1.8 AN ANALYSIS OF THE IMPACT OF DATA INHOMOGENEITIES ON THE SNOW RECORD

Richard R. Heim Jr. *
NOAA / NESDIS / National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

Snow is a significant factor in the economy and water resources of the United States. Its impact can disrupt transportation, cause extensive damage and loss of life, and require considerable snow removal costs. Lack of snow in normally snowy climes can bring economic hardship to the recreation and water management industries. Snow also has an important role in climatology, both reflecting climatic changes and fluctuations as well as exerting an influence on climate.

Unlike other climate variables, snow is inherently difficult to measure accurately and consistently. This is due to three properties of snow: (1) it often melts as it lands or as it lies on the ground, (2) snow settles as it lies on the ground, and (3) snow is easily blown and redistributed. Because of these three factors, the observation time (Groisman et al., 2000), frequency, and location all affect the data (Doesken and Judson, 1997). Inhomogeneities introduced into the data record can affect climatologies and trends computed from the data which would adversely impact economic and scientific users.

Quality control, inventory, and metadata statistics can be used to assess the quality of the station data and, consequently, the accuracy of analyses such as snow climatologies (Heim and Leffler, 1999a, 1999b). Although the importance of the inhomogeneity factor on the snow data record cries out for the development of data adjustment strategies, there have been few analyses done to quantify these impacts and develop data adjustment methodologies (Groisman et al., 2000).

This paper summarizes a case study which examined the impacts of data inhomogeneities on the snow record of a station in Urbana, Illinois (Heim and Angel, 2000), and suggests strategies for coping with the inhomogeneities in climate change studies.

2. DATA

Daily snowfall data from the National Oceanic and Atmospheric Administration's (NOAA's) National Climatic Data Center (NCDC) TD-3200 data base (Cooperative Summary of the Day) were analyzed. Urbana, Illinois (Cooperative station number 118740) was chosen for study because of its long digital record (1903-1996), high degree of completeness (1116 months out of 1128 possible months had no missing days [99% complete]), and small amount of missing data (0.058% of the days were missing).

* *Corresponding author address:* Richard R. Heim Jr., National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801-5001; e-mail: Richard.Heim@noaa.gov.

Extensive quality assurance tests were applied to the snowfall data, including statewide limits checks to identify outliers, internal consistency checks to ensure consistency between snowfall and precipitation, consistency checks with minimum temperature to identify the occurrence of hail rather than snow (flagged snowfall values were set to zero), and factor of 10 errors.

Snowfall data from nearby stations Pana and Hoopeston 1NE were used in a controlled experiment to assess the impact of a known inhomogeneity.

3. METHODOLOGY

The following parameters were computed from the daily snowfall data:

TOTAL SF: total snowfall amount;
MAX 1-DAY: greatest 1-day snowfall amount;
MAX 2-DAY: greatest 2-day snowfall amount, where it snowed both days;
DATE FIRST: date of first snowfall ≥ 0.254 cm. (0.1 in.) during the August-July snow season;
DATE LAST: date of last snowfall ≥ 0.254 cm. (0.1 in.) during the August-July snow season;
LENGTH: length of the snow season (number of days between DATE FIRST and DATE LAST);
NUMDAYS: number of days with snowfall ≥ trace;
MAX MEDIAN: greatest median snowfall amount (for the days with snowfall, the median [MED] snowfall amount was calculated for each month; MAX MEDIAN is the maximum MED); and
AVE MEDIAN: average median snowfall amount (the average of the monthly MED values).

Total snowfall amount (TOTAL SF) was computed for a given year-season only if there were no days missing for the season in question. The other parameters were tolerant of missing data.

The impact of changing station location was examined by creating a "control station" from the data records of stations Pana and Hoopeston 1NE, which are located on either side of Urbana along a southwest to northeast line. A known inhomogeneity (lateral location change of 169 km. [105 mi.]) was forced into the "control" time series by concatenating 1925-1951 data from Pana to 1952-1979 data from Hoopeston. This "control" time series was compared to the corresponding time series for Urbana (the "target" station). This period was chosen as the study period because all three stations have complete data records during the 1924-25 to 1978-79 snow seasons (August-July). A record of known observation time changes at all three stations was obtained from Changon and Boyd (1963) and from examination of station history metadata (see Table 1). Double mass plots ("control" vs. "target") were generated to assess the effects of the location and observation time inhomogeneities on each of the snowfall parameters.

Table 1. Observation time history for Urbana, Pana, and Hoopeston.

Date/Year	Urbana	Pana*/Hoopeston**
1925	7 pm (Oct-Mar), 7 am (Apr-Sep)	5 pm *
1935	7 pm	5 pm *
6-21-1939	7 pm	6 pm *
5-5-1945	7 pm	7 pm *
1952	7 pm	6 pm **
1-1-1956	Midnight	6 pm **
1-1-1966	7 am	6 pm **
7-1-1969	7 am	7 am **
11-8-1972	Midnight	7 am **

4. RESULTS

Plots of the time series for the nine parameters over the period 1903-1996 are shown in Figs. 1-6. Breaks in the TOTAL SF parameter record are apparent due to incomplete data. All of the time series were examined for trends using least squares linear regression, two-phase linear regression, and the nonparametric Wilcoxon signed-rank test, with most of the results indicating no significant trend (Heim and Angel, 2000). The plots of the seasonal and annual (i.e., August-July) TOTAL SF (Figs. 1a and 1b) generally agree with regional and continental-scale studies of snow cover extent (Heim and Angel, 2000).

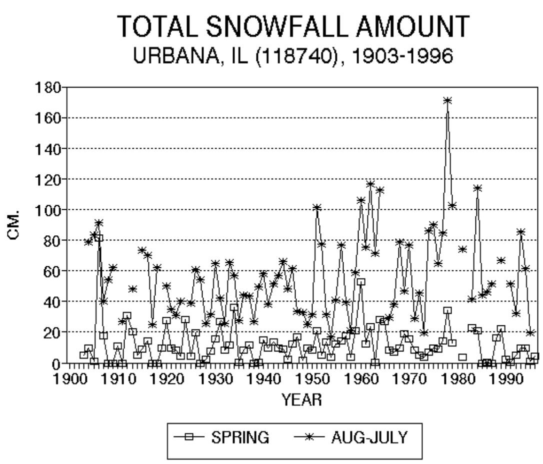

Fig. 1a. Variations in 20th Century total snowfall amount at Urbana, IL, for spring (March-May) and annual (August-July).

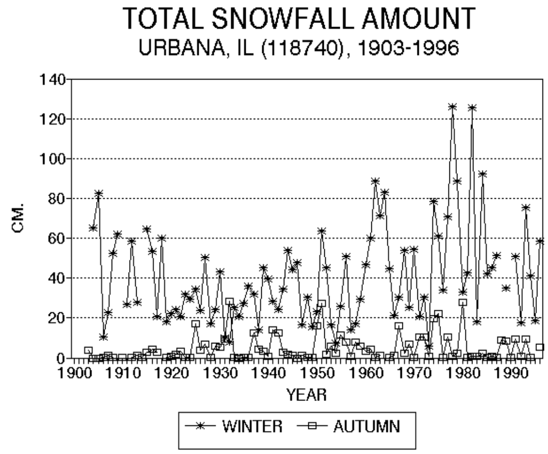

Fig. 1b. Same as Fig. 1a, except for winter (Dec-Feb) and autumn (Sep-Nov).

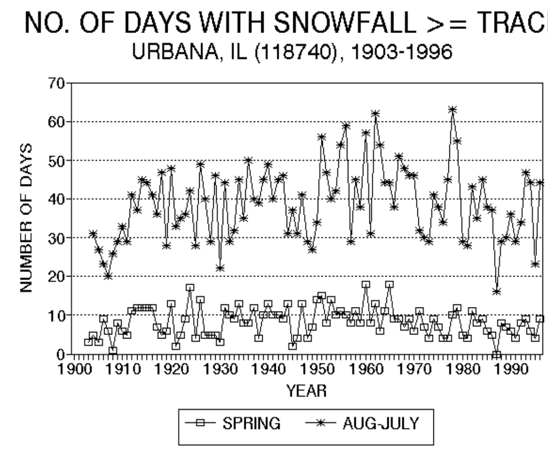

Fig. 2a. Variations in 20th Century number of days with snowfall at Urbana, IL, for spring (Mar-May) and annual (Aug-Jul).

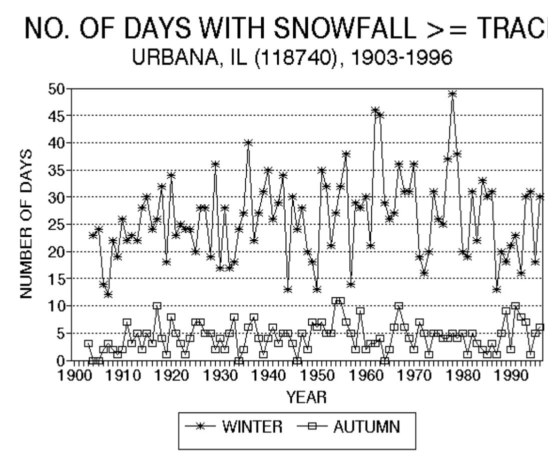

Fig. 2b. Same as Fig. 2a, except for winter (Dec-Feb) and autumn (Sep-Nov).

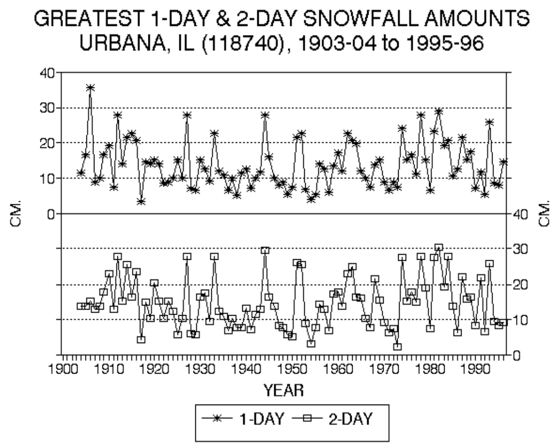

Fig. 3. Top: greatest one-day amount. Bottom: greatest two-day amount, where it snowed both days.

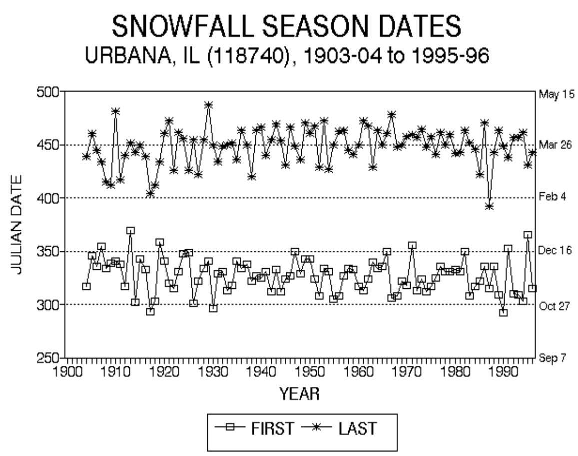

Fig. 4. Variations in 20th Century snowfall season start (bottom) and ending (top) dates for Urbana, IL. Julian date of 213 corresponds to August 1, 365 to December 31, 366 to January 1, and 577 to July 31.

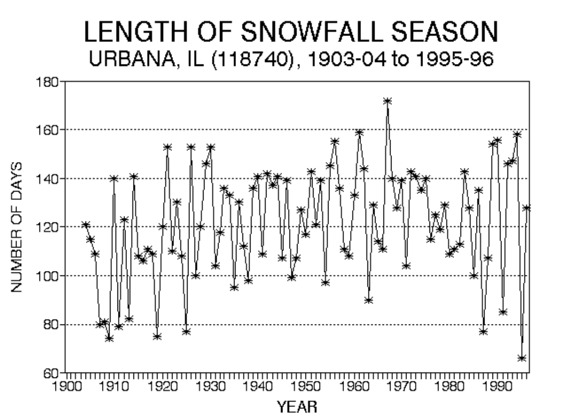

Fig. 5. Variations in 20th Century snowfall season length at Urbana, IL.

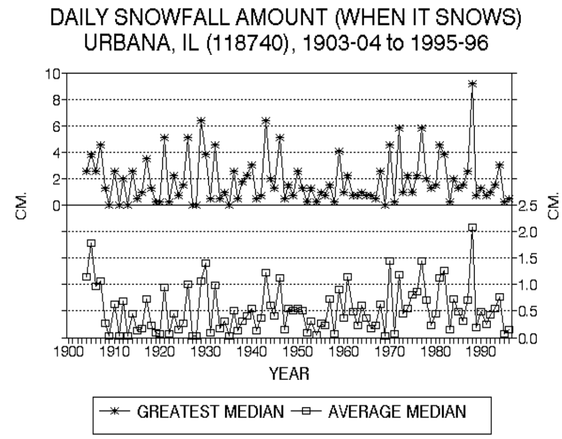

Fig. 6. Variations in 20th Century median daily snowfall amount at Urbana, IL. Monthly median daily snowfall computed for those days with snowfall. Top: greatest of the monthly median values for each "year" (August-July). Bottom: average of the monthly median values for each "year".

The double mass plots of Pana-Hoopeston ("control") vs. Urbana ("target") parameters for the period 1924-1979 are shown in Figs. 7-11. Autumn (Fig. 7a) and spring TOTAL SF have greater deviations from the straight line than winter (Fig. 7b) and annual TOTAL SF. A discontinuity in the late 1930s had a major impact on autumn TOTAL SF, while discontinuities in the 1960s and 1970s adversely impacted both autumn and winter TOTAL SF.

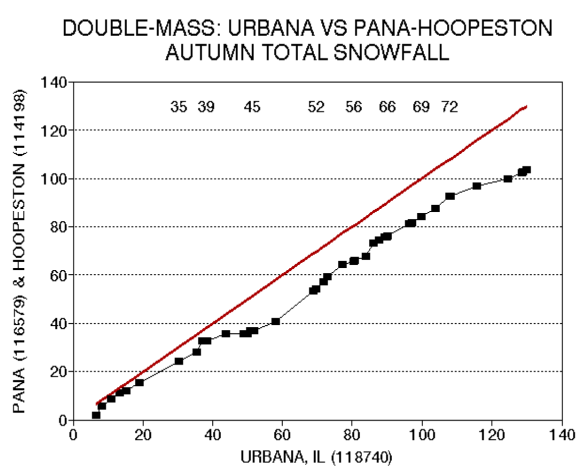

Fig. 7a. Double-mass plot of Urbana vs. "control" station for autumn (Sep-Nov) total snowfall, based on 1924-1979 data. The numbers at the top of the graph identify years of observation time changes indicated in Table 1.

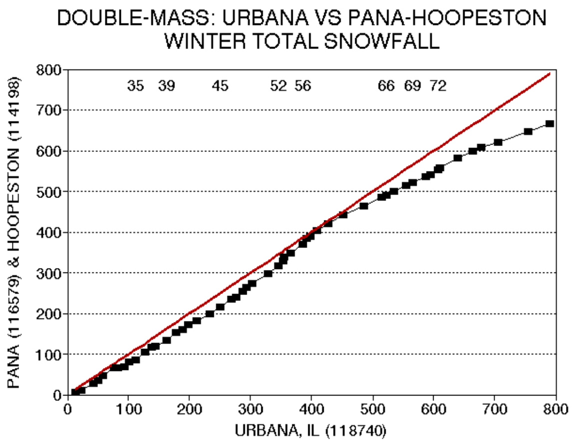

Fig. 7b. Same as in Fig. 7a, except for winter (Dec-Feb) total snowfall.

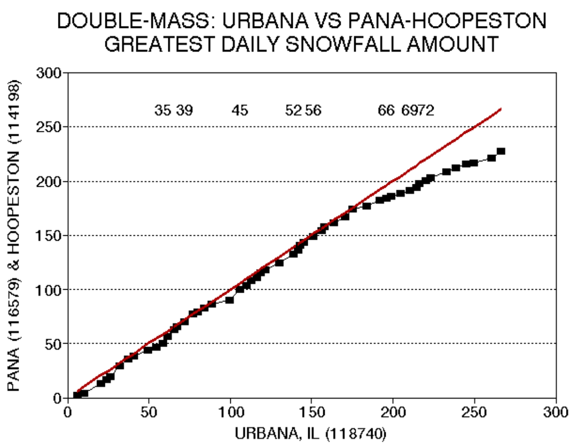

Fig. 8. Same as in Fig. 7a, except for annual (Aug-Jul) greatest daily snowfall amount.

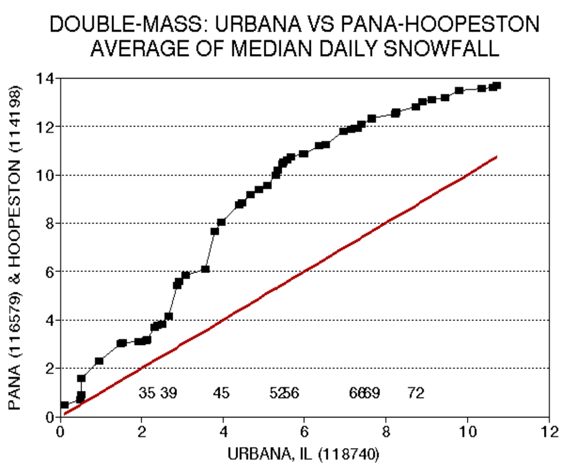

Fig. 9. Same as in Fig. 7a, except for annual (Aug-Jul) average median daily snowfall amount.

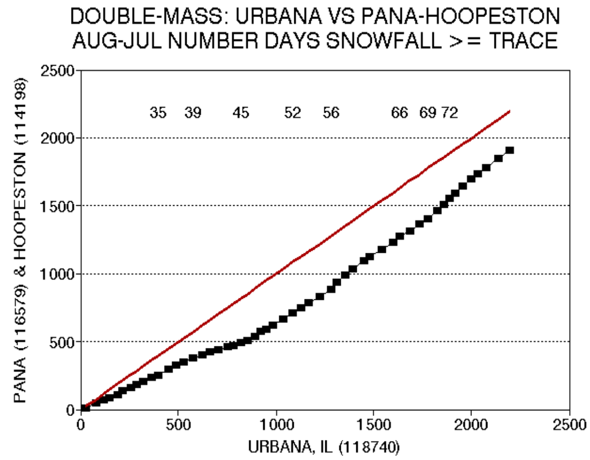

Fig. 10. Same as in Fig. 7a, except for annual (Aug-Jul) total number of days with snowfall.

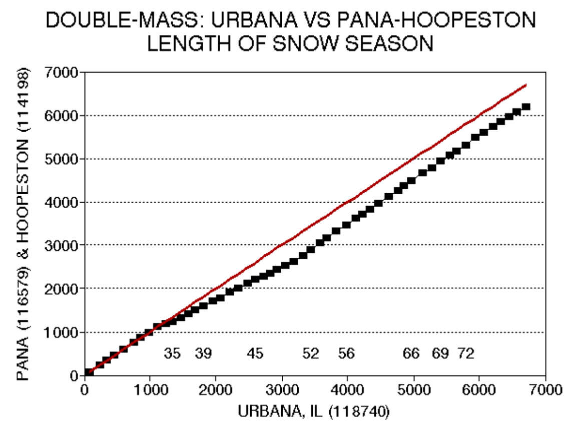

Fig. 11. Same as in Fig. 7a, except for length of snow season.

The plots for MAX 1-DAY (Fig. 8) and MAX 2-DAY are similar to winter TOTAL SF. AVE MEDIAN (Fig. 9) and MAX MEDIAN show the greatest sensitivity to discontinuities, with jumps at almost every marked year. For NDAYS, the seasonal plots are similar to the annual plot (Fig. 10), with a deviation during the first approximately 20 years, and lesser effects afterward. DATE FIRST, DATE LAST, and LENGTH were least sensitive to discontinuities. Season LENGTH (Fig. 11) showed a gradual deviation from approximately 1935 to 1950, with minimal effects before and after.

Previous research suggests changes in observation time should have a significant impact on snowfall amount. A study of hourly surface observations at Springfield, Illinois indicated snowfall was more likely to occur in the late morning to early afternoon hours. This is consistent with Changnon (1969) regarding the onset of heavy (> 15.24 cm) snowstorms in Illinois. Therefore, measurements made around 7 a.m. may record different amounts than those at 6 p.m. after the snow has had time to settle and/or drift.

Some of these deviations in the double mass plots could be due to changes in observer-related measurement practices and exposure. Pana had six observers during 1925-1951, with the greatest station relocation being 0.8 km. (0.5 mi.). Hoopeston had the same observer during 1951-1979, however a station move of 1 km. (0.6 mi.) in February 1962 resulted in a change from a fairly open exposure with no trees to a limited exposure with trees and buildings in the vicinity. This inhomogeneity is likely the explanation for the deviation starting in the early 1960s seen on the double mass plots for winter TOTAL SF (Fig. 7b) and MAX 1-DAY snowfall (Fig. 8).

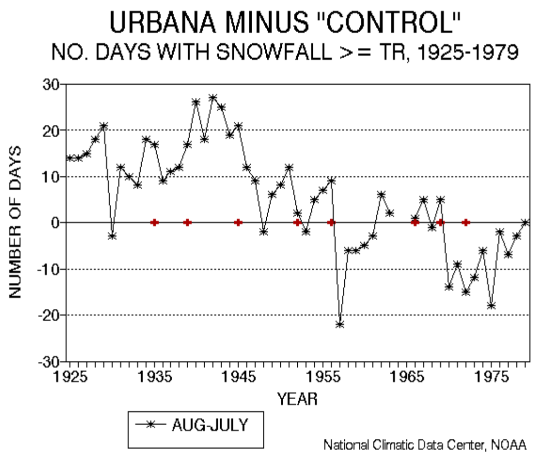

Fig. 12. Annual (Aug-Jul) number of days with snowfall at Urbana minus same parameter at "control" station, 1925-1979.

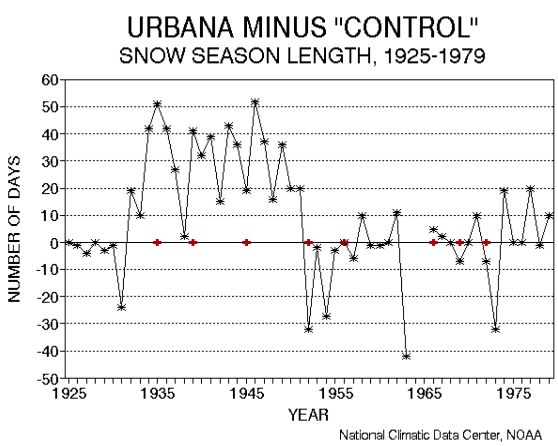

Fig. 13. Same as in Fig. 12, except for snow season length.

Plots of the difference between Urbana's values and the corresponding values for the "control" station for 1924-25 to 1978-79 are shown in Figs. 12-17. The ticks on the zero line in these graphs mark years with known inhomogeneities. Differences that are fairly consistent between ticks indicate periods for which it might be feasible to compute and apply adjustment factors to the data. Only two parameters have difference graphs which suggest adjustment factors may be possible: NUMDAYS (Fig. 12) and snowfall season LENGTH (Fig. 13). The snowfall amount parameters have difference graphs (Figs. 14-17) which indicate that adjustment factors likely cannot be computed.

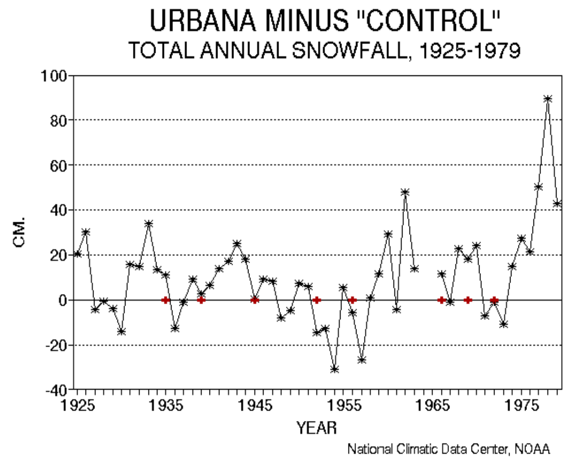

Fig. 14. Same as in Fig. 12, except for total annual (August-July) snowfall.

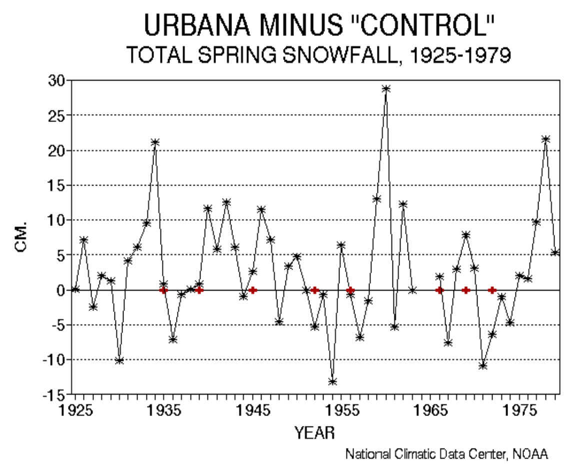

Fig. 15. Same as in Fig. 12, except for total spring (Mar-May) snowfall.

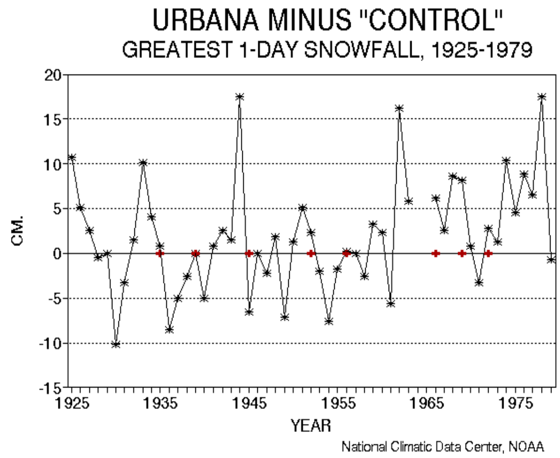

Fig. 16. Same as in Fig. 12, except for annual (Aug-Jul) greatest one-day snowfall amount.

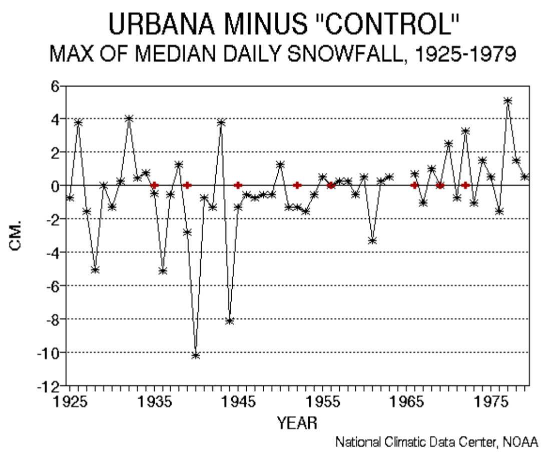

Fig. 17. Same as in Fig. 12, except for annual (Aug-Jul) maximum of median daily snowfall amount.

5. DISCUSSION

Climatic parameters based on daily snowfall were used effectively to show 20th Century variations in the snow climatology at Urbana, Illinois and to illustrate the impacts of inhomogeneities on the data record. Of the parameters examined, snowfall amount was most sensitive, number of days with snow was less sensitive, and snow season length was least affected by inhomogeneities. Changes in observation time appeared to be the most important factor, along with exposure changes.

As suggested by the Pana-Hoopeston "control" experiment, station relocations did not appear to have as large an effect. However, topographic and synoptic considerations need to be carefully considered in such situations. If the move is to a location with similar exposure and snow climatology characteristics, then the impact of the station relocation would be minimized. This is likely the situation in this case study. The Pana-Urbana-Hoopeston line parallels the typical southwest to northeast orientation of Midwest snow storm tracks. Further study is needed to investigate this relationship, including analyses of data from stations in a transect perpendicular to the storm track and from stations in other parts of the country.

It may be possible to compute adjustment factors for station data based on neighbor data for some snowfall parameters (number of days with snowfall and snow season length), but this is unlikely for other parameters (daily and total monthly snowfall amount).

In conclusion, despite the presence of inhomogeneities in the snow data record, parameters can be derived which minimize these impacts and provide useful tools for analyzing snow-related trends and variability.

6. REFERENCES

Changnon, Jr., S.A., 1969: Climatology of Severe Winter Storms in Illinois. ***Bulletin 53***, Illinois State Water Survey, Urbana, 45 pp.

Changnon, Jr., S.A. and G.R. Boyd, 1963: History of the Urbana Weather Station: 1888-1963. ***Circular 88***, Illinois State Water Survey, Urbana, 37 pp.

Doesken, N.J. and A. Judson, 1997: *The Snow Booklet: A Guide to the Science, Climatology, and Measurement of Snow in the United States, 2nd Edition.* Colorado Climate Center, Colorado State University, Fort Collins; 86 pp.

Groisman, P. Ya., B. Sun, and R.R. Heim, Jr., 2000: Trends in spring snow cover retreat over the U.S. and the effect of observation time bias. *Preprints of the Eleventh Symposium on Global Change Studies*, 9-14 January 2000, Long Beach, CA, 54-7.

Heim, Jr., R.R. and J.R. Angel, 2000: Urbana, Illinois: 20th Century snowfall variations and non-climatic influences. *Proceedings of the 56th Annual Eastern Snow Conference*, 2-4 June 1999, Fredericton, NB, Canada, 125-140.

Heim, Jr., R.R. and R.J. Leffler, 1999a: Snow climatologies for NWS Cooperative stations--An overview. *Preprints of the 11th Conference on Applied Climatology*, 10-15 January 1999, Dallas, TX, 225-226.

Heim, Jr., R.R. and R.J. Leffler, 1999b: 1948-1996 snowfall return period statistics computed for the Federal Emergency Management Agency. *Preprints of the 11th Conference on Applied Climatology*, 10-15 January 1999, Dallas, TX, 234-235.