ISTI – Daily data
Support to Monthly Databank and Multi-element datasets
10 October 2014

The Databank WG released Version 1 of the ISTI Monthly Temperature Databank in June 2014. This marked the culmination of more than three years of work in bringing together collections of source data and developing and implementing a merging algorithm, a data provenance methodology, and a framework for dissemination. With the databank established, attention in the coming year will be focused on maintaining the monthly data and releasing an updated version containing new sources of data to address temporal and spatial gaps. Such enhancements will support monitoring and assessment, particularly in areas where climate change is occurring at a rapid pace and there is a need for better in situ coverage. This will also aid the Benchmarking Working Group in their efforts to assess and improve bias correction methodologies.

Although some source data are only available as summary of the month values, the majority of the version 1 monthly databank originates from summary of the day observations. And it is expected that enhancements to the monthly databank in 2015 can be made largely through the addition of new sources of summary of the day data. Of the more than 30 sources of daily data in the version 1 monthly databank, most observations are provided by the Global Historical Climatology Network-Daily (GHCN-Daily) dataset. This dataset contains more than 30,000 stations with maximum and minimum temperature. It also contains more than 90,000 stations with daily precipitation and more than 40,000 with snow observations.

The GHCN-Daily dataset was developed over the past decade using a well-established merge algorithm that is analogous to the databank's monthly merge algorithm in that it uses station metadata and data matching. Unlike the databank, GHCN-Daily makes use of multiple elements in data match assessments and source merging. GHCN-Daily also has a large number of automated update streams. Consequently, the work of establishing an integrated databank of merged summary of the day temperature data has largely been accomplished. Additionally, maintaining the integration of elements in addition to temperature makes the database more broadly useful to the climate community than one containing temperature alone.

Going forward NCDC plans to continue to support the ISTI databank effort through further development of GHCN-Daily, specifically by adding unexploited data sources of daily temperature and elements. Consistent with the GHCN-Daily paradigm, the new sources will be selected to prioritize data sparse regions and sources that are multi-element and that provide automatable updates via web services or ftp. The addition of new daily sources in GHCN-Daily will improve coverage of the monthly stage 3 temperature data with each new monthly databank version release.

Some of the sources are in GHCN-Daily are redundant to the current databank, but most are not. Nevertheless, the monthly merge process will benefit from a closer coordination between the sources incorporated into GHCN-Daily and those used separately in the monthly merge. Moreover, GHCN-Daily can be made to more closely reflect the staged data processing adopted by ISTI. Stage 1 and Stage 2 versions of GHCN-Daily sources already takes place internal to NCDC as part of the GHCN-Daily processing system, but these stages can also be provided to external users. Providing the Stage 1 and Stage 2 daily data permits any other Center to develop an alternative merge algorithm if so desired. For the time being, the merged Stage 3 daily data will exist within the GHCN-Daily dataset. Unless there is a compelling need to separate the temperature data into a unique single element data set, no effort will be made to create a Stage 3 databank for daily temperature data.

Specifically, the proposed sources with the highest priority for addition in GHCN-Daily during the coming year are the following:

1. Antarctica-aws (100+ stations from Automatic Weather Station Program from AMRC)
    a. Is automatable and has multi-element (fulfills the GHCN-Daily model)
    b. Adds data in (obviously) data sparse region
2. Argentina (35 stations provided by Matilde Rusticucci)
    a. The hope is that these can be updated on a routine basis through ftp or web services transfer. The addition of these stations may not be desired if such a system cannot be established.

3. Daily averages derived from hourly data in NCDC's Integrated Surface Daily (ISD) dataset. This will also support the monthly databank by augmenting data received via monthly CLIMAT messages.
    a. This will require the addition of a TAVG element to GHCN-Daily and the exploitation of hourly data to derive daily and monthly averages. There is potential to greatly enhance the number of stations with real-time updates for climate monitoring and avoid delays associated with slow or problematic receipt of CLIMAT messages.

Finally, it should be noted that, ultimately, the work of building an internationally recognized in situ database will likely need to come under a coordination umbrella that extends beyond ISTI to ensure better coordination with other data collection efforts (e.g., precipitation, pressure, snow).