NOAA/National Climatic Data Center
In situ snow observations

Matt Menne
NOAA/National Climatic Data Center
Asheville, North Carolina, USA

ECMWF Snow Archive Workshop
November 17-19, 2014

Outline

- In situ datasets at NCDC
- Sources of snow data
- Near-term plans for data additions
- Reprocessing schedule
- QC checks
- Access

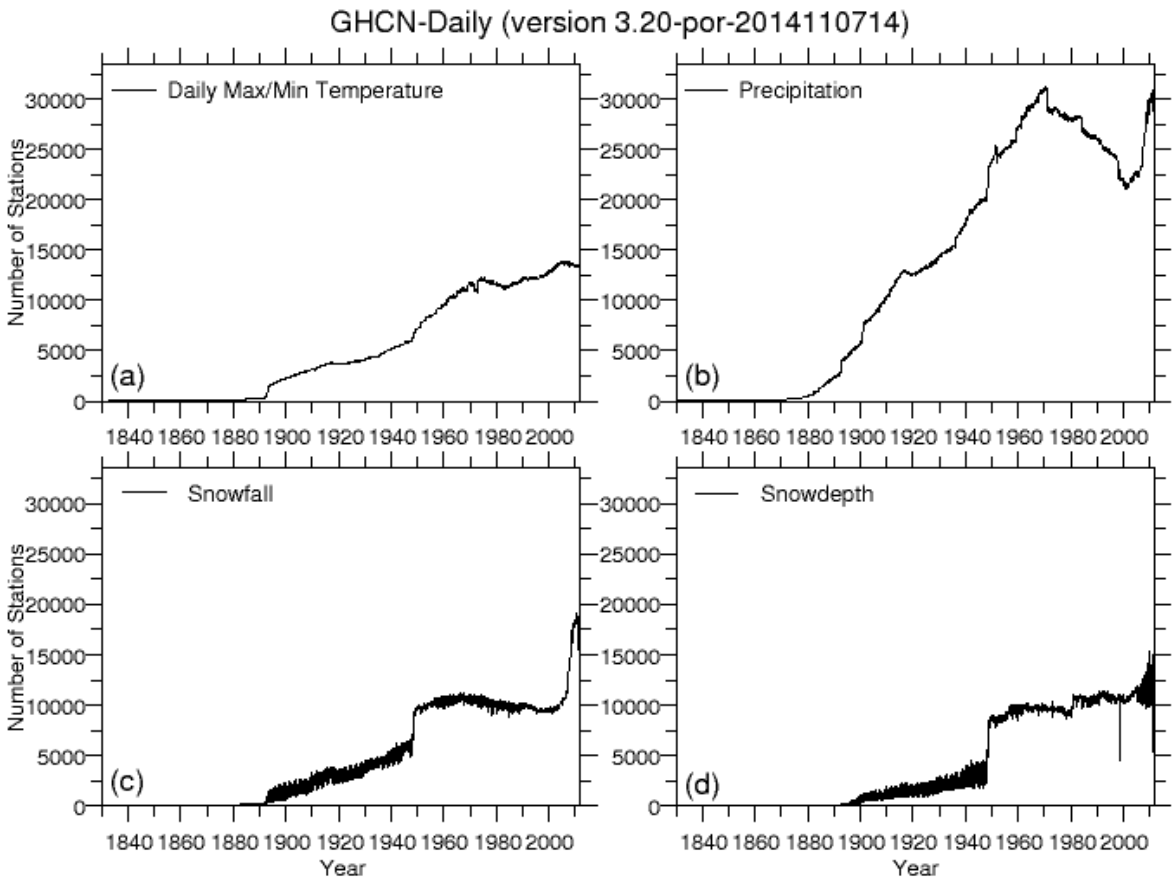
In situ datasets at NCDC: Integrated Surface Dataset

- Contains global hourly and synoptic observations compiled from ~100 sources
- Developed as a joint activity within Asheville's Federal Climate Complex. Data feed and new additions managed by the 14th Weather Squadron of the U.S. Air Force. NOAA/NCDC's ISD is the public facing version of the Air Force database
- Comprises over 20,000 stations worldwide, with data as far back as 1901, though big increases in volume occur in the 1940's and again in the early 1970's
- Currently over 11,000 stations "active" and updated daily in the database.
- Data access includes several methods: FTP, an interactive interface (called Climate Data Online), a GIS services map interface, and various Web Services

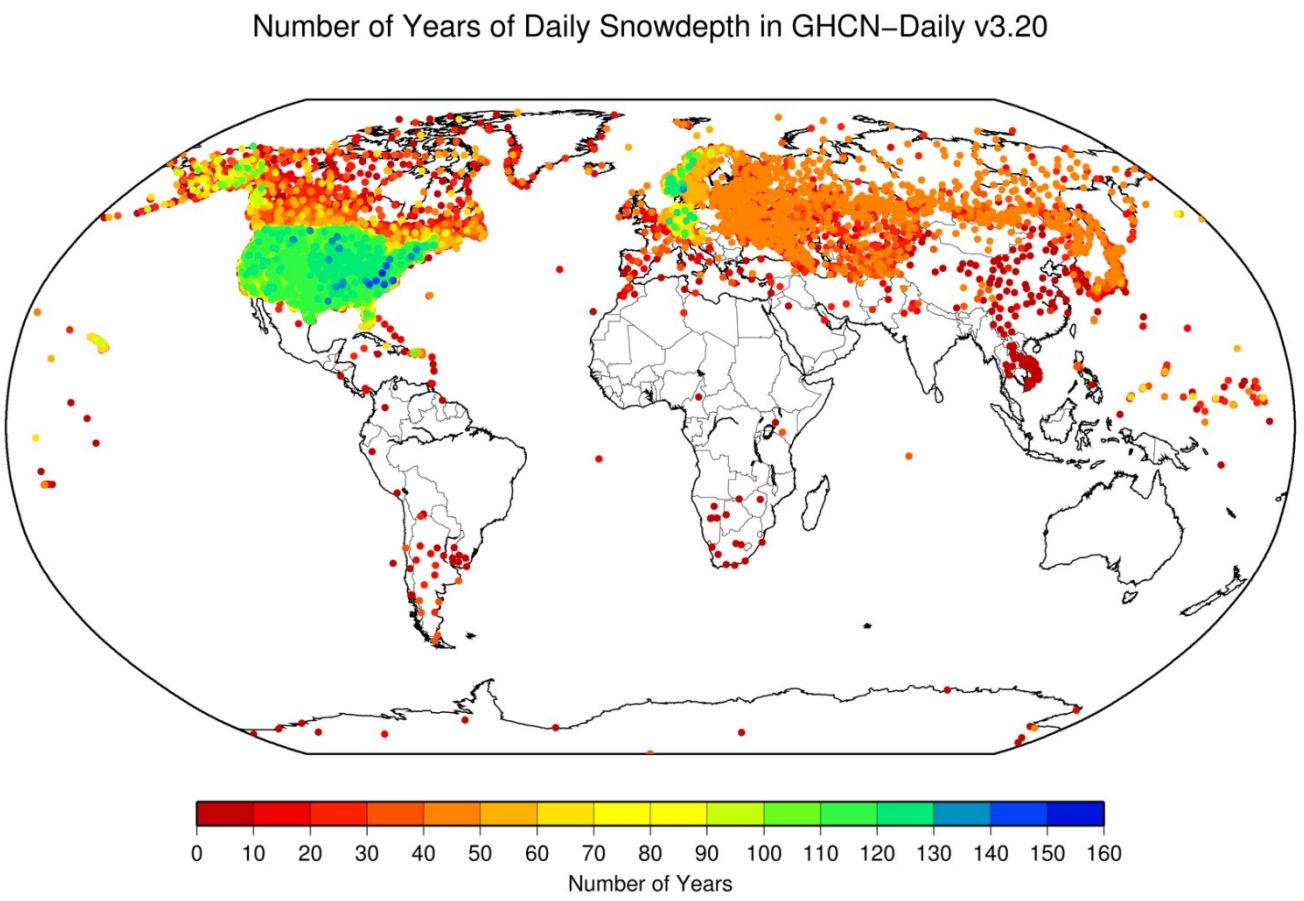
In situ datasets at NCDC: Global Historical Climatology Network-Daily

- U.S. and global daily *in situ* dataset derived from multiple (~30) sources
- Comprehensive daily dataset for the USA with good coverage over many other parts of the world
  - Integrates latest U.S. daily source archives and real-time updates for many U.S. Networks as well as Canada. GTS updates for non-U.S. sites in some cases. Monthly updates for Australia, ECA&D sites.
  - Nexus for U.S. daily data processing at NCDC
- >30,000 temperature stations
- >90,000 precipitation stations
- >40,000 snowfall or snow depth stations
  - Used ISD's "Global Summary of the Day" as a source for snow depth observations outside the U.S. and Canada

Number of Stations by Year by element

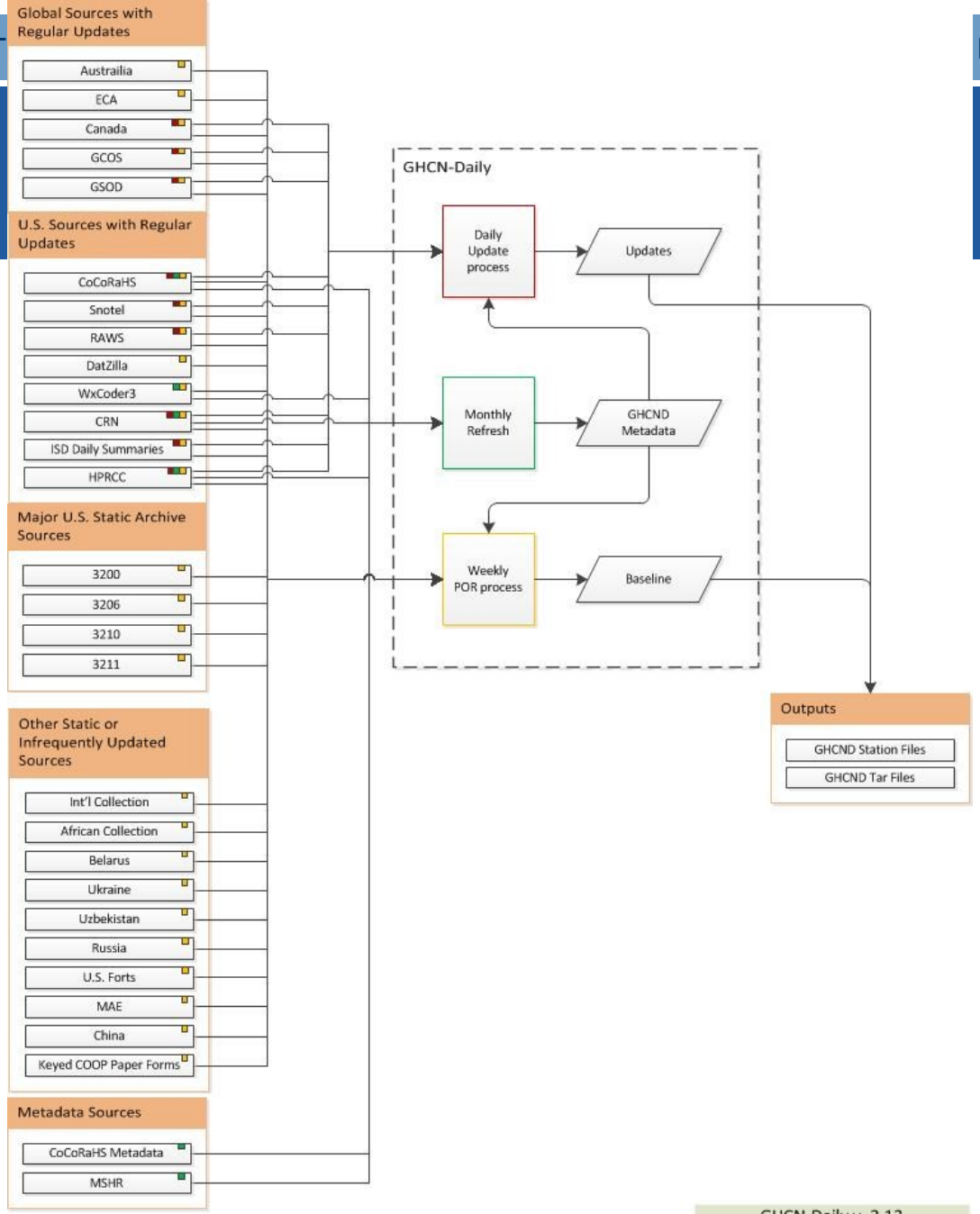


Snow depth record length



Updates and Maintenance: GHCN-Daily

- Data are updated each day
  - for >20,000 stations from several sources
  - ~35,000 stations have data for 2014
- All historic sources are “reprocessed” weekly
  - Dataset is completely reassembled each weekend from primary sources to maintain consistency between each archive source and the integrated dataset
- QC checks are applied to entire period of record with each reprocess
- “Diff” files (deletes, edits, inserts) created each day to help in keeping databases synchronized with GHCN-Daily (inserts, deletes, changes).
- Strict versioning of dataset (all versions are archived)



GHCN-Daily v. 3.12
Overview Level 0 Flow Diagram
Version 1.0
Modified: September 24, 2014

Vertical Integration

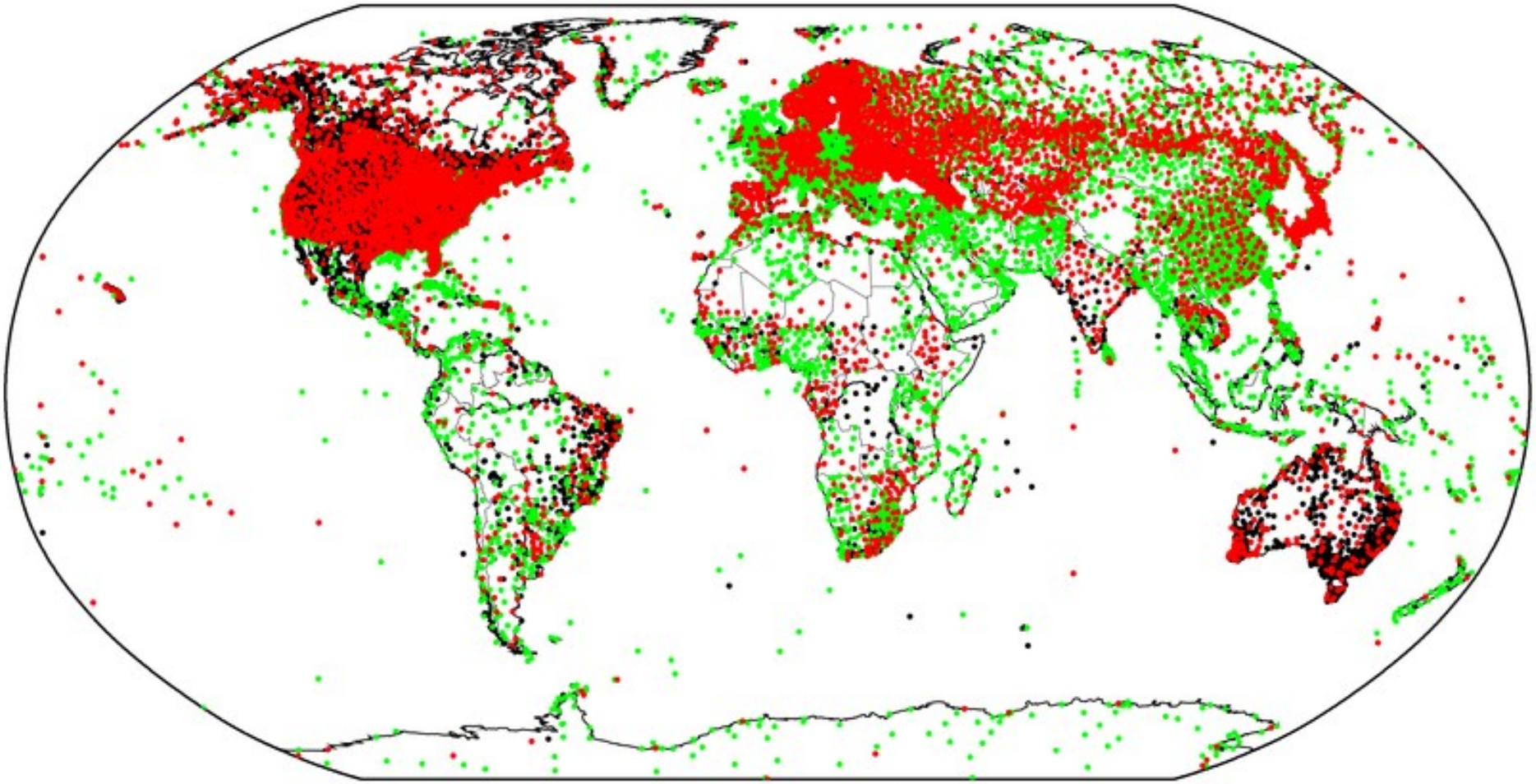
- Progress currently being made to vertically integrate monthly and daily datasets (the two time intervals historically followed separate, disconnected evolutions)
- Better vertical integration also planned for ISD/synoptic dataset
- A likely outcome is to have three vertically integrated datasets (with common id systems etc.) sorted by the time resolution (hourly/synoptic; daily; monthly) perhaps under a broader international umbrella

Plans for data additions

- Russian snow database
- ISD/Global Summary of the Day snow depth observations
- Others identified at this workshop?

- Preferred paradigm is to have access to deep database for regular re-synchronization and to provide automated updates

GHCN–Daily Merged with ISD–Lite



- Existing GHCN-Daily Temperature Station without an ISD-Lite data match
- Existing GHCN-Daily Temperature Station with an ISD-Lite data match
- Potential new GHCN-Station from ISD-Lite

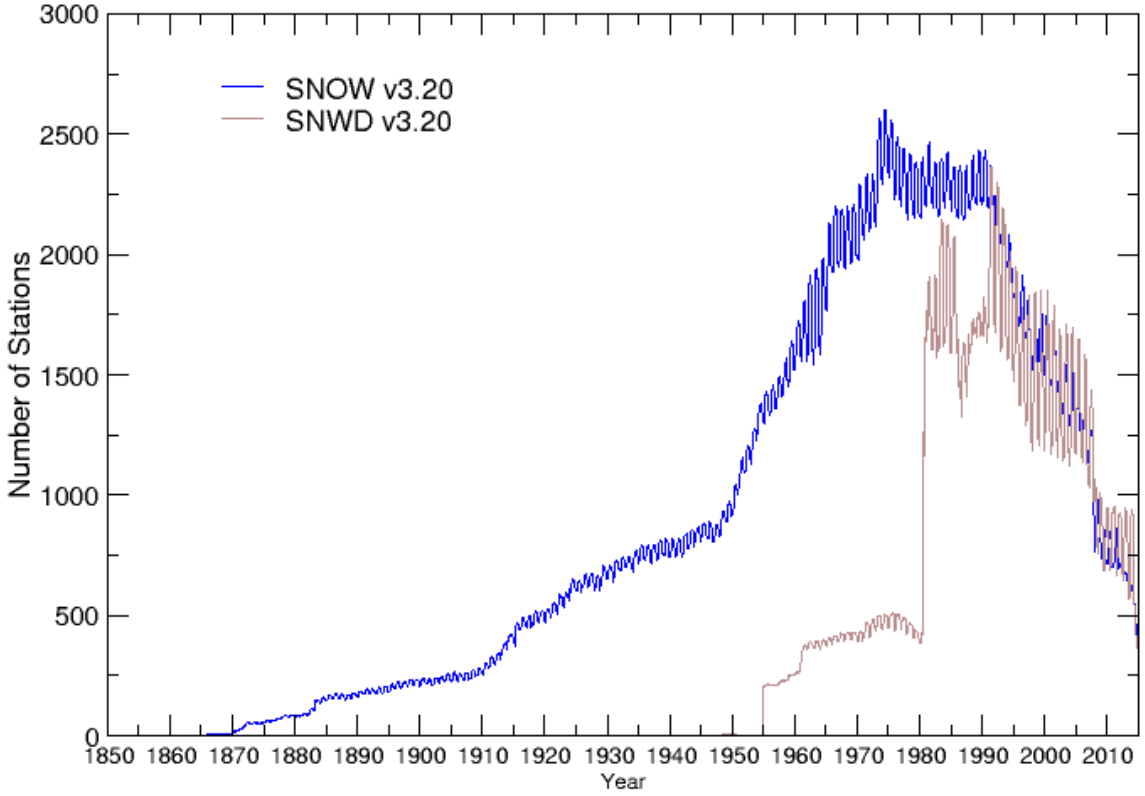
Sources of Snow Data in GHCN-Daily

- U.S. Cooperative Observer Network
  - Volunteer, once per day observations. Dates to 19th Century--many thousands of sites. Snowfall and snow depth reports.
- U.S. City and Airport Sites
  - Met Service or contract personnel--also once per day for snow depth. Hundreds of locations, some dating to 19th Century. Snowfall, snow depth and some snow water equivalent
- SNOTEL
  - Used for water resource management in the Western U.S. Currently about 600 sites in GHCN-Daily with earliest observations dating to 1970. Snow water equivalent rather than snow depth.
- CoCoRaHS
  - New volunteer network in the U.S. Dates to 1998 in Colorado and became nationwide in the 2000's. Recently expanded into Canada. Once per day, includes snowfall, snowdepth, snow water equivalent in some cases. About 10,000 observations per day. Short records.

Sources of Snow Data in GHCN-Daily

- Environment Canada
  - Daily database shared with NOAA as part of a long-standing bi-lateral agreement
  - (See next slide for inventory)
  - GHCN-Daily includes snowfall and snow depth
- European Climate Assessment & Dataset (ECA&D)
  - Provide special feed for GHCN-Daily. Many long records in Nordic countries. Snow depth only.
- Global Summary of the Day
  - Taken from synoptic snow reports transmitted over the GTS. Snow depth only
  - Tends to be incomplete with higher error rates than in other sources

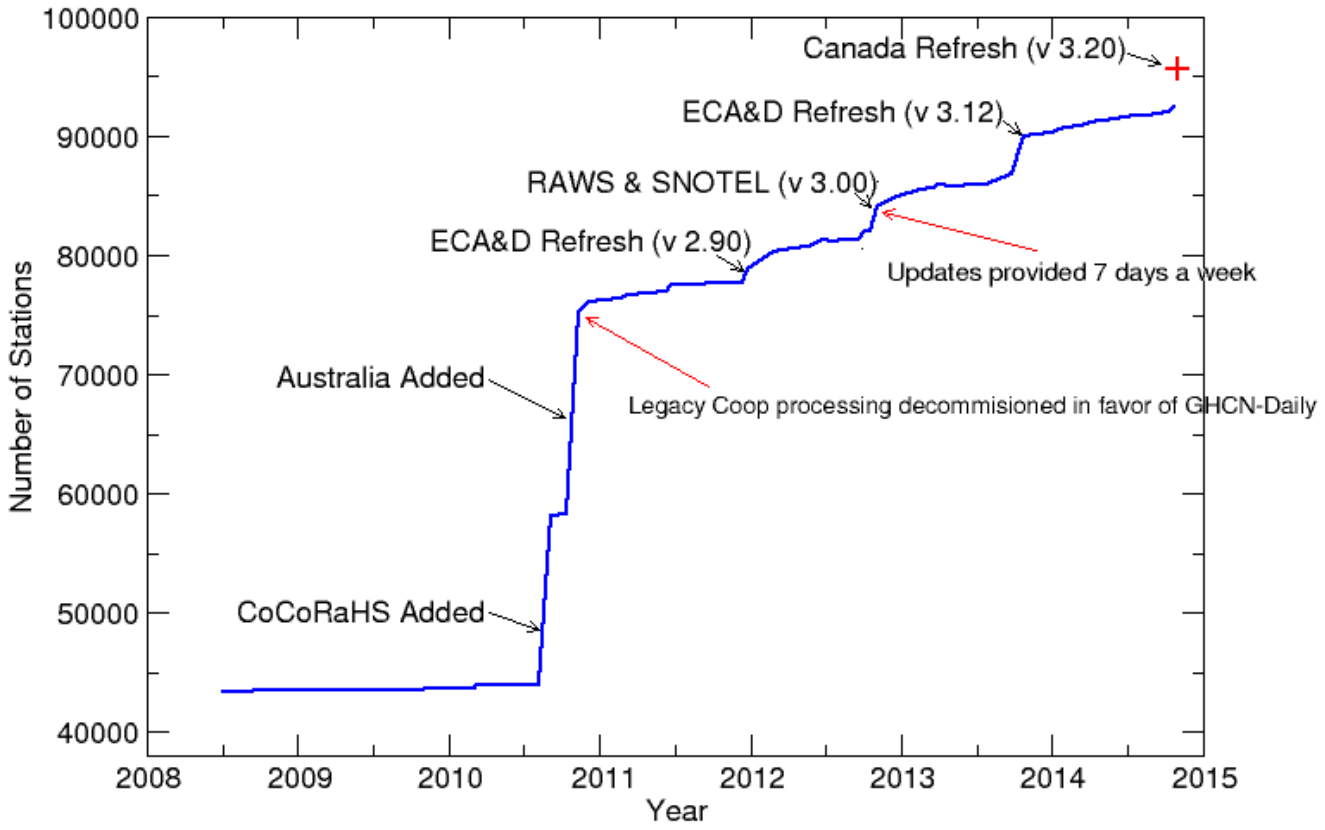
Daily Snow Observations from Environment Canada (GHCN-Daily v3.20 to be release on December 2)



Four facets to GHCN-Daily Processing

1. Daily updates (automated) – Updates Data values
   - Updates values that are new since yesterday's update
2. Weekly reprocessing (automated) – Updates Data Values
   - Reintegrates source databases and reruns quality checks on all values. Helps to ensure that GHCN-Daily is synchronized with its external constituent sources, but does not add new stations
3. Monthly refreshes for select U.S. networks (automated, but requires approval/manual intervention to deploy refreshed list) – Updates Membership in GHCN-Daily
   - Removes and reintegrates active data sources for Coop, CoCoRaHS and CRN. Adds stations that are new since last monthly refresh
4. Periodic adding of new sources or refreshing of large existing data sources (semi-automated) – Updates Membership in GHCN-Daily
   - Removes and reintegrates large data sources to incorporate station additions since last refresh

Historical perspective on major source adds and "refreshes"



Quality Assurance

- 19 different checks most of which can be applied to each of the major five elements (when appropriately tailored)
- Low false positive rate overall (i.e., very limited “collateral damage”)
- Total flag rate equal to approximately 0.24% of all values (highest flag rates for snowfall and snow depth). 1-2% of the flags are estimated to be false positives (i.e., valid values flagged as bad)
  - System can be run “unsupervised”
- Uniform QC for full period of record

Quality Assurance

- **Types of checks:**
  - **Basic integrity**
    - **(Duplicated data, impossible values, etc.)**
  - **Outlier**
    - **(Unrealistic gap in distribution, etc.)**
  - **Internal and Temporal Consistency**
    - **(snow depth increase compared to snowfall/precipitation)**
  - **Spatial Consistency**
    - **(Too warm to snow)**

(Described in *Durre et al. 2010*)

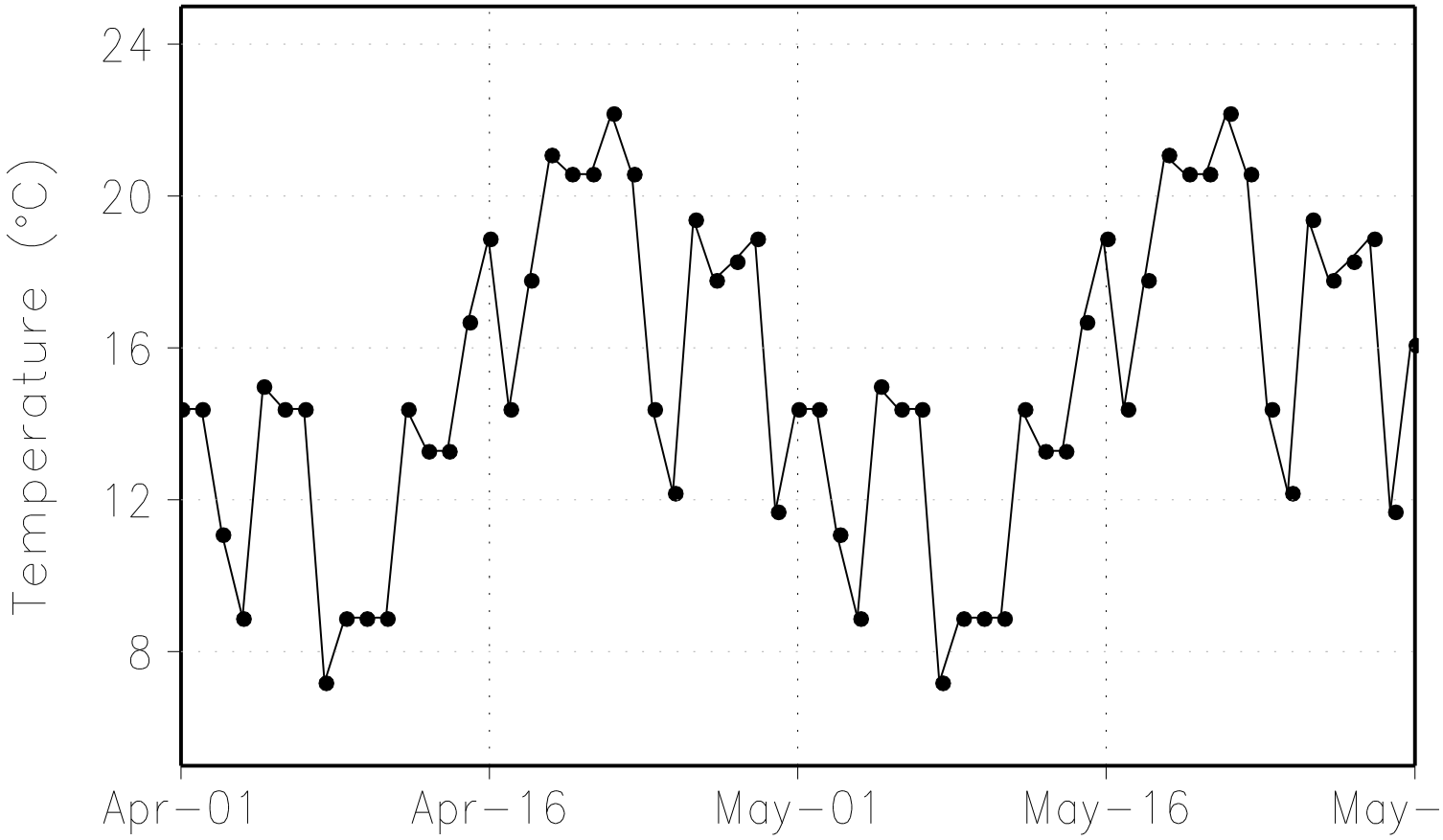
Thank You!

For further information:

- http://www.ncdc.noaa.gov/oa/climate/isd/
- http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/
- Matthew.Menne@noaa.gov

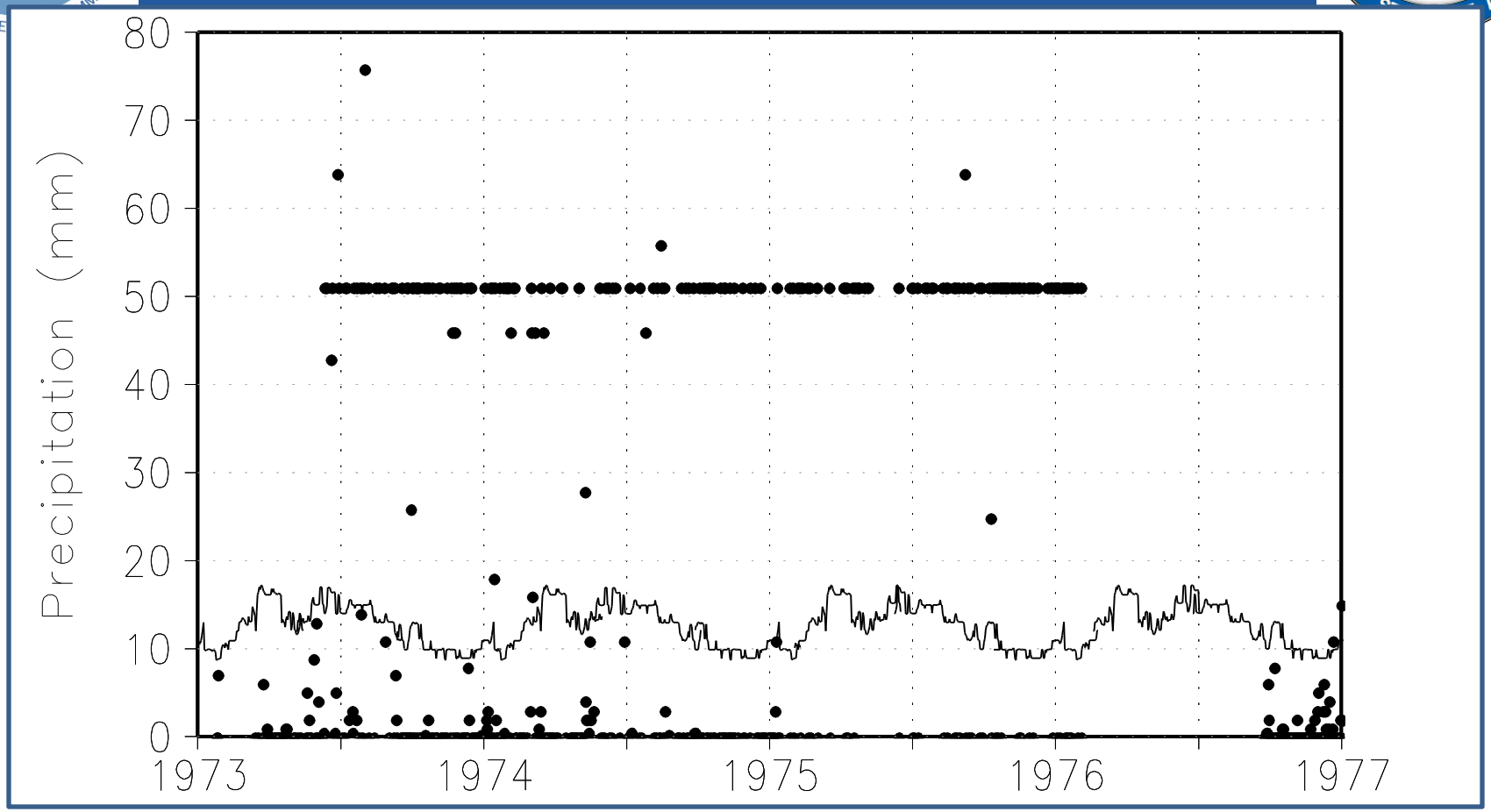
Extra Slides

Example: Duplicated Data



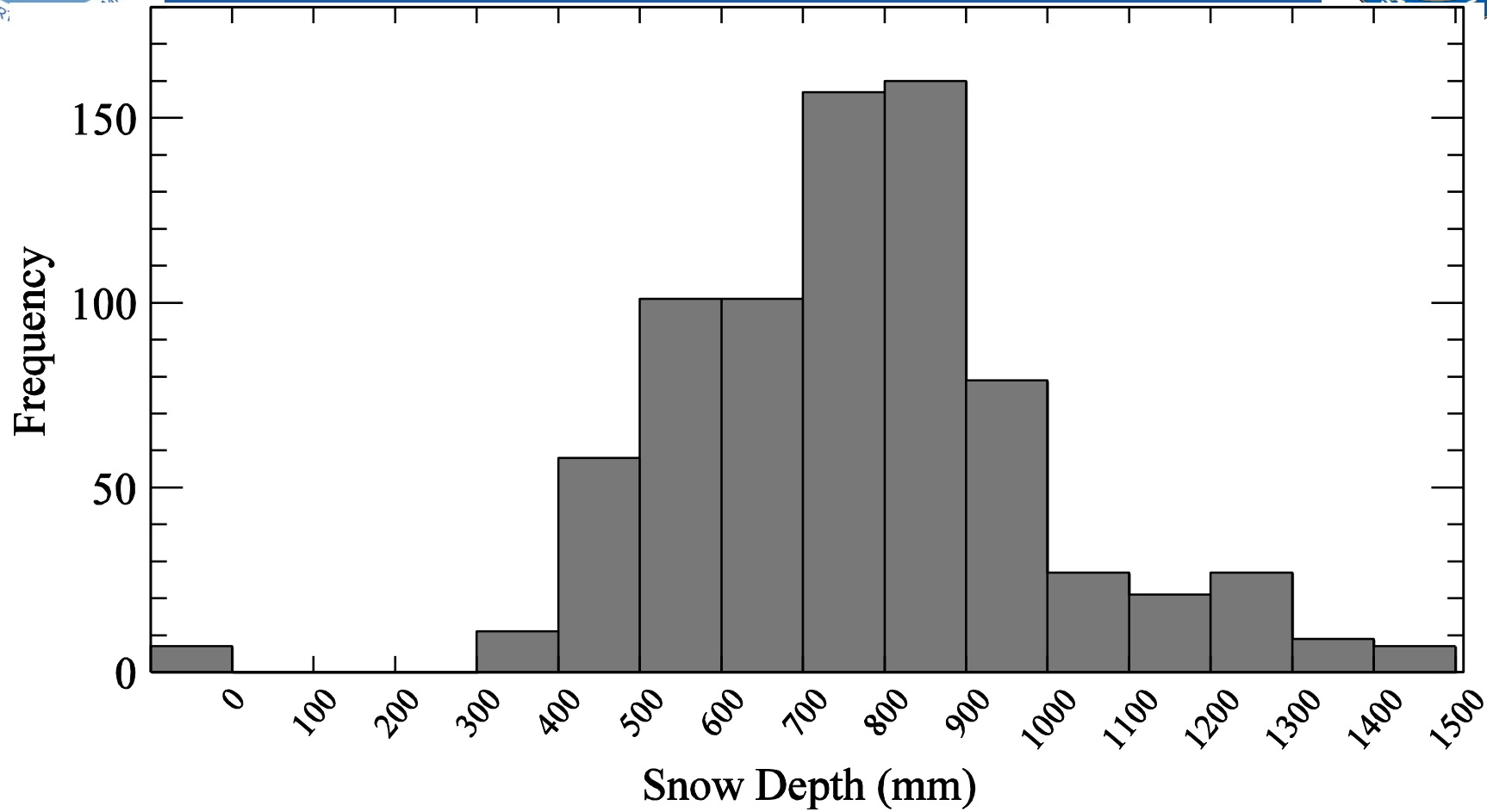
Daily maximum temperatures during April and May 1967 at Lardeau, Canada (GHCN-Daily station ID = CA001144580), showing an example of data duplication identified by the duplicate check comparing data from different months within a year

Example: Frequent Value



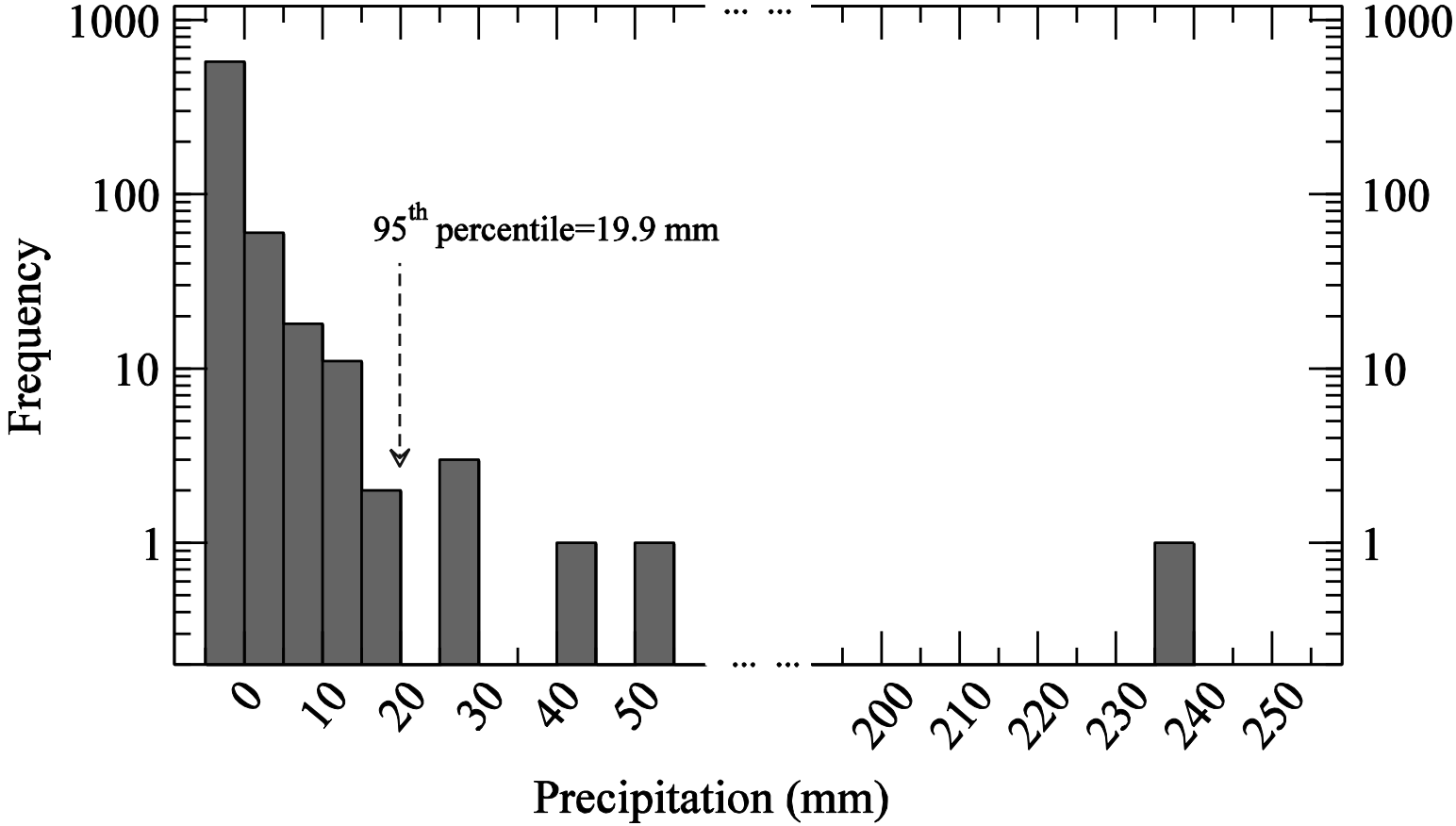
Time series of daily precipitation totals (solid line) during 1973-1976 at Balmaceda, Chile (GHCN-Daily station CI000085874), containing 162 values of 51.1 mm that are flagged by the frequent-value check

Example: Gap in distribution



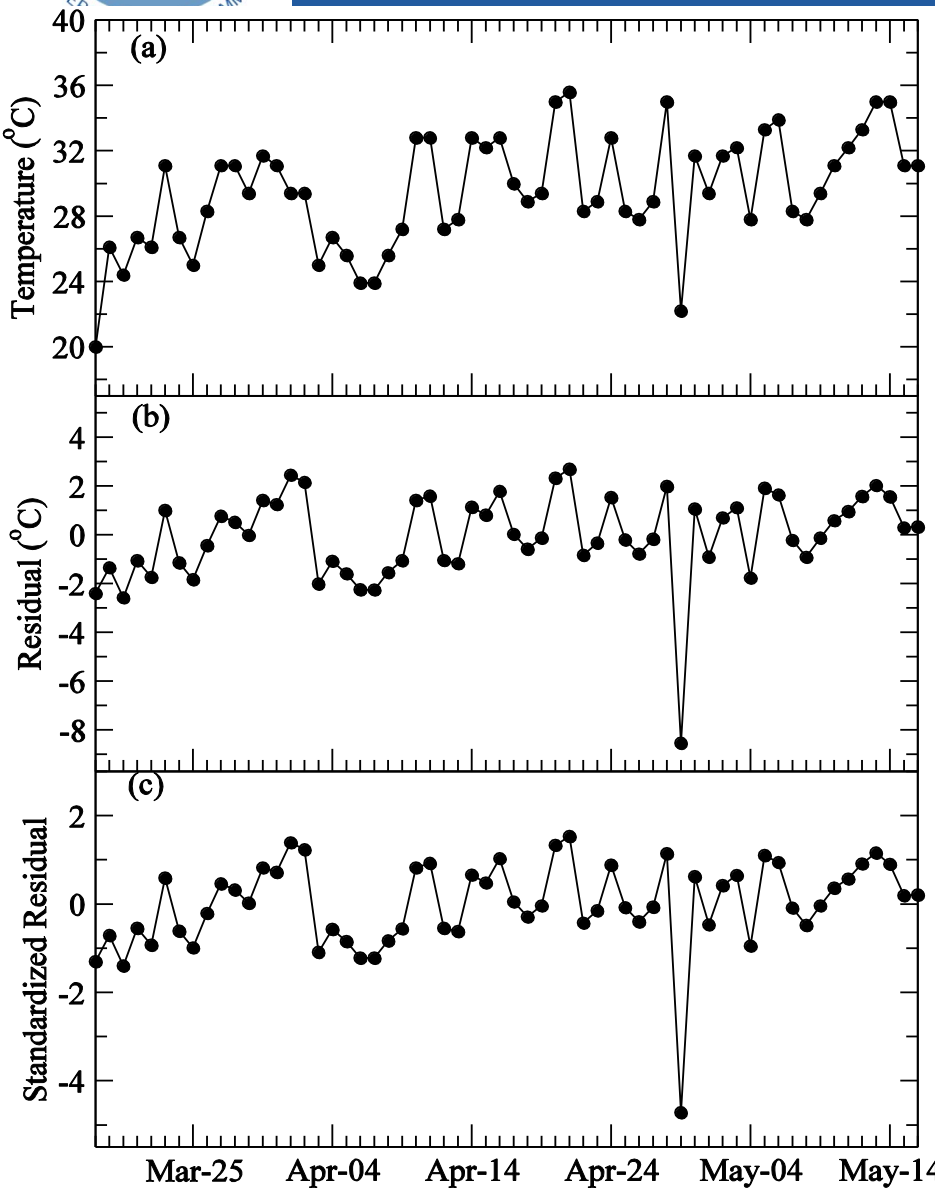
Histogram of all daily snow depths observed in March during the period of record (1975-2008) at Paxson, Alaska (GHCN-Daily station USC00507097), illustrating a data problem identified by the gap check (Table 2). The values of zero (reported in March 1982, 2004, and 2007) are flagged because they differ from the next lowest value ever reported in that calendar month by more than the threshold of 350 mm.

Example: Climatological Outlier



Histogram of daily precipitation totals reported between August 6 and September 3 throughout the 1966-1990 period of record at Gold Hill, Utah (GHCN-Daily station USC00423260), showing an outlier flagged by the percentile-based climatological outlier check

Example: Spatial Outlier

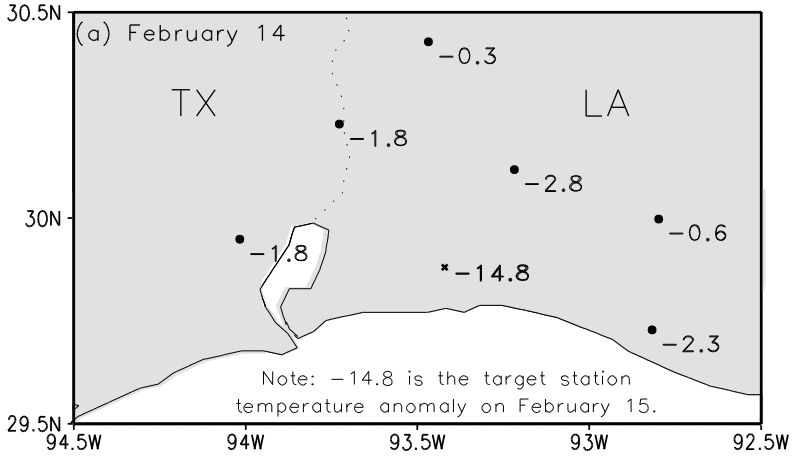


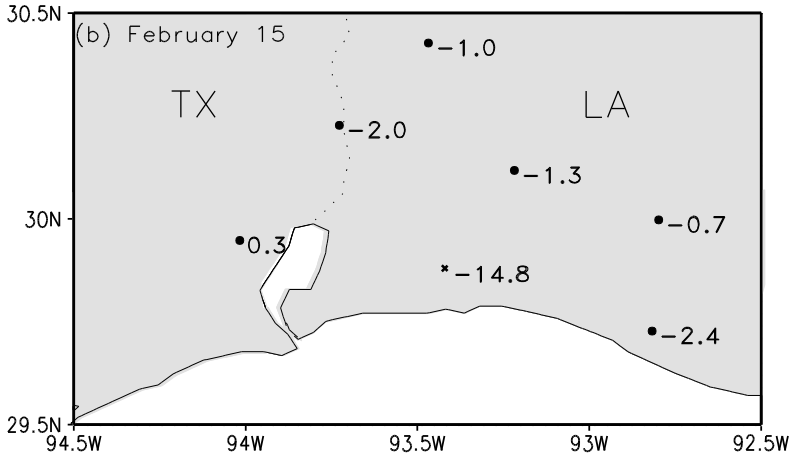
Time series containing a temperature flagged by the spatial regression check

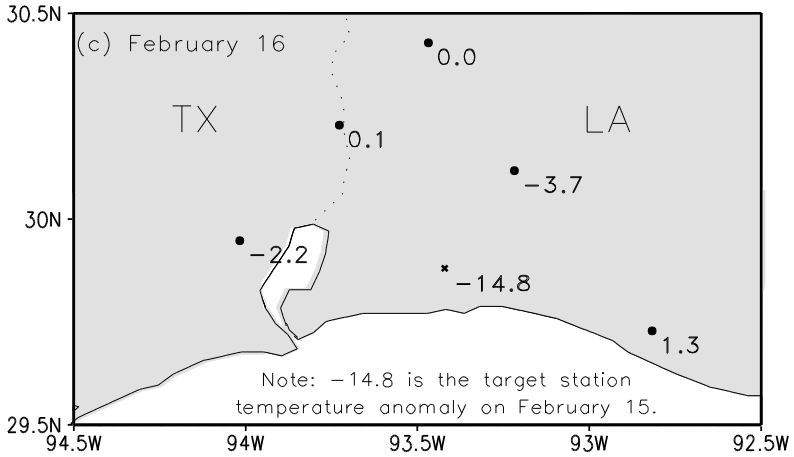
(a) Daily maximum temperatures at Bracketville, Texas (GHCN-Daily station USC00411007) between 17 March and 15 May, 1991;

(b) the corresponding residual time series; and;

(c) the time series of standardized residuals. The temperature of 22.2°C on 28 April is flagged because the residual and standardized residual on that day are greater than 5°C and 4.0 standardized units, respectively





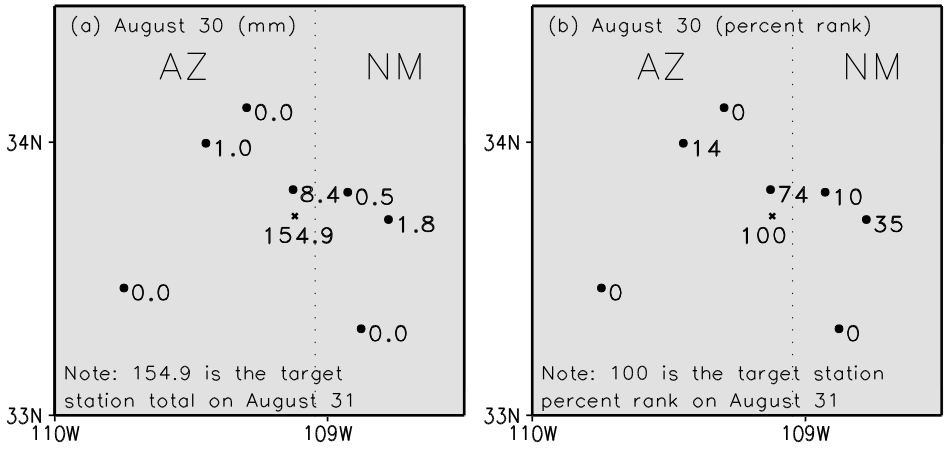


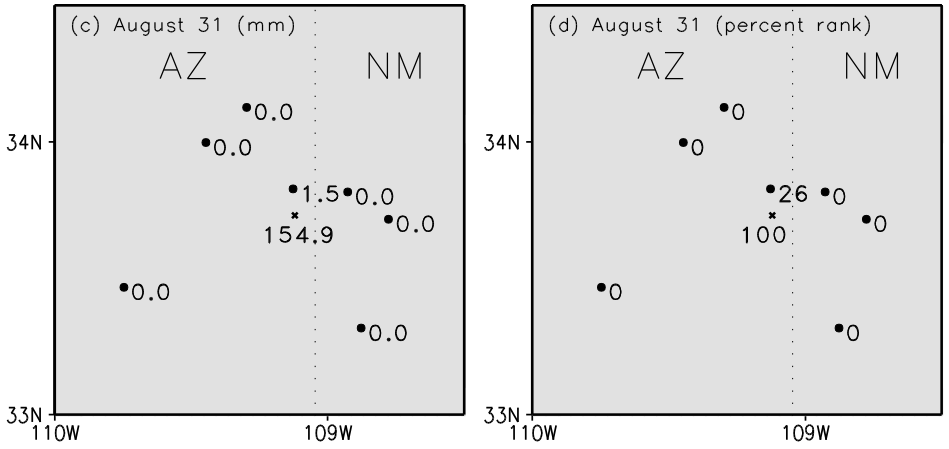
MOSPHERIC ADMINISTRATION

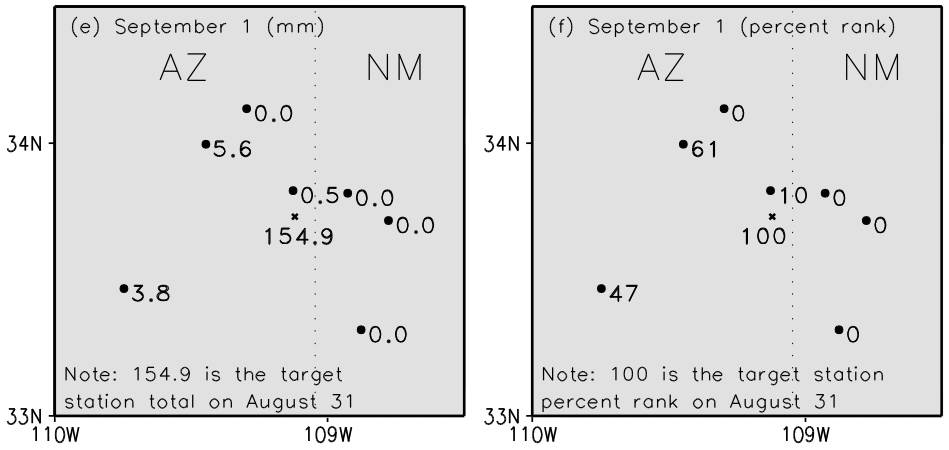
Example: Spatial Corroboration

Maps illustrating the spatial corroboration check on temperature. Shown are the daily minimum temperature anomaly at Hackberry, Louisiana (GHCN-Daily station USC00163979), on 15 February 2002 and the daily minimum temperature anomalies to which this "target value" is compared:

(a)the six available neighbor anomalies on day -1 (14 February);
(b)the six neighbor anomalies available on day 0 (15 February); and
(c)the five neighbor anomalies available on day +1 (16 February). The target value is indicated by an X symbol in each panel, the neighbor values by filled circles. The target anomaly of -14.8°C is flagged because it is 11.1°C lower than the coldest temperature anomaly among the neighbor values within the three-day window.







SPHERIC ADMINISTRATION

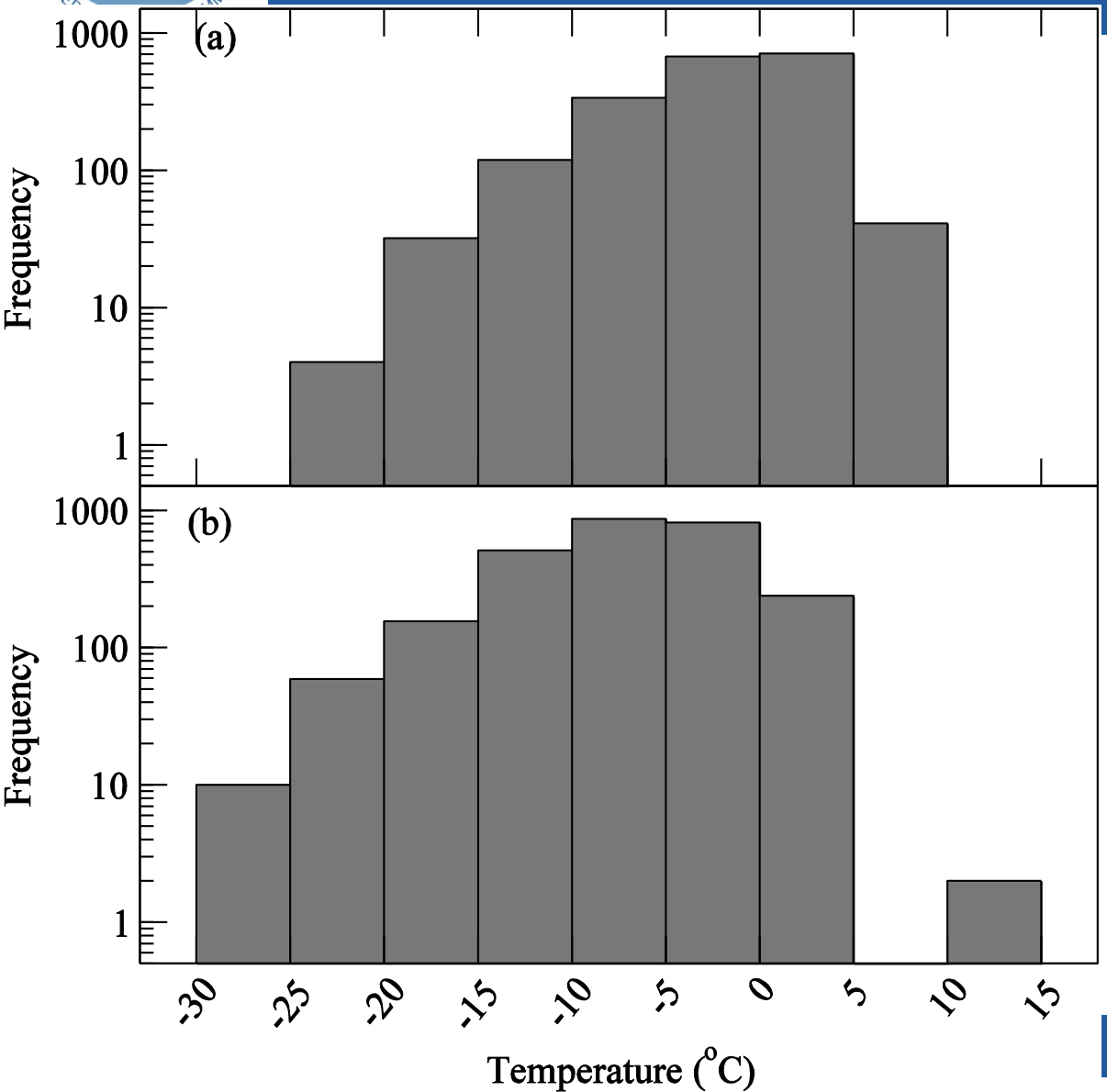
Example: Spatial Corroboration

Maps illustrating the spatial corroboration check (Table 4) applied to a 154.9 mm precipitation total at Alpine, Arizona (GHCN-Daily station USC00020174), on 31 August 1996. In addition to this target total or its percent rank (X symbol), the maps show all neighbor information (filled circles) used in the check:

(a) neighbor totals on day -1;
(b) neighbor percent ranks on day -1;
(c) neighbor precipitation totals on day 0;
(d) neighbor percent ranks on day 0;
(e) neighbor totals on the day +1; and
(f) neighbor percent ranks on day +1.

The minimum absolute target-neighbor percent ranked difference is 26, yielding a test threshold of 120.3 mm (Appendix C). The target value is flagged because the minimum absolute target-neighbor difference among totals is 146.5 mm and therefore exceeds the threshold.

Example: Megaconsistency



Histograms of all January
(a) daily maximum temperatures and
(b) daily minimum temperatures

reported at Jan Mayen, Norway (GHCN Daily station JN000099950), illustrating the extremes megaconsistency check . The 10.3°C and 10.6°C TMINs (both reported in January 1929) are flagged by the check because they exceed the highest unflagged January TMAX (9.5°C) reported during the station's 1921-2009 record.