Berkeley Earth Surface Temperature Analysis

David Brillinger[1], Judith Curry[2], Robert Jacobsen[3], Elizabeth Muller[4], Richard Muller[3,4] (chair), Saul Perlmutter[4], Robert Rohde[5], Arthur Rosenfeld[3], Charlotte Wickham[1], Jonathan Wurtele[3]

([1]Statistics, U. Calif. Berkeley, [2]Earth and Atmospheric Sciences, Georgia Inst. Technology, [3]Physics, U. Calif. Berkeley, [4]Muller & Associates, Berkeley, [5]Novim Group, Santa Barbara)

Project Description

The Berkeley Earth Surface Temperature Study has been organized under the auspices of the non-profit Novim Group (www.novim.org). The project has the following goals:

1) To merge existing surface station temperature data sets into a new comprehensive raw data set with a common format that could be used for weather and climate research
2) To review existing temperature processing algorithms for averaging, homogenization, and error analysis to understand both their advantages and their limitations
3) To develop new approaches and alternative statistical methods that may be able to effectively remove some of the limitations present in existing algorithms
4) To create and publish a new global surface temperature record and associated uncertainty analysis
5) To provide an open platform for further analysis by publishing our complete data and software code as well as tools to aid both professional and amateur exploration of the data

Progress

Preliminary Data Set

The Berkeley Earth Surface Temperature Study has created a preliminary merged data set by combining 1.6 billion temperature reports from 10 preexisting data archives (4 daily and 6 monthly). Whenever possible, we have used raw data rather than previously homogenized or edited data. After eliminating duplicate records, the current archive contains 39,390 unique stations. This is more than five times the 7,280 stations found in the Global Historical Climatology Network Monthly data set (GHCN-M) that has served as the focus of many climate studies. The GHCN-M is limited by strong requirements for record length, completeness, and the need for nearly complete reference intervals used to define baselines. We believe it is possible to design new algorithms that can greatly reduce all of these requirements (see section on “Our Proposed Algorithms”), and as such we have intentionally created a more expansive data set.

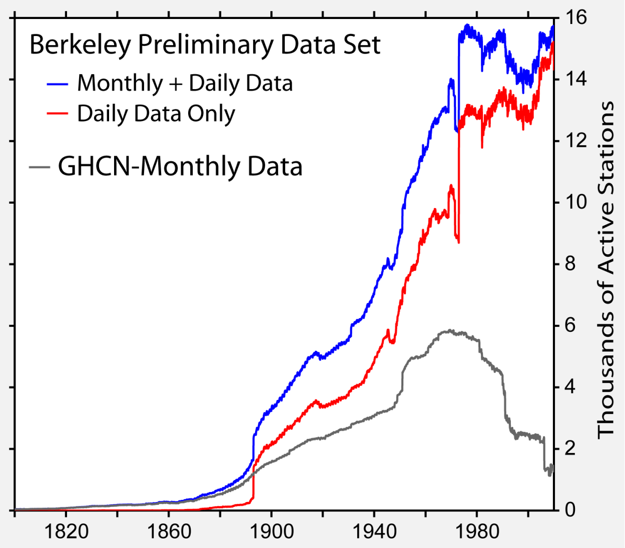

Figure 1: The number of active temperature stations over time in our preliminary data set as compared to that provided by GHCN-M. The decrease in the GHCN-M station count is a consequence of their inclusion criteria, and can be avoided by different methodology.

By taking this approach we are able to double the number of records present during most of the twentieth century and avoid the dramatic falloff in station counts encountered by the GHCN-M in the post-1980 period (Figure 1). It is our belief that the use of additional data can significantly reduce statistical uncertainties and help avoid potential biases in the study of recent climate change. However, we have found very little new data for the pre-1900 period, and believe that the existing analyses already incorporate nearly all of the early records that have been digitized.

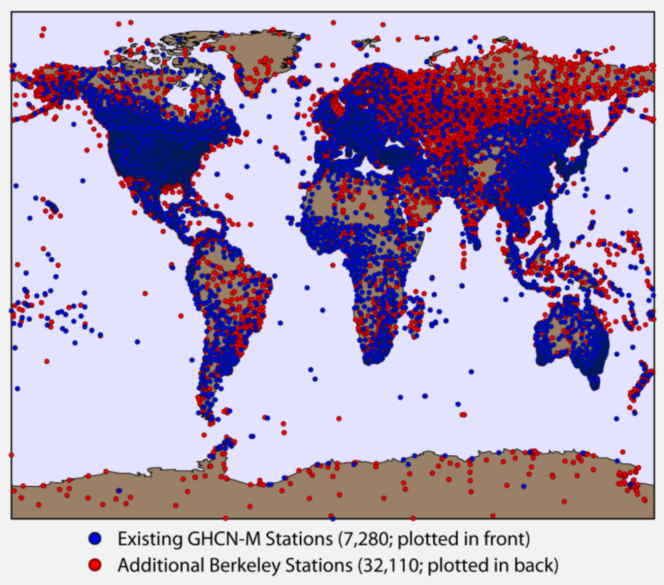
● Existing GHCN-M Stations (7,280; plotted in front)
● Additional Berkeley Stations (32,110; plotted in back)

Figure 2: Comparison of GHCN-M station locations (blue and in front) to the station locations in the Berkeley data set (red and in back). The additional records contribute new coverage most substantially to Asia, South America, and Antarctica. There are also many additional records in the US and Europe but these are obscured due to the already dense coverage in these regions.

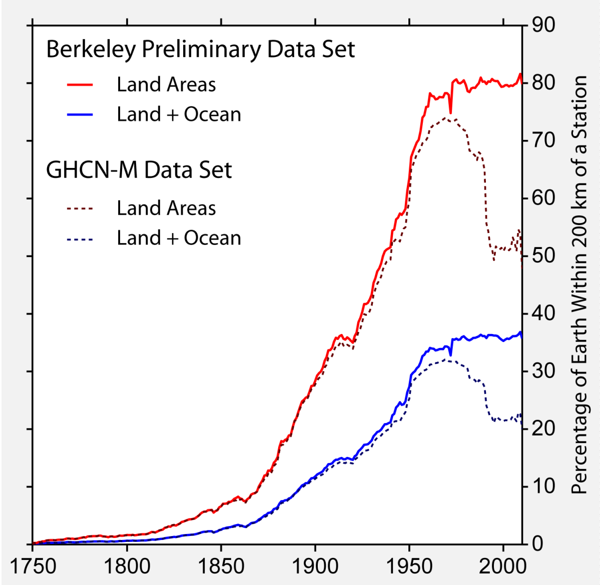

Figure 3: Comparison of the spatial coverage versus time in the preliminary Berkeley data set to that present in the GHCN-M. This is expressed as the fraction of the Earth's surface within 200 km of at least one temperature station. Red lines are calculated using only land areas, while blue lines reflect both land and ocean areas.

As shown in Figure 2, the additional stations in our preliminary data set substantially increase the station density in some regions of the world, such as Asia, South America, Antarctica, and Indonesia. However, while the GHCN-M may be sparse in many parts of the world, it is not entirely absent. Figure 3 shows that until approximately 1980, both GHCN-M and our network sample a similar fraction of the globe's surface area. Some regions, such as the USA, Europe, and China are already densely sampled in the GHCN-M (Figure 2). The primary effect of adding stations is to increase the number of densely sampled regions. We expect that the additional dense sampling can be exploited by homogenization and/or averaging procedures to reduce uncertainties. Lastly, we note that avoiding the GHCN-M decline in station counts (Figure 1) has the ability to avoid similarly abrupt declines in the total spatial coverage (Figure 3).

Our Proposed Algorithms

The central challenge of global temperature reconstruction is to take spatially and temporally diverse data exhibiting varying levels of quality and construct a global index series that is intended to track changes in the mean surface temperature of the Earth. There is no easy answer to this problem, and we believe that there is inherent value in comparing different approaches to this problem as well as understanding the weaknesses intrinsic to any given approach. To this end, we are both studying the existing methodologies for averaging and homogenizing data as well as looking for new approaches whose features seem to incorporate valuable alternatives to the existing methods.

Goals for our algorithms include:

1) Make it possible to exploit relatively short (e.g. a few years) or discontinuous station records. Rather than simply excluding all short records, we prefer to design a system that allow short records to be used with a low – but non-zero – weighting whenever it is practical to do so.
2) Avoid gridding. All three major research groups currently rely on spatial gridding in their averaging algorithms. As a result, the effective averages may dependant on the choice of grid pattern and may be sensitive to effects such as the change in grid cell area with latitude. Our algorithms seek to eliminate explicit gridding entirely.
3) Place empirical homogenization on an equal footing with other averaging. We distinguish empirical homogenization from evidence-based homogenization. Evidence-based adjustments to records occur when secondary data and/or metadata is used to identify problems with a record and propose adjustments. By contrast, empirical homogenization is the process of comparing a record to its neighbors to detect undocumented discontinuities and other changes. This empirical process performs a kind of averaging as local outliers are replaced with the basic behavior of the local group. Rather than regarding empirical homogenization as a separate preprocessing step, we plan to incorporate empirical homogenization as a process that occurs simultaneously with the other averaging steps.
4) Provide uncertainty estimates for the full time series through all steps in the process.

The following equations provide a schematic outline of the approach we are currently pursuing. Our ultimate algorithm will require additional features and modifications to deal with statistical and observational problems not discussed here.

At the core of our methodology, we are expecting to use iteratively weighted least squares to determine effective estimates for the history of global mean temperature. Let $y_i(t_j)$ denote the temperature of the *i*-th station for month t_j after removing seasonality. Further, let $\overline{y_i}$ be the time-invariant baseline temperature for the *i*-th station and $\hat{y}(t_j)$ be the global mean temperature for month t_j. We can then simultaneously estimate $\overline{y_i}$ and $\hat{y}(t_j)$ by minimizing the sum of square differences:

$$SSD = \sum_{i=1}^{stations} \sum_{j=1}^{months} W_i(t_j)\left(\left(y_i(t_j) - \overline{y_i}\right) - \hat{y}(t_j)\right)^2$$

The baseline parameters $\overline{y_i}$ play the role of describing each series as anomalies relative to a standard reference frame, but by making them a part of the fit we avoid the requirement that any single interval of time be present for all temperature records.

In the above sum, $W_i(t_j)$ is a weighting function based on the spatial and statistical characteristics of the data. We consider it to have two natural parts, a spatial part and an uncertainty part, such that $W_i(t_j) = S_i(t_j)U_i(t_j)$.

If we choose any spatial interpolation function $w(\vec{a}, \vec{b})$ such that for all locations on the Earth, $\vec{x}$, we accept that:

$$y(\vec{x}, t_j) = \frac{\sum_{i=1}^{stations} w(\vec{x_i}, \vec{x})\, y_i(t_j)}{\sum_{i=1}^{stations} w(\vec{x_i}, \vec{x})}$$

Defines a "natural" interpolation amongst the stations with positions $\vec{x_i}$, then it follows that a natural estimate of the spatial weighting function becomes:

$$S_i(t_j) = \frac{\displaystyle\int \frac{w(\vec{x_i}, \vec{x})}{\sum_{m=1}^{stations} w(\vec{x_m}, \vec{x})}\, d\vec{x}}{\int d\vec{x}}$$

In the previous two equations, the t_j dependence is implicit in the sums as only stations for which data exists at time t_j should be included. This spatial weighting function thus expresses the relative weight that should be given to individual temperature series based on the number of other stations in the same region. Using a weighting function such as this (or variants derived from it) provides a means of addressing the global average temperature problem without resorting to the gridding of data.

By contrast, the uncertainty term $U_i(t_j)$ would incorporate weights based on the reliability of the station and the changes it exhibits. Such uncertainties could be empirically estimated by looking at the residuals in the sum of square differences and/or by comparing $y_i(t_j)$ to $y(\vec{x_i}, t_j)$. This forms the core of an empirical "homogenization" process, by which low quality data is given reduced weight in the averaging process. The whole process would then be iterated to seek convergence with respect to these empirical estimates of $U_i(t_j)$.

In addition to the uncertainty adjustments detailed above, we might also search for statistically significant discontinuities in the record that could indicate undocumented station moves and similar problems. The impact of discontinuities can be resolved by partitioning such data into two time series with independent baseline estimators. Hence the corrections for such discontinuities can become part of the simultaneous solution to the larger averaging problem rather than a series of local adjustments.

The linear algebra required to determine the fit parameters also gives an immediate estimate of the statistical uncertainty on $\hat{y}(t_j)$, thus completing the main outline of our approach. Addressing additional homogeneity and uncertainty problems are expected to lead to further refinements.

Contact Info

For further information on the Berkeley Earth Surface Temperature Analysis or to provide feedback and suggestions, please contact either:

Dr. Richard A. Muller
ramuller@lbl.gov
+1-510-735-6877

Dr. Robert A. Rohde
robert@robertrohde.com
+1-925-354-4328