P2.18 TRENDS AND VARIABILITY OF SNOWFALL AND SNOW COVER ACROSS NORTH AMERICA AND EURASIA. PART 1: DATA QUALITY AND HOMOGENEITY ANALYSIS

Richard R. Heim Jr. *
NOAA National Climatic Data Center, Asheville, North Carolina
David A. Robinson
Rutgers University, Department of Geography, Piscataway, New Jersey

1. INTRODUCTION

Snow is a significant factor in the national economy and also in the water resources of Northern Hemisphere countries. Snow has an important climatic role reflecting climatic changes and fluctuations, as well as exerting an influence on climate. The availability of satellite monitoring of weather and climate variables enabled scientists to develop and analyze hemispheric snow cover extent using a consistent database. Unfortunately, the satellite snow record goes back only four decades. In situ observations of snow cover as well as snowfall are available for some stations going back to the beginning of the Twentieth Century.

The research presented in this paper and a companion contribution (cf. Robinson and Heim, Trends and variability of snowfall and snow cover across North America and Eurasia. Part 2: What the data say) includes a comprehensive analysis of in situ snow observations from stations in the United States, Canada, and the Former Soviet Union (FSU) using a consistent methodology applied to all of the stations. This paper discusses the first portion of the effort, and includes: 1) data sources and variables analyzed, 2) the quality control that was applied, and 3) the snow indices that were computed from the daily snow observations. The quality control and inventory summary statistics were utilized to identify the best stations to use for Northern Hemisphere snow assessments.

2. DATA

Daily snowfall, snow depth, precipitation, and maximum and minimum temperatures were utilized in the analysis. For the U.S. (US) stations, Cooperative Station Summary of the Day (TD-3200) data were used. For the Canadian (CN) stations, the Daily Climatological Data (DLY02, DLY04) dataset was used. For the FSU stations, snow depth data from the National Snow and Ice Data Center (NSIDC) Historical Soviet Daily Snow Depth-Version 2.0 (HSDSD) dataset were used, supplemented by temperature and precipitation data for the HSDSD stations provided by Pasha Groisman (Groisman, 2005, personal communication). The FSU data did not include daily snowfall observations. Some of the snow indicators created for this study required snowfall, so pseudo-snowfall values were computed for the FSU stations. The pseudo-snowfall value for day X was computed by subtracting the snow depth from day X-1 from the snow depth for day X. If the difference was positive or zero, it was identified as a snowfall amount. If the difference was negative, the snowfall amount was set to zero. These pseudo-snowfall values have their greatest utility as an indicator of whether or not it had snowed on a given day.

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

3. ASSESSMENT OF DATA QUALITY

Examination of the HSDSD dataset revealed the presence of factor-of-ten and other similar data errors. Where possible, errors were corrected, but in cases where a value could not be corrected, it was removed from the analysis. Otherwise, the data from all three sources were put through the same quality assurance routines. These routines include the following:

(1) internal consistency checks between snowfall and snow depth, and between snowfall and precipitation;

(2) nonzero snowfall values were assumed to be hail and were set to zero if the corresponding minimum temperature was greater than or equal to 4.4 C (40 F);

(3) factor of ten checks for snowfall and snow depth; and

(4) limit checks for snowfall and snow depth.

Additional checks were made, but the data values were only flagged, not changed, if they failed. These include:

(5) questionable snowfall-to-precipitation ratios (snowfall values within certain ranges and greater than certain fractional percentages of the daily precipitation); and

(6) questionable decreases in snow depth (which occurred when significant snowfall occurred and temperatures remained below freezing).

The results from the quality assurance checks were combined with an inventory analysis to create quality control and inventory (QCI) statistics. The QCI statistics were used to assess the quality of the stations and identify stations to use in the analysis, and include:

(a) the number of non-missing daily values read from the dataset (NREAD);

(b) the number of daily values that were flagged as suspect by the quality assurance routines (NFLAG), including those flagged values that were adjusted and those values for which no corrective action was taken;

(c) the number of daily values that failed the quality assurance checks and were set to missing (NMISS);

(d) the percent of the non-missing daily values read that were flagged as suspect or set to missing (100%*[NFLAG+NMISS]/NREAD);

(e) the first year with data in the dataset, the last year with data in the dataset, and the number of years between the first and last years;

(f) the number of years in the dataset that had some data (at least one day);

(g) the number of months having complete data (no days missing) (NMONC);

(h) the number of possible months in the data period, i.e., the number of months between the first year first month of data and the last year last month of data (NMONP);

(i) the percent of possible months having complete data (NMONPCT=100%*NMONC/NMONP);

(j) the number of usable daily values processed (NPROC);

(k) the number of daily values missing (NMSG);

(l) the percent of daily values missing (100%*NMSG/[NMSG+NPROC]);

(m) the maximum number of consecutive whole months missing (biggest break);

(n) the total number of breaks.

4. SELECTION OF CANDIDATE STATIONS

In order to be used to assess trends in snow cover and snowfall over a century timescale, a station would need to have a long record that is generally complete. The QCI statistics that best assess these qualities are: first year with data, last year with data, and the percent of possible months having complete data (e and i above). Out of 25,950 US, CN, and FSU stations that were examined, most of them did not meet the selection criteria. A few starting years stand out. For the US-CN stations, a significant number of stations had data beginning in 1948, while a significant number had data starting in the early 1890s. For the FSU stations, about 10% started in the early 1880s, while about a third had data beginning in 1936. But these early (1800s) US-CN-FSU stations had a considerable amount of data missing during the first half of the twentieth century. It quickly became clear that there was insufficient data for a full century-scale analysis, but adequate data were available for an analysis covering the last half of the twentieth century.

There were 10,197 stations with ending year equal to or later than 2000; however, there was a significant number of stations with long and fairly complete records that ended in the 1990s. These latter stations were excluded. Of the 10,197 stations, there were 5574 that had a starting year of 1951 or earlier. These 5574 stations break down as follows:

5034 - US
339 - CN
201 - FSU

The percent of possible months having complete data (NMONPCT) criteria was applied to the post-1947 period instead of the full period in order to exclude the sparsely populated early twentieth century portion of the data record, which would have skewed the results. Of the 5574 stations that met the first two selection criteria, only 1795 stations were 90% or more complete (NMONPCT criteria, applied to snowfall for the US-CN stations and snow depth for the FSU stations). However, this subset of stations included none from Alaska and the FSU, and there were spatial data holes in Canada and the contiguous United States. In order to address these shortfalls, the NMONPCT selection criteria was relaxed to 75% for the FSU stations and 70% for Canada, Alaska, and the contiguous U.S. data holes. This resulted in a subset of 1997 stations which were used for the analysis. These 1997 stations break down as follows:

1553 - US
298 - CN
146 - FSU

5. SNOW PARAMETERS

Time series of the following snow parameters were computed for each station from the daily snowfall and snow depth data:

(1) total monthly and seasonal snowfall amount;

(2) greatest 1-day snowfall amount;

(3) average median daily snowfall amount;

(4)-(6) number of days with snowfall equal to or greater than 0.25 cm. (0.1 in.), 2.54 cm. (1.0 in.), and 12.7 cm. (5.0 in.), respectively;

(7)-(8) number of days with snow depth equal to or greater than 2.54 cm. (1.0 in.) and 12.7 cm. (5.0 in.), respectively; and

(9) length of snowfall season (in number of days, where snow season begins on or after August 1 and ends on or before July 31 and based on snowfall equal to or greater than 0.25 cm.).

6. STATION WEIGHTS

Time series of the above snow parameters were computed for North America (from the US-CN stations) and for Eurasia (from the FSU stations). The stations were weighted by area using a Thiessen polygon method before being averaged into the continental aggregates. The weights for the US-CN stations are shown in Fig. 1 and the weights for the FSU stations are shown in Fig. 2. In addition to an area-averaged continental value for each monthly/seasonal parameter for each year, this methodology also provides an estimate of the percentage areas with data and without data for each year, which can be used as an indicator (confidence value) of how representative the annual values are.

7. SUMMARY

The methodology presented here provided an effective procedure for selecting the best subset of stations with which to compute continental assessments of snow variables over the last half of the twentieth century. The results of the assessments for North America and Eurasia are presented in the companion paper (cf. Robinson and Heim, Trends and variability of snowfall and snow cover across North America and Eurasia. Part 2: What the data say). The research in progress to produce a double-mass analysis to assess the homogeneity of the data was not completed by the deadline for this extended abstract.

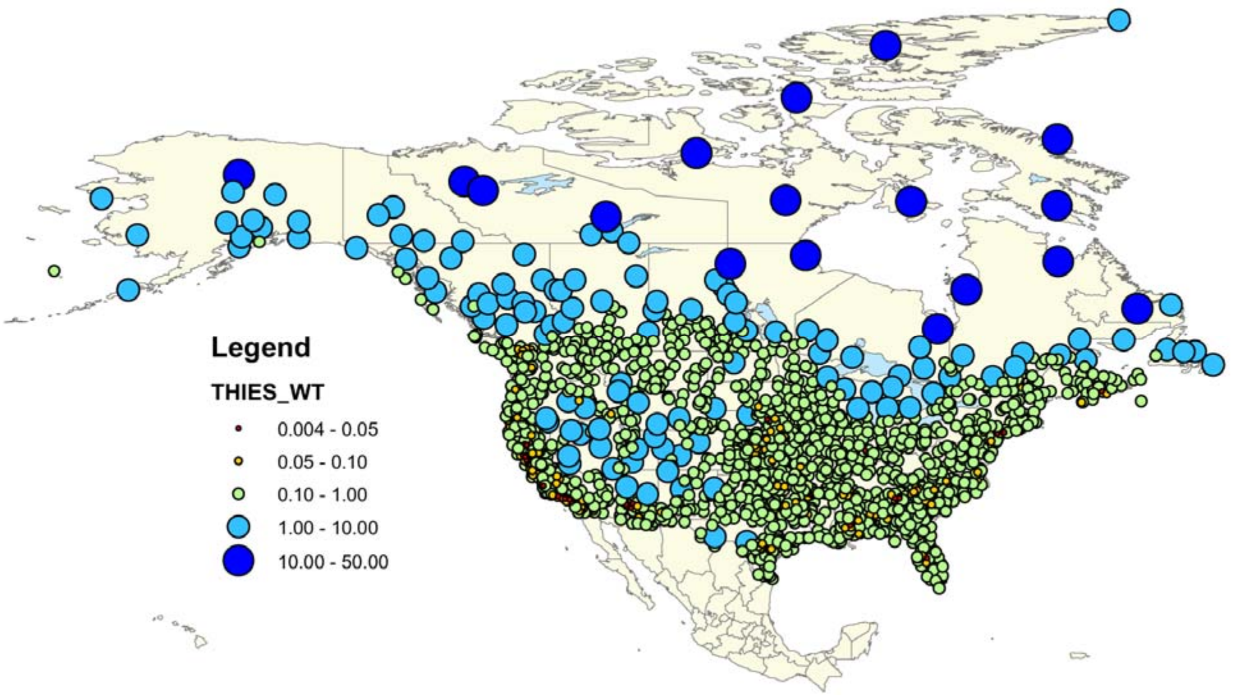

Fig. 1. Location and relative area weights (Thiessen weight X 1000) for US-CN stations.

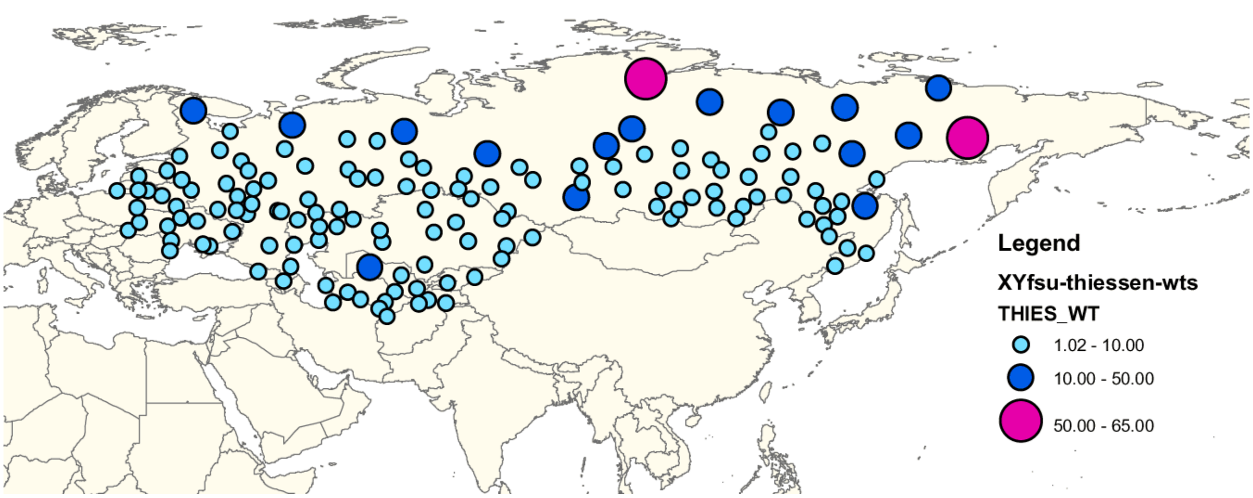

Fig. 2. Location and relative area weights (Thiessen weight X 1000) for FSU stations.