How We Know that the Earth is Warming

Peter Guttorp

There is no doubt that global temperatures are increasing, and that human greenhouse gas emissions largely are to blame, but how do we go about measuring global temperature? It is not just a matter of reading an instrument.

In Figure 1, we see a variety of curves depicting annual global mean temperature. They are not the same, although they all show a strong increase after about 1980. Different groups, using different data and different techniques, have computed the different curves. It would be hoped that the curves would all be measurements of the annual global mean temperature, but global mean temperature is not something that can be measured directly using an instrument. On the other hand, it is the quantity most commonly used to indicate global warming.

Where do the numbers come from? We will go through some issues that are associated with determining surface temperature, and illustrate some of the uses of these temperatures.

Local Daily Mean

The basic measurements that go into the calculation of global mean temperature are readings of thermometers or other instruments determining temperature. For land stations, these instruments are typically kept in some kind of box in an open, flat space covered with grass (see Figure 2). The box keeps direct sunlight from hitting the instrument but allows wind to penetrate the box.

Readings are done at different schedules in different countries. The modern instruments measure continuously, but the measurements are not always recorded. In the United States, daily maximum and minimum temperature are recorded, and their average is the daily mean temperature. In Sweden, three hourly readings throughout the day are combined with the minimum and the maximum to calculate the daily mean temperature. In Iceland, linear combinations of two readings in the morning and afternoon are used.

Modern instruments can compute the daily average automatically, but to compare to historical data, a specific averaging method has to be applied.

Local Annual Mean Temperature

Once you have a daily mean temperature, it is easy to compute an annual mean temperature: Sum all the daily means and divide by the

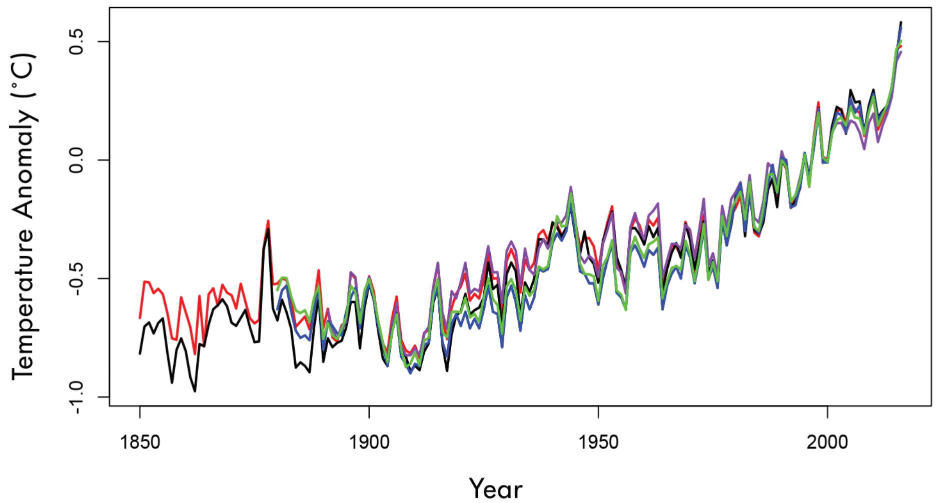

Figure 1. Five estimates of the annual global mean anomalies relative to 1981–2010: Black is from Berkeley Earth, red from the UK Met Office Hadley Center, purple from the Japanese Met Office, blue from the Goddard Institute for Space Science (GISS), and green from the National Oceanic and Atmospheric Administration (NOAA).

Figure 2. Thermometer and other instruments at Stockholm Observatory, where measurements have been made daily since 1756. The station has been moved short distances twice during this time. The box to the left is a Stephenson screen, and was used for the measurements until 2006. The pipe sticking up in the middle contains the modern measurement devise that has been used since then.
Photograph courtesy of Peter Guttorp.

number of days in the year. What is often used instead of the annual mean is something called a mean anomaly: How much did the year deviate from the average over a reference period? This makes it easier to compare sites at different altitudes, for example. A station at a higher elevation always tends to be colder than one at a lower elevation, but anomalies allow us to see if both sites are colder than usual.

The largest collection of land station data, used in the Berkeley Earth global temperature series, has some 39,000 stations and a total of 1.6 billion temperature measurements.

Sea Surface Temperature

Since more than two-thirds of the surface of our planet is water, it is not enough to take temperature measurements on land to compute a global average. Ocean-faring ships have long kept daily logbooks, with measurements of wind, air temperature, and water temperature. The water temperature used to be taken in a bucket of seawater. Later, it would be measured at the cooling water intake for the motor. Of course, ships do not travel everywhere on the oceans and, therefore, there are fairly large areas of ocean

The main groups estimating global mean temperature

- Hadley Center of the UK Met Office with the Climate Research Unit of the University of East Anglia, United Kingdom
- Goddard Institute for Space Science (part of NASA), USA
- National Centers for Environmental Information (part of NOAA), USA
- Japanese Met Office, Japan
- Berkeley Earth Project, USA

A simultaneous confidence band for n normally distributed estimates can be obtained by the Bonferroni inequality

$$P(\bigcup_{I=1}^{n} E_i) \le \sum_{i=1}^{n} P(E_i).$$

In fact, we want the complement— $P(\bigcap_{i=1}^{n} E_i^c) \ge 1 - \sum_{i=1}^{n} P(E_i^c)$.

Let E_i be the event that the true value at time i is not covered by its (pointwise) confidence set. If we let each confidence set have level $1 - \alpha / n$, we see that the probability that all parameters (in our case, the global average temperature for each year) are covered by their respective intervals is at least $1 - n(\alpha / n) = 1 - \alpha$. The confidence band then is $t_i \pm \Phi^{-1}(1 - \alpha / n)\mathrm{se}(t_i)$ where t_i is the estimated global mean temperature for year i, $\mathrm{se}(t_i)$ is the standard error of the estimate, and Φ^{-1} is the normal quantile function (inverse of the cdf).

where we have no sea surface temperature measurements from ships.

In some of these areas, there are buoys that measure the temperature. Over the last several decades, there have been satellite measurements of sea surface temperature; for over a decade, floats that measure the temperature profile of the water have been dropped all over the oceans.

The largest collection of ocean data, the ICOADS 3.0 data set, uses about 1.2 billion different records.

Combining All the Measurements

To combine the many measurements over land and oceans into an average global temperature requires estimating the temperature anomaly where there are no actual measurements, such as on a regular grid, and then averaging the estimates and measurements (if any) over the grid. Such estimation tools are derived in what is called *spatial statistics*, although atmospheric scientists have developed some methods on their own (through what they call objective analysis).

In essence, a statistician would treat the problem as one of regression, with data that are spatially dependent. The process has to take into account the fact that data are on a globe and not in the plane. The average temperature anomaly for land and ocean can be computed separately, and the global mean temperature would then be the area weighted average of the two means.

Uncertainty

There are several sources of uncertainty in the determination of global mean temperature. First of all, each measurement has error associated with it. Second, how to deal with missing areas of measurement causes uncertainty. The choice of measurement stations can also be a source of uncertainty, as can the homogenization of measurements, such as when stations are moved or measurement devices updated. There are other sources of uncertainty as well.

It is important to try to quantify the uncertainty in global mean temperature. Different groups approach this issue in different ways. Figure 3 uses the Hadley series uncertainties to compute a simultaneous Bonferroni-based 95% confidence band for global average temperature. The term simultaneous means that the confidence band covers all the true temperatures at the same time with 95% probability, as opposed to a pointwise confidence interval, which only covers the true temperature at a particular time point with 95% probability.

Ranking

In January 2017, NOAA made the claim that the global mean temperature had set a record for the third straight year. This statement

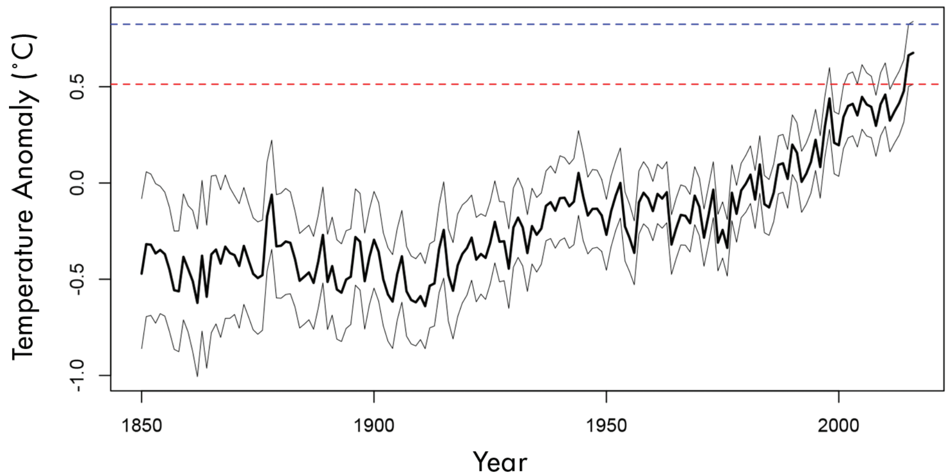

Figure 3. Hadley series with red dashed line being the lower 95% simultaneous confidence bound on the 2016 temperature and blue dashed line the upper bound on the 2015 temperature.

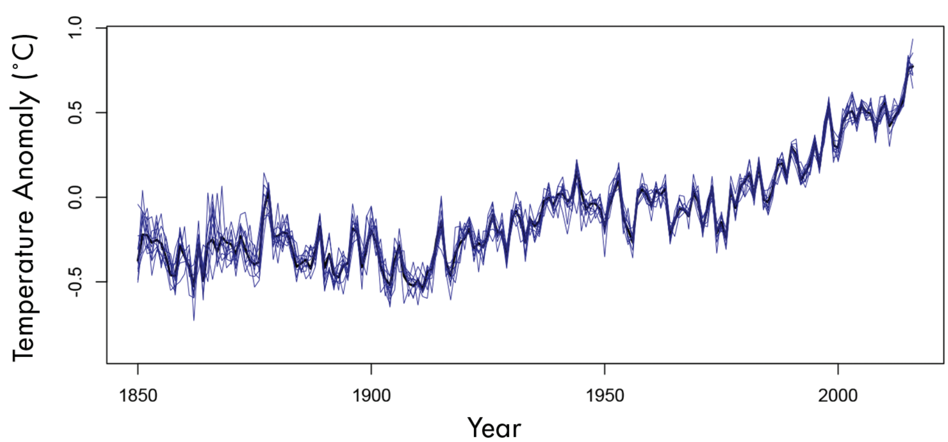

Figure 4. 10 realizations (blue) of possible Hadley temperature series and the Hadley estimate of global mean temperature (black).

is not quite accurate: For the third straight year, the *estimated* global mean temperature had set a record. In fact, four of the five series had this feature, while the Berkeley series showed 2005 as warmer than and 2010 tied with 2014. Only two of the estimates (the Hadley series and the Berkeley series) provide uncertainty estimates.

Figure 3 shows the Hadley series with associated simultaneous 95% confidence bands. If the 2015 actual temperature (which we do not know) were at the high end of its confidence band (blue dashed line), and the 2016 was at the low end of its band (red dashed line), it is quite possible that 2015 could have been substantially warmer than 2016, but that 2016 clearly was warmer than any year before 1998.

How can we say something about the uncertainty in the rankings as opposed to the estimates? One way is to simulate repeated draws from the sampling distribution of the estimates. Since we are averaging a large number of measurements, many of which are nearly uncorrelated, a central limit theorem leads us to treat the estimates as normal, with mean equal to the actual estimate and standard deviation equal to the standard error of the estimate. Figure 4 shows 10 such realizations of the Hadley temperature series.

For each of the realizations, we can calculate the rank of 2016. The distribution of that rank tells us how likely 2016 is to be the warmest year on record: It is warmest

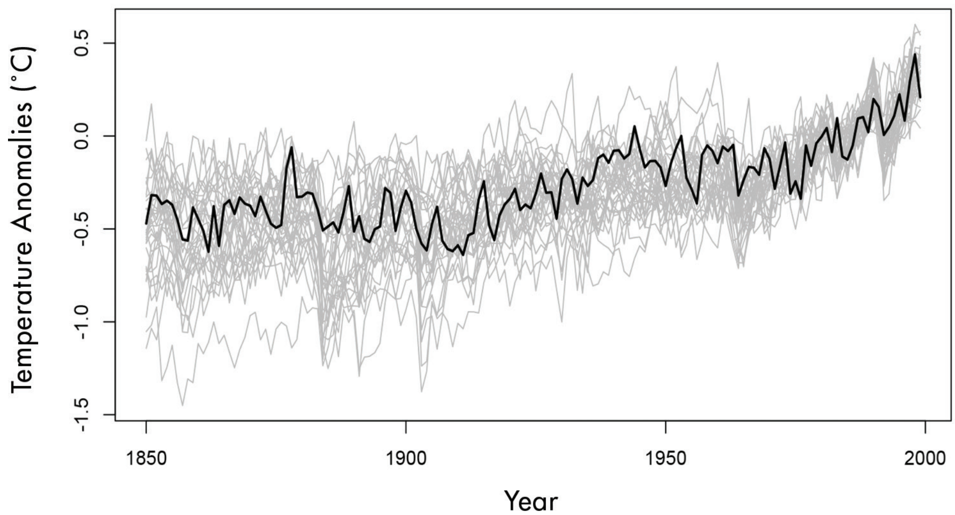

Figure 5. Global annual mean temperature anomalies from 32 CMIP5 models with historical simulations (gray), and the Hadley Center data series (black). Reference period is 1970–1999.

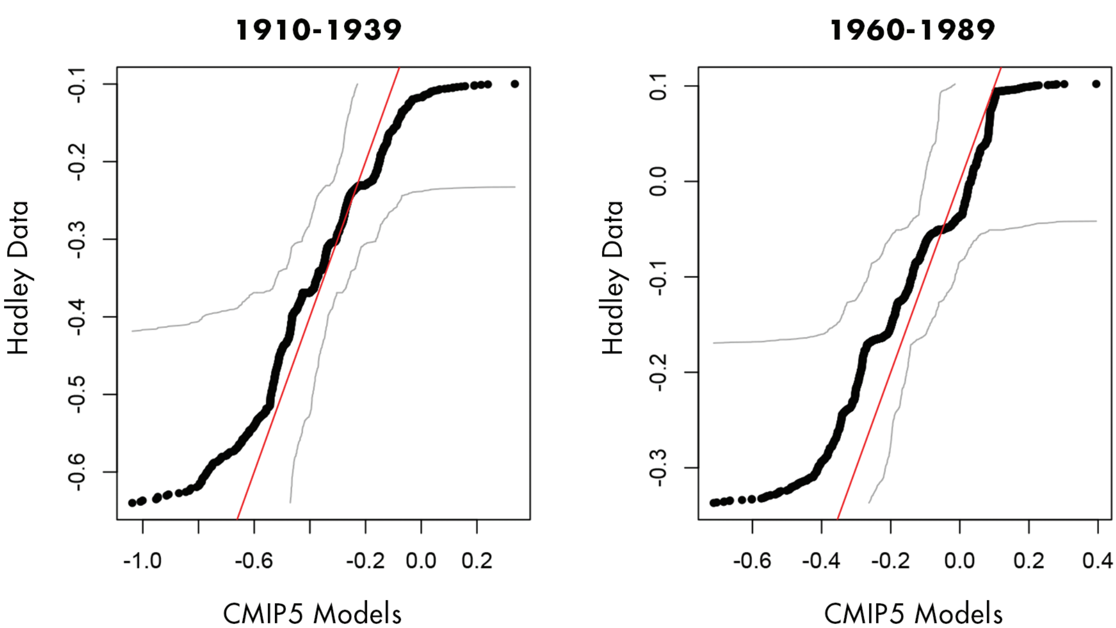

Figure 6. QQ-plots of historical climate model simulations against Hadley Center data or two 30-year periods. The gray lines are simultaneous 95% confidence bands, and the red lines are lines of equal distributions.

in 58% of the simulations, while 2015 is warmest in 42%. In 10,000 simulations, 2016 was as low as the eighth-warmest in one of them.

How about all three years—2014–2016—being record-breakers? That happened in 21% of the simulations, and in the actual Hadley estimates, of course.

Many tests have been developed to compare some aspects of distributions, such as means or medians. To compare two entire distributions, we can plot the quantiles of one against the other (called a quantile-quantile plot or QQ-plot). An advantage of this plot is that if the distributions are the same, then the plot will be a straight line. Of course, we will be estimating the quantiles from data, so there will be uncertainty. Another advantage of the QQ-plot is being able to develop simultaneous confidence bands, enabling a simple test of equal distributions: Does the line *y=x* fit inside the confidence band?

Models and Data

Climate, from a statistical point of view, is the distribution of weather. Climate change means that this distribution is changing over time. The World Meteorological Organization recommends using 30 years to estimate climate. This definition indicates, for example, that it does not make sense to look at shorter stretches of data to try to assess questions such as "Is global warming slowing down?"

What is a Climate Model?

A climate model is a deterministic model describing the atmosphere, sometimes the oceans, and sometimes also the biosphere. It is based on a numerical solution of coupled partial differential equations on a grid. In fact, the equations for the atmosphere are essentially the same as for weather prediction, but the latter is an initial value problem (we use today's weather to forecast tomorrow's) of a chaotic system, while the climate models has to show long-term stability. Many processes, such as hurricanes or thunder storms, are important in transferring heat between different layers in the model, but often take place at a scale that is at most similar to a grid square, and sometimes much smaller.

Different climate models deal with this subgrid variability differently and, as a consequence, the detailed outputs are different. CMIP5 is a large collection of model runs, using the same input variables (solar radiation, volcanic eruptions, greenhouse gas concentrations, etc.). These model outputs were used for the latest IPCC report in 2013. Figure 5 shows the global mean temperature anomalies (with respect to 1970–1999) with the corresponding Hadley Center series.

Comparing Distributions

It is not trivial to compare climate model output to data. Remember, the climate model represents the *distribution* of the data. The observations in Figure 5 are, therefore, not directly comparable to the model runs. Instead, we need to compare the distributions of model output and data, respectively.

Figure 6 compares these distributions using QQ-plots for two different 30-year stretches. In both cases, the distribution of the data fit the distribution of the ensemble of model outputs quite well, in that the red *y=x* line falls inside the simultaneous 95% confidence bands. Since we have 32 x 30 observations of the models, and only 30 of the data, the empirical tails of the model distribution are much longer than the tails of the data, but the confidence band is quite wide in the tails, meaning that we are very uncertain there. Thus, for these two time intervals and for the global mean temperature variable, the ensemble of CMIP5 models and the Hadley Center data seem to have the same distribution—they are describing the same climate. C

Further Reading

Arguez, A., Karl, T.R., Squires, M.F., and Vose, R.S. 2013. Uncertainty in annual rankings from NOAA's global temperature time series. *Geophysical Research Letters 40*:5,965–5,969.

Doksum, K. 1974. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist. 2*:267–277.

Guttorp, P. 2014. Statistics and Climate. *Annual Reviews of Statistics and its Applications 1*, 87–101.

Katz, R.W., Craigmile, P.F., Guttorp, P., Haran, M., Sanso, B., and Stein, M.L. 2013. Uncertainty analysis in climate change assessments. *Nature Climate Change 3*, 769–771.

About the Author

Peter Guttorp is a professor at the Norwegian Computing Center in Oslo and professor emeritus at the University of Washington in Seattle. He has worked on stochastic models in a variety of scientific applications, such as hydrology, climatology, and hematology. He has published six books and about 200 scientific papers.