## Appendix

### The Berkeley Earth Minimization Process

In this appendix we discuss mathematical aspects of our method for deriving Earth's land surface temperature average $T_{avg}(t)$ from the thermometer data.

Given a set of temperature measurements at location $x_i$ and time $t$, we can interpolate and extrapolate to find the temperature at any other location $T(x,t)$ by using the method of Gaussian process regression (Krige [1]; Cressie [2]; Stein [3]), a method widely used in academia and business. In the simplest approach the average surface land temperature estimate $T_{avg}(t)$ is then defined by averaging $T(x,t)$ over the Earth land area $A$:

$$T_{avg}(t) \equiv \left(\frac{1}{A}\right) \int T(x,t)\, dA \qquad (1)$$

We define the local climate $C(x)$ by

$$C(x) \equiv \langle T(x,t) - T_{avg}(t) \rangle_t \qquad (2)$$

Here, the subscript $t$ indicates that the average is done over time for a given location. The fundamental difference in mean temperature between the North Pole and the Equator will be contained in $C(x)$. The local "weather" $W(x,t)$ is defined as the remainder, that is, the difference between the actual temperature record at a given location and what you would estimate from $T_{avg}(t)$ and $C(x)$ alone:

$$W(x,t) \equiv T(x,t) - T_{avg}(t) - C(x) \qquad (3)$$

The term $W(x,t)$ will include local variations in climate, such as the effects of El Nino. If every location on the Earth had a temperature anomaly (difference from the average) that followed the global record $T_{avg}(t)$, then the weather term $W(x,t)$ would be zero. $W(x,t)$ can be thought of as the *residual*, that part of the record that is not fit by a simple model of global land temperature change $T_{avg}(t)$, a function of time alone, and locally stable climate $C(x)$, a function of position alone.

To get the best estimates of $T_{avg}(t)$, the values that give the fullest explanation of the worldwide data, we can adjust parameters to minimize the square of the residuals; in other words, we minimize

$$F(t) = \int W^2(x,t)\, dA \qquad (4)$$

for each time $t$ (i.e. typically for each month) in the record. Conceptually, want the weather term to include only that temperature variation that cannot be accounted for by the temporally constant climate term $C(x)$ and the global average $T_{avg}(t)$. Minimization of $F$ provides the constraint that allows us to calculate $T_{avg}(t)$.

To minimize $F$ at a given time $t$ we adjust the following parameters:

1. $T_{avg}(t)$ There is one parameter for each time $t$, 12 values per year, and 1200 values per century.
2. A baseline temperature $b_i$ for every temperature station $i$; this baseline temperature is calculated in our optimization routine and then subtracted from each station prior to the Gaussian interpolation to intermediate points (locations on the surface which have no temperature station). This converts the temperature observations to a set of anomaly observations with an expected mean of zero. These parameters don't depend on time, and they represent the average temperature for that station above and beyond the climate $C(x)$.

3. Constants that define the slowly varying function for the local climate function $C(x)$. We use splines and polynomials to take into account variations with latitude and altitude, and these contain typically 18 parameters total (2 for temperature vs. altitude, and 16 to capture temperature vs. latitude). The key point here is that the number of parameters for this function is small.
4. Weights applied to the stations for the Gaussian interpolation. We follow the methods of robust data analysis developed by Tukey [4,5]. We begin with equal weights for all stations and calculate the $T_{avg}(t)$ that minimizes *F*. An automated routine then identifies outliers, points that give a particularly poor fit compared to the fit of nearby points. A smaller weight is then applied to such stations to reduce their contributions to the Gaussian interpolation. Then the solution to minimize *F* is recalculated; the procedure is repeated until the weights applied to each station converge. No station is omitted, but a poor station could receive a weight as low as 1/26 that of a trusted station. Note again that although the weights affect the interpolated temperature estimate for a given location, all square kilometers of land temperature contribute equally to $T_{avg}$.
5. Although the minimization process involves a large number of parameters, it is relatively rapid for a given set of weights, since it can be done by the equivalent of matrix inversion. When a solution is calculated, the weights are adjusted (to lower the weights of statistical outliers) and the equations are once again inverted. If the new weights do not change, the iteration process is complete. The computer program that accomplishes this minimization is available for inspection online at www.BerkeleyEarth.org.

## References


1. Krige DG (1951) A statistical approach to some mine valuations and allied problems at the Witwatersrand, Master's thesis of the University of Witwatersrand.
2. Cressie N (1990) The Origins of Kriging. Math Geol 22: 239-252.
3. Stein ML (1999) Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York, USA.
4. Tukey J (1977) Exploratory Data Analysis. Addison-Wesley, New York, USA.
5. Tukey JW (1958) Bias and confidence in not quite large samples. The Annals of Mathematical Statistics 29: 614.