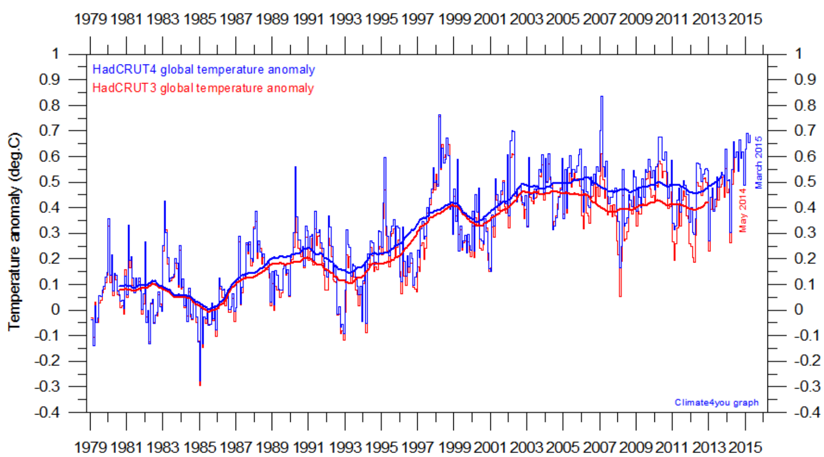
A First Look at "Possible artifacts of data biases in the recent global surface warming hiatus" by Karl et al., *Science* 4 June 2015

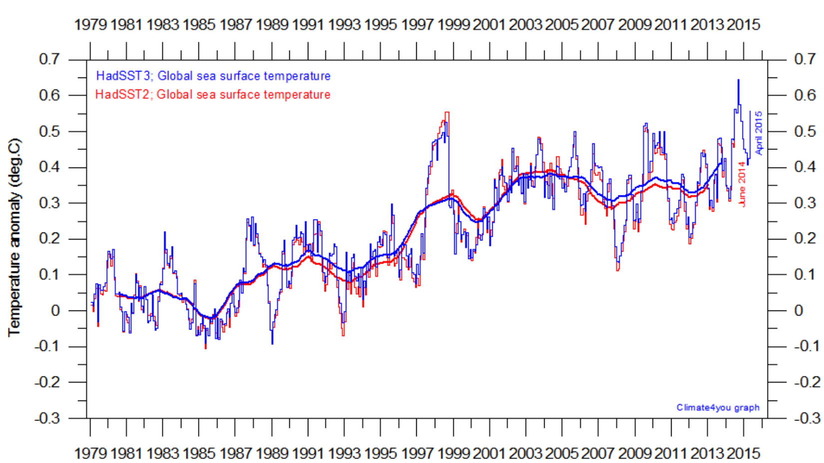
Ross McKitrick
University of Guelph
June 4, 2015

Background

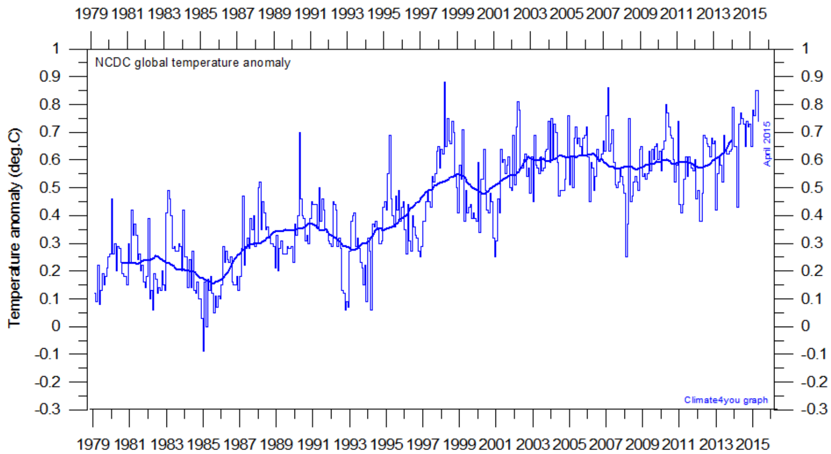
The idea that there has been a hiatus in global warming since the late 1990s comes from examination of several different data sets:



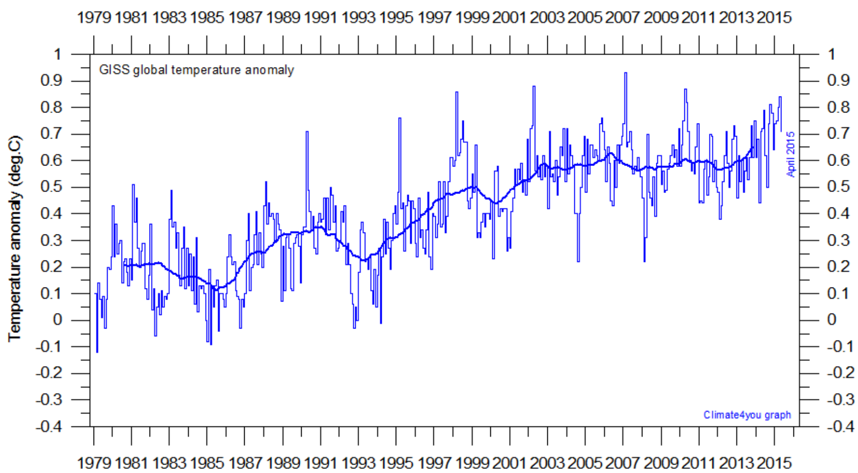
HadCRUT
(land surface + ocean)



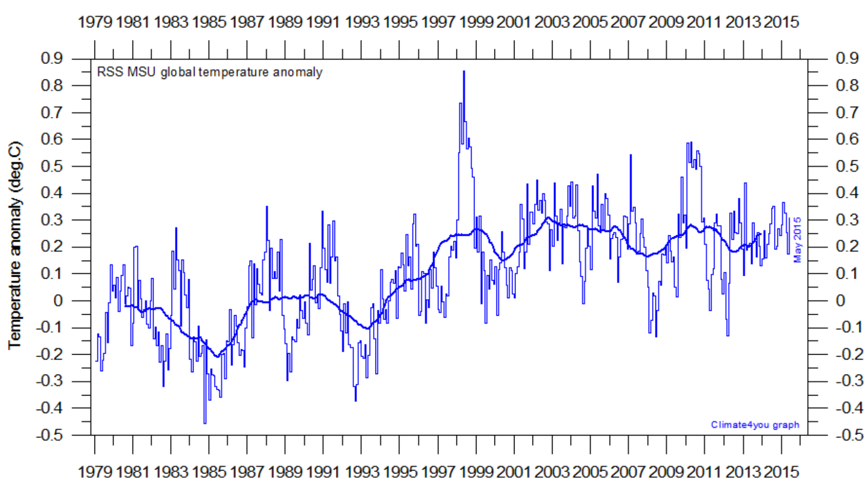
HadSST
(ocean surface only)



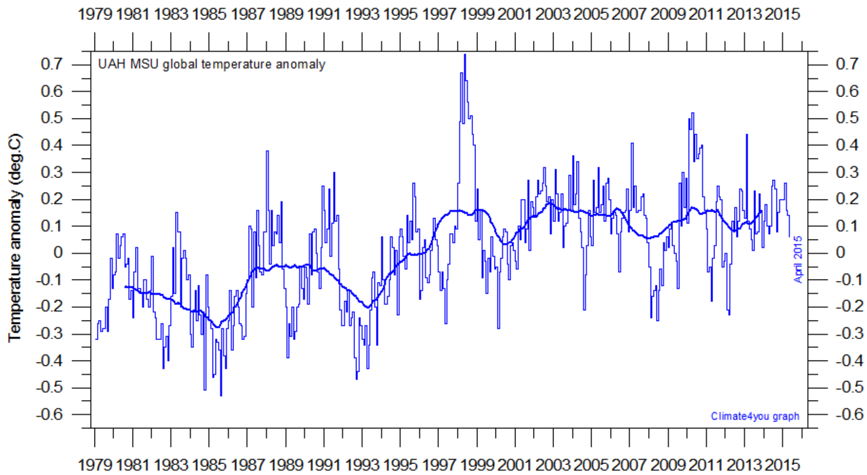
NCDC
(land surface + ocean)



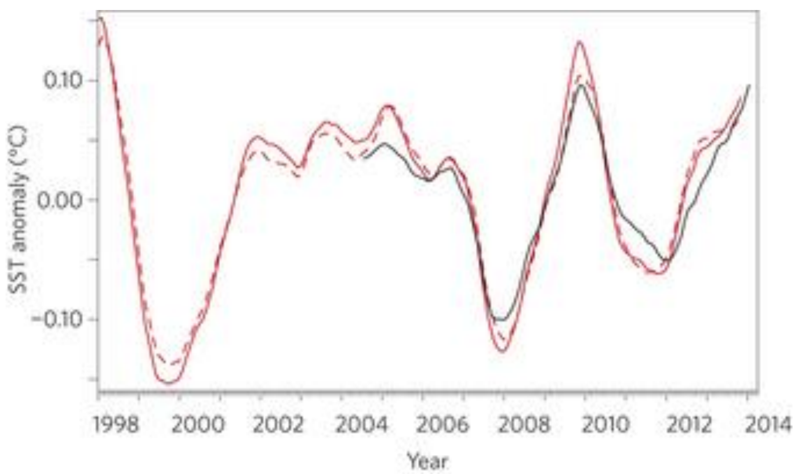
GISS
(land surface + ocean)



RSS
(lower troposphere)



UAH
(lower troposphere)



Ocean Heat Content (0-2000m)
Argo floats (black line)
NOAA SST est's (red solid and dashed lines)

Sources: all data accessed through http://www.climate4you.com/GlobalTemperatures.htm except last one, taken from http://www.nature.com/nclimate/journal/v5/n3/full/nclimate2513.html.

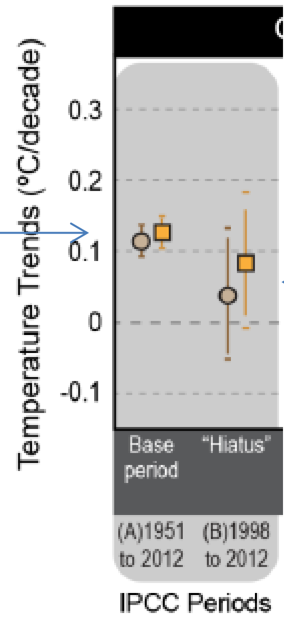
The IPCC's recent report identified this hiatus and commented as follows (Working Group I, Chapter 9, Box 9.2):

> The observed global-mean surface temperature (GMST) has shown a much smaller increasing linear trend over the past 15 years than over the past 30 to 60 years… Depending on the observational data set, the GMST trend over 1998–2012 is estimated to be around one-third to one-half of the trend over 1951–2012.

K15 New Estimates

Karl et al. (2015, which I'll call K15) have struck a very different note, saying that the post-1998 trend is much higher than previously thought, and is in fact about the same as that of the post-1951 interval. Their trend estimate revisions are as follows:

1951-2012 trend magnitude
Old (circle) 0.12 °C/decade
New (square) 0.13 °C/decade
not much different

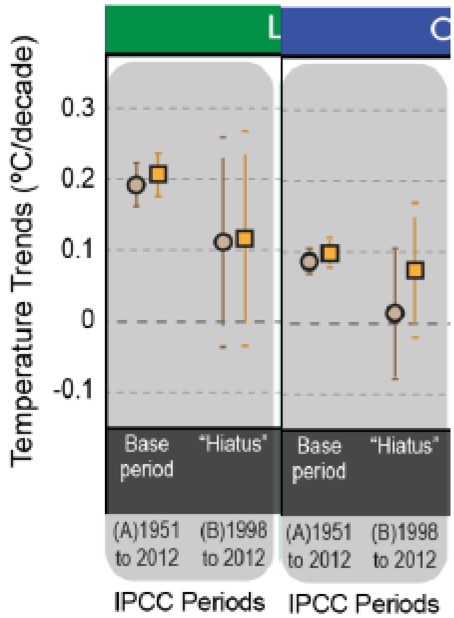


1998-2012 trends
Old (circle) 0.04 °C/decade
New (square) 0.09 °C/decade
very different

	Analysis	Global	Land	Ocean
1951-2012	New	**0.129** +/- 0.020 (0.022)	**0.207** +/- 0.031 (0.032)	**0.100** +/- 0.017 (0.020)
	Old	**0.117** +/- 0.021 (0.021)	**0.194** +/- 0.031 (0.032)	**0.088** +/- 0.017 (0.017)
1998-2012	New	**0.086*** +/- 0.075 (0.100)	0.117 +/- 0.119 (0.153)	0.075 +/- 0.075 (0.097)
	Old	0.039 +/- 0.082 (0.094)	0.112 +/- 0.119 (0.150)	0.014 +/- 0.090 (0.092)

The big source of the change is an upward revision (+0.06 °C /decade) to the global post-1998 Sea Surface Temperature (SST) trend, with only a small change to the land trend:

LAND



OCEAN

So what changed in the SST records? Bear in mind that there are very few records of air temperatures over the oceans, especially prior to 1950. So to get long term climate estimates, scientists use SST (i.e. water temperature) data, which have been collected since the 1800s by ships. The long term SST records were never collected for climate analysis and they are notoriously difficult to work with. Many judgments need to be made to yield a final record, and as the K15article shows, changes in some of those assumptions yield major changes in the final results.

A Primer on SST Data

There is a large literature on methods to derive a consistent climate record from the SST archives. The contribution of K15 is to take one such record, called the Extended Reconstructed Sea Surface Temperature version 4 (ERSSTv4), and use it to compute a new global climate record. The difference in recent trends they report is due to the changes between ERSST versions 3b and 4.

Almost all historical SST products are derived from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, http://icoads.noaa.gov/) or one of its predecessors. ICOADS combines about 125 million SST records from ship logs and a further 60 million readings from buoys and other sources.[1] A large contributor to the ICOADS archive is the UK Marine Data Bank. Other historical sources include navies, merchant marines, container shipping firms, buoy networks, etc.

SST data have historically been collected using different methods:[2]

- Wooden buckets were thrown over the side, filled with seawater and hauled on deck, then a thermometer was placed in the water;
- Same, using canvas buckets;

[1] Woodruff, S.D., H. F. Diaz, S. J. Worley, R. W. Reynolds, and S. J. Lubker, (2005). "Early ship observational data and ICOADS." Climatic Change, 73, 169–194.
[2] See http://www.metoffice.gov.uk/hadobs/hadsst3/Kennedy_2013_submitted.pdf for a review.

- Same, using insulated buckets;
- Automated temperature readings of Engine Room Intake (ERI) water drawn in to cool the ship engines;
- Ship hull temperature sensors;
- Drifting and moored buoys.

In addition, there are archives of Marine Air Temperature (MAT) taken by ships that have meteorological equipment on deck.

Here are some of the problems that scientists have to grapple with to construct consistent temperature records from these collections:

- Ships mainly travel in shipping lanes, and vast areas of the oceans (especially in the Southern Hemisphere) have never[3] been monitored;
- Sailors are not inclined to take bucket readings during storms or perilous conditions;
- Readings were not necessarily taken at the same time each day;
- During the process of hauling the water up to the deck the temperature of the sample may change;
- The change will be different depending on how tall the ship is, whether the bucket is wood or canvas, whether it is insulated, and how quickly the reading is taken;
- The ERI intake may be just below the surface in a small ship or as much as 15 m below the surface in a large ship;
- Similarly the hull sensors may be at widely-varying depths and may be subject to temperature effects over time as the engines heat up the hull;
- MAT readings are taken at the height of the deck, and modern ships are much taller than older ones, so the instruments are not at the same height above sea level;
- Buoys tend to provide readings closer to the water surface than ERI data;
- There were not many surface buoys in the world's oceans prior to the 1970s, but there are many more now being averaged in to the mix.

Now add to these challenges that when data is placed in the archive, in about half the cases people did not record which method was used to take the sample (Hirahari et al. 2014). In some cases they noted that, for example, ERI readings were obtained but they not indicate the depth. Or they might not record the height of the ship when the MAT reading is taken. And so forth.

Ships and buoys are referred to as *in situ* measurements. Since *in situ* data have never covered the entire ocean, most groups use satellite records, which are available after 1978, to interpolate over unmonitored regions. Infrared data from the Advanced Very High Resolution Radiometer (AVHRR) system can measure SST accurately but need to be calibrated to existing SST records, and can be unreliable in the presence of low cloud cover or heavy aerosol levels. In the past few years, new satellite platforms (Tropical Rainfall Measuring Mission or TRMM, and the Advanced Microwave Scanning Radiometer or AMSR-E) have enabled more accurate data collection through cloud and aerosol conditions.

[3] Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, (2003): Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Res*earch, 108(D14), 4407, doi:10.1029/2002JD002670.

Hadley, GISS and Hirahara et al. (2014)[4] all use satellite data to improve interpolation estimates over data-sparse regions. The ERSST team (i.e. K15) did prior to version 3b but doesn't anymore, due to their concerns about its accuracy.

The Three Main ERSSTv4 Adjustments

The measurement problems mentioned above all well-known. A great deal of work has been done in recent decades both to try and recover some of the metadata for *in situ* temperature readings, and also to estimate corrections in order to overcome biases that affect the raw data. K15 have made some relatively minor changes to the bias correction methods, and the result is a large increase in the post-1998 trend.

A. They added 0.12 °C to readings collected by buoys, ostensibly to make them comparable to readings collected by ships. As the authors note, buoy readings represent a rising fraction of observations over recent decades, so this boosts the apparent warming trend.
B. They also gave buoy data extra weight in the computations.
C. They also made adjustments to post-1941 data collected from ships, in particular a large cooling adjustment applied to readings over 1998-2000.

Taken together these changes largely explain the enhanced trend over the past 15 years. So now everybody needs to decide if they think these adjustments are valid.

Perhaps they are. The main problem for us observers is that other teams have looked at the same issues and come to different conclusions. And the post-1998 K15 data don't match that from other independent sources, including weather satellites.

A. Looking at the first adjustment, K15 take the buoy data and add 0.12 °C to each observation. They computed that number by looking at places where both buoy data and ship data were collected in the same places, and they found the ship data on average was warmer by 0.12 °C. So they added that to the buoy data. This is similar to the amount estimate found by another teams, though the bias is usually attributed to ships rather than buoys:

> Recent SST observations are conducted primarily by drifting buoys deployed in the global oceans (Figs. 1, 2). The buoys measure SST directly without moving seawater onto deck or to the inside of a ship. Therefore, buoy observations are thought to be more accurate than either bucket or ERI data… In the present study, we regard this difference as a bias in the ERI measurements, and no biases in drifting buoy observations are assumed. The mean ERI bias of +0.13 °C is obtained and is within the range for the global region listed in Table 5 of Kennedy et al. (2011).
> *(quote from Hirahari et al. 2014 p. 61)*

That quote refers to a paper by Kennedy et al. (2011 Table 5)[5] which reports a mean bias of +0.12 °C. However, Kennedy et al. also note that the estimate is *very* uncertain: it is 0.12 ± 1.7°C ! Also, the bias varies by region. This is a key difference between the method of K15 and that of others. K15 added 0.12 °C to all buoy data, but the Hadley group and the Hirahari group use region-specific adjustments.

---

[4] Hirahara, S. et al. Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty, *Journal of Climate* Vol 27 DOI: 10.1175/JCLI-D-12-00837.1

[5] http://www.metoffice.gov.uk/hadobs/hadsst3/part_2_figinline.pdf

B. There is not much detail about this step. K15 simply say that because the buoy data are believed to be more reliable, they were given more weight in the statistical procedure, and "This resulted in more warming." Steps A + B accounted for just under half of the additional warming.

C. It has been noted by others previously that SST data from ships shows a more rapid warming trend than nearby air temperature collected by buoys (Christy et al. 2001).[6] K15 compute an adjustment to SST data based on comparisons to Nighttime MAT (NMAT) records from a data set called HadNMAT2. This step entailed making a large cooling adjustment to the ship records in the years 1998-2000. K15 say that this accounts for about half the new warming in their data set. They defended it by saying that it brought the ship records in line with the NMAT data. However, this particular step has been considered before by Kennedy et al. and Hirahara et al., who opted for alternative methods in part because, as Kennedy et al. and others have pointed out, the NMAT data have their own "pervasive systematic errors",[7] some of which were mentioned above. So rather than using a mechanical formula based on NMAT data, other teams have gone into great detail to look at available metadata for each measurement type and have made corrections based on the specific systems and sites involved.

Numerical Example

Here is a simple numerical example to show how these assumptions can cause important changes to the results. Suppose we have SST data from two sources: ships and buoys. Suppose also that ships always overestimate temperature by exactly 1 degree C and buoys always underestimate it by exactly 1 degree C. We have one set of readings every 10 years, and we are not sure what fraction is from ships versus buoys. Both ships and buoys accurately measure the underlying trend, which is a warming of 0.1 ^{o}C /decade from 1900 to 1990 then no trend thereafter.

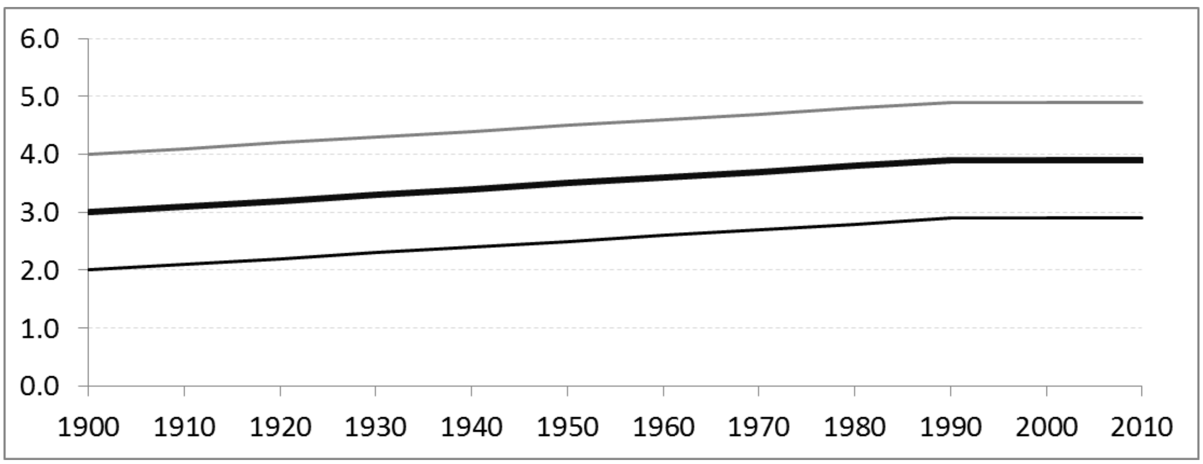
The Table below shows the simulated numbers. Suppose the true fraction of ships in the sample starts at 95% in 1900 and goes down by 8% every decade, ending at 7% in 2010.

[6] Christy, John R., David E. Parker, Simon J. Brown, et al. 2001 Differential trends in tropical sea surface and atmospheric temperatures since 1979. *Geophysical Research Letters* 28, no. 1

[7] http://www.metoffice.gov.uk/hadobs/hadsst3/Kennedy_2013_submitted.pdf page 28.

Year	Buoy	Ship	True Ship %	True Avg
1900	**2.00**	**4.00**	0.95	3.00
1910	**2.10**	**4.10**	0.87	3.10
1920	**2.20**	**4.20**	0.79	3.20
1930	**2.30**	**4.30**	0.71	3.30
1940	**2.40**	**4.40**	0.63	3.40
1950	**2.50**	**4.50**	0.55	3.50
1960	**2.60**	**4.60**	0.47	3.60
1970	**2.70**	**4.70**	0.39	3.70
1980	**2.80**	**4.80**	0.31	3.80
1990	**2.90**	**4.90**	0.23	3.90
2000	**2.90**	**4.90**	0.15	3.90
2010	**2.90**	**4.90**	0.07	3.90

The true average is calculated using the weight in the True Ship % column, adding 1 ^{o}C to the buoy data and subtracting 1 ^{o}C from the ship data. The result is shown in the graph:



The thin black and gray lines are the ship (top) and buoy (bottom) data, while the thick black line in the middle is the true average.

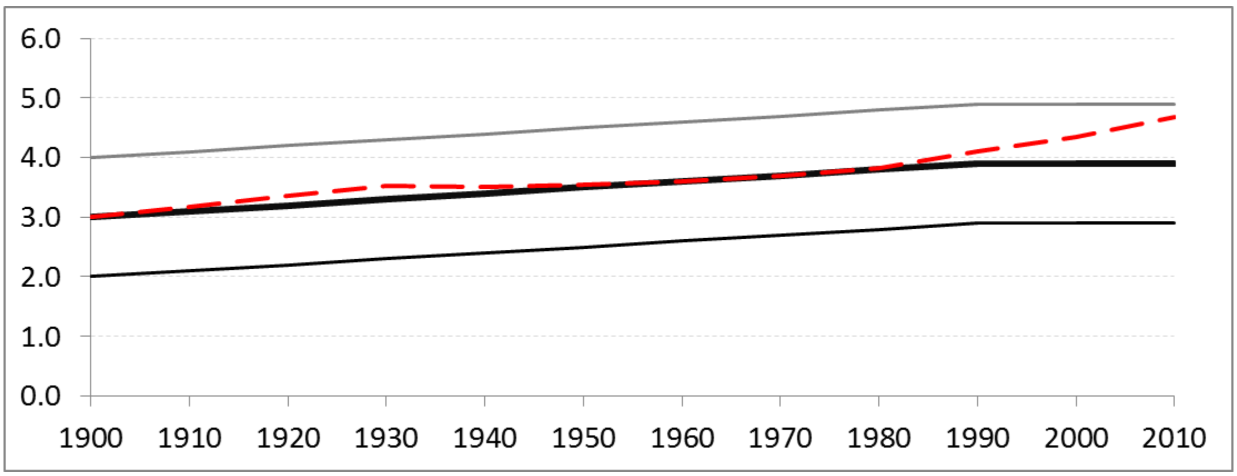
But now suppose we don't know what the correct adjustment is for the buoy data or the ship data, and we don't know the True Ship % figures either. We will estimate the global average as follows:

- Adjust the buoy data up by +2 ^{o}C every year (a bit too much)
- Adjust the ship data down by 1 ^{o}C every year (the right amount)
- After 1940 we will also apply a cooling adjustment to the ship data that starts at -0.25 ^{o}C and goes up by that amount every decade
- We further cool the ship data by 1 ^{o}C in 1990 and 2000 only
- We estimate the ship %, starting it at 99% in 1900 (a bit high) and reducing that by 7% every decade (a bit too little) up to 1990, at which point we observe the True Ship % and follow it exactly thereafter.

Before looking at the results, ask yourself if you think these adjustments will make much difference.

Year	Buoy	Buoy adj	Ship	Ship Adj	True Ship%	True Avg	Est Ship %	Est Avg
1900	**2.00**	2.00	**4.00**	-1.00	0.95	3.00	0.99	3.01
1910	**2.10**	2.00	**4.10**	-1.00	0.87	3.10	0.92	3.18
1920	**2.20**	2.00	**4.20**	-1.00	0.79	3.20	0.85	3.35
1930	**2.30**	2.00	**4.30**	-1.00	0.71	3.30	0.78	3.52
1940	**2.40**	2.00	**4.40**	-1.25	0.63	3.40	0.71	3.51
1950	**2.50**	2.00	**4.50**	-1.50	0.55	3.50	0.64	3.54
1960	**2.60**	2.00	**4.60**	-1.75	0.47	3.60	0.57	3.60
1970	**2.70**	2.00	**4.70**	-2.00	0.39	3.70	0.50	3.70
1980	**2.80**	2.00	**4.80**	-2.25	0.31	3.80	0.43	3.83
1990	**2.90**	2.00	**4.90**	-3.50	0.23	3.90	0.23	4.10
2000	**2.90**	2.00	**4.90**	-3.75	0.15	3.90	0.15	4.34
2010	**2.90**	2.00	**4.90**	-3.00	0.07	3.90	0.07	4.69

The new estimated average is the red dashed line.



The fit is not bad up to 1990, but the accumulated effect of all the small mistakes is the artificial trend introduced at the end of the series. At this point we would hope to have some independent data on the post-1990 trend to compare the result to in order to decide if our methods and assumptions were reasonable.

This example proves nothing about K15, of course, except that small changes in assumptions about how to deal with uncertainties in the data can have a large effect on the final results. But that was already clear because the K15 themselves explain that their new assumptions—not new observations—are what introduced the warming trend at the end of their data set.

Conclusion

Are the new K15 adjustments correct? Obviously it is not for me to say – this is something that needs to be debated by specialists in the field. But I make the following observations:

- All the underlying data (NMAT, ship, buoy, etc) have inherent problems and many teams have struggled with how to work with them over the years

- The HadNMAT2 data are sparse and incomplete. K15 take the position that forcing the ship data to line up with this dataset makes them more reliable. This is not a position other teams have adopted, including the group that developed the HadNMAT2 data itself.

- It is very odd that a cooling adjustment to SST records in 1998-2000 should have such a big effect on the global trend, namely wiping out a hiatus that is seen in so many other data sets, especially since other teams have not found reason to make such an adjustment.

- The outlier results in the K15 data might mean everyone else is missing something, or it might simply mean that the new K15 adjustments are invalid.

It will be interesting to watch the specialists in the field sort this question out in the coming months.

Ross McKitrick

Department of Economics
University of Guelph
ross.mckitrick@uoguelph.ca
rossmckitrick.com