An Open and Transparent Databank of Global Land Surface Temperature

Jared Rennie[1], Peter Thorne[2], Jay Lawrimore[3], Byron Gleason[3], Matt Menne[3], Claude Williams[3]

[1] Cooperative Institute for Climate and Satellites – NC, Asheville, NC, USA | [2] Nansen Environmental and Remote Sensing Center, Bergen, Norway | [3] NOAA's National Climatic Data Center, Asheville, NC, USA

Introduction

The instrumental record of temperature has its roots in the development of universal temperature scales in the early 18th century. By the 1900s, measurements expanded across the entire globe. In the 1980s and 1990s, major efforts were made to collect consolidated global datasets, and that became the foundation for understanding trends in surface temperature:

	Land	Land + Ocean
NCDC	GHCN-M	MLOST
UK Met Office	CRUTEM	HADCRUT
NASA	GISS	GISS

While these and other datasets met needs for our understanding of global climate change, many data limitations and metadata deficiencies still exist, including:

- Insufficient global coverage prior to 1950
- Poor data provenance
- Limited data accessibility
- Additional sources of data not digitized
- Incomplete metadata outside of the U.S.

In 2010, the International Surface Temperature Initiative (ISTI) was created to address these known issues. Current activities include data rescue, digitization of imaged records, data provenance, version control, and benchmarking. ***The first milestone of this initiative included creating a single, comprehensive global databank of surface meteorological observations.***

The databank has been constructed and is available in six stages. The initial focus is on temperature data on daily and monthly time scales, although other elements and time scales will be added later.

More Info / Contact

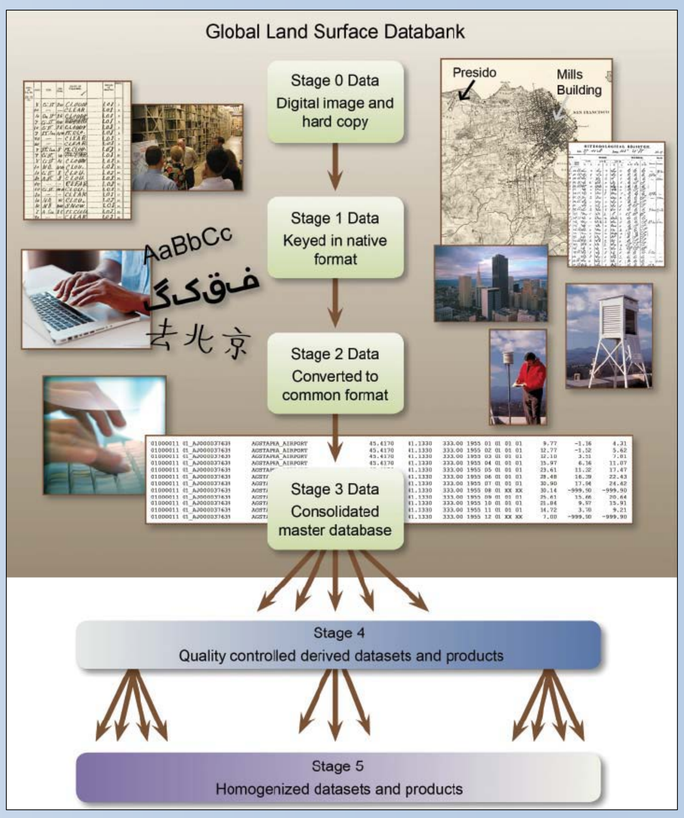
WEBSITE
www.surfacetemperatures.org

DATABANK LOCATION
ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/

GENERAL COMMENT?
general.enquiries@surfacetemperatures.org

HAVE A LEAD ON DATA?
data.submission@surfacetemperatures.org

Databank Design

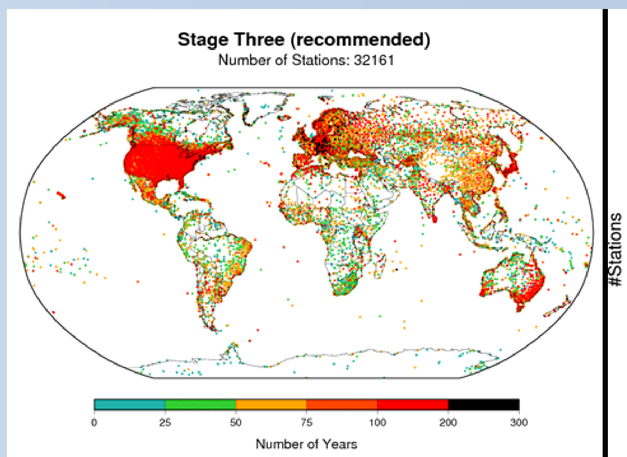


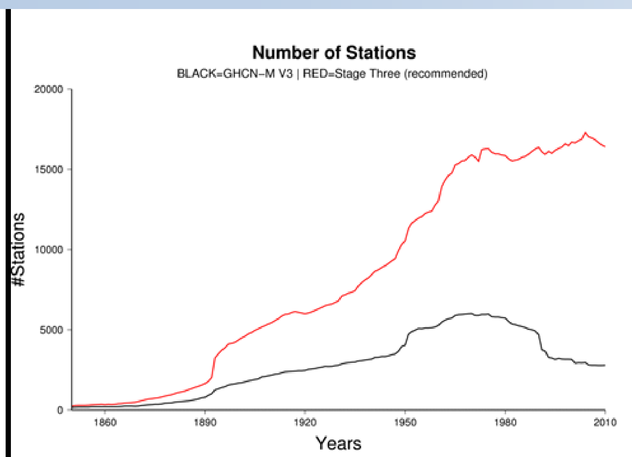
Over 50 sources of data are gathered in hard copy (Stage Zero) and/or its native digitized format (Stage One) and converted to a common format (Stage Two). Provenance tracking flags are provided to document as much of the history of each observation as possible. These Stage Two sources then go through an algorithm to merge stations together and remove duplicate information (Stage Three). Other organizations are encouraged to start with the recommended merged product to create their own suite of quality controlled (Stage Four) and bias corrected (Stage Five) datasets.

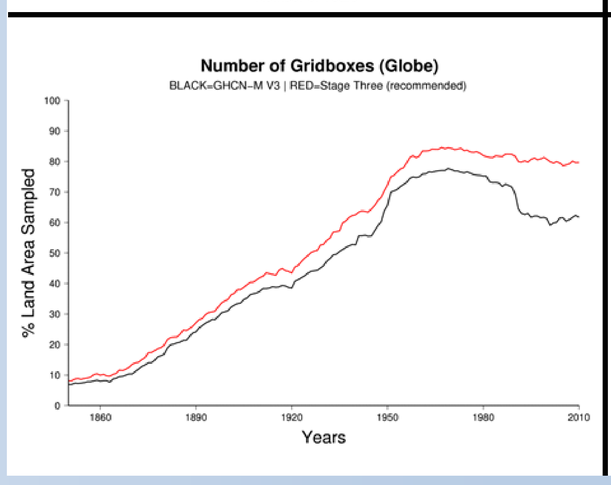
The merging algorithm is probabilistic in approach and contains metadata matching and data equivalence criteria. It is an iterative process between a target source and all other candidate sources. Once a prioritized list of sources is generated, each candidate station is compared to all target stations, and one of three possible decisions is made:

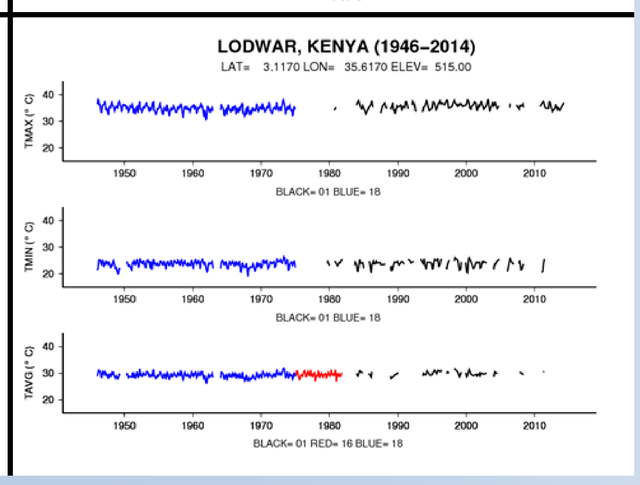
- Candidate station should ***merge*** with a target station
- Candidate station is ***unique*** and is added to target source
- There is not enough information, and the station is ***withheld***

Recommended Stage Three Product









The final Stage Three product, recommended and endorsed by ISTI, contains nearly 32,000 stations. More than four times the amount currently in GHCN-M (7,280).

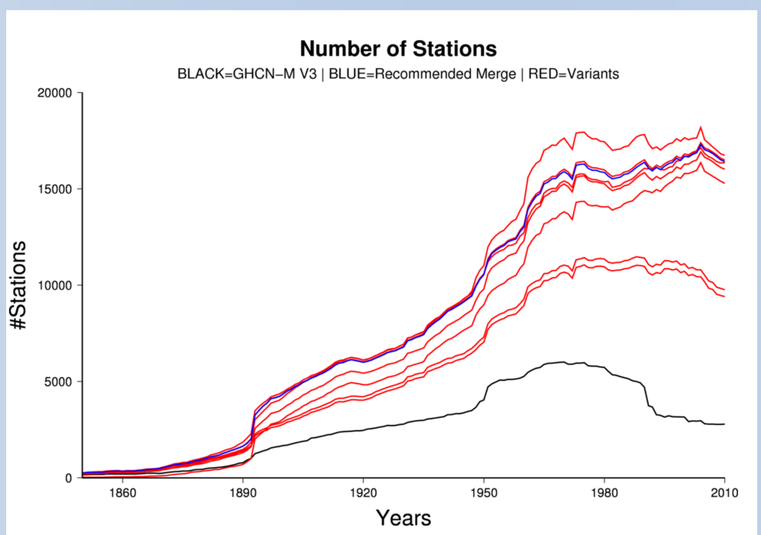
Not only do we see an increase in the amount of stations after 1850, there is a net increase of the available 5° X 5° grid boxes that contain land, especially past 1990.

ISTI aims to be open and transparent in all their processing. All data and code are provided, and users are encouraged to evaluate the methods that we have tried, establish their own methods, run the source code available online, and make results available for comparison.

Stage Three Variants

A lot of subjective decisions go into the merge algorithm. In order to characterize the uncertainty, seven variants were created along with the recommended merge:

Variant	Source Deck Change?	Threshold Change?	Code Change?
1	Prioritize sources from NMA's	no	no
2	NMA's with TMAX and TMIN given highest priority	overlap_threshold changed from 60 months to 24 months	no
3	No TAVG sources used, rest ranked by order of longest station record present	Thresholds to merge and unique stations are lowered to merge more stations	metadata_probability weighted to favor distance_probability over all others
4	no	no	During data comparisons, candidate station only merged or unique
5	All homogenized sources removed	no	no
6	no	All thresholds adjusted to make more candidate stations unique	no
7	no	All thresholds adjusted to make more candidate stations merge with target stations	no



Version 1 release

The first version of the databank holdings will be released on June 30th at 10 EDT. This version will form the basis for the benchmarking exercise under ISTI.

Investigators are strongly encouraged to download and analyze this first version release and feedback any issues as well as taking part in the broader ISTI activities.

cicsnc.org
ncdc.noaa.gov
Climate.gov

CICS - NC
Cooperative Institute for Climate & Satellites

NOAA's NATIONAL CLIMATIC DATA CENTER
ASHEVILLE, NORTH CAROLINA, USA