This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Statistics used for gridded temperature analysis

Ronan Connolly and Michael Connolly

1 "Global" temperature trends

It is debatable if it is even possible to define a "global temperature" for the planet[1]. The Earth has a very wide and variable temperature field - temperatures vary with latitude, altitude/depth, season, time-of-day, etc. However, there is considerable evidence that over periods of decades/centuries/millennia, the Earth can alternate between periods of "global warming" and "global cooling", during which the globally-averaged near ground temperature slightly increases or decreases, e.g., the alternation between glacial maxima and interglacial periods during ice ages. One approach to studying global temperature *trends* without defining an absolute "global temperature" is to use *"temperature anomaly records"*

2 Temperature anomalies

The idea of a *temperature anomaly* record is quite simple. The mean temperature for a given weather station record, $\overline{T}$ is calculated over a specific time period - the "anomaly period". The anomaly record is then determined by subtracting this mean value from each of the values in the original record.

In this way, positive temperature anomalies correspond to temperatures that are "above-average", while negative anomalies are "below-average". This means that the relative temperature *trends* at different locations can be compared, even if the absolute temperatures at those locations are completely different.

Ideally, the anomaly period should be as long as possible to improve the accuracy of $\overline{T}$, and it should also be fairly consistent between all records being studied, so that the records can be directly compared.

The length of a weather record typically varies from a few decades to a few centuries. A 30 year period is often chosen as a length that is long enough to yield a reasonably accurate mean, but short enough to allow most records to be analysed[1]. The anomaly period is often chosen to correspond to a period when most of the records have data, e.g., the NASA GISS group have used the 1951-1980 period[2], while the CRU group have used the 1961-1990 period[3]. In our analysis, we typically use the 1961-1990 period.

3 Gridded means

A related problem with calculating "global temperature trends" from weather station records is that weather stations are not evenly distributed across the globe. There are spatial differences, e.g., some regions (e.g., U.S. or Europe) have higher station densities than others (e.g., central Africa or

[1] E.g., the World Meteorological Organization defines "climate" as being the 30 year average weather conditions: http://www.wmo.int/pages/prog/wcp/ccl/faqs.html

Antarctica), and $\sim 70\%$ of the Earth's surface is water rather than land. There are also temporal differences, e.g., the number of stations was much smaller in the 18th and 19th centuries than in the 20th century, the early 20th century had fewer stations than the later 20th century, and in NOAA NCDC's GHCN dataset and similar datasets, there is a dramatic decrease in station numbers after 1990. Finally, there are spatio-temporal differences, e.g., the vast majority of available mid-19th century stations were European or American.

For this reason, a simple averaging of all available station anomalies in each year is inappropriate, since it would over-represent regions of high station density and under-represent regions of low station density. Hence, a number of different approaches have been taken to overcome this problem. Initial approaches attempted to select a subset of the available station records which **are** reasonably evenly distributed and of a similar length[4, 5]. However, this often means discarding a large amount of data, which is undesirable considering the data is already quite limited as it is. Also, as the decision to discard records is somewhat subjective, it could lead to unintentional cherry-picking, potentially increasing the possibility of the researcher succumbing to confirmation bias.

To include all data, without biasing the "global" temperatures towards regions with higher spatial densities, some groups have used spatial interpolation techniques[6, 7], e.g., kriging. A simpler, though less sophisticated approach is to simply divide the globe into discrete grids, calculate the mean temperature anomalies of the grids by averaging together all the station anomalies that are in a given grid, and then average together the gridded anomalies to yield a "global" (or "regional") temperature anomaly.

NASA GISS adopt this approach by constructing the grid boxes so that they are all of uniform area[2]. However, station locations are reported using (latitude, longitude) coordinates. So a simpler approach taken by other groups[3] (including us) is to divide the globe into grid boxes according to latitude and longitude. A common grid box size in this approach is 5° latitude ×5° longitude.

The problem with the latitude/longitude gridding approach is that the surface area of a latitude × longitude box varies with latitude. High latitude grid boxes have a smaller surface area than low latitude grid boxes. This can be easily visualised by comparing the relative area of the polar regions on a globe to their relative area in a flat 2-D Mercator map of the world. Hence, if this approach is taken, the temperature anomalies of the grid boxes are usually weighted by the area of the grid boxes.

If we represent latitude by ϕ and longitude by λ (both given in radians), define ϕ_1 and ϕ_2 as the lowest and highest latitudes (respectively) of a box (in a given hemisphere), and λ_1 and λ_2 as the most westerly and easterly (respectively) longitudes of the box, then the surface area (S) of the box is[2]:

$$S = \int_{\phi_1}^{\phi 2} R \left(\int_{\lambda_1}^{\lambda 2} R \cos\phi d\lambda \right) d\phi \tag{1}$$

$$= R^2(\lambda_2 - \lambda_1)(\sin\phi_2 - \sin\phi_1) \tag{2}$$

Where $R \simeq 6371\ km$ is the radius of the Earth. The *relative* weight of the anomaly for each box, $w_{\text{box } i}$, to the global mean anomaly is then,

$$w_{\text{box } i} = \frac{S_{\text{box } i}}{\sum_{k=1}^{N} S_{\text{box } k}} \tag{3}$$

[2]For a graphical representation, see British Atmospheric Data Centre Help pages in the "Coordinates" section, for example.

$$w_{\text{box } i} = \frac{[R^2(\lambda_2 - \lambda_1)(\sin\phi_2 - \sin\phi_1)]_{\text{box } i}}{\sum_{k=1}^{N}[R^2(\lambda_2 - \lambda_1)(\sin\phi_2 - \sin\phi_1)]_{\text{box } k}} \tag{4}$$

Where N is the number of grid boxes with data for that year.

One might argue that N should be the total number of all possible grid boxes (i.e., 2592 for $5^\circ \times 5^\circ$ boxes). That is a reasonable argument. If a station (or set of stations) from a region with a relatively large climatic variability are only available for some years of the analysis, then it would create the impression that those years showed greater variability. E.g., when Jones et al., 2012[8] updated Brohan et al., 2006[3]'s analysis, they introduced more stations for the Arctic regions (e.g., Siberia) for recent years. This made the global temperature anomalies for recent years appear slightly warmer in the new analysis[8]. But, perhaps earlier years would also have appeared warmer if those stations had also been available then.

On the other hand, if "empty" grid boxes are included for a given year, then this has the effect of implying that the grid box had a temperature anomaly of $0^\circ C$ for that year, artificially dampening years with low station coverage towards the anomaly period mean, i.e., "cooling" positive anomalies, and "warming" negative anomalies. Hence, while NASA GISS took this approach in an early version of their analysis[2], most gridding approaches favour the approach of only including boxes with available data, i.e., N = number of grid boxes with data for a given year.

If the grid boxes are all of the same size in degrees longitude, i.e., $\lambda_2 - \lambda_1$ = a constant, e.g., 5°, then this can be simplified,

$$w_{\text{box } i} = \frac{(R^2(\lambda_2 - \lambda_1))\,(\sin\phi_2 - \sin\phi_1)]_{\text{box } i}}{(R^2(\lambda_2 - \lambda_1))\sum_{k=1}^{N}[(\sin\phi_2 - \sin\phi_1)]_{\text{box } k}} \tag{5}$$

$$= \frac{[\sin\phi_2 - \sin\phi_1]_{\text{box } i}}{\sum_{k=1}^{N}[\sin\phi_2 - \sin\phi_1]_{\text{box } k}} \tag{6}$$

Eqn. 6 could therefore be used for weighting different grid boxes by their relative surface area. However, another approach is to weight grid boxes by the cosine of the mid-latitude of the box[8]. This approach can be justified from Eqn. 6 in the following way:

We know that for two angles, A and B,

$$\sin A - \sin B = 2\cos\left[\frac{(A+B)}{2}\right]\sin\left[\frac{(A-B)}{2}\right] \tag{7}$$

Hence,

$$\sin\phi_2 - \sin\phi_1 = 2\cos\left[\frac{(\phi_2+\phi_1)}{2}\right]\sin\left[\frac{(\phi_2-\phi_1)}{2}\right] \tag{8}$$

Therefore,

$$w_{\text{box } i} = \frac{2\left[\cos\left(\frac{\phi_2+\phi_1}{2}\right)\sin\frac{(\phi_2-\phi_1)}{2}\right]_{\text{box } i}}{\sum_{k=1}^{N}\left[2\cos\left(\frac{\phi_2+\phi_1}{2}\right)\sin\frac{(\phi_2-\phi_1)}{2}\right]_{\text{box } k}} \tag{9}$$

If $\phi_2 - \phi_1$ = a constant for all boxes, e.g., 5°, then $\sin(\phi_2 - \phi_1)$ is also a constant. Therefore,

$$w_{\text{box } i} = \frac{\left(2\sin\frac{(\phi_2-\phi_1)}{2}\right)\left[\cos\left(\frac{\phi_2+\phi_1}{2}\right)\right]_{\text{box } i}}{\left(2\sin\frac{(\phi_2-\phi_1)}{2}\right)\sum_{k=1}^{N}\left[\cos\left(\frac{\phi_2+\phi_1}{2}\right)\right]_{\text{box } k}} \tag{10}$$

$$= \frac{\left[\cos\left(\frac{\phi_2+\phi_1}{2}\right)\right]_{\text{box } i}}{\sum_{k=1}^{N}\left[\cos\left(\frac{\phi_2+\phi_1}{2}\right)\right]_{\text{box } k}} \tag{11}$$

In other words, the weight of a grid box is proportional to the cosine of the mid-latitude of the box. The computational advantage of Eqn. 11 over Eqn. 6 is that it only requires the calculation of one cosine, instead of two sines. However, Eqn. 6 is probably more straightforward, if you are also interested in calculating the actual surface area of the grid (rather than just its ratio), as it is simply related to Eqn. 2.

In our analysis, we used the weighting from the above equation (Eqn. 11).

Whichever approach is taken, a further step is necessary to calculate the weighting of an individual station. Let x_i be the anomaly for station i in a given year. Then, the weight that station i's anomaly contributes to the global mean,

$$w_i = \frac{w_{\text{box } i}}{\text{number of stations in grid box}} \tag{12}$$

Then, the global mean temperature anomaly for that year, $\bar{x}$ is a weighted mean:

$$\bar{x} = \frac{\sum_{i=1}^{N}(x_i w_i)}{\sum_{k=1}^{N} w_k} \tag{13}$$

There are several ways which could be used to estimate the confidence intervals (CI) for a given year's temperature anomaly, e.g., Rohde et al. (submitted 2011)[7] applied sub-sampling techniques such as the "jackknife method". Most approaches rely on the assumption that the temperature data being averaged is relatively unbiased, or the biases are at least relatively "stationary" (i.e., constant over time), in which case they will be reduced or eliminated by the anomaly process. In other words, temperature variability is mostly due to either local weather fluctuations or long-term climatic change. However, in the case of weather station data, that assumption is highly problematic[9], as a number of non-climatic biases are known to affect station records, e.g., changes in station location, observation practice, instrumentation, instrument surroundings ("microclimate"), urbanization or change in the land use of the surrounding area.

Brohan et al., 2006 attempt to quantify the magnitude of the uncertainty introduced by some of these biases[3]. But, as we discuss in several articles[10–13], this is a non-trivial problem, which Brohan et al. appear to have substantially underestimated. Hence, we do not attempt to calculate the confidence intervals of the "true" global temperature anomalies, but rather the confidence intervals for the globally-averaged temperature anomalies of the available weather station data, whether they may happen to be biased or not. In other words, it is the *statistical* significance of the data which is being considered, rather than the *climatological* significance.

With that in mind, one approach is to first estimate the standard deviation of the mean, $\sigma(\bar{x})$. Researchers unfamiliar with sampling statistics should note that the "standard deviation of the mean" or "standard error of the mean" (often abbreviated to "standard error") is different from the "standard deviation of the sample". The former represents the precision to which the estimated mean value of a sample is known, whereas the latter represents the distribution of values in the sample[14, 15].

Since the global temperature anomaly is calculated as a weighted mean (Eqn. 13),

$$\sigma(\bar{x}) = \sqrt{\sum_{i=1}^{N}\left(\frac{w_i}{\sum_{k=1}^{N} w_k}\right)^2 \sigma_i^2} \tag{14}$$

Where σ_i is the standard deviation of the temperature anomalies for station i. If all weights were equal, $w_i = w$, then

$$\frac{w}{\sum_{k=1}^{N} w} = \frac{1}{N} \tag{15}$$

$$\sigma(\bar{x}) = \sqrt{\sum_{i=1}^{N} \frac{\sigma_i^2}{N^2}} \tag{16}$$

If all standard deviations were equal, then $\sigma_i = \sigma$, so

$$\sigma(\bar{x}) = \sqrt{N\left(\frac{\sigma^2}{N^2}\right)} \tag{17}$$

$$= \frac{\sigma}{\sqrt{N}} \tag{18}$$

However, as different stations have different weights and standard deviations, Eqn. 14 is necessary. For a given year, each station has only one annual value, so σ_i is unknown. One approximation to overcome this would be to assume that the standard deviations of all stations are the same, i.e., $\sigma_i = \sigma_{\text{all stations}}$, which could be calculated from the anomalies for all stations in that year. But, the temperature variability can vary significantly from region to region (e.g., equator to the poles, or inland to coastal), so that approximation is rather crude. Hence, a less crude approximation would be to calculate the standard deviation for a given grid box, i.e., from the temperature anomalies for a given year in all stations in that box. However, if a particular grid box has a low station density (perhaps only one or two stations), the standard deviation may be incalculable or unrealistic.

For these reasons, we calculate σ_i for each station over the entire temperature record for that station, and then use those values for calculating the standard deviation of the global mean for a given year. Note that this conflates the temporal variability of the temperature anomalies at each station with the spatial variability of temperature anomalies in a given year. This is potentially problematic as the estimates for the confidence intervals for the annual global temperatures are no longer independent of temperature variability over time, but the confidence intervals are being calculated specifically to assess that, i.e., global temperature trends over time.

As an aside, since the lengths of station records vary dramatically from a decade or two to a few centuries, with different start and end dates, for the sake of standardization, one could argue for calculating σ_i for all stations over the same period (e.g., the period used to calculate the station anomalies). However, we use the entire record for each station.

Finally, once the standard deviations of the means are calculated for the global temperatures, they are assumed to be normally distributed. In that case, the true value of the mean is $\sim 68.3\%$ likely to be within $\pm\sigma(\bar{x})$ of the estimated mean, $\bar{x}$, while it is $\sim 95.4\%$ likely to be within $\pm 2\sigma(\bar{x})$ of the estimated mean. Both of these ranges are sometimes used as confidence intervals. But, some researchers prefer to define the 95.0% confidence interval, i.e., $\pm 1.96\sigma(\bar{x})$. All of these ranges are valid values for CI, provided it is noted which criteria is being applied [14, 15]. Hence, we use the middle option,

$$CI = \pm 2\sigma(\bar{x}) \tag{19}$$

4 "Smoothing" data

Fundamentally, weather seems to be quite chaotic. Hence, much of the temperature variability in a weather station record might include random fluctuations which are only temporary in nature. Hence, in climate data analysis, it is quite common to apply some sort of averaging, or "smoothing", to the data series in order to remove the random short-term variability[3].

Unfortunately, it is well-known in statistics that stochastic (i.e., random) processes can occasionally produce spurious "trends", e.g., see Wunsch, 1999[16]. So, if you are looking to see if there are any "climatic trends" in a temperature series, then it is important to make sure you are detecting "signal" rather than "noise". That is, make sure the apparent trend is genuine, and not just a spurious random trend, *or* a statistical artefact of the smoothing process.

The monthly weather records we use in our analysis already have been subjected to considerable averaging/smoothing - each monthly temperature is the mean of all the daily temperatures for that month, and each "daily temperature" is itself an estimate of the mean temperature for that day.

Most of our analysis is concerned with annual trends, so we apply a further smoothing, i.e., the average annual temperature is the mean of the 12 monthly averages. However, there is considerable year-to-year variability in most temperature records. So, if you are trying to identify long-term trends, it can sometimes be useful to apply further smoothing to the records.

A common smoothing technique is to calculate a "running mean"[4] using a fixed number of years, e.g., a five year-running mean. Running means are calculated by replacing the value for a given year with the mean value over the period starting a fixed number of years before the given year and ending that same fixed number of years after the given year. This has the effect of making consecutive years seem quite similar to each other, i.e., it reduces the inter-annual variability. Hence, long-term trends are more apparent.

One problem with running means is that they can *artificially* create the appearance of long-term trends which might not exist. For instance, if a few years had anomalously high (or low) mean temperature, then this would increase (or decrease) the values of the temperatures for several years before and after this anomalous period, creating the impression of a gradual trend over a long period. In order to reduce the magnitude of this statistical artefact, while still reducing the inter-annual variability enough to consider long-term trends, one approach is to apply *"binomial smoothing"*. Like a running-mean, this approach also involves averaging all values over a fixed period. However, in the binomial mean, the weights the neighbouring years contribute to the average are reduced the further away in time they are from the target year. This approach accentuates long-term trends (low-frequency information) without losing all the inter-annual variability (high-frequency information).

The relative weights of each year are determined from the binomial distribution[5] when the probability of an event = 0.5. This particular binomial distribution is often referred to as "Pascal's triangle".

For our analysis in this paper, we often consider the 11-point binomial means of temperature trends. This uses weighting coefficients of $[1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1]$, i.e., the central value receiving 252 times the weight of the outer values.

[3] By the way, the smoothing of data series is quite analogous to the filtering of data series in the field of signal analysis to improve the "signal-to-noise ratio".

[4] This technique is sometimes called a "boxcar average"

[5] e.g., see http://en.wikipedia.org/wiki/Binomial_distribution for a summary.

References

[1] C. Essex and R. McKitrick. *Taken by storm: The troubled science, policy and politics of global warming*. 1st. Key Porter Books, 2003, 320 pp. ISBN: 978-1552632123.

[2] J. Hansen and S. Lebedeff. "Global trends of measured surface air temperature". *J. Geophys. Res. D* 92 (1987), pp. 13345–13372. DOI: 10.1029/JD092iD11p13345.

[3] P. Brohan et al. "Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850". *J. Geophys. Res.* 111 (2006), p. D12106. DOI: 10.1029/2005JD006548.

[4] G. S. Callendar. "The artificial production of carbon dioxide and its influence on temperature". *Quart. J. Roy. Meteor. Soc.* 64 (1938), pp. 223–240. DOI: 10.1002/qj.49706427503.

[5] K. Ya. Vinnikov, P. Ya. Groisman, and K. M. Lugina. "Empirical data on contemporary global climate changes (temperature and precipitation)". *J. Clim.* 3 (1990), pp. 662–677. DOI: 10.1175/1520-0442(1990)003<0662:EDOCGC>2.0.CO;2.

[6] T. M. Smith et al. "Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006)". *J. Clim.* 21 (2008), pp. 2283–2296. DOI: 10.1175/2007JCLI2100.1.

[7] R. Rohde et al. "Berkeley Earth temperature averaging process". *Submitted to journal* (2011). URL: http://berkeleyearth.org/.

[8] P. D. Jones et al. "Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010". *J. Geophys. Res.* 117 (2012), p. D05127. DOI: 10.1029/2011JD017139.

[9] K. E. Runnalls and T. R. Oke. "A technique to detect microclimatic inhomogeneities in historical records of screen-level air temperature". *J. Clim.* 19 (2006), pp. 959–978. DOI: 10.1175/JCLI3663.1.

[10] R. Connolly and M. Connolly. "Urbanization bias I. Is it a negligible problem for global temperature estimates?" 28 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/28.

[11] R. Connolly and M. Connolly. "Urbanization bias II. An assessment of the NASA GISS urbanization adjustment method". 31 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/31.

[12] R. Connolly and M. Connolly. "Urbanization bias III. Estimating the extent of bias in the Historical Climatology Network datasets". 34 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/34.

[13] R. Connolly and M. Connolly. "Has poor station quality biased U.S. temperature trend estimates?" 11 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/11.

[14] D. L. Streiner. "Maintaining standards: Differences between the standard deviation and standard error, and when to use each". *Can. J. Psychiatry* 41 (1996), pp. 498–502. URL: https://ww1.cpa-apc.org/Publications/Archives/PDF/1996/Oct/cjp-octpdf-96.html.

[15] P. Nagele. "Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals". *Br. J. Anaesth.* 90 (2003), pp. 514–516. DOI: 10.1093/bja/aeg087.

[16] C. Wunsch. "The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations". *Bull. Amer. Meteor. Soc.* 80 (1999), pp. 245–255. DOI: 10.1175/1520-0477(1999)080<0245:TIOSCR>2.0.CO;2.