temperature range conform to the expected physical link between atmospheric humidity and temperature. This should be confirmed by a tendency of positive precipitation trends to be collocated with negative trends in the temperature range.

2) If increasing precipitation is the only cause of the decreasing temperature range, then a large portion of stations with positive precipitation trends should be detected, especially in areas where the temperature range is decreasing. If this is not the case, factors other than precipitation are likely to be responsible for the decreased temperature range.

2. Procedures

a. Field significance tests

Monthly precipitation data were readily available for 123 (about 95%) of the original 130 stations. We computed the linear precipitation trends over the 1941–80 period and investigated their relationship with temperature trends, particularly the diurnal temperature range. Similar to Karl et al. (1984), correlation coefficients were used as a measure of association between temperature or precipitation and time (in 40 equal time steps) for all stations that reported both temperature and precipitation (never less than 122 stations in any given month). The elements analyzed included mean monthly maximum, minimum, average and diurnal temperature range as well as total monthly precipitation. Monte Carlo procedures were used to ascertain both the 90% level of locally significant correlation coefficients (that is, the magnitude of a correlation coefficient at a specific location that would occur due to chance only 10% of the time) and the field significance of the number of stations in the network with locally significant correlations with time (trends). All reported significance levels and probabilities of occurrence throughout this study are based on 1000 simulations. Livezey and Chen (1983) discuss and present the general concepts and motivation of field significance testing.

In the present study, one substantial difference from our previous study of the decreased range was required with respect to the precipitation field significance tests. Because there were very few locally significant trends of precipitation, the field significance tests were based on the percent frequency of stations in the network with either positive or negative trends as opposed to *locally significant* trends. In this regard, weak signals in the data would not be overlooked.

b. The in situ relationship of precipitation and temperature trends

A Monte Carlo method was used to determine whether there is an in situ relationship between the trends of precipitation and the trends of temperature. The procedure consisted of randomizing the 40 time steps (related to the 40 years of data) for both monthly mean temperature and total monthly precipitation, and subsequently calculating the correlation coefficient between temperature and the randomized time variate, as well as between precipitation and the same randomized time variate at each station. For each simulation, i , the percent frequency of stations with positive (or negative) precipitation collocated with locally significant temperature trends was calculated ($RAND_i$). This quantity was compared to the observed percent frequency of stations with positive (or negative) precipitation trends collocated with locally significant temperature trends (OBS). The proportion of simulations

where $OBS > RAND_i$ is a direct measure of the statistical probability (or significance) that the quantity OBS is larger than would be expected due to chance alone.

c. Field significance of the percent of locally significant positive versus negative temperature trends

Because it is possible to have a significant number of stations (more than would be expected by chance) associated with locally significant trends of one sign and, at the same time, a significant number of stations associated with an opposite locally significant trend, another test was developed to determine whether the difference between these two quantities is greater than would be expected due to chance. In this regard, the significance of a *dominant* trend can be determined. This is important because the diurnal temperature range is potentially sensitive to minor station relocations (Pittock et al., 1978) that may introduce spurious locally significant trends.

This new test consisted of calculating the observed difference between the percent of stations with locally significant positive correlations and the percent with locally significant negative correlations of temperature with time (OBSDIF). The observed quantity OBSDIF was then compared to the difference obtained by its Monte Carlo counterpart by randomizing the 40 time steps, recalculating the correlation with temperature, and computing the difference between the percent of simulations with locally significant positive and the percent with locally significant negative correlation coefficients (RANDIF). The number of simulations in which $OBSDIF > RANDIF_i$ is a direct measure of the probability that the quantity OBSDIF is larger than would be expected due to chance.

It is desirable to neutralize the potential impact of precipitation trends in the field significance tests, especially for temperature range by accounting for the difference between the number of stations with increasing and the number with decreasing precipitation trends, particularly if the in situ tests indicate a relation between temperature and precipitation trends, as has been hypothesized in section 1. For this reason, every station whose trend was the same sign as the predominant precipitation trend in the network was assigned a weighting factor of 1. Stations with the opposite sign were assigned a weight equal to the ratio of the number of stations with the dominant precipitation trend to the number of stations with the opposite trend. In this way, the number of stations with the predominant sign of the precipitation trend were given less weight, and as a result the two groups with opposite trends were each given equal weight in the field significance tests.

d. The in situ versus the field significance tests

It is important that we elaborate on the use and purpose of the in situ tests of significance and the field

significance tests described in sections 2b and 2c. The in situ tests can provide us with information about the relationship between significant decreases (or increases) of the temperature range and trends of total precipitation. If a significant relationship is found, it implies that the two trends may be linked. Although precipitation may influence temperature range, it would not necessarily follow that precipitation trends were responsible for an unusually large number of stations with significant positive or negative trends of temperature range. In order to address this point, the weighted field significance tests are used to remove the effects of any disparity in the number of stations with a given precipitation trend. Furthermore, using the difference between the percent of stations with significant positive and the percent with significant negative trends of temperature, we derive the probability of occurrence reported as from both the excess and the deficiency of stations with increasing or decreasing temperature trends. For the decreased temperature range, this test estimates the probability that a network of stations will have an unusually large number of stations with significant decreasing temperature range or a much smaller number of stations with opposite trends of temperature range but similar magnitude.

3. Results

a. Test of possible urban influences on temperature trends

In light of recent work related to urban heat islands in the United States (Cayan and Douglas, 1984; Kukla et al., 1986), the presence of several stations in the network with populations in excess of 20 000, and with growth rates during the 1941–80 period in excess of 100%, prompted us to determine if the urban heat is-

lands in these more densely populated, fast growing rural stations could be responsible for the decreased temperature range. We define the growth rate as the difference between the population in 1980 and 1940 divided by the 1940 population. Using these criteria, 24 of the original 130 stations were removed, and field significance tests of the locally significant temperature trends were conducted on this reduced network of 106 stations. The results are given in Table 1. They are, for all practical purposes, identical to those reported in Karl et al. (1984) for the full network. The significance of the trends of the decreased temperature range remains virtually unchanged. For this reason, the full set of stations is used in our remaining analyses.

b. Local and field significance of the precipitation trends

The percent of stations with either positive or negative precipitation trends are depicted in Fig. 1. The field significance tests indicated that only in June and September were there more stations with increasing or decreasing precipitation trends than would be expected by chance at the 90% level. Although the field significance is lacking, it is apparent from Fig. 1 that precipitation in North America is tending to increase, primarily in late summer, fall and early winter.

Figure 2 depicts the geographic patterns of precipitation trends compared to trends of the diurnal temperature range at each station. It should be emphasized that the patterns are based on our limited network of stations, and they only depict large, spatially coherent patterns of precipitation trends. A close inspection of Fig. 2, particularly during the months of June through October, suggests that there may be a relationship between the precipitation trends and the location of the

TABLE 1. Probabilities ($\times 100\%$) that the percent of stations with locally significant trends of the monthly mean maximum (MAX), minimum (MIN), average (AVG) and diurnal temperature range (RNG) are not due to chance alone. Probabilities relate to the network of stations with population $< 20\ 000$ and growth rates from 1941 to 1980 $< 100\%$. Only probabilities ≥ 0.90 are given.

Month	Increasing				Decreasing			
	MAX	MIN	AVG	RNG	MAX	MIN	AVG	RNG
Jan					99	99	99	99
Feb				93	94	97	94	98
Mar								97
Apr								96
May		92		90	92			94
Jun	99	98	99	99	94	94	94	99
Jul		98			93		90	99
Aug		96			94	93		99
Sep		93			97			99
Oct					99	94	95	97
Nov								96
Dec				90				96
					<i>Collective probability</i>			
	90	97		99	99	99	99	99

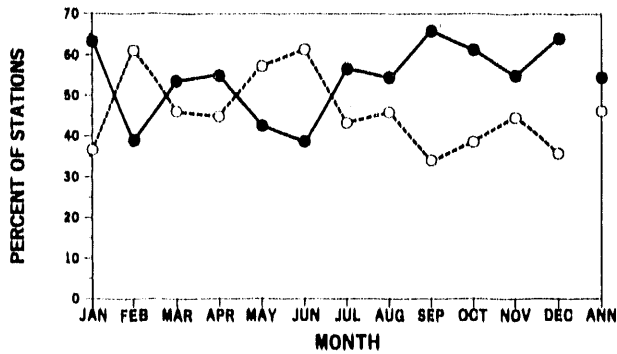


FIG. 1. Percentage of stations with either positive (solid line, solid circles) or negative (dashed line, open circles) precipitation trends.

station with decreasing (increasing) trends of temperature range. Furthermore, although the field significance tests indicate that September is the only month when the number of positive precipitation trends is greater than expected by chance, the relatively high number of stations with increasing precipitation trends,

particularly during the fall (i.e., more clouds, precipitable water, etc.), raises the possibility that a portion of the decreased range may still be due to the precipitation trends. The need for in situ Monte Carlo tests is apparent.

c. The in situ relationship of temperature and precipitation trends

The Monte Carlo procedures described in section 2b were used to objectively determine the in situ relation between locally significant trends of temperature, in particular the diurnal temperature range, and precipitation. Table 2 summarizes the results. The in situ inverse relationship of precipitation trends to trends of diurnal temperature range is greater than would be expected by chance throughout much of the year. Decreased trends of the diurnal temperature range tend to be located at stations with increasing precipitation and vice versa for increased temperature range. The collective significance of this result is at the 99% level. The link is further supported by the probabilities associated with decreased maxima and increased minima

TABLE 2. Probabilities ($\times 100\%$) that the locally significant trends of the maximum (MAX), minimum (MIN), average (AVG) or diurnal temperature range (RNG) will be collocated with either positive or negative precipitation trends more often than is expected by chance. Only probabilities ≥ 0.90 are given.

Month	Temperature increasing				Temperature decreasing			
	MAX	MIN	AVG	RNG	MAX	MIN	AVG	RNG
<i>a. Precipitation decreasing</i>								
Jan								
Feb								
Mar								
Apr								
May				93				
Jun	99		98	99				
Jul								
Aug				92		90		
Sep								
Oct								
Nov								
Dec								
Collective probability	99		97	98				
<i>b. Precipitation increasing</i>								
Jan		90						97
Feb								
Mar								
Apr								95
May								
Jun								
Jul		99			92			98
Aug					97			95
Sep		94			99			98
Oct								97
Nov					96			
Dec								91
Collective probability		97			99			99

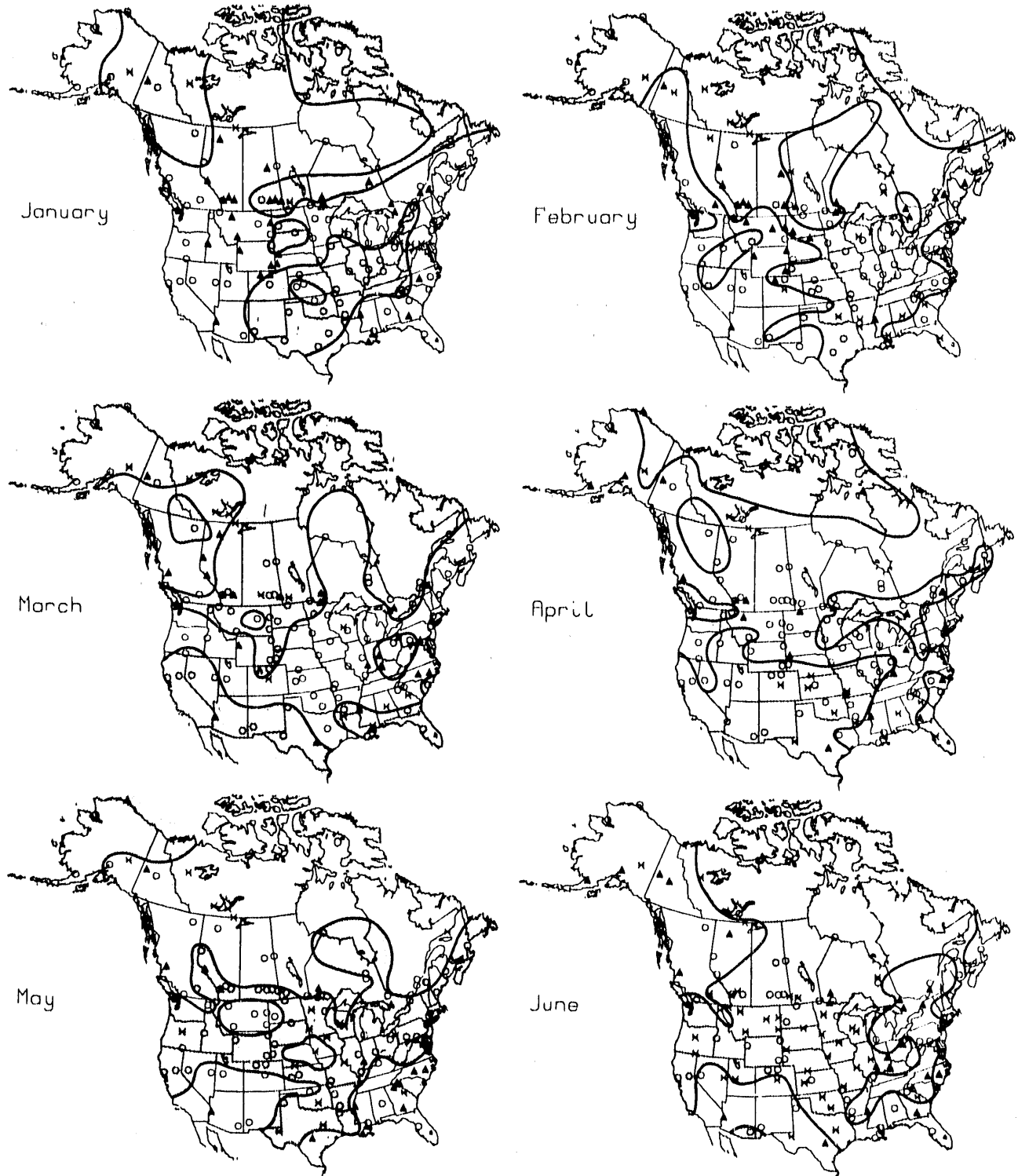


FIG. 2. Spatial patterns of positive (unshaded) and negative (shaded) trends of precipitation for each month (1941-80). Locally insignificant trends of the diurnal temperature range are depicted by open circles, solid triangles depict locally significant (95% one-tail) decreasing temperature range, and asterisks represent locally significant (95% one-tail) increasing temperature range.

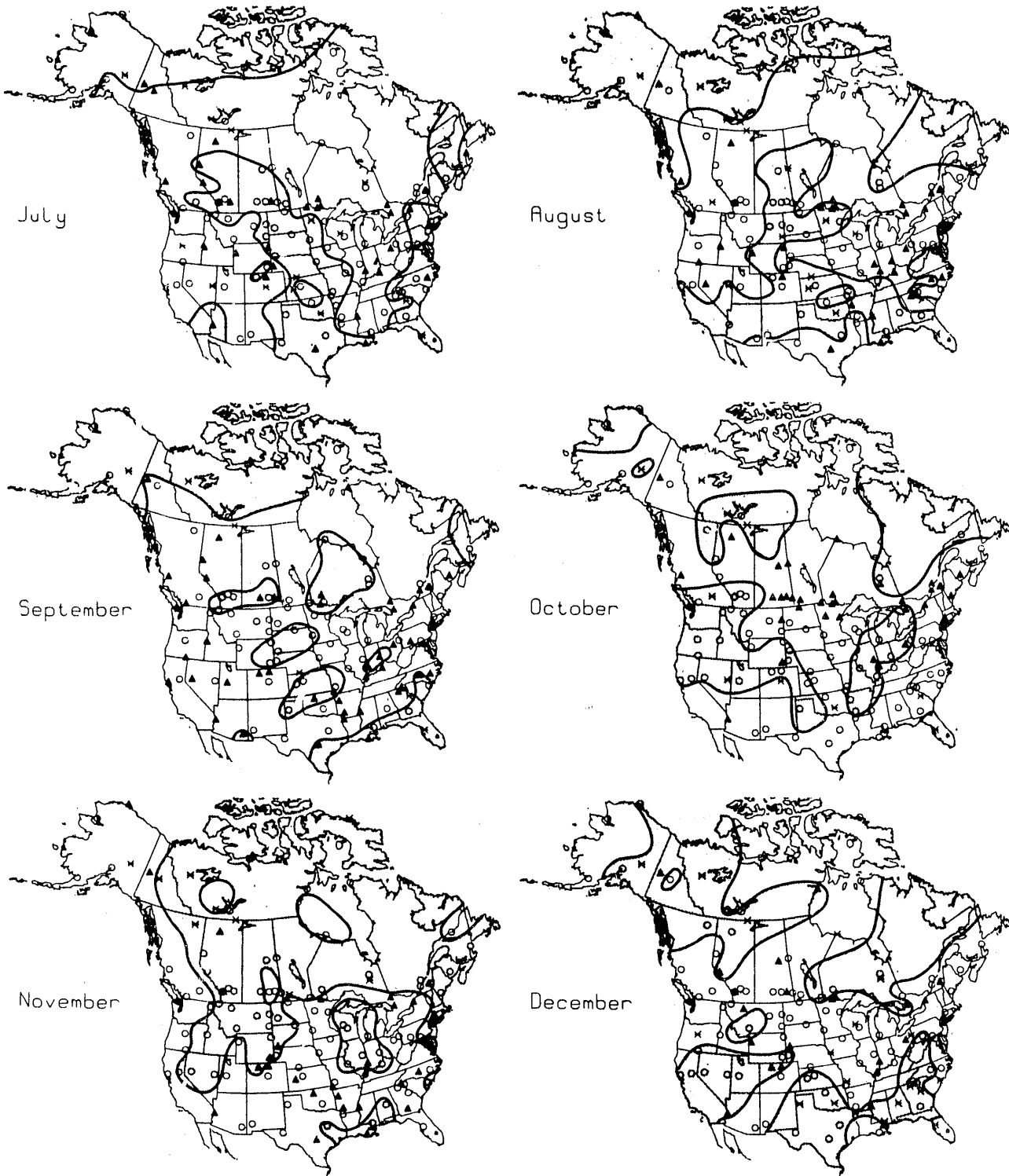


FIG. 2. (Continued)

temperatures (cf. Table 2). As hypothesized, decreased maxima and/or increased minima tend to be located more often at stations with increasing precipitation trends, whereas increased maxima and/or decreased

minima tend to be located at stations with decreasing precipitation trends.

The in situ Monte Carlo tests indicate that there is some relation between the trends of locally significant

temperature range and precipitation, but the tests still do not establish why so many stations have decreased compared to increased temperature range. To address this point, field significance tests that equalize the disproportionate number of stations with increasing and decreasing trends of precipitation are required (cf. Fig. 1).

d. Field significance of the percent of locally significant positive versus negative temperature trends

The results of the Monte Carlo simulations given in Table 3 indicate the statistical significance of the difference between the percent of stations with locally significant positive and negative temperature trends. Two kinds of simulations are presented: Table 3a gives results for equally weighted stations, and Table 3b for unequally weighted. Each of these sets contains simulations for the percent of stations with increasing minus decreasing trends, as well as the percent of stations with decreasing minus increasing trends.

Although there is obviously an in situ relation be-

tween the temperature range and the trends of precipitation (cf. Table 2), Table 3b indicates that this relation cannot explain the high field significance of the decreased diurnal temperature range when the disparity of the number of stations with positive or negative precipitation trends is removed. Note the high significance levels from midsummer through autumn and during midwinter. The significance levels reported for the unweighted (Table 3a) versus the weighted set (Table 3b) are virtually identical. The few exceptions occur during May and June, when there are proportionately more stations associated with decreasing trends of precipitation than with increasing trends. After standardizing for the differences of precipitation trends, the statistical significance of the decreasing diurnal temperature range (RNG % DECRG) is emphasized at the expense of the increasing diurnal temperature range (RNG % INCRG), e.g., June RNG % INCRG - %DECRG and May RNG % DECRG - %INCRG.

A comparison of Table 3 to Table 1 reveals an overall decrease in the significance of the decreased temperature range, probably attributable to the sensitivity of

TABLE 3. The probability ($\times 100\%$) that the magnitude of the difference between the percent of decreasing (DCRG) and increasing (INCRG) locally significant trends of temperature is greater than that due to chance alone. Probabilities are given for both equally weighted stations and stations weighted by the ratio of the number of positive to negative precipitation trends. Probabilities less than 0.85 are blank and those between 0.85 and 0.89 are denoted by an asterisk.

Month	%INCRG - %DECRG				%DECRG - %INCRG			
	MAX	MIN	AVG	RNG	MAX	MIN	AVG	RNG
<i>a. Equal weights</i>								
Jan					99	99	99	98
Feb					94	96	95	92
Mar		*						*
Apr					*	*		*
May					*		*	
Jun	97	91	*	*				
Jul		94			*			99
Aug					95			99
Sep		*			99			99
Oct					99	*	95	97
Nov								97
Dec								
Collective significance					96		*	98
<i>b. Unequal weights</i>								
Jan					99	99	99	98
Feb					92	96	94	91
Mar		*						*
Apr					*	*		*
May					91			*
Jun	*		*					
Jul		93			*			99
Aug					95			99
Sep		*			98			99
Oct					98	*	95	95
Nov								96
Dec						*		
Collective probability					97		90	98

the range to minor station changes. In spite of this reduction, the decreased temperature range continues to be an unlikely chance event. The significance level of this conclusion equals 99% for three months, at or above 95% for six months, and at or above 85% for ten months. The collective probability across all months is 98%. As for the other measures of temperature, the results in Table 3 indicate that, with respect to the mean temperature, the greater number of stations with locally significant decreases of temperature compared to the number of stations with increases has about a 1 in 10 probability of occurrence due to random chance. The results suggest that mean temperature decreases at more stations in the network than would be expected by chance. Additionally, the statistical significance of the decreased temperature range is strong enough as to suggest that some mechanism is operating other than random chance. The decrease of the temperature range appears to be unrelated to changes in the mean temperature, as some months have highly significant decreases in the range when the mean temperature does not change significantly and vice versa.

4. Discussion

Our investigation has shown that the range of daily temperature over much of North America is linked to changes in precipitation and, therefore, probably to increased atmospheric humidity. In spite of this link, it has also been shown that this connection cannot explain the decreasing range over North America, and it is likely that other factors unrelated to precipitation are involved. For example, our analysis does not rule out the possibility of increased upper-level cloudiness, which may not necessarily be reflected in changes of low-level humidity or precipitation. On the other hand, inasmuch as precipitation trends are the result of circulation changes, this would infer that circulation features may not be the predominant cause of the decreased temperature range.

The increase in greenhouse gases, such as CO_2 , CH_4 , NO_x , etc., is also among the potential causes of the trend. In fact, since the surface air temperature across the United States and Canada during 1941–80 has been shown to decrease, the temperature range may be one of the few climatic variables whose secular change conforms with the expected greenhouse effect. It should be realized, however, that on an empirical basis a direct link between increased CO_2 and other trace greenhouse gases and decreased daily temperature range has not been established.

Another possible cause of the decreasing daily temperature range is the change in the composition and concentration of atmospheric aerosols. The relatively strong statistical significance of the decreased range from midsummer through autumn is compatible with such an explanation. During this time of the year, the dispersion characteristics of the lower troposphere are poor across much of the United States (Holzworth,

1972; Korshover, 1976). Recent studies (Sloane, 1983; Sloane, 1984; Vinzani and Lamb, 1985; Husar et al., 1978) have confirmed a general decrease in visibility in the United States, particularly strong during summer and autumn. The visibility degradation is probably related to an increase of or change in the composition of atmospheric aerosol concentration, particularly weakly absorbing aerosols such as sulfates, nitrates and organic compounds. The long-range transport of these anthropogenic emissions out of their source regions and into rural areas is now widely accepted (Charlson et al., 1975; Pack et al., 1978; Karl, 1978; Rosen et al., 1981). Therefore, it is feasible that aerosol concentrations even at the rural sites used in this study, have changed significantly over the past 40 years.

The impact of increasing atmospheric aerosols on the climate is uncertain because the composition of tropospheric aerosols (as well as stratospheric aerosols) is quite variable and poorly known (Deepak and Gerber, 1983), but a decrease of maximum temperature, due to increased backscattering, and an increase of nighttime minimum temperature, due to absorption of longwave radiation, would be consistent with the known properties of anthropogenic aerosols. The seasonal cycle in the decreased temperature range in midsummer and autumn may be linked to the seasonal changes in the dispersion characteristics of the atmosphere and to trends of aerosol concentration and distribution.

5. Conclusion

In a network of mostly nonurban stations in North America, decreasing diurnal temperature range has been found to occur more often than expected by chance, even after consideration of the effects of precipitation trends in North America. The trends of diurnal temperature range were related to precipitation trends, and a link between these two variables was established. Despite this expected link, no evidence could be found to indicate that precipitation was responsible for the decreased range.

The magnitude of the net difference between the percent of stations with locally significant decreases of temperature range and the percent of stations with increased temperature range was found to have an appreciable seasonal cycle. The months of July through September had far more stations with decreased range compared to increased range, but other months also had a significant proportion of stations with decreased range versus increased range. During several months, a significantly greater number of stations were found to have locally significant decreases of average temperature compared to the number with locally significant increases. In light of this decrease of average temperature, the decreasing temperature range is perhaps one of the few signals in North America over recent decades consistent with increases in greenhouse gases,

and/or it may be corroborative evidence for the impact of changes in tropospheric anthropogenic aerosol composition of the climate.

Acknowledgments. We want to thank Robert Quayle of the National Climatic Data Center for his critical review of this article. This work has been supported by the Department of Energy, Carbon Dioxide Research Division, under Grant DE-FG02-85ER60372.

REFERENCES

- Baker, S. E., 1969: A statistical study of the depth of precipitation water in western Texas and eastern New Mexico. Texas Water Department Board, Austin, 79 pp.
- Cayan, D. R., and A. V. Douglas, 1984: Urban influences on surface temperatures in the southwestern United States during recent decades. *J. Climate Appl. Meteor.*, **23**, 1520-1530.
- Charlson, R. J., A. T. Vanderpol, D. S. Covert, A. P. Waggoner and N. C. Ahlquist, 1975: $\text{H}_2\text{SO}_4/(\text{NH}_4)\text{SO}_4$ background aerosol: Optical detection in the St. Louis region. *Science*, **184**, p. 156.
- Deepak, A., and H. E. Gerber, Eds., 1983: Report of the experts meeting on aerosols and their climatic effects. Williamsburg, World Climate Programme Series WCP-55.
- Holzworth, G. C., 1972: Mixing heights, wind speeds, and potential for urban air pollution throughout the contiguous United States. U.S. Environmental Protection Agency, Office of Air Programs Publication No. AP-101 118 pp. [Available from Superintendent of Documents, U.S. Government Printing Office, Washington, DC.]
- Husar, R. B., D. E. Patterson, J. M. Holloway, W. E. Wilson and T. G. Ellestad, 1978: Trends of eastern U.S. haziness since 1948. Fourth Symp. on Turbulence, Diffusion, and Air Pollution. Reno, Amer. Meteor. Soc., 249-256.
- Karl, T. R., 1978: Ozone transport in the St. Louis area. *Atmos. Environ.*, **12**, 1421-1430.
- , G. Kukla and J. Gavin, 1984: Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980. *J. Climate Appl. Meteor.*, **23**, 1489-1504.
- Korshover, J., 1976: Climatology of stagnating anticyclones east of the Rocky Mountains, 1936-1975. NOAA Tech. Memo. ERL ARL-55, 26 pp.
- Kukla, G., J. Gavin and T. R. Karl, 1986: Urban warming. *J. Climate Appl. Meteor.*, **25**, 1265-1270.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46-59.
- Lowry, D. A., and H. R. Glahn, 1969: Relationships between integrated atmospheric moisture and surface weather. *J. Appl. Meteor.*, **8**, 762-768.
- Pack, D. H., G. J. Ferber, J. L. Heffter, K. Telegados, J. K. Angell, W. H. Hoecker and L. Machta, 1978: Meteorology of long-range transport. *Atmos. Environ.*, **12**, 425-444.
- Pittock, A. B., L. A. Frahes, D. Jenssen, J. A. Peterson and J. W. Zillman, Eds., 1978: Climatic change and variability: A southern perspective. Cambridge University Press, 455 pp. [ISBN 0 521 21562 5.]
- Rosen, H., T. Novakov and B. A. Bodhaine, 1981: Soot in the Arctic. *Atmos. Environ.*, **15**, 1371-1374.
- Sloane, C. S., 1983: Summertime visibility declines: meteorological influences. *Atmos. Environ.*, **17**, 763-774.
- , 1984: Meteorologically adjusted air quality trends: Visibility. *Atmos. Environ.*, **18**, 1217-1229.
- Spar, J., 1953: A suggested technique for quantitative precipitation forecasting. *Mon. Wea. Rev.*, **81**, 217-221.
- Vinzani, P. G., and P. J. Lamb, 1985: Temporal and spatial visibility variations in the Illinois vicinity during 1949-80. *J. Climate Appl. Meteor.*, **24**, 435-451.
- Younkin, R. J., J. A. LaRue and F. Sanders, 1965: The objective prediction of clouds and precipitation using vertically integrated moisture and adiabatic vertical motions. *J. Appl. Meteor.*, **4**, 3-17.

Climatic Temperature Normals

NATHANIEL B. GUTTMAN AND MARC S. PLANTICO

National Climatic Data Center, Asheville, NC 28801

(Manuscript received 30 January 1987, in final form 1 May 1987)

ABSTRACT

The published 1951–80 daily normals of maximum and minimum temperatures were prepared by interpolating between average monthly values. This study compares the published normal and 30-yr average daily temperatures in the eastern half of the United States. It was determined that the published normals statistically differ from the series created by using daily data. It was also determined that 1-day persistence is a feature of the daily data. The possibility of climatic singularities as evidenced from the analysis of 30 temperatures (1951–80) on selected dates became apparent and warrants further investigation.

1. Introduction

The National Climatic Data Center has the responsibility to fulfill the mandate of Congress “. . . to establish and record the climatic conditions of the United States (15 U.S.C. 313).” The World Meteorological Organization (WMO) recommends that member countries should compute normals for representative stations. Normals are defined as “period averages computed for a uniform and relatively long period comprising at least three consecutive 10-yr periods (WMO, 1979).”

Combining the WMO recommendations with the Congressional mandate, the National Climatic Data Center prepares daily normals of maximum and minimum temperatures. Daily normals are not explicitly treated in the WMO regulations, but period averages are defined as the “arithmetical mean of climatological data . . . (WMO, 1979).”

The published 1951–80 daily normals of maximum and minimum temperatures were prepared by interpolating between average monthly values. The interpolation scheme was a cubic spline fit following the procedures described by Greville (1967). The series of daily values resulting from the cubic spline yields a smooth curve throughout the year that represents the annual temperature cycle for a station. Each series was edited to remove spurious inflection points caused by rounding, and to ensure functional relationships and consistencies among the several variables for which normals were prepared.

These published normals are synthetic in the sense that they are interpolated from monthly values. The WMO (1983) discusses synthesized data from the viewpoint of data reconstruction and estimation of missing values, but not from the viewpoint of substitution for available data. Of importance, however, is the WMO guideline that synthesizing be based on

sound physical reasoning and indicate the probable error involved. This paper examines the errors in the synthetic, published normals.

In order to examine the error, 30-yr averages of serially complete daily maximum and minimum temperatures observed at 74 National Weather Service first-order stations from 1951–80 in the eastern half of the United States were computed. Figure 1 shows the station locations. These averages were then arranged into series of 365 daily maximum and 365 daily minimum temperatures for each station and are called 30-yr average daily temperatures. The station network contains stations that did not experience significant moves during the 30-yr period.

This study compares the published normal and 30-yr average daily temperatures with the purpose of determining how well the published normals describe the daily temperature climate. If significant departures are found, they should be identified and described since energy users and planners, farmers and others could be impacted. If the published normals present an adequate description, then the cubic spline interpolation through mean monthly values is a cost-effective procedure for computing normals. The cost effectiveness results from reduced processing when working with 12 monthly values instead of 365 daily values.

2. Comparative analysis

The initial hypothesis in the analysis was that if the annual series of normals and 30-yr average daily values represent nearly the same curve, then the series of differences between the two curves should be random and uncorrelated. The difference series of both maximum and minimum temperatures were subjected to linear trend, runs up and down, runs above and below the mean, and serial correlation tests.

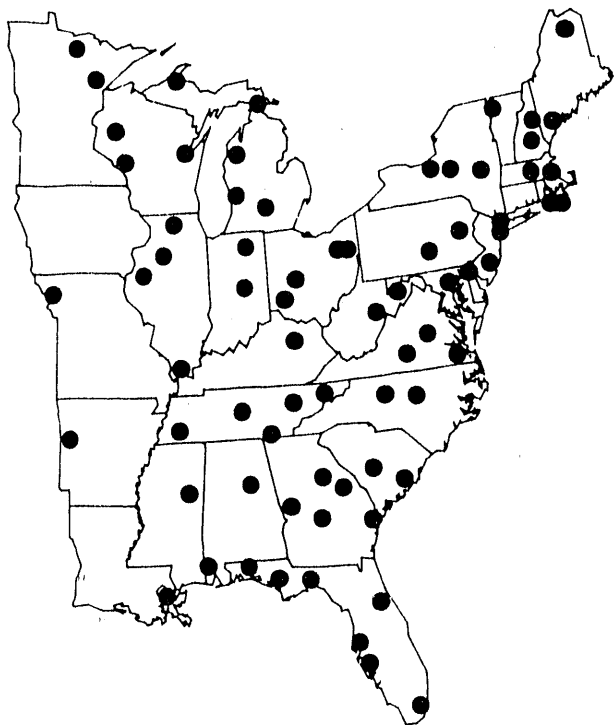


FIG. 1. Station locator map.

The tests were performed using the STATLIB computer package (Tryon and Donaldson, 1978). The linear trend test fits a least-squares line through the time-ordered data points X_i , $i = 1, N$. If no long-term trend is evident, the slope of the fitted line should be zero. This hypothesis is checked by Student's *t*-test. The presence of short-term drifts or cycles in the time-ordered data is indicated by the runs-up-and-down test.

The algebraic signs of the differences between adjacent numbers, i.e., the signs of $(X_{i+1} - X_i)$, form a sequence of $N - 1$ "+" or "-" signs. A sequence of consecutive "+" signs is called a run up, and a sequence of consecutive "-" signs is called a run down. Assuming that all $N!$ possible arrangements of the N values are equally likely, the expected means and variances of the number of runs can be computed and the number of runs can be regarded as normally distributed. A similar test for randomness looks at the distribution of runs of "+" and "-" signs of the sequential departures from the mean $(X_i - \bar{X})$. Serial correlation between adjacent data values was examined by computing the ratio of the mean of the square of successive differences $(X_{i+1} - X_i)$ to the variance of the data. If the data are randomly distributed, the ratio will be within appropriate confidence limits defined by the normal distribution. More detailed descriptions of the randomness tests are in Ku (1973), Hald (1952) and Brownlee (1960).

The hypothesis was rejected at the 0.05 confidence level for all stations for both maximum and minimum temperatures. Test results indicated highly significant nonrandomness from the runs tests and a highly significant serial correlation. Linear trends were not apparent at any station. Selected time plots of the annual difference series indicated that the nonrandomness could possibly be attributed to mixing of different seasonal synoptic environments. Note in Fig. 2, for example, the greater variation in the 30-yr average daily temperature series during the winter season compared to the summer season.

The tests were repeated on seasonal (e.g., January through March, October through December) difference series. Similar results were obtained; all stations failed, but the magnitude of the test statistics was generally reduced by half. Knowing that the serial correlation

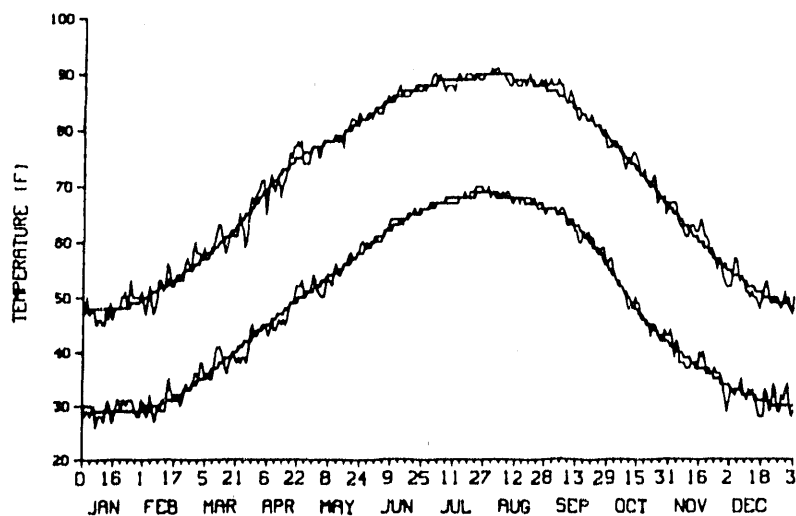


FIG. 2. Annual time series of Chattanooga, Tennessee published normals and 30-yr average daily maximum (top curves) and minimum (bottom curves) temperatures.

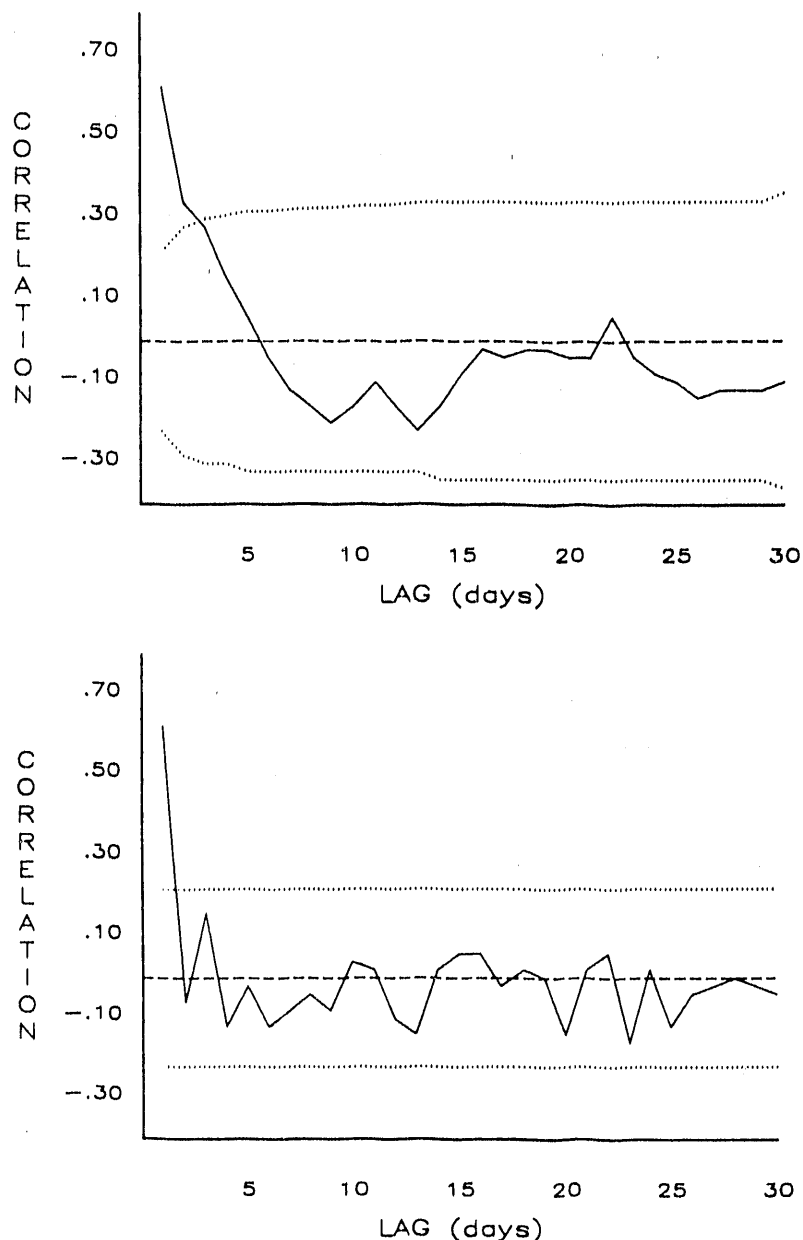


FIG. 3. (a) Autocorrelation and (b) partial autocorrelation for Green Bay, Wisconsin winter season maximum temperature differences. Dotted lines indicate 0.05 significance limits.

could influence the runs tests, the seasonal series were modeled by a first-order autoregressive process using the BMDP time series computer package (Dixon, 1981). Examination of autocorrelation (1-day lag correlation) and partial autocorrelation (2-day lag correlation adjusted for the autocorrelation) functions (Fig. 3, for example) justified the model in a statistical sense. Physically, a 1-day persistence or lag in the 30-yr average daily temperatures was considered reasonable in terms of the instability of temperature distributions as discussed by AFCRL (1967; 1968a-d). It was also considered reasonable in terms of multiday persistence that

is common in the daily weather associated with the passage of pressure centers over a station.

Residuals (original difference series minus modeled series) from the first-order autoregressive model were passed through the STATLIB randomness tests. Results indicated that the serial correlation was removed from the seasonal maximum and minimum temperature difference series by the model. The runs tests, however, showed nonrandomness in the seasonal series at many more stations than could be expected by chance. In all cases, fewer runs were observed than expected. Table 1 summarizes the tests results.

TABLE 1. Test results for quarterly residual maximum and minimum

	Runs up and down		Runs above and below the mean		Serial correlation	
	Max	Min	Max	Min	Max	Min
Jan.-Mar.						
Accept	48	41	73	67	74	72
Reject	26	33	1	7	0	2
Apr.-June						
Accept	49	34	72	71	74	74
Reject	25	40	2	3	0	0
July-Sept.						
Accept	24	10	67	59	74	74
Reject	50	64	7	15	0	0
Oct.-Dec.						
Accept	46	42	72	70	74	72
Reject	28	32	2	4	0	2

The areal extent of nonrandomness in the residual series was mapped by season to try to isolate geographical similarities. Unfortunately, no clear-cut pattern emerged.

The differences between the published normals and the 30-yr average daily temperatures were then examined quarterly for several stations to try to identify any consistent temporal problem that could be contributing to the nonrandomness. A subjective analysis of the time series revealed the existence of a warm spell

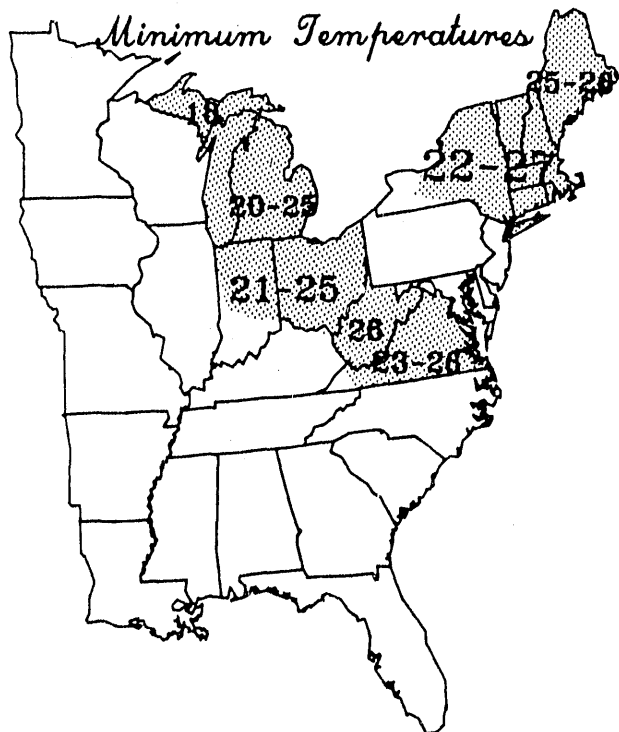


FIG. 4. Dates and area of significant January warming.

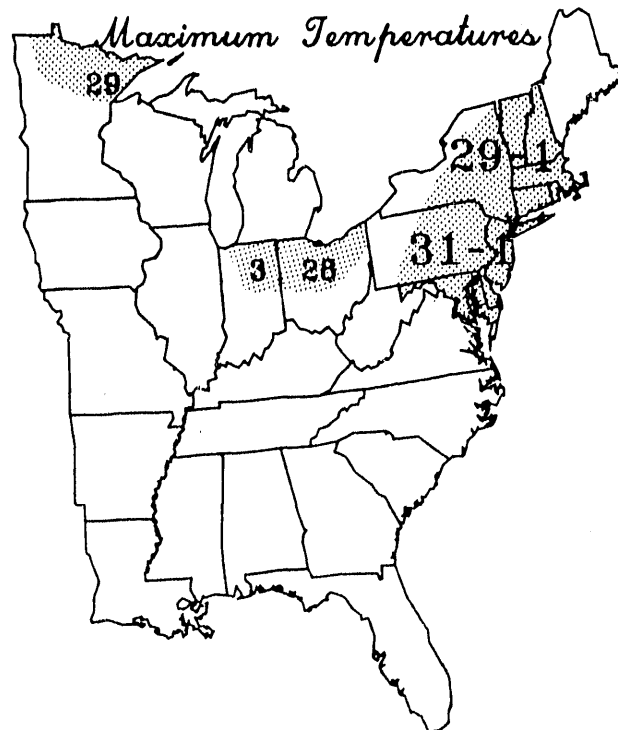
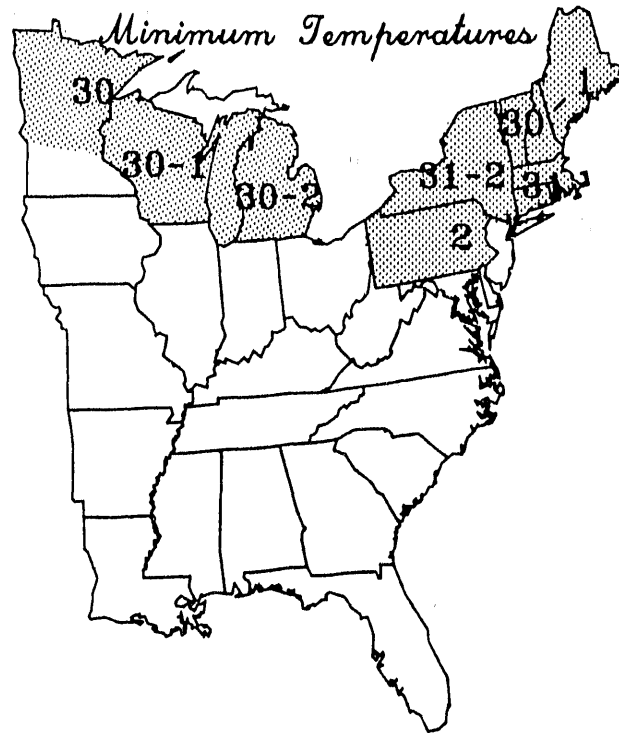


FIG. 5. Dates and area of significant January-February cooling.

in January, a cool period at the end of March, a warm period in April, a late summer warming, and a November warming followed by a cold spell.

These periods were then examined in more detail. The 30 values (1951-80) for a calendar date of maxi-

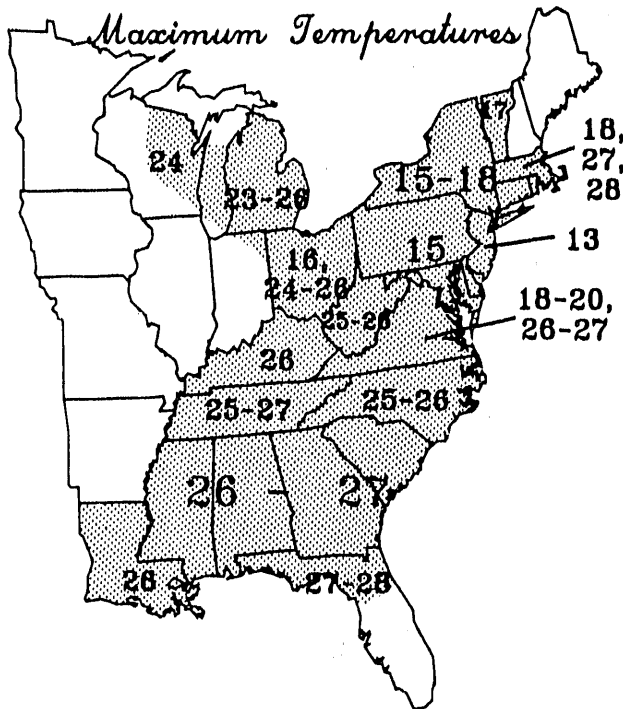


FIG. 6. Dates and area of significant March cooling.

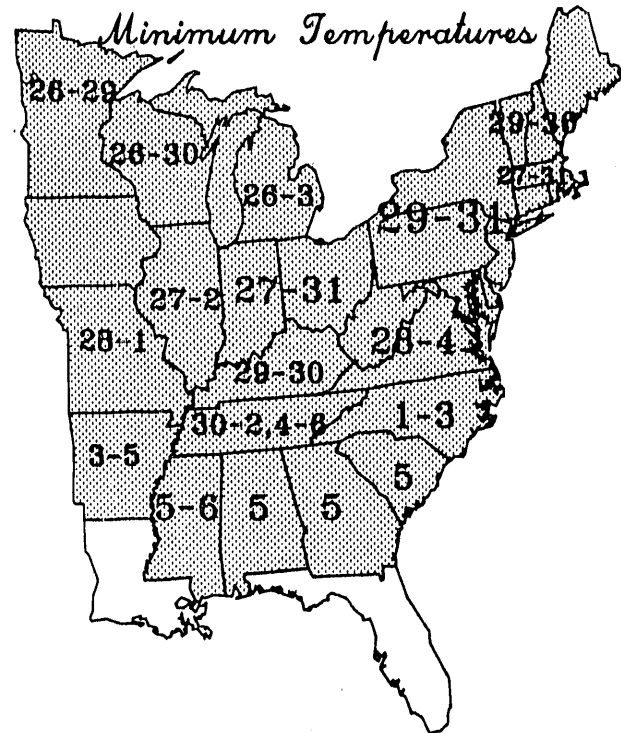


FIG. 8. Dates and area of significant August-September warming.

mum and minimum temperatures at all 74 stations were extracted from the database for January 16-February 4, March 12-30, April 12-30, August 26-Sep-

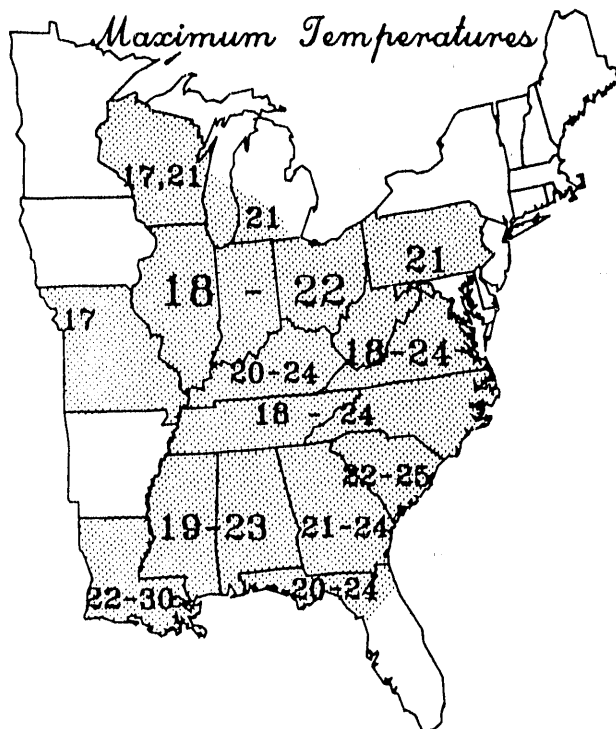


FIG. 7. As in Fig. 4 except for April.

tember 14 and November 10-December 4. The departures of the temperatures from the published normals were computed, and the number of positive and negative departures for each date were tabulated.

Hypothesizing that if the published normals adequately represent the climate, it was concluded that the number of positive departures for a date should equal the number of negative departures (i.e., the median value of the departures should equal zero). Using a sign test described by van der Waerden (1969), the confidence limits for the median were determined. The departure distributions were then tested at the 0.05 confidence level for the null hypothesis that the observed medians were statistically equal to zero.

Significant positive departures of about 3-6 deg of both the maximum and minimum temperatures occur in January in the northeast part of the study area. Figure 4 shows the significant dates for the minimum temperatures. (The map for the maximum temperatures is essentially the same.) This feature is commonly known as the January thaw (American Meteorological Society, 1970). It lasts about 3 to 5 days. Figure 5 shows that it is followed by a cold spell. Temperatures during this cold spell are generally 4°-6°F below the published normals.

Figure 6 depicts the dates of significant below-normal maximum temperatures in March. Although not shown, the minimum temperatures show the same pattern. The northern areas indicate two periods of cold around March 15-17 and 24-26. The southern

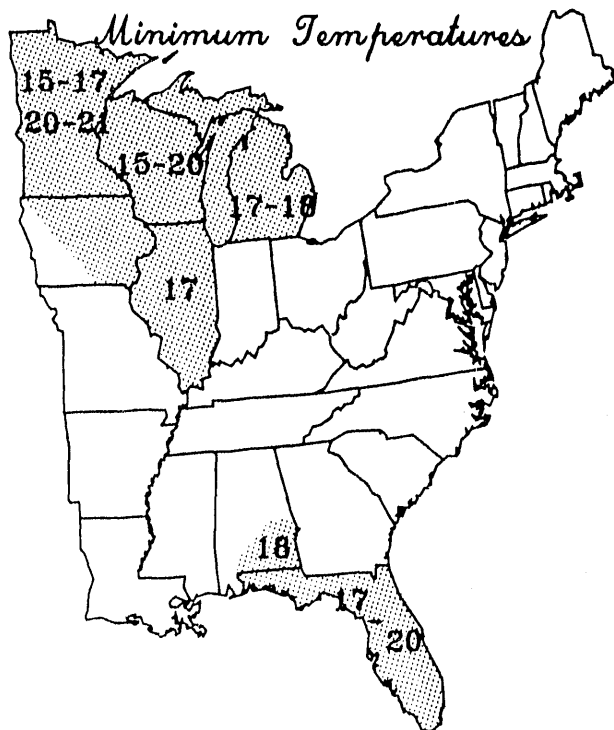
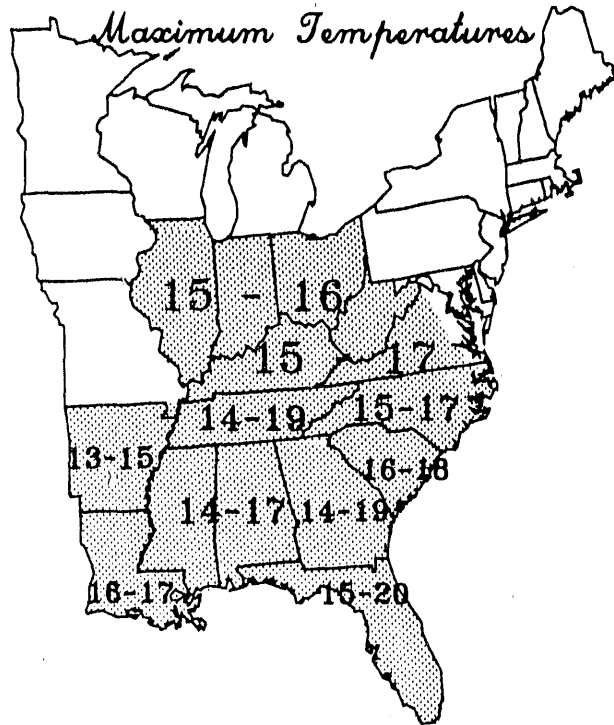


FIG. 9. As in Fig. 4 except for November.

areas indicate one cold period around March 26-27. The magnitude of the departures from the published normals ranges from 4° to 7°F. About one month later (Fig. 7), temperatures about 4° to 5°F warmer than normal are common south of New England and Mich-

igan. The maximum temperature pattern in this figure is the same as the minimum temperature pattern (not shown).

Late summer warming of a magnitude of more than 3°F above normal occurs during the last week in August in the northern states. A week later the southern states appear to be warmer than normal, but the magnitude is only about 1-2°F. Figure 8 shows the minimum temperature pattern; maximum temperatures exhibit a similar pattern.

Maximum temperatures in the southern two-thirds of the study area are about 5°-6°F warmer than normal in mid-November (Fig. 9a). Above normal minimum temperatures do not follow the same pattern. Two separate geographical areas (Fig. 9b), the Great Lakes and Florida, experience mid-November temperatures of about 4°F above normal. At the end of the month (Fig. 10), minimum temperatures of about 4°-6°F below normal prevail throughout the eastern United States. Maximum temperatures (not shown) exhibit a similar pattern.

The singular events previously described are based on the analysis of the difference series without consideration of the autoregressive model. Figure 11 shows examples of quarterly time series comparing the residuals from the autoregressive model to the difference series. These examples were selected because they most graphically illustrate singularities, variability of the data, and the coherence of the two curves. Similar graphs for other stations for all four seasons for both

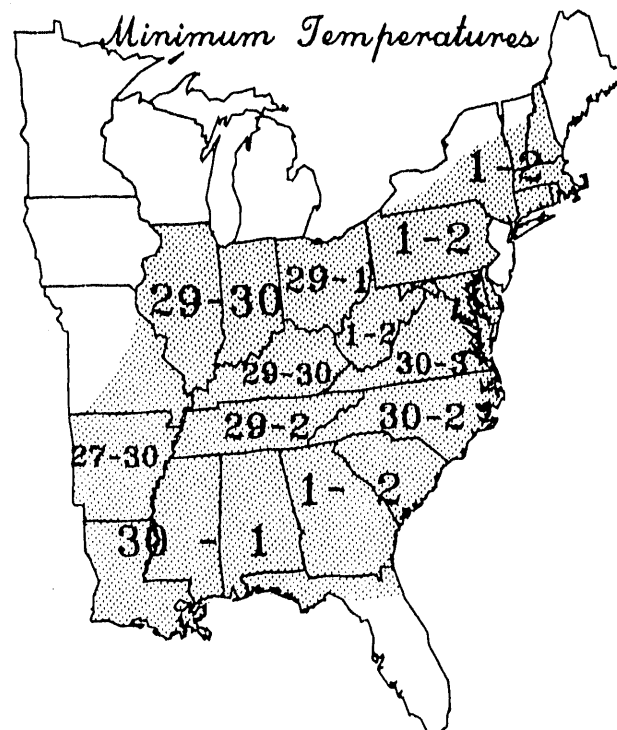


FIG. 10. As in Fig. 5 except for November-December.

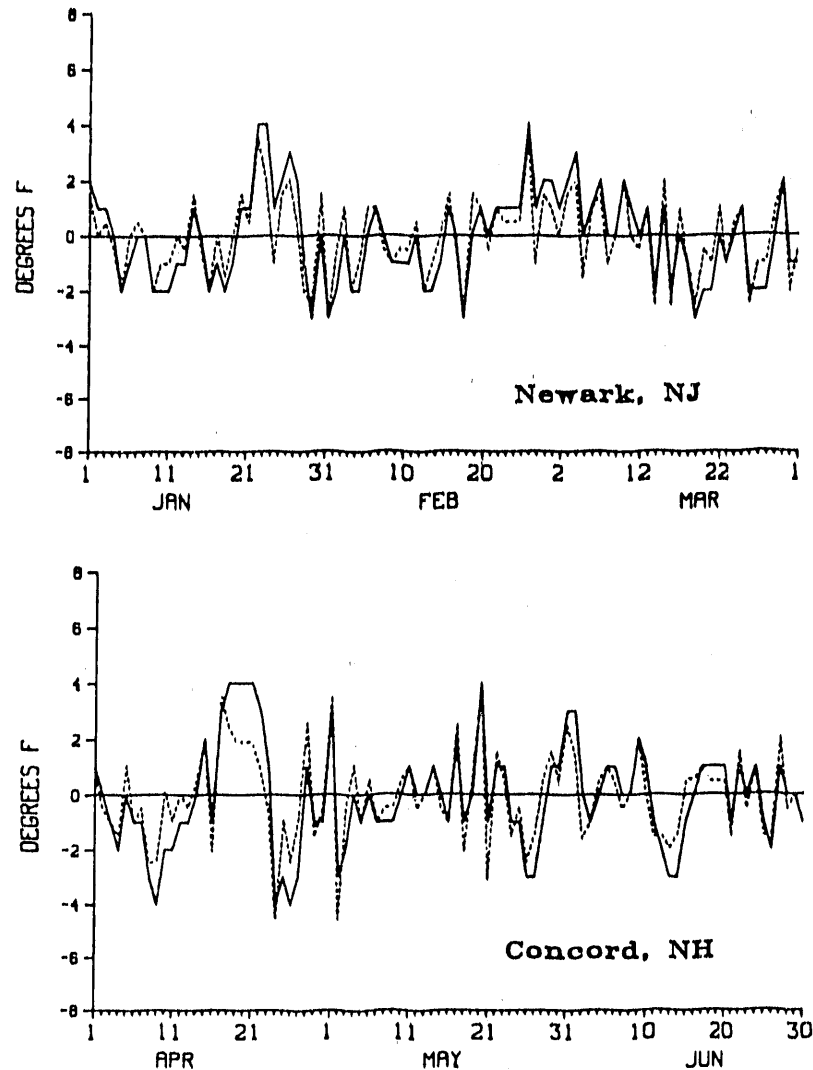


FIG. 11. Maximum temperature differences (solid curve) and residuals from persistence model (dotted curve).

maximum and minimum temperatures exhibited the same coherence as those in Fig. 11. The singularities apparently exist even after consideration of the autoregressive model.

3. Conclusions

Based on the analysis of data from 74 stations in the eastern half of the United States, the published normals and 30-yr average daily temperatures were determined to be different. The magnitude and variability of the differences, which are measures of the errors in the synthetic, published normals, are sufficient to question the use of the cubic spline fit through monthly values to describe a daily climate.

The investigation into the structure of the differences should be continued. It was determined from this study that 1-day persistence is a feature when working with 30 yr of daily data. The residual series created after

modeling the serial correlation still shows nonrandomness for many stations. The causes should be identified. The possibility of singularities as evidenced from the analysis of the 30 temperatures (1951-80) on selected dates also needs to be investigated in more detail.

These additional studies would, hopefully, lead to a predominantly deterministic representation of the temperature series. The model would consist of additive components. It is anticipated that the first two components would be the cubic spline fit through monthly values to account for the effects of the earth's revolution and the first order autoregression to account for day-to-day persistence. Other components, such as singularities, would arise from the further investigations. The residual from the model should be noise.

Modeling would not be necessary if 30-yr, serially complete, homogeneous daily datasets were available. Unfortunately, station moves, instrument changes, procedural changes in observing and/or missing data

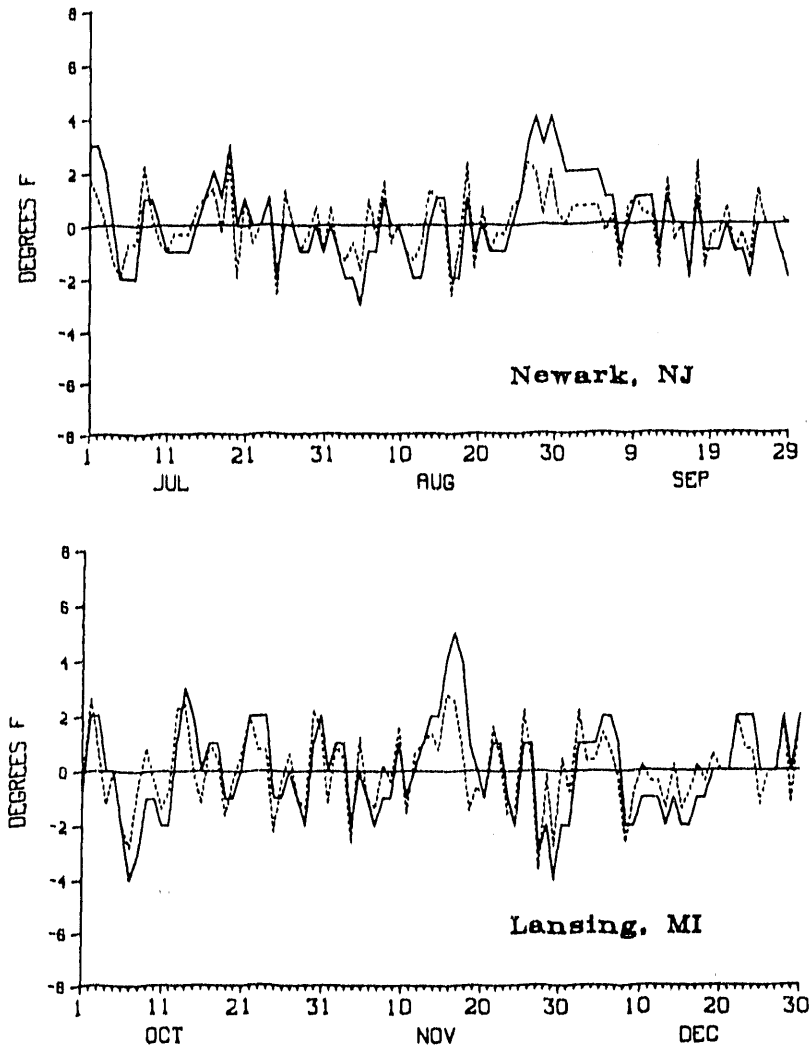


FIG. 11. (Continued)

have caused the records at many stations to depart from the ideal. The deterministic components of a model become helpful in adjusting data and fitting in missing observations so that a serially complete, homogeneous dataset can be created. This new dataset would be synthetic yet representative of the climate. Averaging the calendar data from the new set over the 30 yr would provide a better description of the daily temperature climate than the existing normals.

REFERENCES

- Air Force Cambridge Research Laboratories, 1967: Climate normals as predictors, Part 1: Background. *AFCRL-67-0313*, Bedford, MA.
- , 1968a: Climatic normals as predictors, Part 2: Extension. *AFCRL-68-0673*, Bedford, MA.
- , 1968b: Climatic normals as predictors, Part 3: Median vs. mean. *AFCRL-68-0255*, Bedford, MA.
- , 1968c: Climatic normals as predictors, Part 4: Verification. *AFCRL-69-0002*, Bedford, MA.
- , 1968d: Climatic normals as predictors, Part 5: Conclusion. *AFCRL-69-0003*, Bedford, MA.
- American Meteorological Society, 1970: *Glossary of Meteorology*. Boston, MA.
- Brownlee, K. A., 1960: *Statistical Theory and Methodology in Science and Engineering*. John Wiley and Sons.
- Dixon, W. J., Ed., 1981: *BMDP Statistical Software 1981*. Univ. of California Press.
- Greville, T. N. E., 1967: Spline functions, interpolation, and numerical quadrature. *Mathematical Methods for Digital Computers, Vol. II*. A. Ralston and H. S. Wilf, Eds., Wiley and Sons.
- Hald, A., 1952: *Statistical Theory with Engineering Applications*. John Wiley and Sons.
- Ku, H. H., 1973: A users' guide to the OMNITAB command "Statistical Analysis," Tech. Note 756, Nat. Bur. Standards, Washington, DC.
- Tryon, P. V., and J. R. Donaldson, 1978: *STATLIB, A Library of FORTRAN Subroutines for Statistical Analysis of Experimental Data*. Nat. Bur. Standards Center for Applied Mathematics, Boulder.
- van der Waerden, B. L., 1969: *Mathematical Statistics*. Springer-Verlag.
- World Meteorological Organization, 1979: Technical Regulations, Vol. 1, WMO-No. 49, Geneva.
- , 1983: Guide to climatological practices. WMO No. 100, Geneva.

On an Additive Model of Daily Temperature Climates

NATHANIEL B. GUTTMAN AND MARC S. PLANTICO

National Climatic Data Center, Asheville, North Carolina

21 April 1988 and 13 December 1988

ABSTRACT

Guttman and Plantico reported on an additive model to describe daily temperature climates. This note reports on spectral analyses of the nonrandom residuals from the model. We concluded that quasi-periodic features are not present in the 1951–80 residual maximum and minimum temperature data.

Two areas of search for model components are suggested for future research. First, apparent singularities should be investigated and mathematically described. Second, nonstationary patterns over the period of record that may be linked to long term climatic variability should be understood and modeled.

1. Introduction

In a recent paper (Guttman and Plantico 1987) we examined the difference between the sets of daily maximum and daily minimum temperature normals and the sets of the 1951–80 30-year averages of serially complete daily maximum and minimum temperatures observed at 74 National Weather Service stations in the eastern United States. The normals were prepared by interpolating along a cubic spline fit to the average monthly values, while the 30-year averages were computed for each calendar date from daily data. We concluded that the normals do not adequately portray the daily temperature climate. We also proposed an additive model to describe the daily temperature climate T_i :

$$T_i = \sum_j D_{ij} + e_i \quad i = 1, \dots, 365 \quad (1)$$

where D_{ij} is the j th deterministic component for day i and e is a random component. If the e_i are assumed to be independent and identically normally distributed with mean $\mu = 0$ and variance σ^2 , then the normal maximum or minimum temperature N_i becomes

$$N_i = \sum_j D_{ij}. \quad (2)$$

Two deterministic components were identified. The first is the cubic spline fit S through mean monthly temperatures. This curve represents the smooth annual march of temperature. The second component is a first order autoregressive process AR that models a 1-day persistence. Residuals from the spline and autoregression

$$R_i = T_i - S_i - AR_i \quad (3)$$

were found to exhibit nonrandomness.

This note reports on continuing efforts to explain the residual nonrandomness. Using the techniques of exploratory data analysis, we are attempting to isolate independent, deterministic components that additively contribute to the observed time series. Section 2 describes an examination of the data for periodicities. Section 3 outlines directions for additional work.

2. Search for periodicities

Figure 1 is an example of a residual time series R_i . While subjectively examining this and other R_i , we noticed a similarity to time series of water waves (e.g., see Kinsman 1965). Wave records show patterns that result from the interaction among and addition of individual waves and these patterns can be decomposed into individual, independent, periodic components. Because of the apparent similarities between R_i and water wave time series, we applied spectral analyses to the R_i to try to identify any periodic components. If they exist, these components would then become part of the additive temperature model.

The existence of periodicities in the daily temperature record has been questioned for over 100 years. Clayton's (1943) collection of papers summarizes the thinking about cycles from the late 1800s through the early and mid 1900's. Although the time series of the residuals from the 1951–80 data suggested possible cyclical behavior, we did not have any preconceptions about the structure of any underlying periodic process. An exploratory analysis approach was used (e.g., see Tukey 1977 and Andrews 1978) with the intent of determining whether the study of periodicities as a temperature model component warrants further investigation.

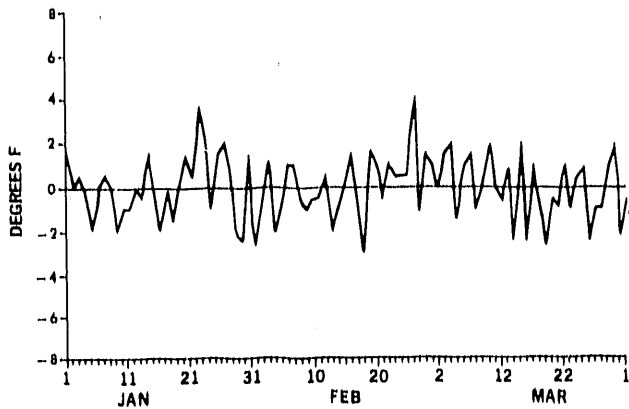


FIG. 1. Example of residuals R_t from an annual spline and 1-day persistence model of 1951–80 average daily maximum temperatures at Newark, New Jersey.

Data for 16 geographically scattered stations in the eastern United States were processed via the BMDP Univariate Spectral Analysis program (BMDP 1985). The spectral estimates are weighted averages of adjacent periodograms where the weighting function is cosinusoidal. For each dataset, spectra were computed for bandwidths yielding estimates with 8 , $3n^{1/3}$ and $n^{2/3}$ degrees of freedom, where n is the number of observations in the dataset. The spectra were computed for calendar quarter and annual time series of the maximum and minimum temperature residuals.

Most of the spectra exhibited primary peaks at periods of two to five days. Figure 2 shows the Syracuse, New York, first quarter maximum temperature spectrum estimated from periodograms averaged over a frequency bandwidth of 0.12. Although the frequency of the highest peak shifts slightly, the shape of this spectrum is typical of all the spectra that were computed.

To gain insight into the nature of the cycle represented by the maximum spectral peaks, the BMDP complex demodulation program was used to trace over

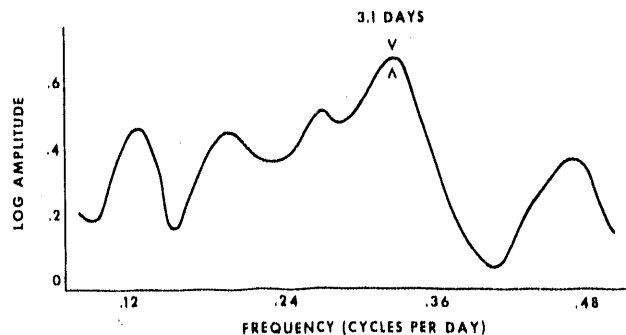


FIG. 2. Spectral density of first quarter (January–March) residuals from an annual spline and 1-day persistence model of 1951–80 average daily maximum temperatures at Syracuse, New York. The peak spectral density occurs at a frequency of 0.3222 cycles per day and a period of 3.1 days.

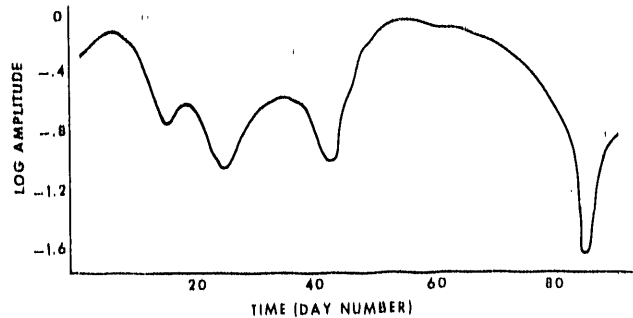


FIG. 3. Log amplitude time series of the spectral density shown in Fig. 2 demodulated at the peak frequency of 0.3222 cycles per day.

time the frequency band component corresponding to the peaks. The aim of complex demodulation is to extract information about a perturbed periodic component with slowly varying amplitude and phase. The analysis is similar to harmonic analysis in that the amplitude and phase are described, but different in that the amplitude and phase are determined only by the data in a local time interval rather than by the whole series. An excellent description of complex demodulation is given by Bloomfield (1976).

Figures 3 and 4 show the amplitude and phase, respectively, of the demodulated 3.1 day cycle identified in Fig. 2. The amplitude varies substantially; the plotted curve is quite ragged. The phase plot also shows substantial variation. The erratic behavior of the curves in Fig. 3 and 4 suggests that the residual data do not, in fact, contain a consistent cycle with a period near 3.1 days.

For comparison Fig. 5 and 6 conceptually depict an amplitude and phase plot for a demodulated cycle that would lend support to the existence of a periodic cycle. The consistency of the plots indicates that the demodulated cycle is strong and persists through time. Consistent changes in the amplitude and phase would indicate the existence of a cycle near, but not at, that of the demodulated cycle.

The demodulation example for Syracuse is typical of the plots for all other stations for all seasons as well as for the annual time series. We therefore conclude

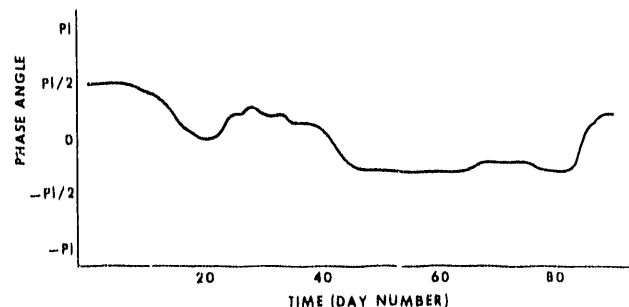


FIG. 4. Phase time series of the spectral density shown in Fig. 2 demodulated at the peak frequency of 0.3222 cycles per day.

that the exploratory analysis did not reveal any quasi-periodic components in the residual maximum and minimum temperature data for the years 1951–80.

3. Discussion

The lack of cycles during 1951–80 indicates that further study of periodicities is unwarranted and that other exploratory avenues should be pursued to explain the residual nonrandomness. An anonymous reviewer suggested that a moving average autoregressive model or the inclusion of lags out to about five days may result in a white noise (random) residual time series. The use of a simpler, first order autoregressive model by Guttman and Plantico (1987) was based, however, on the statistical analysis that showed only a 1-day persistence in temperature series after removal of the annual cycle. The analysis also showed that serial correlation was removed by the first order autoregressive model. At this point in the exploratory process, modeling higher order lags does not appear to be fruitful. The use of moving average models is discounted at the present time because the smoothing effects of the averaging process may mask potentially important day to day variations.

We suggest two areas of search for model components. First, the apparent singularities in temperature described by Guttman and Plantico (1987) may possibly be modeled by a deterministic perturbation or impulsive force component P_i so that

$$R_i = T_i - S_i - AR_i - P_i. \quad (4)$$

The nature of P_i should be investigated in terms of its mathematical form as well as in terms of its physical justification. The relationship between singularities and other atmospheric singularities identified by, e.g., Lanzante (1983) and Kalnicky (1987) needs to be described.

The second suggested area is the search for nonstationary patterns that may be linked to long term cli-

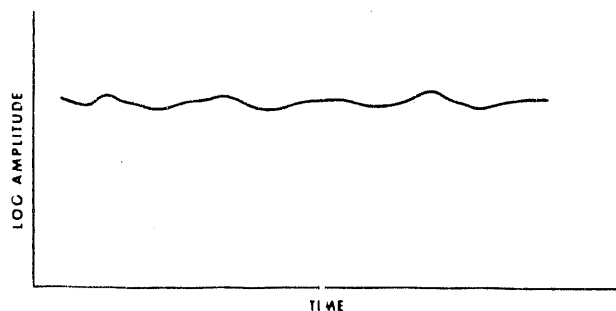


FIG. 5. Conceptual log amplitude time series of a persistent cycle with frequency λ .

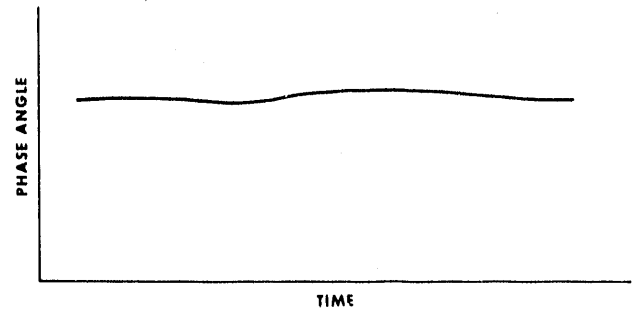


FIG. 6. Conceptual phase time series of a persistent cycle with frequency λ .

matic variability. If long term variations exist, a temperature model should be based on all available data rather than on a 30-year subset spanning the normals period. The use of additional data should enable the researcher to gain a better understanding of anomalous events such as singularities of climate trends, and of other components that may be either independent or interactive. Understanding these components should lead to the development of a model that describes the observed temperatures over the entire period of record. Subsets of the model values, such as for any desired normals period, could then be easily extracted.

The advantage of a modeling approach lies in the fact that a model can be evaluated in probabilistic terms so that measures of confidence may be placed on the mathematical representation of the real world climate. Decisions on the reality of hypothesized signals in the data can therefore be made. Additionally, the noise can be statistically evaluated to provide confidence information about means, extremes and variances.

REFERENCES

- Andrews, D. F., 1978: Data analysis, exploratory. *International Encyclopedia of Statistics*, W. H. Kruskal, and J. M. Tanner, Eds., Free Press, 97–107.
- Bloomfield, P., 1976: *Fourier Analysis of Time Series: An Introduction*, Wiley, 118–146.
- BMDP Statistical Software, 1985: University of California Press.
- Clayton, H. H., 1943: *Solar Relations to Weather and Life, Vol. II*. Clayton Weather Service, 439 pp.
- Guttman, N. B., and M. S. Plantico, 1987: Climatic temperature normals. *J. Climate Appl. Meteor.*, **26**, 1428–1435.
- Kalnicky, R. A., 1987: Seasons, singularities, and climatic changes over the midlatitudes of the Northern Hemisphere during 1899–1969. *J. Climate Appl. Meteor.*, **26**, 1496–1510.
- Kinsman, B., 1965: *Wind Waves*. Prentice-Hall, **43**, 183–236.
- Lanzante, J. R., 1983: Some singularities and irregularities in the seasonal progression of the 700 mb height field. *J. Climate Appl. Meteor.*, **22**, 967–981.
- Tukey, J. W., 1977: *Exploratory Data Analysis*. Addison-Wesley, 688 pp.

Statistical Descriptors of Climate

Nathaniel B. Guttman
National Climate Data Center
Asheville, NC 28801

Abstract

An adequate description of climate is required to meet the informational needs of planners and policy-makers who use climate as a factor in their decision-making processes. Because normals have become firmly entrenched as a descriptor of climate, their history and their perception by the public are discussed. An "exploratory data analysis" approach is suggested.

1. Introduction

In his classic handbook, Hann (1903) described *climate* as the sum total of the meteorological phenomena that characterize the average condition of the atmosphere at any one place on the earth's surface. Although Hann recognized the importance of departures from average conditions, he emphasized that climate should be described by average weather conditions. During the 1800s and the first half of the 1900s, this view dominated the work in climatology. Smith (1975) states that during this time period, description and classification of climate, often based on means, prevailed. The concept of normals also evolved and matured during this time period.

A gradual change in thinking has taken place during the past few decades. Information about average climatic conditions is still important, but descriptions of climatic variability have become prominent. The importance of climatic variability lies in the impacts of climate on soil, water, food, energy, and shelter—the basic ingredients of survival. Planners and policy-makers who use climate as a factor in their decision-making processes require an adequate description or assessment of climate.

The history of normals and perception of normals by the public are discussed first; the inadequacies of normals in meeting the climatic information requirements of the users are then outlined, and an "exploratory data analysis" approach to describing and modeling climate is presented. The results of the model should give the planner practical information about the underlying structure of the climate system of interest.

2. Background of normals

The term "normal" first appeared in the meteorological literature in 1840 in a monograph by Dove on temperature variations over the globe (Landsberg 1955). The word was used by Dove in several papers in three different contexts: (1) as a reference value obtained for a given latitude belt by averaging all observations in that belt, (2) as a reference locality with a long observational record for comparisons in time and space, and (3) as equivalent to the average or mean of a long series of observations.

The third context survived during the latter half of the 19th century. The International Meteorological Committee in 1872 resolved to compile mean values over a uniform period in order to assure comparability between data collected at various stations. According to Landsberg (1955, 1972, 1975), the doctrine gradually developed that climate is essentially constant during intervals that are long compared to human experience. It was assumed that long-term averages would converge to this stable value or normal. International agreements eventually led to the compromise that the appropriate interval for computing a normal would be 30 years. In 1935 it was also agreed that 1901 to 1930 would be the period for computing normals.

One of the problems of these normals is that many places in the world did not begin observations in 1901. Another problem is that the notion of a "stable climate" has, over the last few decades, become obsolete. The World Meteorological Organization, which eventually succeeded the International Meteorological Committee, has attempted to rectify these problems by using a sliding time scale to compute normals (Jagannathan et al. 1967). *Normals* are now defined as "period averages computed for a uniform and relatively long period comprising at least three consecutive 10-year periods" (WMO 1984). *Climatological standard normals* are "averages of climatological data computed for the following consecutive periods of 30 years: January 1, 1901 to December 31, 1930, January 1, 1931 to December 31, 1960, etc." (WMO 1984). Normals are computed every decade in an attempt to keep up with any climatic changes that may take place.

Climatologists generally understand that a normal is simply an average of a climatic element over thirty

years, and that departures from normal are the difference between currently observed values and that particular 30-year average. It is also generally recognized by climatologists that a "normal" value is usually not the most frequent value (mode) nor the value above which half the cases fall (median). The general public, however, has a tendency to perceive the normal as ordinary or frequent (Landsberg 1955, 1972). The perception probably arises from the common usage of the word normal: "(1) conforming, adhering to, or constituting a usual or typical standard, level or type, (2) the usual or expected state, form, amount or degree" (DeVenne 1982).

The improper interpretation of normals by the public has a profound effect on the results of applying the values that are published each decade. Any large deviations from the temporal averages are thought to be indicative of climatic change, even though they may be only or part of the "noise" pattern (Landsberg 1975). As stated by McKay (1975), planners make decisions based on the assumption that a 30-year average has predictive value. Since the assumption is doubtful, McKay claims that those decisions must ultimately be in error and may lead to unfortunate consequences.

3. Normals and needs

Historically, normals have been used for two purposes—comparison and prediction. The former usage allows the analyst to evaluate and assess the deviation of a value of a climatic element from that for a reference period (Jagannathan et al. 1967; Saxton 1979). Normals also allow comparison in synoptic analysis, i.e., spatial comparisons at a given time. For these purposes, the period of record should be uniform for all stations. The normals as they have been previously defined and published meet the needs of those making these kinds of comparisons. It is emphasized, however, that these comparisons imply very little about climatic change, non-random fluctuations, or extremes. They are simply an assessment of deviations from a reference value.

The second use of normals is prediction from a 30-year average. *Prediction* is defined, for our purposes here, as planning or assessing risks based on the likelihood of recurring climatic scenarios. It is instructive to examine what is being predicted in several areas of activity so that commonalities can be drawn.

In 1983 a workshop was convened to discuss the interactions of climate and energy. Landsberg (1984) in his keynote paper stressed that industries are not only concerned with moderate weather and climate fluctuations, but also with extremes of the climatic elements. The oil industry (Leavitt 1984) needs pre-

cise estimates of the range of climatic variation and, in particular, extremes. This climatic information is used in planning for oil exploration and in estimating energy use. The gas industry requires a long-term weather history to plan future supply needs, facilities, and storage, and also requires means and extremes to determine price structures (Laurmann 1984). Climatic descriptions including averages, variability, extremes, and special events are needed by the coal industry for planning reclamation activities and avoiding weather-related shutdowns (Wilson 1984; McKee and Doesken 1984). The electric utilities require predictions of means and extremes for determining rate structures, baseline and peak loads, and facility siting (Rotty 1984).

The climatic information needs of agribusiness are summarized by Lamb et al. (1985). Historical data such as normals are used by the industry in assessing frost risks, estimating crop yields, analyzing supply and demand relationships, developing marketing strategies, planning planting schedules and herbicide applications, evaluating chemical-product trials, and scheduling of financial borrowing and investment activity. Most of the information is used for general background as opposed to making specific decisions. The needs, however, include predictions of climate, detailed information for input to crop yield models, and estimates of variability for design and planning of operations. The need for analytical measures of climatic variability were also recently mentioned by many authors, e.g., Lehman (1987), Todorov (1985), Smolander and Lappi (1985), Russell (1984), and Stern and Coe (1982). An extensive annotated bibliography on the use of climatic variables that are agriculturally sensitive is given by McQuigg (1975).

Other areas where climatic assessments are made are the building industry and architecture. Architectural design criteria include temperature means, extremes and variability; solar radiation variability; average and extreme wind (Simiu and Lozier 1975); and precipitation averages and extremes (Watson and Labs 1983; Marsh 1977; Olgyay 1963; Aronin 1953). These climatic factors enter into the placement of building openings such as windows, directional orientation, landscaping, building shape, etc. The materials used in construction as well as the mechanical systems used to heat and cool a building are determined partially by the risk of exceeding values of specified climatic elements (Ecodyne 1980; Lunde 1980; ASHRAE 1985). The need for climatic information is based on the goal of designing and building safe and comfortable structures with a predetermined level of risk of collapse or discomfort.

Applying climatic information to the prediction of future events in virtually all disciplines is similar to the examples just described. The common feature is

the assessment of risk or prediction error. The predictive value of normals was extensively studied over two decades ago by Enger (1959), Court (1967, 1968a-c), and Slusser (1968). After his 3-year study, Court (1968d) stated, "Climatic normals. . . are extremely inefficient for the primary use to which they are put: estimating future conditions" and ". . . the concept of climatic normal should be abandoned in practical climatology." Normals provide one measure of "central tendency" for a specified period of record, but users require a complete statistical description of a climatic element, i.e., central tendencies, extremes, variability, and durations, in order to assess risk.

4. A satisfactory description of climate

The limited applicability of normals in meeting the needs of users who are interested in predicting future atmospheric events was outlined in the previous section. The following discussion defines what might constitute an adequate portrayal of climate and proposes an approach for obtaining the description.

According to Jagannathan et al. (1967), a satisfactory specification of the climate of a region or locality includes statistical measures such as the mean, range, frequency, and variations of several climatic elements, as well as the sequential character of the elements. The usually important climatic elements are wind, temperature, moisture, and pressure; phenomena observed within the atmosphere, such as rainfall, snow, and cloud amount; and sunshine and radiation balance. Jagannathan et al. (1967) further state that "climatology should provide information, as accurately as the available observations warrant, about the state and behavior of the atmosphere." The aim is not only to compute normals, but also to abstract, from the past observational record, the underlying patterns characterizing the atmospheric environment. These patterns become the basis for risk assessments and prognostications discussed in the previous section.

The collection of historical weather data, i.e., the observational record, is an incomplete log of complex temporal and spatial interactions in the global environment. Obviously, data do not exist for every geographical point for all time. In addition, the data are subject to errors of instrumentation, observing, and processing. The chronological series of data can therefore be considered as a realization or sample of the underlying physical processes that control climate.

A basic approach in analyzing the record of observations for the purpose of describing climate includes three steps and is known as exploratory data

analysis (Tukey 1977). Andrews (1978) defines the approach as the manipulation, summarization, and display of data to uncover the underlying structure in the data. Judging from the literature, the procedures have only recently been formally applied to atmospheric data (Snijders 1986; Flueck et al. 1986; Zeger 1985), although consulting meteorologists and climatologists have probably used the analytical techniques informally for decades.

The first step of the analysis is examination. It is designed to uncover systematic errors, to exhibit overall patterns, and to show departures from an a priori contemplated structure. Graphical tools are appropriate for this phase. "Stem-and-leaf" displays, histograms, weather diagrams, quantile plots, cumulative-frequency distributions, maps, and time-series plots are some of the graphical tools that allow an evaluation of the structure of the data. A good knowledge of the data acquisition and measurement processes as well as the physics underlying the climatic element is invaluable in interpreting the graphics.

It is appropriate during this step to perform non-parametric tests to determine randomness, symmetry, trends, correlations, and differences among samples. Several of these tests are discussed by Siegel (1956). Descriptive statistics—those that compactly express the salient features of the observations—should also be computed. These statistics include the sample average, median, and variance, i.e., measures of "location" and dispersion.

The data examination gives the analyst information about inhomogeneities within the observational record. At this point the data should be adjusted to remove known systematic errors such as instrument and observer biases and gradual changes in site exposures (Brooks and Carruthers 1953); inhomogeneity caused by discontinuities in the observations, such as instrument changes and recalibrations, location changes of the observing site, and changes in observing methods; changes in computational procedures; environmental changes at the site such as urbanization; and data processing errors in calculating, coding, transcribing, and transmitting. (An excellent review of errors in meteorological data is given by Filippov 1968.) Adjustment methods that compensate for some of these problems are described by, among others, Brooks and Carruthers (1953), Landsberg (1958), Conrad and Pollak (1950), Thom (1966), and Craddock (1981). After the known errors have been removed from the data, the corrected observational record should once again be examined and descriptive statistics should be recomputed.

Of importance is that the examination step makes no assumptions about statistical models. It provides information that gives direction to "cleaning up" the

data as well as allowing the analyst to postulate reasonable models.

The second step, estimation, relies on the formal procedure of inferential statistical analysis and is concerned with the process of using data to make decisions about general situations on the basis of incomplete information (Mood and Graybill 1963). The analysis involves selecting a statistical model, checking the reasonableness of the model, and then drawing conclusions from the model (Hoel 1962). The model chosen assumes the existence of a signal such as a trend, oscillation, persistence, dependence, or frequency distribution in the observations. Specification of the model is based on the descriptive statistics and patterns that are identified and interpreted during the examination step.

Evaluation of the model is based on its "goodness-of-fit" to the real-world pattern being described, by the fulfillment of assumptions upon which the model is based, and by the degree of confidence that can be placed on the statistical inferences drawn from the model. If the model incorrectly describes a feature inherent in the data, then another model should be chosen, applied, and evaluated.

A statistical model that is deemed reasonable merely describes what is a feature within data. It does not explain why the feature occurs. The "what" becomes the raw material for developing physical models that explain "why." Statistical models describe patterns in a given data set. Physical models, on the other hand, explain the underlying physics of observed patterns and are not dependent upon any one data set.

Once a statistical model is determined to reasonably describe a data feature, values calculated from the model are removed from the data. The removal process, which is the third step in the exploratory data analysis, typically involves a subtraction or other transformation based on the statistical model. With the first gross component of the structure removed, the reduced or residual data may be examined for finer structure. This finer structure may again be modeled, estimated, and removed. The exploratory data analysis process may be repeated until all of the apparent structure has been removed and the residual data appear to be random, or patternless.

An example based on Guttman and Plantico (1987) illustrates the exploratory data analysis concept. Examination of 1951–1980 average daily maximum temperatures at stations in the eastern United States showed an annual cycle in the data. A smooth curve through average monthly maximum temperatures at each station was chosen as a model to represent the observed temperature patterns. Evaluation of the model led to the conclusion that the smooth curve is reasonable.

For each station, daily values of the smooth curve were subtracted from the observed data, i.e., the modeled values were removed from the data. The residual data were examined and found to exhibit a one-day persistence. A first-order, autoregressive model was chosen to describe the persistence in the data. Evaluation of this model showed that the first-order autoregression did describe the persistence. Note that the autoregression describes, but does not explain, the persistence; "what," but not "why," is determined.

The values obtained using the autoregressive model were subtracted from the residual data yielding a second, residual data set. Examination revealed nonrandomness and apparent quasi-periodic fluctuations in this second set. These fluctuations are sometimes referred to as singularities, i.e., certain times of the year when temperatures are considered anomalous. (An example is the January thaw.) A spectral-analysis model was chosen to describe the nonrandomness; evaluation showed the model was inappropriate. The search for a reasonable description of the second, residual data set is continuing.

The observed temperature data have been partially described by the sum of two components, the annual cycle and the autoregression. A "complete" description requires additional iterations of the exploratory data analysis approach. The advantage of the approach is the isolation of independent components upon which statistical inferences can be made and upon which research efforts to develop physical models can be focused.

A complete statistical model of an observational record is composed of all the components used to describe the systematic portion of the record, as well as the statistical description of the residual portion of the record. If based on sound statistical principles, the model represents, but does not explain, the underlying statistical processes that the data measure. Within a probabilistic framework, the model gives information appropriate for estimating the likelihood of occurrence of future events and for estimating risk.

5. Supplementing normals

A complete exploratory data analysis is a difficult, time consuming, expensive process. Results from a partial analysis will, however, provide the user of climatic data with more information than is currently readily available. Current routinely published, summarized information includes averages, extremes, normals, and departures from normals. Limited histories of station location and instrumentation are also available.

The examination step of an exploratory data anal-

ysis yields descriptions of climatic data that supplement normals. It identifies distributional characteristics of climatic data such as symmetry, independence of observations, and variability. Publication of descriptors such as standard deviations for symmetric distributions, medians and quantiles for asymmetric distributions, and correlation coefficients for dependent data would provide users with more information than is conveyed by publishing only one measure of central tendency, i.e., an average.

Station histories could be improved from the knowledge gained from the examination step. Detailed historical documentation of non-climatic effects such as computation procedures, instrument response, site exposure, changes in recording times of observations, and data-correction procedures would help an analyst assess the homogeneity of a climatological data series.

Publication of descriptive statistics, observing-system characteristics, and processing procedures will not solve the user's problem of predicting future events from the climatological record. It will, however, lead to a better understanding of the nature of climatic parameters. This understanding becomes the basis for developing a set of characteristics from which predictions can be made.

6. Summary

Normals have become firmly entrenched as a descriptor of climate. They were developed for comparative purposes. For predictive purposes, a more complete description of climate is required than is provided by normals. An exploratory data analysis approach is suggested to provide the climatic descriptions necessary for developing predictive relationships. The approach is based on the modeling of systematic phenomena and of probabilistic estimates of random, patternless phenomena. It necessitates validation and clean-up of data, utilizes the information available in the observational record, and results in descriptions for specific user applications.

References

- Andrews, D. F. 1978. Data analysis, exploratory. *International Encyclopedia of Statistics*, ed. W. H. Kruskal and J. M. Tanner. New York: Free Press.
- Aronin, J. E. 1953. *Climate and Architecture*. New York: Reinhold.
- ASHRAE. 1985. *ASHRAE Handbook, 1985 Fundamentals*. Atlanta: Amer. Soc. Heating, Refrigerating and Air-Conditioning Engineers.
- Brooks, C. E. P., and N. Carruthers. 1953. *Handbook of Statistical Methods in Meteorology*. London: Her Majesty's Stationery Office.
- Conrad, V., and L. W. Pollack. 1950. *Methods in Climatology*. Cambridge: Harvard Univ. Press.
- Court, A. 1967. Climatic normals as predictors, Part 1: Background. AFCRL 67-0313. Air Force Cambridge Research Laboratories, Bedford, MA.
- 1968a. Climatic normals as predictors, Part 2: Extension. AFCRL 68-0673. Air Force Cambridge Research Laboratories, Bedford, MA.
- 1968b. Climatic normals as predictors, Part 4: Verification. AFCRL 69-0002. Air Force Cambridge Research Laboratories, Bedford, MA.
- 1968c. Climatic normals as predictors, Part 5: Conclusion. AFCRL 69-0003. Air Force Cambridge Research Laboratories, Bedford, MA.
- 1968d. Climatic normals are inefficient. AMS Conference and Workshop on Applied Climatology, October 29–31, 1968, Asheville, NC.
- Craddock, J. M. 1981. Monitoring a changing climate series. *J. Clim.* **1**: 333–343.
- DeVenne, P. B., ed. 1982. *The American Heritage Dictionary, 2nd Coll. Ed.* Boston: Houghton Mifflin.
- Ecodyne. 1980. *Weather Data Handbook*. New York: McGraw-Hill.
- Enger, I. 1959. Optimum length of record for climatological estimates of temperature. *J. Geophys. Res.* **64**: 779–787.
- Filippov, V. V. 1968. Quality control procedures for meteorological data. World Weather Watch Planning Report No. 26. World Meteorological Organization, Geneva.
- Flueck, J. A., W. L. Woodley, A. G. Barston, and T. J. Brown. 1986. A further assessment of treatment effects in the Florida Area Cumulus Experiment through guided linear modeling. *J. Clim. Appl. Meteor.* **25**: 546–564.
- Guttman, N. B., and M. S. Plantico. 1987. Climatic temperature normals. *J. Clim. Appl. Meteor.* **26**: 1428–1435.
- Hann, J. 1903. *Handbook of Climatology, Part I*. R. DeC. Ward, trans. London: Macmillan.
- Hoel, P. G. 1962. *Introduction to Mathematical Statistics*. New York: Wiley and Sons.
- Jagannathan, P., R. Arley, H. ten Kate, and M. V. Savarina. 1967. A note on climatological normals. Tech. Note No. 84, WMO-No. 208.TP.108. World Meteorological Society, Geneva, Switzerland.
- Lamb, P. J., S. T. Sonka, and S. A. Changnon. 1985. Use of climate information by U.S. agribusiness. NOAA Tech. Rpt. (NCP0-001). U.S. Department of Commerce, NOAA, Washington, D.C.
- Landsberg, H. E. 1955. Weather "normals" and normal weather. *Weekly Wea. Crop Bull.* **42**: 7–8.
- Landsberg, H. E. 1958. *Physical Climatology (2nd ed.)*. DuBois, PA: Gray Printing.
- 1972. Weather "normals" and normal weather. *Environmental Data Service*. October 1972: 8–13. National Oceanic and Atmospheric Administration.
- 1975. The definition and determination of climatic changes, fluctuations and outlooks. *2nd Carolina Geographical Symposium, Atmospheric Quality and Climatic Change*. Ed. R. J. Kopeck. Chapel Hill: U. North Carolina.
- 1984. Overview: Climate and energy interactions. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.
- Laumann, J. A. 1984. Summary report of the panel on climate and the gas industry. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.
- Leavitt, P. R. 1984. Improving the utility of climatic data to the oil industry. Proc. of the DOE/Industry Workshop on the Interaction

- of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.
- Lehman, R. L. 1987. Probability distributions of monthly degree day variables at U.S. Stations, Part I: Estimating the mean value and variance from temperature data. *J. Clim. Appl. Meteor.* **26**: 329–340.
- Lunde, P. 1980. *Solar Thermal Engineering*. New York: Wiley and Sons.
- Marsh, P. 1977. *Air and Rain Penetration of Buildings*. Lancaster, England: Construction Press, Ltd.
- McKay, G. A. 1975. Future climate and decision-making. *2nd Carolina Geographical Symposium, Atmospheric Quality and Climate Change*. Ed. R. J. Kopec. Chapel Hill: U. of North Carolina.
- McKee, T. B., and N. J. Doesken. 1984. An assessment of climate information to serve the energy industry in Colorado. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.
- McQuigg, J. D. 1975. *Economic Impacts of Weather Variability*. Department of Atmospheric Science, University of Missouri-Columbia.
- Mood, A. M., and F. A. Graybill. 1963. *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- Olgay, V. 1963. *Design with Climate*. Princeton: Princeton University Press.
- Rotty, R. M. 1984. Summary report of the panel on climate and the electric utilities. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.
- Russell, J. S. 1984. Mathematical representation of related period variables using Fourier series and projections of vector functions: Application to the biological sciences. *Agric. For. Meteor.* **33**: 215–224.
- Saxton, D. W. 1979. Climate normals in the Greenland Sea area. *J. Appl. Meteor.* **18**: 1222–1223.
- Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- Simiu, E., and D. W. Lozier. 1975. The buffeting of tall structures by strong winds. NBS Building Ser. No. 74. U.S. Department of Commerce, National Bureau of Standards, Washington, D.C.
- Slusser, W. F. 1968. Climatic normals as predictors, Part 3: Median vs. Mean. AFCRL 68-0255. Air Force Cambridge Research Laboratories, Bedford, MA.
- Smith, K. 1975. *Principles of Applied Climatology*. New York: Wiley.
- Smolander, H., and J. Lappi. 1985. Integration of a nonlinear function in a changing environment: Estimating photosynthesis using mean and variance of radiation. *Agric. For. Meteor.* **34**: 83–91.
- Snijders, T. A. 1986. Interstation correlations and nonstationarity of Burkina Faso rainfall. *J. Clim. Appl. Meteor.* **25**: 524–531.
- Stern, R. D., and R. Coe. 1982. The use of rainfall models in agricultural planning. *Agric. For. Meteor.* **26**: 35–50.
- Thom, H. C. S. 1966. Some methods of climatological analysis. WMO No. 199. TP. 103. World Meteorological Organization, Geneva.
- Todorov, A. V. 1985. Sahel: The changing rainfall regime and the "normals" used for its assessment. *J. Clim. Appl. Meteor.* **24**: 97–107.
- Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Watson, D., and K. Labs. 1983. *Climatic Design*. New York: McGraw-Hill.
- Wilson, J. S. 1984. Summary report of the panel on climate and the coal industry. Proc. of the DOE/Industry Workshop on the Interaction of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.
- World Meteorological Organization. 1984. Technical regulations, Vol. I. WMO-NO. 49. Geneva, Switzerland.
- Zeger, S. L. 1985. Exploring an ozone spatial time series in the frequency domain. *J. Am. Stat. Assn.* **80**: 323–331. ●

**DETECTING CLIMATE VARIATIONS AND CHANGE: NEW CHALLENGES FOR OBSERVING
AND DATA MANAGEMENT SYSTEMS**

**Thomas R. Karl
Robert G. Quayle
Pavel Ya. Groisman*
Global Climate Lab.
NOAA/NESDIS/NCDC
Federal Building
Asheville NC, 28801**

***Present affiliation: State Hydrological Institute,
Department of Climatic Change,
23, Second Line,
St. Petersburg, 199053, USSR**

Abstract

Several essential aspects of weather observing and the management of these data are discussed as related to improving our knowledge of climate variations and change in the surface boundary layer and the resultant consequences on socio-economic and biogeophysical systems. The issues include: (1) long-term *homogeneous* time series of routine weather observations; (2) time and space-scale resolution of data sets derived from the observations; (3) information about observing systems, data collection systems, and data reduction algorithms; and (4) enhancing weather observing systems to serve as *climate observing systems*.

Although much has been learned from existing weather networks and methods of data management, the system is far from perfect. There are several vital areas which have not received adequate attention. Particular improvements are needed in: (1) the interaction between network designers and climatologists; (2) operational analyses which focus on detecting and documenting outliers *and time-dependent biases* within data sets; (3) developing the means to cope with and minimize potential inhomogeneities in weather observing systems; and (4) authoritative documentation of how various aspects of climate have *or have not* changed. In this latter area, close attention must be given to the time and space resolution of the data. In many instances the time and space resolution requirements for understanding why the climate changed are not synonymous with understanding how it has changed or varied. This is particularly true within the surface boundary layer. A standard global daily/monthly climate message should also be introduced to supplement current Global Telecommunication System's (GTS's) CLIMAT data. Overall, a call is made for improvements in routine weather observing, data management, and analysis systems. Routine observations have provided (and will continue to provide) us with most of the information we possess regarding how the climate has changed during the last 100 years where we live, work, and grow our food.

1. Introduction

There are several broad areas of interest associated with data management for climate change: (1) data bases derived from the output of computer models; (2) data bases derived from process studies aimed at improving our understanding of a particular aspect of the climate system, e.g., severe storms, surface evapotranspiration, cloud microphysics, etc.; (3) data bases derived from long-term measurements which can be used to document and help understand ongoing and past climate variations and change; and (4) data bases derived from long-term measurements which can provide information to assess the biogeophysical and socio-economic impacts of climate change. Discussion in this paper focuses on the latter two issues for data collected during the instrumented era.

Long-term climate monitoring is critical to both a better understanding of climate change, and its interaction and impact upon socio-economic and biogeophysical systems. Of particular interest in this paper are several issues associated with end-to-end data management¹ from long-term observing systems within the surface boundary layer or "bioatmosphere." This is where climate variations and change have direct and wide-ranging impacts.

The concept of "data management" is not well defined as it relates to developing data sets that can be used to document climate variations and change. For example, in the huge Earth Observing System of NASA, a Data Information System (EOSDIS) is being developed for a suite of instruments planned for future satellites. As part of this 15-year research program, a number of pathfinder spaced-based data sets have been identified to break new ground in the use of high-performance computing power required to fully develop and exploit those data sets. Table 1 outlines some of the current and planned pathfinder activities in EOSDIS. Programs such as EOSDIS are expected to lead the way in the application of computer technology to the management of data, and much of that

¹The term "end-to-end" is used here to denote aspects of data management which range from the planning and design of observing systems to the delivery of these data from long-term archives to the retrospective data user.

Table 1. Elements of EOSDIS requiring prototype development (National Research Council, 1990). Underscored items are directly relevant to rigorous documentation of climate variation and change.

1	Data visualization and the user interface
2	Browsing capability
3	Data formats and media
4	Accessibility of data and information
5	Cataloging
6	Search and query capabilities
<u>7</u>	<u>Model and data interaction</u>
<u>8</u>	<u>Metadata and data structures</u>
<u>9</u>	<u>Data reduction algorithms</u>
10	Networking

technology should be transferable to the data management for routine measurement systems.

Nonetheless, it will certainly take more than a state-of-the-art data delivery system from a computer-based Data Base Management System to produce adequate data bases to address many of the issues associated with documenting climate variations and change. For this reason, in addition to items 7 through 9 of Table 1, a number of other issues will be discussed which are associated with data requirements for documenting climate change and variations.

In Table 1 it should be understood that metadata is defined as "important information about the data". So for climate change we could presume that this is vital information which will be used to assess and correct any biases and time-dependent inhomogeneities inherent in the data. EOSDIS may be able to provide such information for space-based observations, but at present, no such formal Data and Information System (DIS) is envisioned for the routine weather observing systems upon which we must depend for basic understanding of how climate is changing in the bioatmosphere.

Sound management of the past, present, and future data from weather measurement systems is critical to documenting and detecting climate variations and change (IPCC 1990). EOSDIS will be complementary to the long-term data sets derived from weather observing systems. For these reasons we focus on management of climate data from weather observing systems, particularly with regard to:

- (1) Long-term *homogeneous* data bases,
- (2) Resolution of data sets for various time and space scales,
- (3) Information about the observing systems, data collection systems, and data reduction algorithms, broadly defined here as "metadata," and
- (4) Enhancing weather observing systems to reduce the uncertainties about how *climate* has (or is) changed and varied.

2. Long-Term Homogeneous Data Bases

A primary concern of many climate change scientists focuses on the homogeneity of the data sets and data bases that they use, especially when it relates to quantifying climate change relative to some long-term baseline. Inhomogeneous data sets are the result of changes in the biases associated with the data. Homogeneous data sets provide the means to be confident that any changes and variations identified in the data are climate induced, not artificial. Important climate changes and variations can be significantly smaller than the year-to-year variability of the climate system (Karl and Riebsame 1984). Inhomogeneous data sets can arise from a variety of reasons, but in general they can be classified into four groups:

- (1) More (less) precise or accurate measurements,
- (2) Changes of temporal and/or spatial sampling,
- (3) Changes in data processing, and
- (4) Micro-climatic changes of the local sampling environment.

Changes in the accuracy and precision of observing systems can lead to serious inhomogeneities. Perhaps the best examples of how this can occur pertain to the measurements of precipitation where efforts have been made to increase the efficiency of precipitation catch of the measuring devices. Even without changing the measurement system however, biases can still creep into the climate record when the sampling frequency or spatial averaging algorithms of the measuring and processing systems change. Perhaps the most intractable inhomogeneities occur when local or micro-climatic changes around the sampling environment introduce nonrepresentative (but nonetheless locally correct) changes in the record. Examples of this include the growth of urban heat islands, changes of land-use, and station relocations, all of which tend to mask larger-scale climate variations or change. These latter inhomogeneities are related to the representativeness of a sampling site.

Few, if any, long-term data sets are absolutely homogeneous, but the term has come to be applied in a practical sense. The data set is "practically homogeneous" when the changes in the biases are small relative to the detected signal of climate variation or change. This assessment focuses on ways in which changes in biases can be minimized and identified in the climate record.

2.1 Inhomogeneities Caused By Changes in Precision and Accuracy

Precision and accuracy of the measurements relate to the observing system, not just sensor response. Some of the best examples of large biases introduced into climate records due to changes in accuracy relate to the method of precipitation measurement. It has been well documented that current and past measurement practices associated with precipitation measurement underestimate the true precipitation (Sevruk 1982, 1986, 1989a; Karl et al. 1989; IPCC 1990; Groisman et al. 1991). Over the last 100 years this has led scientists, operational network managers, and system designers to strive for improved methods of measuring precipitation. Of particular importance is the change from unshielded to shielded precipitation gages. Figure 1 depicts the types of changes that have affected the measurement of liquid and solid precipitation for a number of countries in the Northern Hemisphere. Only rough estimates of the effect of the changes are given in Fig. 1, but it is apparent that the magnitudes of these changes are such that they will make the detection of climate change tenuous at many stations within these countries.

A word of caution with respect to the interpretation of Fig. 1 is pertinent. These changes do not necessarily reflect the potential bias throughout the countries listed. For example, in the USA there are many secondary stations which have not switched to shielded gages or moved to airport locations. Nonetheless, it points out some problems associated with some of our longest professionally-staffed hourly weather observing stations when used to detect climate variations and changes. Many of these primary stations are routinely transmitted internationally and published in

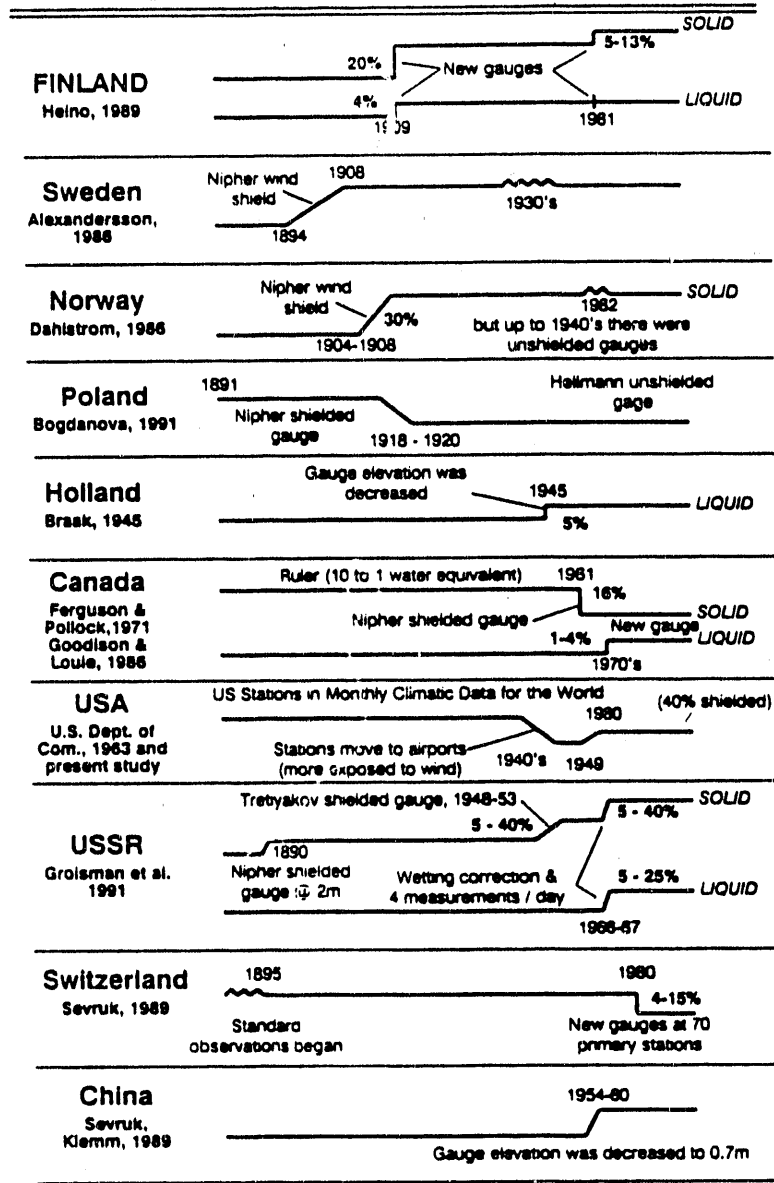


Figure 1

Some important precipitation discontinuities over the last 100 years at many primary observing stations within various mid and high-latitude northern hemisphere countries.

World Weather Records or Monthly Climatic Data for the World, two of the basic sources of long-term data.

The inhomogeneities depicted in Fig. 1 have already occurred, and there will be a continuation of such inhomogeneities in the future, as we strive for better methods of measuring in-situ precipitation. As many countries begin to automate precipitation measurements, the undercatch of precipitation may again be increasing (Routledge 1991; Carr 1991; Sevruk 1989b). To minimize these inhomogeneities when changes in operational networks are proposed and implemented, old systems should be concurrently maintained and operated with new systems. When this is not possible, simultaneous observations (located at both old and new locations) from identical well-calibrated portable automated stations with known precipitation collection efficiencies may provide a partial solution. This requires operation of three instruments: two portable automated stations and an operational permanent station. These observations would be required to continue until appropriate transfer functions can be developed to account for the differences in micro- or local-climatic variations between the two sites, including exposure changes. These observations would not be able to ascertain any inhomogeneities introduced into the record produced by changes in the type of instrumentation.

Another example related to instrument accuracy involves the changes in the measurement of sunshine in the United States. Three major changes in instrumentation occurred during the Twentieth Century, one at the turn of the Century when the "photographic" sunshine recorder (similar in concept to the Campbell-Stokes recorder used in many countries of the world) was replaced with the Maring-Marvin thermometric sunshine recorder. The thermometric sunshine recorder operated on the principle of differential heating between blackened and clear thermometer bulbs exposed to sunlight. As a result, the instrument responded to both direct and diffuse sky radiation, and had a significant lag in its response. Beginning in the 1950s and continuing until the mid-1960s these instruments were replaced with the Foster Photoelectric Sunshine Switch (Karl and Steurer 1990). This switch was

insensitive to diffuse sky radiation and had a faster response than the Maring-Marvin thermometric sunshine recorder (Foster and Foskett 1953). So, although the new device would seem to have improved the accuracy of the sunshine measurement by responding only to direct sunshine with fast response, it created a number of problems associated with the interpretation of changes in sunshine. Analysis of the Twentieth Century changes of sunshine have been made very difficult, and it has been the topic of many papers and countless hours of research (Karl and Steurer 1990; Cerveny and Balling 1990; Hameed and Pittalwala 1989; Quinlan 1985; Angell et al. 1984; Powell 1983; Doehring and Karl 1981, etc.). In several of these papers researchers have lamented over the absence of side-by-side observations at a number of the sites in order to establish the difference in sensitivity of the instruments. Some authors have even called for the reintroduction of the Maring-Marvin thermometric sunshine recorder at a sufficient number of sites to test the differences in sensitivities of the instruments (Cerveny and Balling 1990; Hameed and Pittalwala 1989). Clearly, the value of previous measurements would be greatly enhanced if we were to ascertain the effects of the changes in instruments. Ironically, the proper interpretation of the cause(s) of secular temperature change in the United States during the Twentieth Century may be directly related to changes of cloud cover (Plantico et al. 1990). So the sunshine measurements take on added importance.

One of the ways in which the sunshine climate record can be inspected for homogeneity is comparison with daytime cloudiness. Karl and Steurer (1990) have performed such a comparison, but the cloud observations have their own inhomogeneities associated with changes in observing procedures until 1948. Since 1948 cloud observing practices have not changed, but comparisons with the sunshine record during the 1950s suggest a changing relationship between sunshine and daytime cloud cover which is probably due to the change to the Foster Photoelectric Sunshine Switch. Nonetheless, since the 1960s the relationship between cloud cover and sunshine has remained relatively constant. This is a good example of the need for some redundancy in an operational

observing system whereby one set of measurements can be used to help verify the robustness of any trends or changes detected in other measurements.

Another very important climatological element which has been subject to biases associated with poor calibration and instrument drift is the measurement of solar radiation at the earth's surface. In the United States these data begin during the 1950s, but calibration and maintenance problems have degraded the data during much of the period of record. Jenne and McKee (1985) suggest that the calibration in the national solar radiation network was inadequate to maintain data suitable for climatological purposes (let alone climate change studies). In 1976 a DOE-sponsored program was initiated to correct solar radiation instrument bias in the archived record. Although this has helped the data base for many climatological applications (NOAA 1979), the data are not recommended for trend analysis. DeLuisi (1991) indicates that calibration procedures since the mid 1970s have been quite rigorous, and the data could be useful for documentation of climate change if the instruments were better maintained. Today the WMO is actively coordinating a Global Baseline Radiation Network which includes a significant quality control effort.

An example of the impact of changes in precision that can lead to biases in data sets is associated with the changeover from liquid-in-glass thermometers used in cotton region shelters at US National Weather Service Stations to an electronic Maximum-Minimum Temperature System (MMTS). The digital thermometer provides output on a display to the nearest tenth of a degree Fahrenheit. The observer is supposed to mentally round to the appropriate whole degree Fahrenheit. In fact, about 10% of the observers reported data to tenths of degrees. It took a few years before the problem was recognized and software changes were implemented to prevent the truncation of the tenths digit (Goodge 1991). Even after such changes in the software, it is still possible that a significant fraction of the observers do not routinely round their readings in an appropriate manner (based on the number of observers who seem to have trouble with decimal points in reporting

precipitation, Goodge 1991). In addition to the change from a liquid-in-glass thermometer to an electronic thermometer, the instrument shelter used to house the thermometer changed from a wooden box-like shelter to a much smaller cylindrical plastic type of shelter. Based on thousands of comparisons of monthly mean temperatures from stations with and without an MMTS, Quayle et al. (1991) find that the new system produces maximum temperatures about 0.3°C lower and minimum temperatures about 0.4°C higher than the old system. Unfortunately, because large sample of side-by-side overlapping measurements are not available, site specific corrections cannot yet be derived, and only large-scale temperature changes can be corrected. Furthermore, daily biases, which are likely to be dependent on synoptic conditions, are unlikely to be the same from day-to-day. The biases found by Quayle et al. (1991) are of the same magnitude as the changes of global and USA mean temperatures since the turn of the Century. The magnitude of the changes of the global mean temperatures have been one of the cornerstones of the greenhouse warming issue.

An example of suspected inhomogeneities in the thermometric time series of northern hemisphere seasonal temperatures is provided in Fig. 2. The Nineteenth Century warmth of the summer temperatures is likely the result of inadequate shading of the thermometer from direct and/or indirect radiation. Parker (1991) discusses this in great detail, but the exact corrections required are unknown. If the record is taken literally, there is an absence of northern hemisphere summertime warming since the Nineteenth Century.

Changes of instrument shelters have also affected accuracy and precision of temperature records at a number of tropical locations in ways which vary by region and synoptic regime. Table 2 indicates large positive differences of temperature at Sri Lanka between a thermometer suspended inside a felt shed and one located inside a double-louvered Stevenson Screen, but a smaller and opposite difference between a similar thatched shed and a Stevenson Screen at Hong Kong. This

NORTHERN HEMISPHERE LAND Seasonal Temperature Anomalies

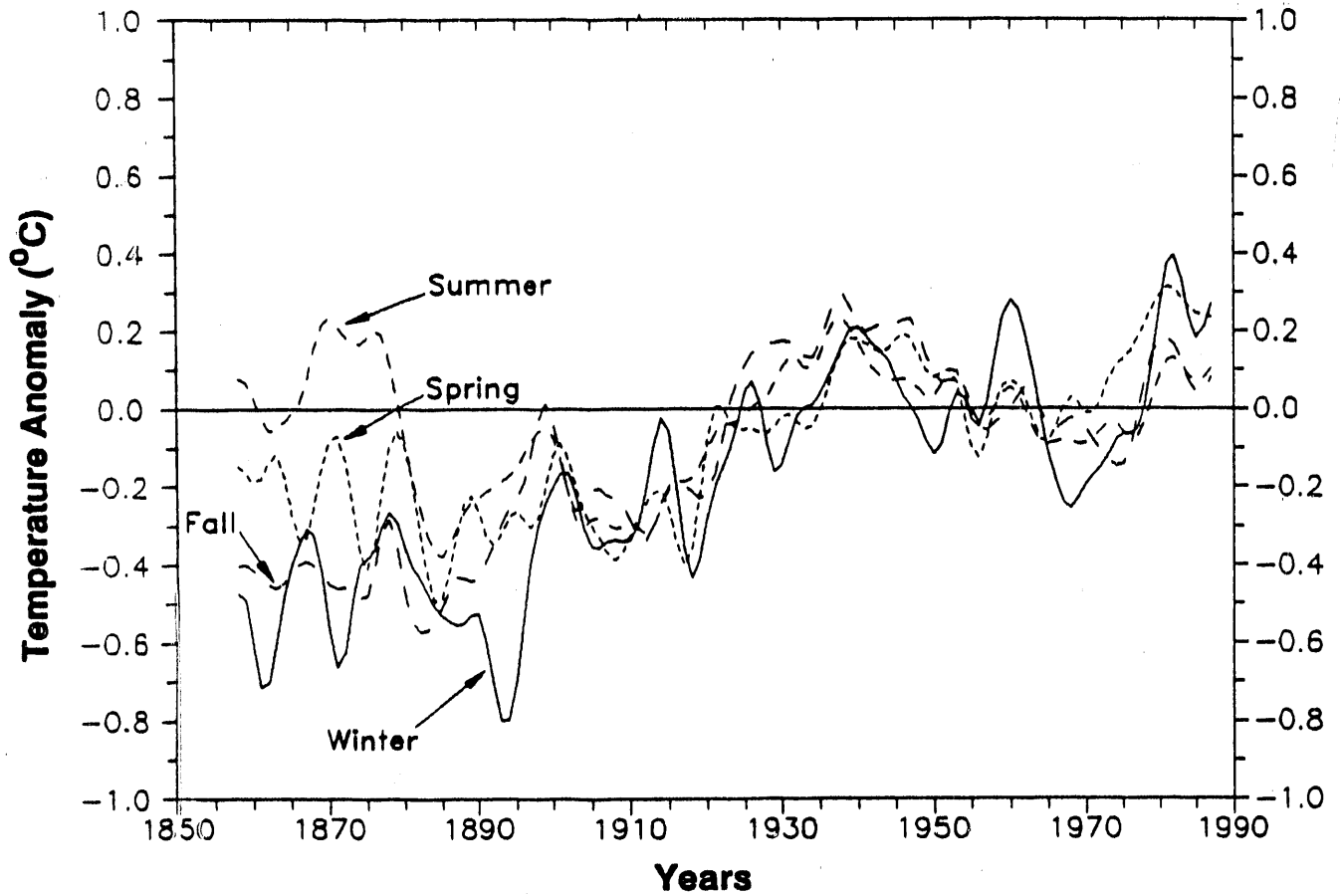


Figure 2 Time series of temperature anomaly (from the 1951-80 mean) for Northern Hemisphere land areas based on the data of Jones et al. (1986) as provided in IPCC (1990).

Table 2. Temperature differences ($^{\circ}\text{C}$) at Sri Lanka and Hong Kong of a liquid-in-glass thermometer suspended beneath a felted and thatched shed respectively compared with a thermometer in a nearby Stevenson Screen. An asterisk(*) implies means derived from days with more than 1 hour of sunshine, and a pound(#) implies means derived from days with less than 1 hour of sunshine. Data derived from Bamford (1928) and Chen (1979).

	Sri Lanka	Hong Kong	
Average Maximum	-0.6	0.3*	0.1#
Average Minimum	-0.5	0.0*	0.0#

emphasizes the extreme difficulty in adapting general correction factors without side-by-side comparisons. It also hints at the desirability of stratifying corrections by synoptic regimes.

Frequently, changes in instrumentation are also associated with small, but sometimes significant stations relocations. When new instruments are introduced, scientific debates are often manifested as to the accuracy of new observing systems, but this does not necessarily address the issue of climate variations and change, i.e., identifying bias between old and new systems before the old system is abandoned. For example, in the early 1950s a new precipitation gage was installed in the USSR. It was known that it would yield significantly more accurate measurements of solid precipitation, resulting in a greater total precipitation catch. It was not until a decade later however, when Shver (1965) quantitatively established that with the information at hand, the bias between old and new systems over the country could not be properly evaluated for Soviet Arctic stations. As a result, the Century-long precipitation measurements for this part of the globe cannot be reliably used in climate change studies.

2.2 Inhomogeneities Associated With Changes of Temporal and Spatial Sampling

One of the most insidious problems associated with the measurement of temperature is the derivation of a consistent technique for calculating mean daily temperatures. The problem is not related to the exact method chosen, but rather changes in the method which evolve over the course of time. Virtually every country in the world has a set procedure, but these procedures differ for the type of station and they often change with time. Such changes cause serious problems of interpretation of the thermometric record. Figure 3 shows how changes in the time of observation have impacted the United States area-averaged temperatures using the model developed by Karl et al. (1986a) and the history of observation time in the United States. Many observers who previously recorded the 24-hour daily maximum and minimum temperature around sunset, now go out shortly after dawn. On a station-by-station basis the model of Karl et al. (1986a) corrects only for monthly

NATIONAL ANNUAL 1931 - 1985

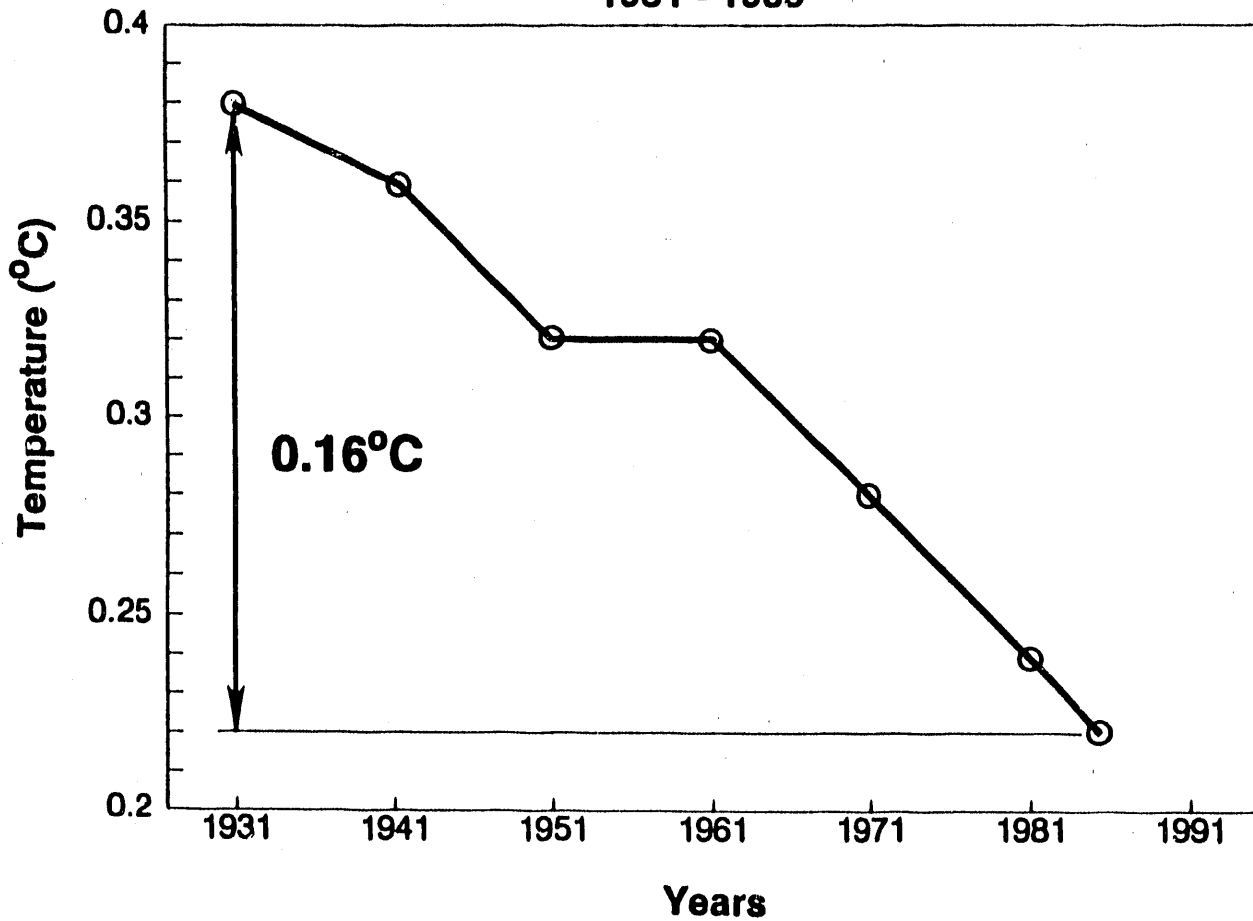


Figure 3 Estimated effects of changes in observation time on the contiguous United States area-average temperature based on all stations which measure temperature relative to a calendar day observing schedule.

biases (which can be as large as 2°C). Unfortunately, on a daily basis the problem is more complex, and as yet unsolved. For example, observers who report the minimum temperature ending at 0700 Local Standard Time (LST) can have twice as many days with temperatures below freezing under certain climate regimes than if they were to observe the 24-hour minimum at 1700 LST. Changes in the frequency of various threshold temperatures can be a very important aspect of climate impacts and adjustments. It is easy to see the dilemma a scientist quickly confronts when faced with using data from such heterogeneous observation systems.

2.3 Inhomogeneities Due to Changes in Data Processing

A recent example of inhomogeneities introduced into the climate record due to changes in the data processing algorithm for northern hemisphere snow cover is well documented by Robinson et al (1990). It is relevant here because it exemplifies the importance of information about data reduction algorithms. Snow cover charts are digitized by trained meteorologists on a northern hemisphere grid from visible satellite imagery. Figure 4a depicts two time series of the annual average snow cover based on a homogeneous and an inhomogeneous data reduction algorithm. Data derived from the latter data reduction procedures were used in the IPCC (1990) scientific assessment of climate change and other reports (Robinson and Dewey 1990). The inhomogeneity arose primarily because prior to 1981 continental areas were calculated from monthly summary charts which consider a grid cell to be snow covered if two or more weeks in the month were snow covered (Dewey and Heim 1982). Since 1981, monthly snow cover was calculated by averaging snow/no-snow information from the weekly charts. Robinson et al. (1990) found that the latter approach tended to produce reduced snow cover in all months except August. As a result, a significant bias in snow cover results as depicted in Fig. 4b. As more and more data become dependent on processing algorithms, problems such as this are likely to grow rapidly unless special care is taken to avoid or at least document changes in processing algorithms. For example, traditional direct measurements of precipitation derived from stick

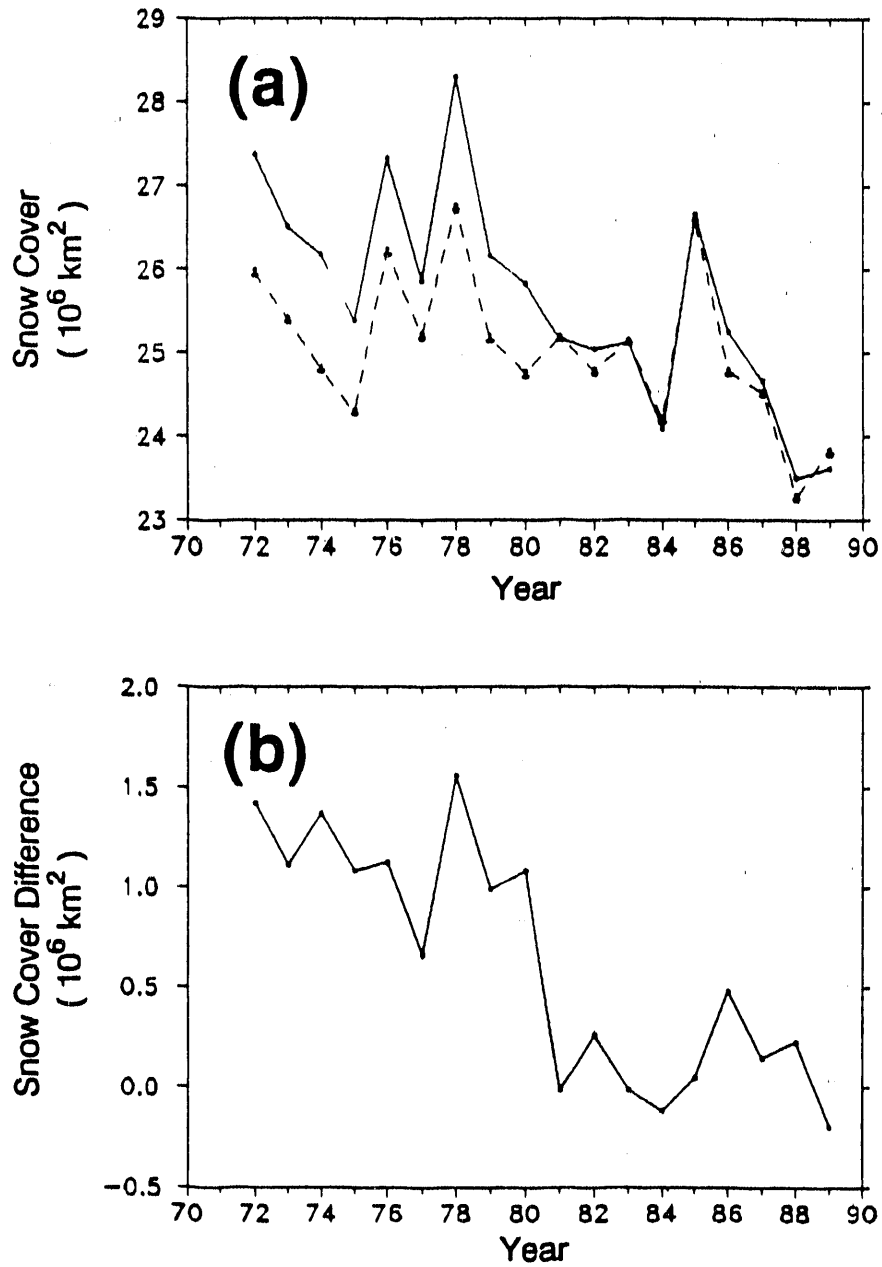


Figure 4

a) Northern hemisphere snow cover from NOAA polar orbiters using a consistent data reduction algorithm (dashed line, Robinson, et al. 1990) versus the resulting data from earlier algorithms (solid line).

b) Difference between the two time series in a).

measurements or weighing are being replaced by tipping devices with built-in conversion processing software. Similarly, liquid-in-glass thermometers are being replaced by electronic systems. The future may include optical precipitation measurement devices. Automation will require indirect sensing of all our climate variables. Special procedures are required to archive these measurements in their proper and most basic units so that changes in either external software or internal built-in microprocessing chips will evaluate homogeneous reprocessing of the data when the inevitable improvements in the system occur.

2.4 Micro-climatic Changes of the Local Sampling Environment

Changes in the micro-climatic environment around a sampling site can occur abruptly, as in the case of station relocations or rapid changes in land use, but it can also occur more gradually as sometimes is the case with the transition from agricultural to urban environments. Many papers have focused on the problems of interpreting climate change and variations when stations change their location and are affected by increased urbanization (Jones et al. 1986; Karl and Williams 1987; Karl et al. 1988; Karl and Jones 1990; Jones et al. 1990). Sometimes there is little that can be done to prevent stations relocations and changes in land use around stations, but these studies indicate that if there is a sufficient number of nearby stations, adequate information about the local land use, or especially if there is an overlap period of simultaneous observations from a nearby homogeneous site, inhomogeneities in the climate record can be rectified. In the transformation of weather observing networks to climate observing networks these goals should be given a high priority.

It is important to realize that small details and seemingly innocuous changes associated with the micro-climate around an instrument can often introduce large discontinuities in the climate record. One such example is provided in Fig. 5 where the orifice of the precipitation gage was elevated from 1 to 2 m in 1963. During winter most of the precipitation that is received at International Falls is frozen, and the wind speed is often relatively high causing considerable blowing and drifting of snow.

International Falls / Minnesota, Division 2

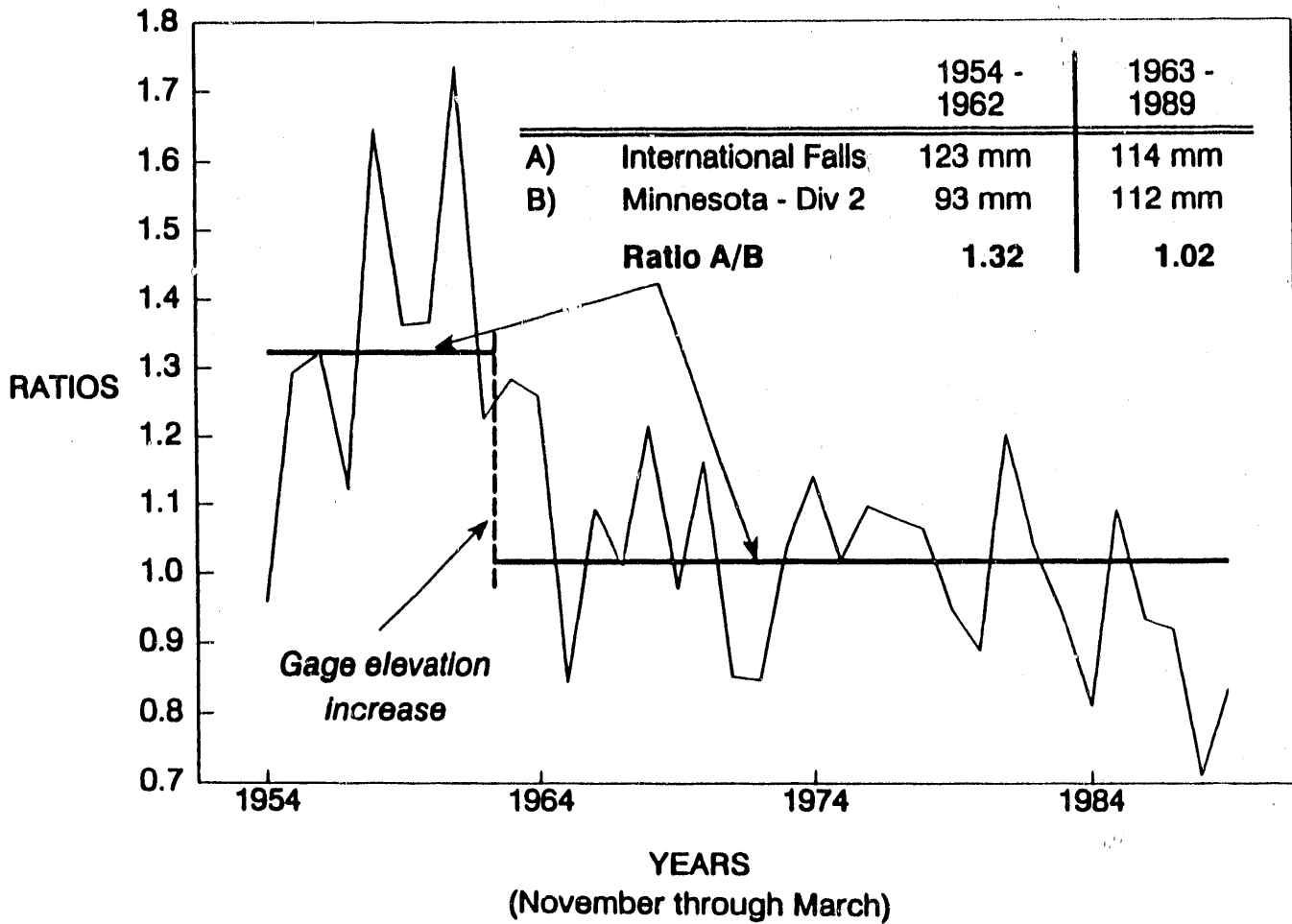


Figure 5 Ratio of precipitation at International Falls, Minnesota to the precipitation in the North Central Climate Division of Minnesota during the snow season (November through March).

Prior to 1963 the gage was located 1m above the ground (the recommended height), about 10m from a snow fence. Observers sometimes had to dig down through the drifting snow to read the gage chart (Kline 1991). The bias was much reduced in 1963 when the gage was raised in height (Fig. 5). The decrease in catch at International Falls was certainly due to improved exposure, but it may also be partly related to a loss caused by increased wind (from increased elevation) at the orifice. The precise reasons may never be known.

3. Resolution of Data Sets for Various Time and Space Scales

Providing information about climate change and variability at the appropriate time and space scales is often an important aspect of climate change and variations. In the IPCC (1990) scientific assessment of climate change, the establishment of a comprehensive system for climate monitoring was singled out as one of the critical areas that require immediate attention. Observations are often taken at climatically-appropriate time and space scales by existing weather observing networks, but as pointed out in the IPCC report, "For most climate variables, the spatial and temporal resolution of the exchanged data (between countries) is inadequate, precluding world-wide analyses of extremes." A major reason why such data are not readily available is associated with the manner in which the data are processed and packaged for distribution. Many important aspects of climate change and its associated human dimensions cannot be answered by consideration of averages of statistical distributions alone.

Examples of the changes in the United States, Canada, the USSR, and the People's Republic of China which have occurred and could not have been anticipated by the use of means alone have been well documented for the mean monthly maximum and minimum temperature (Karl et al. 1984, 1986b, 1991). The point is further illustrated by inspection of the time series in Fig. 6. These time series are a measure of the changes in the anomalies of monthly extreme high and low temperatures averaged throughout the year across the USA and the USSR. The data indicate that monthly extreme

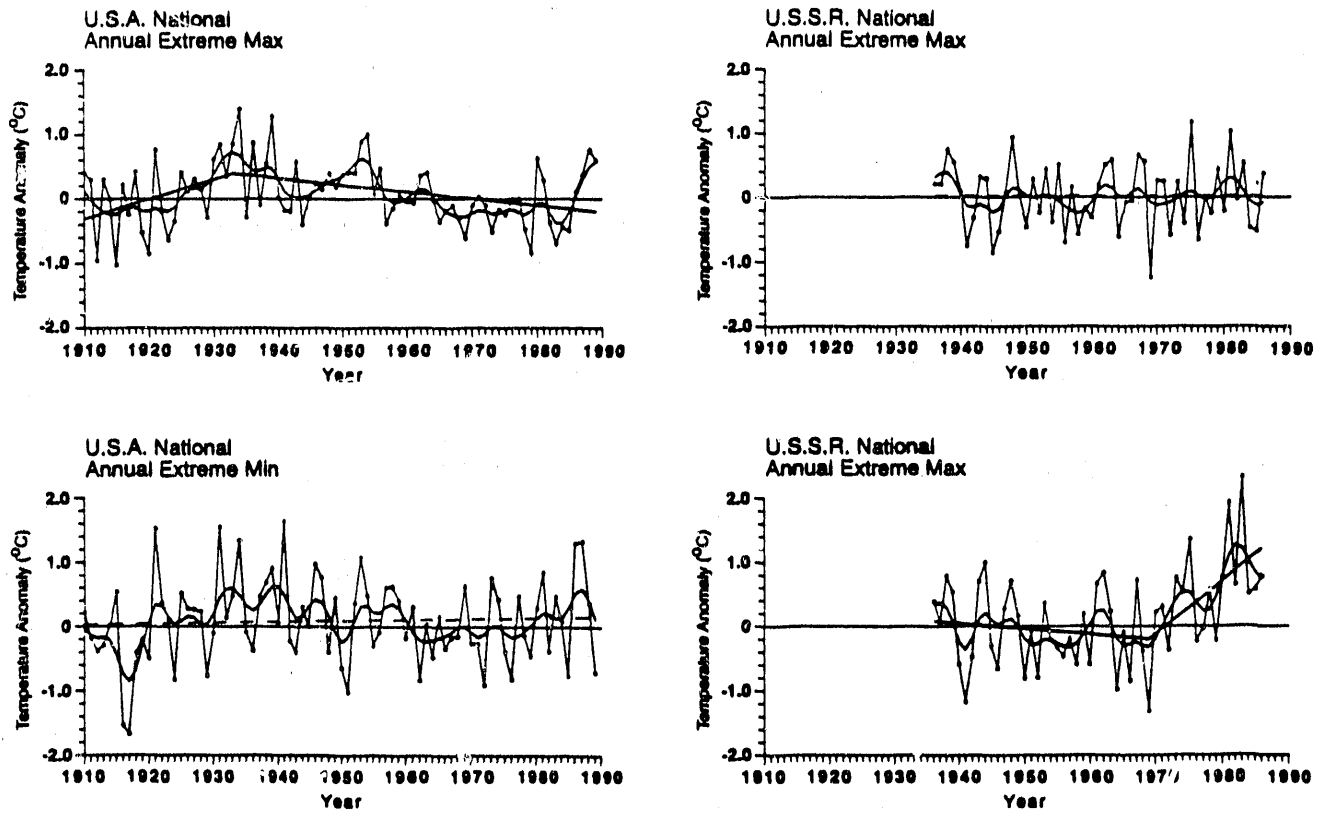


Figure 6 Variations and trends of the annual mean of the monthly extreme maximum and minimum temperature anomalies for the USA and the USSR. Smooth curves are derived from a nine-point binomial filter and two-phase or single-phase linear trends are also depicted (Solow 1987).

high temperatures have tended to decrease over the past several decades and extreme low temperatures have tended to increase. The time series are derived from over 190 continuously reporting rural stations across the Soviet Union and 370 stations in the contiguous United States. The time series consist of the average of the one-day extreme temperature anomalies from the 1951-80 monthly mean maximum and minimum temperature. These anomalies are then area-averaged (23 regions) over the USSR and USA separately. The time series can be interpreted to imply a reduction in climate variability, as the range of the extremes has become narrower. (The MMTS bias described earlier is not operative here because of the period of record and choice of stations.) Many of our socio-economic and biophysical systems are most sensitive to changes in extremes as opposed to changes in the mean. Observing systems must be able to readily provide reliable information of this nature. As already indicated however, there has been very little work to identify inhomogeneities in the climate record for data with high temporal resolution, such as daily data.

Presuming we can enhance existing weather observing networks and data management systems, they are likely to be our only reliable source of information about changes in extremes of climate for years to come. Given the important of extreme climate events, a high priority must be placed on developing data bases with the appropriate time resolution. In addition, research should be encouraged and supported which addressed the special problems associated with changes in biases on daily time scales.

An often-overlooked aspect of documenting climate change and variations concerns spatial averaging. In long-term data sets many of the inhomogeneities of the data tend to cancel as many stations are averaged together in large-scale spatial averaging processes. In some instances, even effects such as the urban heat island bias (Karl and Jones, 1990) can be offset to a significant degree. On regional space-scales (hundreds to thousands of square kilometers), these biases do not readily cancel, and they can smear real and artificial climate variations and change. It is precisely these

space-scales, however, which are often very important for understanding the impacts of climate change or variations. Furthermore, as Barnett and Schlesinger (1987) argue, it is the patterns of climate change that are important in unequivocally identifying cause(s) of observed climate change and variation. For these reasons, in a long-term climate observing system, where changes are inevitable, e.g. stations close, new instruments are installed, etc., it is important to avoid over-reliance on a small number of stations, no matter how well maintained. In fact, it is the abundance of in-situ observations compared to the spatial degrees of freedom associated with specific climate variables that enables assessments of inhomogeneities in the climate record.

We have been fortunate up to now that reasonably useful results have been obtained from our rather haphazard approach to monitoring climate change and variations, but it would be unwise to imply that we can continue to take such a passive attitude about climate monitoring in the future. Depending upon the hope that "errors will cancel" is not an appropriate strategy upon which to build science or policy. Today, it is very difficult to reconstruct reliable time series of regional and local climate variations, even from our most reliable instrumentation, in-situ thermometers.

Two examples are provided to demonstrate the seriousness of the temperature problem. First, the annual average differences between two different data sources were calculated for area-average temperature over the USSR. One source is a 223-station network compiled by the All-Union Research Institute of Hydrometeorological Information in Obninsk, USSR (World Data Center-B for Meteorology) and the other is a set of data compiled by Jones et al. (1986). The Jones et al. (1986) data set is arguably the most authoritative and comprehensive long-term global surface land thermometric data set yet produced. Figure 7a shows that there are some year-to-year differences of annual temperature anomalies between the two data sets, but the trend of the differences is near zero. A closer evaluation of the differences however, reveals some unsettling characteristics with respect to the seasonal trends. A negative bias in the USSR data set relative to the Jones et al. (1986) data set in

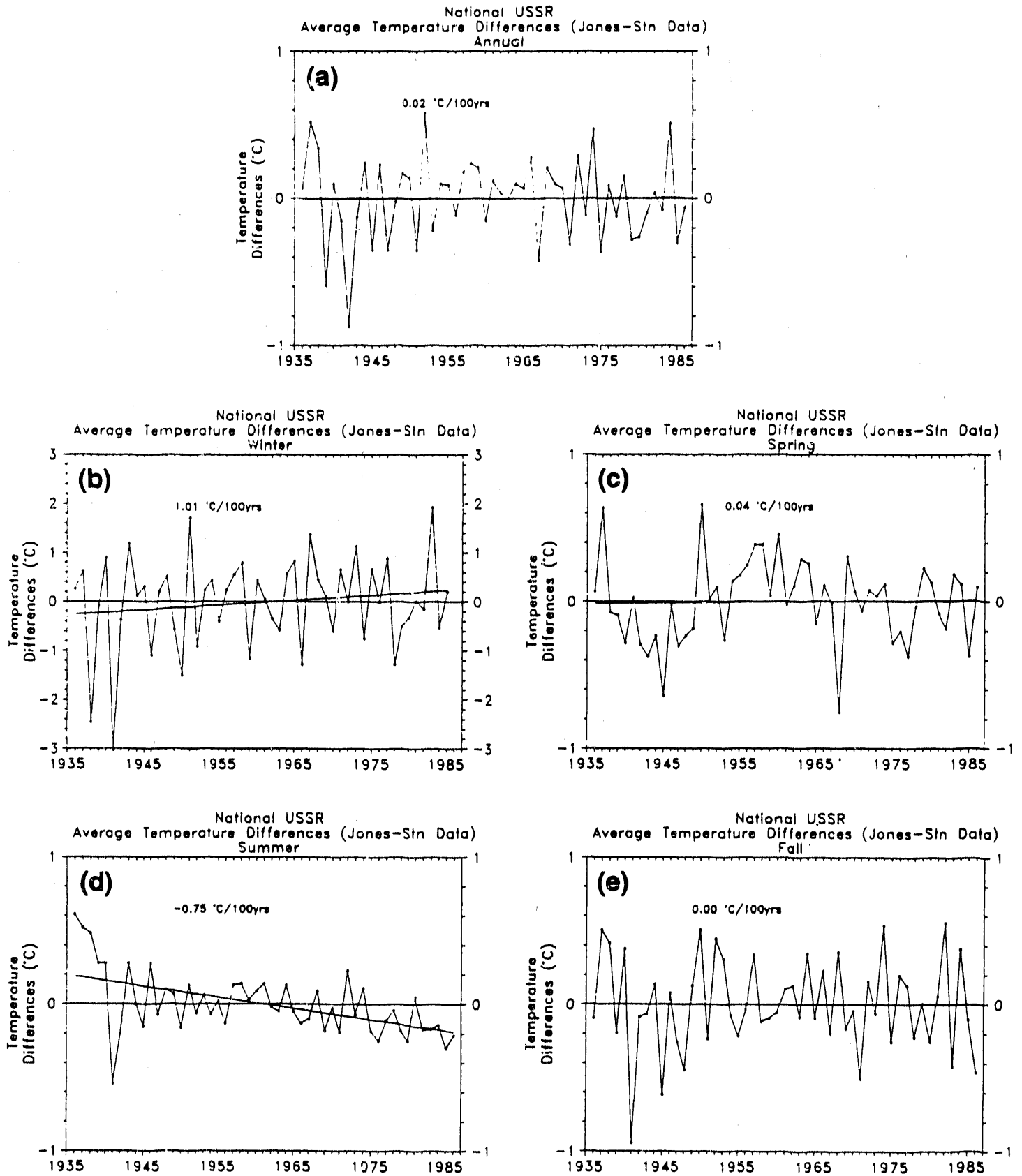


Figure 7 Difference of temperature anomalies between the Jones et al. (1986) data set and a recently-compiled long-term data set from the USSR World Data Center-B for Meteorology (station data), and associated linear trends .

winter (Fig. 7b) is offset by a positive bias during summer (Fig. 7d). So although on an annual basis the two data sets agree fairly well, it is only because of offsetting seasonal biases. Biases as large as 1°C per 100 years are quite significant relative to the detection and impact of greenhouse-induced climate warming.

As another example, Klestchenko et al. (1990) calculate (Fig. 8) the correlation of the temperature anomalies from a hemispheric thermometric data set they have updated from Gruza and Rankova (1979) with those derived from the data of Jones et al. (1986). Many areas of the Northern Hemisphere have poor agreement between the two data sets, or inadequate data to perform a correlation, especially outside of North America, eastern Asia, and Europe. Based on these analyses, and others, it is clear that our global data sets still have trouble resolving regional and local-scale climate variations in some portions of the globe.

4.0 Metadata

There are a number of methods already developed (Kohler 1949, Conrad and Pollak 1950, Mitchell 1961, Maronna and Yohai 1978, Potter 1981, Alexandersson 1986) to detect inhomogeneities in climate records with an absence of any information about the history of the observations (differences, ratios, log ratios, etc., of a given variable between stations). These methods, although necessary and extremely useful, should be considered only as a last resort, or as a final check of the homogeneity of a station's records. The primary reason relates to the additional information provided by knowledge of the timing of changes in the observing system or micro-climate. Much of this information is commonly available in what has come to be known as station histories. Station histories can easily be ignored in maintaining weather observing networks because stations can function without them, they are not routinely disseminated, and they have very few users relative to the data themselves. Over the past several decades, an increasing number of climatologists have begun to assemble station histories because of their interest in climate change.

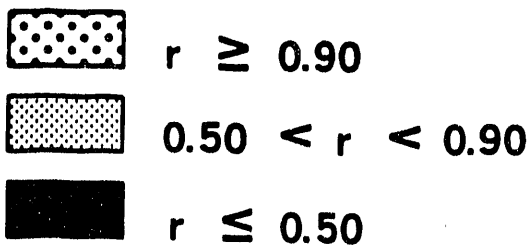
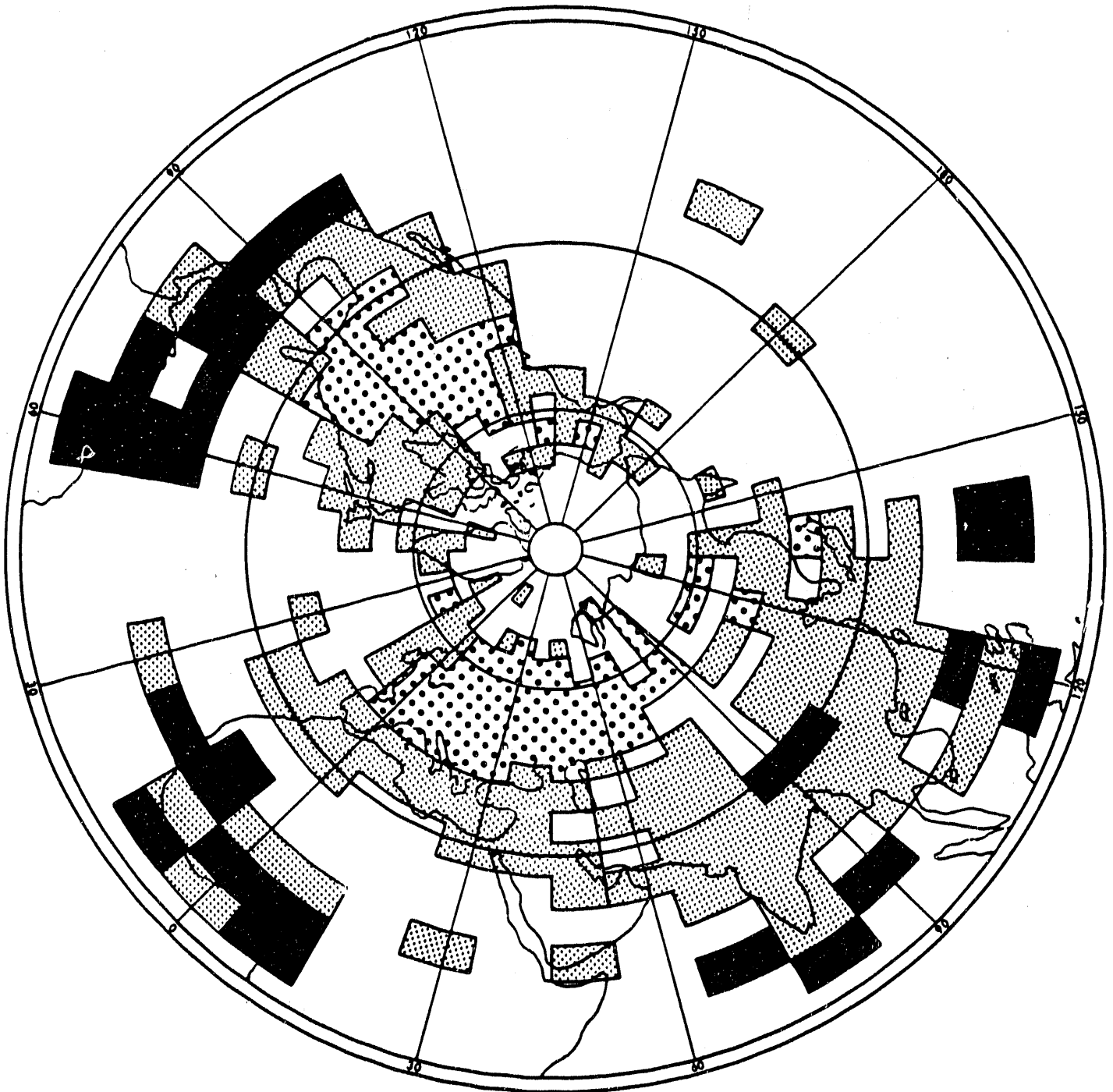


Figure 8

Correlation of the annual temperature anomalies between the Jones et al. (1986) data set with an updated hemispheric land based data set reported by Klestchenko et al. (1990). Unshaded areas represent missing data.

Station histories contain information about station location, the types of instruments used, their elevations above ground and exposure, information about the local surroundings, observing schedules, maintenance procedures, etc. Extant station histories have been very useful. Without such information many of the inhomogeneities already discussed could not have been identified or corrected. Unfortunately, current station histories are far from perfect, and thousands of hours have been spent poring over old records to attempt resolution of incomplete or conflicting information. In the United States a portion of the station history information for surface observations is available on computer-compatible media at the National Climatic Data Center, but a sizable fraction remains only on hardcopy and much is nonexistent. The situation is considerably worse in many other countries of the world. In many instances information about potentially important micro-climatic details is fragmentary, at best. For example, information about an obstacle located near a precipitation gage may be noticed only because some reference is made to it in the station history notes years after it was built or grew.

The importance of this type of information cannot be overemphasized. For example, researchers using the precipitation data from the North American professionally-staffed hourly observing station records published in *Monthly Climatic Data for the World (MCDW)*, one of the primary sources of global climate data, must address the fact that over the past three decades about 40% of all the contiguous US precipitation gages reported in MCDW had shields installed around the orifice of the gage. Prior to 1948, shields were absent. This information is available only in the detailed manuscript station histories, and is neither computerized nor uniform throughout the 100-year existence of a National Weather Service.

Ancillary information about changes in land use could also be very important, especially as related to the effect of urban heat islands. The use of high resolution (1km X 1km) space-based products to document land use and changes of land use may help identify and track man-made changes

(United States Congress, Office of Technology and Assessments, 1991). Given these potential costs alone, it seems prudent to invest a tiny fraction of such an enormous expenditure in strengthening the operation of weather networks and data management practices throughout the globe so they will provide robust information about climate change throughout the rest of this Century and the next. The question is, "What would be needed and how much would it cost?"

It is feasible to outline many of the essential improvements required in operational surface observations and data management practices. The cost of these improvements in the United States should be modest compared to programs such as EOS, but it should be emphasized that such a system must be more than any one nation's commitment. An international organization such as the World Meteorological Organization (WMO) would have to take the leadership. These changes should be a cornerstone of the Global Climate Baseline Data Set Development, the Global Climate Change Detection, and the Global Climate Observing System Projects now in various stages of development within the WMO.

A number of basic principles of observation and data management must be adopted in order to broaden the mission of existing observing networks to include monitoring for climate change and variations. Such a mission is consistent with the call made by the IPCC (1990) to narrow the uncertainties related to documenting and understanding climate change and variation. These principles include:

- 1) Establish standard procedures for collecting overlapping measurements for all significant changes made in instrumentation, observing procedures, or the location of the instruments;
- 2) Make routine assessments of ongoing calibration, maintenance, or homogeneity problems for the purpose of taking corrective actions when necessary;
- 3) Develop standard station histories with routine monthly dissemination of changes (or initial status for new stations);

- 4) Develop standard data packages for important climate variables at various time and space-scales, which include a full discussion of processing procedures and algorithms used to reduce data;
- 5) Raw data sensed from the instruments prior to transformation into standard atmospheric variables or products should be archived along with the processed data and processing algorithms;
- 6) Closer working relationships between network designers and climatologists must exist at the outset of network design, implementation, and changes; and
- 7) Maintenance of some level of redundancy in observing networks is required to add robustness to data analyses.

Ideally, when changes occur in the network, side-by-side simultaneous measurements need to be conducted. The simultaneous measurements could be discontinued when the impact of the change can be quantified. This is not a new concept. The US National Weather Service Operations Manual (Section B-11) recommends overlapping observations for a period of 1 to 3 years. At present, infrastructure is lacking within weather services to support such procedures in the USA, and much of the world. These procedures need to be established and supported at least for a representative subset of the existing global weather observing stations. In the future, portable automated climate stations may be used to help ensure that suitable transfer functions between new and old observing sites can be developed to account for changes in microclimate. Dual observations could be made with identical portable automated stations at both the new and old observing sites, while the more permanent station operates side-by-side with one of the portable automated stations.

In order to avoid uncertainties regarding changes in instrument biases, it would be desirable to establish a working museum of old decommissioned instrument types. In the USSR, one such museum of hydrological instruments exists in the Valdai branch of The State Hydrological Institute. It

stores and tests scores of hydrological instruments (some from the 19th century). The extension of this museum for precipitation gauges has been supported by the International Organizing Committee for the WMO Solid Precipitation Intercomparison Project (WMO 1991). Old gauges from various countries will be put into operation to gain knowledge about the biases of historical precipitation measurements. The creation of such museums could be an important asset toward resolving present and future in-situ measurement problems.

In an operational environment it may not always be possible to allow for simultaneous side-by-side measurements. For this reason it is advisable to operate a dense network of weather stations, scientifically designed so that occasional station losses will not badly degrade climatic analyses. This will allow comparisons among and between stations when simultaneous observations are not possible.

An aggressive program is required for near-real-time homogeneity assessment. This can best be accomplished by using both data and metadata. If the information about changes in biases can be relayed to network managers in a timely manner, then this should significantly increase the quality of the data for early warnings of climate change and variations. In addition, real-time analysis fields (00 hr forecast) of numerical weather prediction models should be used to identify and investigate stations which consistently report values outside of an expected range. Mechanisms for archiving and effectively utilizing this information should be developed and implemented. It should be clearly understood, however, even the best long-term data assimilation and re-analysis projects are not substitutes for long-term homogeneous observations.

Just as the WMO has developed standard practices for observing and reporting synoptic weather information, a similar scheme needs to be developed and implemented for the exchange of climate data (e.g. a uniform summary message for daily data and monthly extremes); and information on each station's history (current observing practices, instrumentation, sensor locations, exposure, the surrounding environment, etc). The advantages of a standard report are many: facilitating the

exchange of information among countries, lowering the probability of neglecting important information, and increasing the probability of that the information is correctly recorded and distributed.

Computer software and hardware for on-line Data Base Management Systems will be an important aspect of the data delivery system in the years to come. Just what that delivery system provides the scientist is dependent on adequate metadata. This is particularly important with respect to documenting the processing procedures associated with data reduction. All algorithms, averaging procedures, quality control, homogeneity checks, and corrections must be well documented and packaged to allow straightforward scientific analysis. Furthermore, those responsible for data reduction algorithms must work toward ready access of important climate characteristics beyond the usually reported mean anomalies integrated over large time and space scales. All this can be a time-consuming task, but its importance to the proper interpretation of the data cannot be overemphasized.

Fewer and fewer observations are direct measurements of basic climatological variables. Pre-processing of electronic measurements is rapidly becoming the standard. Not only must an effort be made to maintain adequate metadata, but data archives should be striving to preserve data in its most basic form. This would enable reprocessing should more appropriate transfer functions be developed (and they will if we use remote sensing as an analogy).

Finally, over the past several decades observational networks have been designed for a variety of purposes, but rarely have they been designed to detect and monitor climate change and variations. This requirement can best be incorporated into network design by inclusion of climate change monitoring requirements in the earliest stages of network design. Climatologists and other scientists cannot afford to take a passive interest in this area, as has so often been the situation in the past. Failure in this area will mean that future generations will still have to contend with many of the data uncertainties associated with the homogeneity issue we are addressing today. An institutional climate

requirements infrastructure is needed to quantitatively, scientifically, and authoritatively deal with climate issues on an equal footing with weather reporting and weather prediction interests. These steps will allow us to adequately resolve many of the important aspects of climate change. Knowledge of the precision and biases of measurement and processing can mean the difference between science and speculation.

Acknowledgements. Support for this work was provided the United States Department of Energy through interagency agreement number DE-AI05-90ER60952 and contract number ACO2-DE-FG02-88ER60372, and the NOAA Climate and Global Change Program. Dr. Groisman was supported by a National Research Council Senior Research Associateship from the National Climatic Data Center. This manuscript benefited from two excellent reviews and discussions with the National Academy of Sciences Climate Research Committee.

REFERENCES

- Alexandersson, H. 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, 6, 661-675.
- Angell, J., Korshover, J., and G.F. Cotton 1984: Variation in United States cloudiness and sunshine, 1950-82. *J. Climate Appl. Meteor.*, 23, 752-761.
- Bamford, A.J. 1928: On the exposure of thermometers in Ceylon. *Ceylon J. Sci.*, Section E, Vol. I, 153-167 plus 3 plates.
- Barnett, T.P. and M.E. Schlesinger. 1987: Detecting changes in global climate induced by greenhouse gases. *J. of Geophys. Res.*, 92, 14722-14780.
- Bogdanova, E.G., 1991: Personal Communication, Main Geophysical Observatory, St. Petersburg, USSR.
- Braak, C. 1945: Invloed van den vind op regnwaarnemingen. Koninklijk Nederlandsch. Meteorol. Inst. No. 102. Mededeelingen en verhandelingen, No. 48 (102), 7-74.
- Carr, D.A. 1991: Ratios of Fischer and Porter precipitation gage data versus standard climatological measurements. Atmospheric Environment Service, Downsview, Ontario, Canada, 20 pp plus figures.
- Cervený, R.S., and R.C. Balling, Jr. 1990: Inhomogeneities in the long-term United States' sunshine record. *J. of Clim.*, 3, 1045-1048.
- Chen, T.Y. 1979: Comparison of air temperatures taken from a thermometer screen, a thatched shed, and a whirling thermometer. Royal Observatory, Hong Kong Technical Note No. 49.
- Conrad, V., and L.W. Pollak. 1950; *Methods in Climatology*. Harvard University Press, 459pp.
- Dahlström, B. 1986: The improvement of point precipitation data on an operational basis. Nordic Hydrological Programme, NHP-Report No. 17, 86 pp.
- DeLuisi, J. 1991: Personal Communication, Chief, Aerosols and Radiation Monitoring Group, Geophysical Monitoring for Climate Change, ERL. NOAA.
- Dewey, K.F., and R.R. Heim, Jr. 1982: A digital archive of Northern Hemisphere snow cover, November 1966 through December 1980, *Bull. A.M.S.*, 63, 1132-1141.
- Doehring, F., and T. Karl. 1981: A history of sunshine data in the United States 1891-1980. NOAA Historical Climate Series, No. 2-2, 40 pp. [NCC, Asheville, NC 28801].
- Ferguson, M.L., and D.M. Pollock. 1971: Estimating snowpack accumulation for runoff prediction. Proc. of Canadian Hydrology Symposium No. 8 "Runoff from snow and ice," Quebec City, 1971, 7-27.

- Foster, N.B., and L.W. Foskett. 1953: A photoelectric sunshine recorder. *Bull. Amer. Meteor. Soc.*, **34**, 212-215.
- Gallo, K. 1991: Personal Communication, Land Sciences Branch, ORA, NESDIS, NOAA.
- Goode, G. 1991: Personal Communication, Section Chief, Data Operations Branch, National Climatic Data Center, NESDIS, NOAA.
- Goodison, B.E., and P. Y.T. Louie. 1986: Canadian methods for precipitation measurement and correction. WMO/TD-No. 104. Instruments and observing methods. Report No. 25. Papers presented at the Workshop on the correction of precipitation measurements, Zürich, Switzerland, 1-3 April 1985, 141-145.
- Groisman, P.Y., V.V. Koknaeva, T.A. Belokrylova, and T.R. Karl. 1991: Overcoming biases of precipitation measurement: A history of the USSR experience. *Bull. Amer. Meteor. Soc.*, **72**, 1725-1733.
- Gruza, G.V., and E. Ya. Rankova. 1979: Data on climate structure and variability: Sea level air temperature of the Northern Hemisphere. Obninsk: All-Union Research Institute for Hydrometeorological Information, 203pp.
- Hameed, S., and I. Pittalwala. 1989: An investigation of the instrumental effects on the historical sunshine record of the United States. *J. Climate*, **2**, 101-104.
- Heino, R. 1989: Changes of precipitation in Finland. Proc. of conference on climate and water, Helsinki, Finland, 11-15 September, 1989. Ed., Vation Painatuskeskus, 111-120.
- Inter-governmental Panel on Climate Change (IPCC), 1990: Climate Change: The IPCC Scientific Assessment. Ed. J.T. Houghton, G.I. Jenkins, and J.J. Ephraums, Cambridge Press, 40 West 20th Street, NY, NY 10011, 362pp.
- Jenne, R.L., and T.B. McKee. 1985: Solar Data Sets. *Handbook of Appl. Meteor.*, D.D. Houghton, Ed., John Wiley and Sons, 1227-1230.
- Jones, P.D., S.C.B. Raper, R.S. Bradley, H.F. Diaz, P.M. Kelly, and T.M.L. Wigley. 1986: Northern hemisphere surface air temperature variations: 1851-1984. *J. Clim. Appl. Meteor.*, **25**, 161-179.
- Jones, P.D., P. Ya Groisman, M. Coughlan, N. Plummer, W.-C. Wang, and T.R. Karl. 1990: How large is the urbanization bias in large-area-averaged surface air temperature trends? *Nature*, **347**, 169-172.
- Karl, T.R., and W.E. Riebsame. 1984: The identification of 10- to 20-year temperature and precipitation fluctuations in the contiguous United States. *J. Clim. Appl. Meteor.*, **23**, 950-966.
- Karl, T.R., G. Kukla, and J. Gavin. 1984: Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980. *J. Climate Appl. Meteor.*, **23**, 1489-1504.

- Karl, T.R., C.N. Williams, Jr., P.J. Young, and W.M. Wendland. 1986a: A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. *J. Clim Appl. Meteor.*, 25, 145-160.
- Karl, T.R., G. Kukla, and J. Gavin. 1986b: Relationship between decreased temperature range and precipitation trends in the United States and Canada, 1941-80. *J. Clim. Appl. Meteor.*, 25, 1878-1886.
- Karl, T.R., and C.N. Williams, Jr. 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Clim Appl. Meteor.*, 26, 1744-1763.
- Karl, T.R., H. Diaz, and G. Kukla, 1988: Urbanization: Its detection in the U.S. climate record. *J. Clim.*, 1, 1099-1123.
- Karl, T.R., D. Tarpley, R.G. Quayle, H.F. Diaz, D.A. Robinson, and R.S. Bradley. 1989: The recent climate record: What it can and cannot tell us. *Revs. of Geophys.*, 27, 405-430.
- Karl, T.R., and P.D. Jones. 1990: Reply to comments on "Urban bias in area-averaged surface temperature trends." *Bull. Amer. Met. Soc.*, 71, 572-574.
- Karl, T.R., and P.M. Steurer. 1990: Increased cloudiness in the United States during the first half of the twentieth century: Fact or fiction? *Geophys. Res. Lett.*, 17, 1925-1928.
- Karl, T.R., G. Kukla, V. Razuvaev, M. Changery, R. Quayle, R. Heim, Jr., D. Easterling, and Congbin Fu: Global Warming: Evidence for asymmetric diurnal temperature change. In Press. *Geophys. Res. Letter*.
- Klestchenko, L.K., V.T. Radiuhin, and R.N. Khvostova. 1990: On the comparison of Northern Hemisphere grid point mean monthly air temperature archives. Proceedings of All-Union Research Institute of Hydrometeorological Information, World data Center, ISSVE 153 (in Russian).
- Kline, J. 1991: Personal Communication, Electronics Technician, National Weather Service, International Falls, NOAA.
- Kohler, M.A. 1949: Double-mass analysis for testing the consistency of the records and for making required adjustments. *Bull. Amer. Meteor. Soc.*, 30, 188-189.
- Maronna, R., and V.J. Yohai. 1978: A bivariate test for the detection of a systematic change in mean. *J. Amer. Statist. Assoc.*, 73, 640-645.
- Mitchell, J.M., Jr. 1961: The measurement of secular temperature change in the eastern United States. Research Paper No. 43, U.S. Weather Bureau, Washington, DC.
- National Oceanic and Atmospheric Administration (NOAA). 1979: Hourly solar radiation - Surface meteorological observations. SOLMET, Vol. 2: Final Report, TD-9724, National Climatic Data Center, 184pp.

- National Research Council. 1990: The U.S. Global Change Research Program: An Assessment of FY 1991 Plans. *National Academy Press*, Washington D.C. 107pp.
- Parker, D.E. 1991: Effects of changing exposure of thermometers at land stations. *Observed Climate Variations and Change: Contributions in support of section 7 of the 1990 IPCC Scientific Assessment WMO/UNEP*, Chapter XVIII.
- Plantico, M.S., T.R. Karl, G. Kukla, and J. Gavin. 1990: Are recent changes of temperature, cloudiness, sunshine, and precipitation across the United States related to rising levels of anthropogenic greenhouse gases? *J. Geophys. Res.*, **95**, 16,617-16,632.
- Potter, K.W. 1981: Illustration of a new test for detecting a shift in mean precipitation series. *Mon. Wea. Rev.*, **109**, 2040-2045.
- Powell, G.L. 1983: Instrumental effects on the sunshine record. *J. Climate Appl. Meteor.*, **22**, 962-963.
- Quayle, R.G., D.R. Easterling, T.R. Karl, and P.J. Hughes, 1991: Effects of recent thermometer changes in the cooperative station network. *Bull. Amer. Meteor. Soc.*, **72**, 1718-1723.
- Quinlan, F.T. 1985: A history of sunshine data in the United States. *Handbook of Applied Meteorology*, D.D. Houghton, Ed., John Wiley and Sons, 1199-1201.
- Robinson, D.A., and K.F. Dewey. 1990: Recent secular variations in the extent of northern hemisphere snow cover. *Geophys. Res. Let.*, **17**, 1557-1560.
- Robinson, D.A., F.T. Keinig, and K.F. Dewey. 1990: Recent variations in northern hemisphere snow cover. *Proceedings of the Fifteenth Annual Diagnostics Workshop, NWS, Climate Analysis Center/NMC, Washington, DC, NOAA.*
- Routledge, B., 1991: A study of tipping bucket correction factors. *Atmospheric Environment Service, Downsview, Ontario, Canada, 12pp plus figures.*
- Sevruk, B. 1982: Methods of correction for systematic error in point precipitation measurement for operational use. *Oper. Hydrol. Rep.*, **21**, Publ. 589, 91pp., World Meteorol. Organ., Geneva, Switzerland.
- Sevruk, B. 1986: Correction of precipitation measurements. *Proceedings of the Zurich Workshop on the Correction of Precipitation Measurements. Zurcher Geogr. Schr.*, **23**, pp. 13-23, Eidenöss. Tech. Hochsch., Zurich.
- Sevruk, B. 1989a (ed.): *Precipitation measurement. Swiss Federal Institute of Technology, WMO/IAHS/ETH Workshop on Precipitation Measurement, St. Moritz, 3-7, December 1989, ETH-Zentrum, CH-8092, Zurich, Switzerland, 584pp.*
- Sevruk, B. 1989b: Inhomogeneities in precipitation time series. In: T. Sansom (ed.), *Proc. 4th International Meeting on statistical Climatology, Rotorua, New Zealand Meteorol. Service, Wellington 1989, 24-27.*

- Sevruk, B., and S. Klerim, 1989: Catalogue of national standard precipitation gauges. Instruments and Observing Methods, Report No. 39, 50pp., Geneva, Switzerland.
- Shver, Ts. A. 1965: Study of the results of the precipitation measurements by the gage with Nipher and Tretyakov wind shields. Leningrad: Gidrometeoizdat, 170pp. (in Russian).
- Solow, A.R., 1987: Testing for climate change: An application of the two-phase regression model. *J. Clim. Appl. Meteor.*, 26, 1401-1405.
- United States Congress, Office of Technology Assessment (OTA). 1991: Changing by degrees: Steps to reduce greenhouse gases. USGPO, Washington, DC 20402-9325.
- U.S. Department of Commerce, 1963: History of Weather Bureau Precipitation Measurements. Weather Bureau, Key to meteorological records documentation No. 3.082, Washington, DC, 19pp.
- World Meteorological Organization (WMO), 1991: Final Report of the Fifth Session of International Organizing Committee for the WMO Solid Precipitation Measurement Intercomparison. World Meteorol. Organ., Geneva, Switzerland, 19pp.

INTERNAL DISTRIBUTION

1. L. D. Bates
2. T. A. Boden
3. J. H. Cushman
4. R. M. Cushman
5. M. P. Farrell
6. D. E. Fowler
7. C. W. Gehrs
8. S. G. Hildebrand
9. D. P. Kaiser
10. P. Kanciruk
11. D. E. Reichle
12. F. E. Sharples
13. D. S. Shriner
14. S. H. Stow
15. R. I. Van Hook
- 16-115. CDIAC
116. Central Research Library
- 117-121. ESD Library
122. Information Analysis Library
- 123-124. Laboratory Records Department
125. Laboratory Records, RC
126. ORNL Patent Office
127. ORNL Y-12 Technical Library

EXTERNAL DISTRIBUTION

- 128-132. William A. Brower, National Climatic Data Center, Federal Building, Asheville, NC 28801
133. Roger C. Dahlman, Carbon Dioxide Research Program, Environmental Sciences Division, Office of Health and Environmental Research, ER-74, U.S. Department of Energy, Washington, DC 20585
134. David Easterling, National Climatic Data Center, Federal Building, Asheville, NC 28801
135. Jerry F. Franklin, Bloedel Professor of Ecosystem Analysis, College of Forest Resources, University of Washington, Anderson Hall (AR-10), Seattle, WA 98195
136. Fu Congbin, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

137. David J. Galas, Office of Health and Environmental Research, ER-70, U.S. Department of Energy, Washington, DC 20585
138. Thomas J. Gross, Carbon Dioxide Research Program, Environmental Sciences Division, Office of Health and Environmental Research, ER-74, U.S. Department of Energy, Washington, DC 20585
139. Sultan Hameed, Laboratory for Planetary Atmospheres Research, State University of New York, Stony Brook, NY 11794-2300
140. George M. Hornberger, Department of Environmental Sciences, Clark Hall, University of Virginia, Charlottesville, VA 22903
- 141-145. Pamela Y. Hughes, National Climatic Data Center, Federal Building, Asheville, NC 28801
146. Roy L. Jenne, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000
147. G. Y. Jordy, Director, Office of Program Analysis, Office of Energy Research, ER-30, G-226, U.S. Department of Energy, Washington, DC 20585
- 148-152. Thomas R. Karl, National Climatic Data Center, Federal Building, Asheville, NC 28801
153. F. A. Koomanoff, Office of Basic Energy Sciences, ER-10, U.S. Department of Energy, Washington, DC 20585
154. T. K. Lau, Program Officer, Office of International Affairs, IE-12, U.S. Department of Energy, Forrestal Building, Washington, DC 20585
- 155-159. Elaine H. Mason, National Climatic Data Center, Federal Building, Asheville, NC 28801
160. H. M. McCammon, Acting Deputy Director, Environmental Sciences Division, Office of Health and Environmental Research, Office of Energy Research, ER-74, U.S. Department of Energy, Washington, DC 20585
161. R. H. Olsen, Vice President for Research, University of Michigan, Medical Science Building II, #5605, 1301 East Catherine Street, Ann Arbor, MI 48109-0620
162. Ari Patrinos, Acting Director, Environmental Sciences Division, Office of Health and Environmental Research, ER-74, U.S. Department of Energy, Washington, DC 20585
163. David Portman, Atmospheric and Environmental Research, Incorporated, 840 Memorial Drive, Cambridge, MA 02139

164. Michael R. Riches, Carbon Dioxide Research Program, Environmental Sciences Division, Office of Health and Environmental Research, ER-74, U.S. Department of Energy, Washington, DC 20585
165. Peter M. Steurer, National Climatic Data Center, Federal Building, Asheville, NC 28801
166. Tao Shiyan, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
167. Wei-Chyung Wang, Professor and Head, Climate System Sciences, Atmospheric Sciences Research Center, University of Albany, State University of New York, 100 Fuller Road, Albany, NY 12205
168. F. J. Wobber, Environmental Sciences Division, Office of Health and Environmental Research, Office of Energy Research, ER-74, U.S. Department of Energy, Washington, DC 20585
169. Zeng Zhaomei, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
170. Zhang Qingyun, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
171. Office of Assistant Manager for Energy Research and Development, U.S. Department of Energy Oak Ridge Field Office, P. O. Box 2001, Oak Ridge, TN 37831-8600
- 172-181. Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, TN 37831

END

**DATE
FILMED**

7 / 9 / 92

