Extended Summary of the

Climate Dialogue

on

the (missing) tropical hot spot

Author: Marcel Crok
with input from Bart Strengers (PBL) and Bart Verheggen

October 2014

Content

Introduction

Based on theoretical considerations and simulations with General Circulation Models (GCMs), it is expected that any warming at the surface will be amplified in the upper troposphere. More warming at the surface means more evaporation and more convection. Higher in the troposphere the (extra) water vapour condenses and heat is released. The following figure shows this in a very simple way:

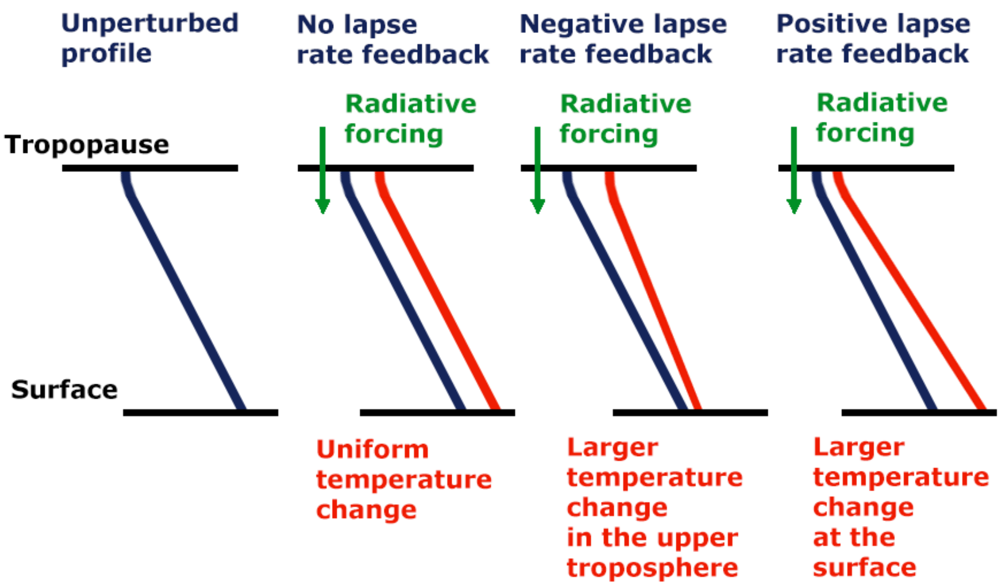

Figure 1: Schematic representation of the lapse rate feedback.

The first picture on the left shows the unperturbed temperature profile in the atmosphere. If you go up the temperature drops about 6°C per kilometre. This decrease in temperature with increasing altitude is the so-called lapse rate.

Now what will happen in the perturbed situation where the increased concentration of greenhouse gases exerts a positive forcing on the climate system? Climate scientists think that the third picture is most likely to happen: the lapse rate will become slightly steeper, causing more warming in the upper troposphere. This is called a negative lapse rate feedback because warming at the surface is less than expected based on a uniform temperature change (picture 2). So this process reduces the surface warming influence of greenhouse gases, but higher up the warming effect is amplified.

This effect is independent of the cause of the warming. So an increase in solar radiation should also show this pattern of warming in the troposphere. Calculations with climate models (GCMs) show that globally the lower troposphere should warm about 1.2 times faster than the surface. For the tropics, where most of the moisture is, the amplification is larger, about 1.4.

IPCC published the following figure in its fourth report (AR4) in 2007:

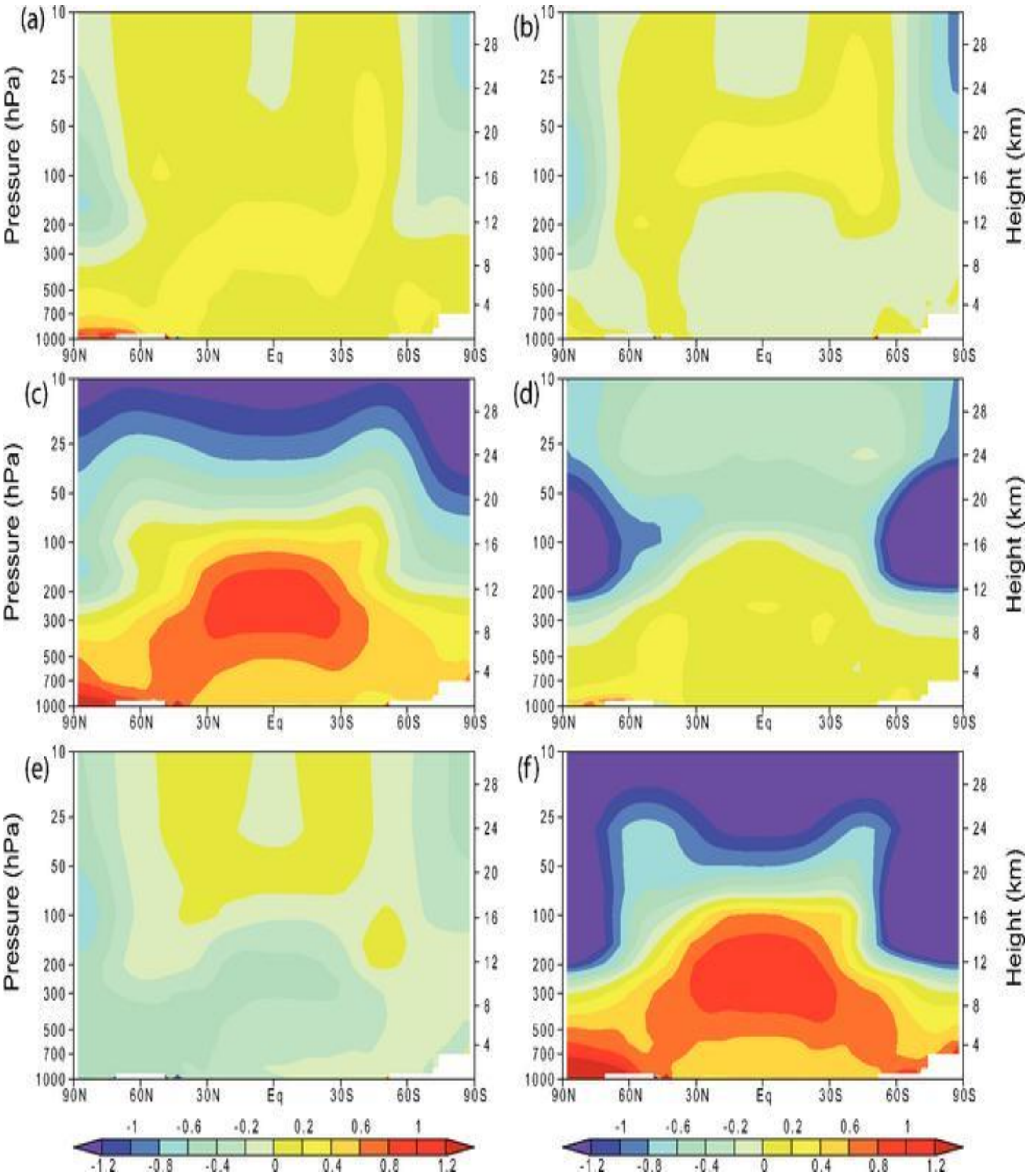

Figure 2. Repeat of figure 9.1 from AR4. Zonal mean atmospheric temperature change from 1890 to 1999 (°C per century) as simulated by the PCM model from (a) solar forcing, (b) volcanoes, (c) well-mixed greenhouse gases, (d) tropospheric and stratospheric ozone changes, (e) direct sulphate aerosol forcing and (f) the sum of all forcings. Plot is from 1,000 hPa to 10 hPa (shown on left scale) and from 0 km to 30 km (shown on right). Source: http://www.ipcc.ch/publications_and_data/ar4/wg1/en/figure-9-1.html

The figure shows the response of the atmosphere to different forcings (warming or cooling influences on the climate) in a specific GCM. The IPCC report adds that the "[t]he major features shown in Figure 9.1 are robust to using different climate models". As one can see, over the past century, the greenhouse forcing was expected to dominate all other forcings. The expected warming is highest in the tropical troposphere. Only in recent years this warm area in the tropical troposphere has been dubbed the tropical "hot spot".

Temperature datasets for the (tropical) troposphere start in 1958 (based on weather balloons also called radiosondes) and based on satellite measurements (since 1979). So we now have several decades to see if the theoretical and model expectations hold up in the observations. The issue became controversial when US scientists John Christy and Roy Spencer started to build their satellite dataset in the 90-ies[i], because originally it showed no warming at all in the global troposphere. Later several deficiencies were found and corrected in their dataset and a second group (RSS, Carl Mears and Frank Wentz) also prepared a temperature time series.

However some of the controversy remained because both satellite and radiosonde datasets still show (much) less warming than the models do. John Christy and Fred Singer pointed this out in a 2008 article in the International Journal of Climatology[ii], but this was criticized in the same issue of the journal by another article co-authored by a large group of climate scientists[iii]. The original article claimed models and observations differed significantly. The critique on the article was that the authors underestimated the uncertainties in both the models and the observations and that when you take these uncertainties into account the ranges for models and observations overlap. Ergo models and observations were consistent with each other. The models were not "falsified".

We were really glad that three well-known players in this controversy accepted our invitation to participate: Steven Sherwood of the University of New South Wales in Sydney, Carl Mears of Remote Sensing Systems and John Christy of the University of Alabama in Huntsville. Sherwood and Mears were co-authors of the 2008 article which criticized the article by Christy and Singer. So with Mears, Sherwood and Christy as participants in this Climate Dialogue we had three scientists who are all very familiar with this issue.

In this summary we will discuss all the questions that we posed in our introductory article and some more, because it turned out that sometimes our questions were not totally clear or complete.

Do the discussants agree that amplified warming in the tropical troposphere (i.e. the hot spot) is expected?

The discussants **agree** on this topic.

Sherwood writes that lapse-rate changes differing significantly from those expected from basic thermodynamic arguments would be very interesting. In 2005 he even dropped the 'hot spot' as a research topic because based on his own work and others' plus a better understanding of the basic challenges, he concluded there was no credible evidence for any unexpected changes in atmospheric temperature structure.

Mears explains that in the deep tropics, in the troposphere, atmospheric processes cause the lapse rate to be largely controlled by the moist adiabatic lapse rate or MALR[iv]. Because the MALR increases with decreasing temperature, any temperature increase at the surface becomes even larger high in the troposphere. This causes the so called hot spot, a region high in the troposphere that shows more warming (or cooling) than the surface.

Christy writes that the hot spot is expected via the traditional view that the lapse rate feedback[**Fout! Bladwijzer niet gedefinieerd.**] operates on both short and long time scales.

However in the course of the dialogue it turned out that different definitions of the "hot spot" were being used. The agreement noted above is limited to the "strict" definition of the "hot spot" which is just the fact that warming at the surface will be amplified higher up in the tropical troposphere. So the key word here is "amplification". A second, broader definition of the hot spot is the fact that models – given the known greenhouse forcing (see figure 1) – expect a lot of warming in the tropical troposphere. So the key words here are "the magnitude of the trend".

When we brought this confusion to the attention of the participants Sherwood gave the following reaction:

> Marcel asks whether by “hot spot” one means the warming aloft, or the difference (or ratio) between warming aloft and at the surface. The problem here is that the “hot spot” concept was not created by scientists (as far as I know) but is a term coined by climate skeptic bloggers. If one looks at the problem from the point of view of climate physics it decomposes naturally into one on lapse rates (which are governed by atmospheric convective processes) and global surface temperature (which is controlled by top-of-atmosphere radiative balance and ocean heat uptake). For this reason the focus in the scientific literature (as opposed to the internet) has been on either lapse rates, or surface temperatures, and this is the focus I prefer. Obviously it is fair enough to ask whether warming in any particular location is consistent with models or not, if one’s only goal is to falsify models. But if one is trying to understand the system it is better to ask first what is happening at the surface, and then, given that, what is happening in the atmosphere.

First, Sherwood notes that the term “hot spot” was not created by scientists but by sceptical bloggers. A little searching using Google (Scholar) indicates this might be true. The term “hot spot” is mainly used on sceptic websites. It starts popping up after the publication of the 2007 paper of Douglas, Singer and Christy[v] which was presented as a falsification of the models. Google Scholar gives no further relevant peer reviewed papers although Douglas and Christy use “hot spot” in their 2013 paper[vi].

With regard to the definition of the hot spot Sherwood clearly is in favour of the strict definition, i.e. the one focusing on the “amplification”.

However in a public comment Canadian economist Ross McKitrick, who published several papers about this topic, stated that the “uniqueness of the hot spot” has more to do with the magnitude of the warming (our bold):

> Regarding the uniqueness of the tropical “hotspot”, the uniqueness arises from the **magnitude of the trend**, not the amplification with respect to the surface. While it is true that amplification would be observed in response also to increased solar forcing, it’s clear from comparing panels (a) (solar), (c) (GHG) and (f) (all) in the IPCC figure that only GHG’s are expected to have had a sufficiently strong effect to yield the level of warming projected overall. Were there to be a lack of warming, it would be most inconsistent with the GHG simulation.

Christy also emphasizes both “aspects” of the hot spot. In his guest blog he wrote:

> So, what has the extra CO2 and other greenhouse gases done to the climate as of today? Climate model simulations indicate that a prominent and robust response to extra greenhouse gases is the warming of the tropical troposphere, a layer of air from the surface to about 16 km altitude in the region of the globe from 20°S to 20°N. A particularly obvious feature of this expected warming, and is a key focus of this blog post, is that this warming increases with altitude where the rate of warming at 10 km altitude is over twice that of the rate at the surface.

Here he mentions the amplification but also the fact that models show “a prominent and robust response to extra greenhouse gases” in the tropical troposphere. He adds:

> Thus, there are two ideas to test in the tropics, (1) the overall magnitude of the layer-average temperature rise and (2) the magnification or amplification of the surface temperature change with height.

Mears showed the magnitude of the trends in both models and observations in figure 2 of his guest blog showing a larger trend in the models than in the observations. He notes:

> Looking at Fig. 2., it is obvious that the observed trends in both temperature datasets are at the extreme low end of the model predictions. This problem has grown over time as the length of the measured data grows. (...) For the time being, I am tabling the discussion of this problem and focusing in the discussion of the hot spot. In my mind, the problem of the **trend magnitude is more interesting than the argument about the hotspot**, and I hope to return to it later in this process. But for now I will stay focused on the hotspot.

Here he clearly distinguishes between the "magnitude of the trend" and the "hotspot", by which he means "amplification". Mears focuses on the amplification aspect because for him this is the topic of the Climate Dialogue, although for him "the problem of the trend magnitude is more interesting".

As we will see later on in this summary, remarks about the hot spot always have to be seen in the context of the definition used. We will make this explicit as much as possible.

Table 1

	Sherwood	**Christy**	**Mears**
Is amplified warming in the tropical troposphere expected?	Yes	Yes	Yes
Is the magnitude of the warming trend in the tropical troposphere included in your definition of the hotspot?	No	Yes	No

Can the hot spot in the tropics be regarded as a fingerprint of greenhouse warming?

The short and simple answer to this question is a "no" from all three participants. They agree that any positive forcing – natural or anthropogenic – should give warming at the surface and amplified warming aloft and therefore a hotspot in the tropics. As such it is not a specific fingerprint for greenhouse gases.

However, given the fact that as far as we know greenhouse forcing was the dominant forcing in the 20th century and therefore we expect lots of warming in the tropical troposphere, Christy does regard the hot spot as "a" fingerprint, as he explained in an email exchange with the moderators:

> The question implies that if the only new forcing in the system were enhanced GHGs, then the hot spot would appear. Thus it is indeed "a" fingerprint if the atmosphere behaves according to the traditional theory. This is demonstrated in model experiments by comparing runs with (experiment) and without (control) enhanced GHGs. An observed hot spot, however, does not imply GHGs are its cause since any forcing can generate the signal. Some might suppose that the GHG-forced hot spot is somehow mitigated by an unknown negative forcing in the past 34 years in the real world. This is speculation, but it is conceivable. Thus the question is not well-posed as one can have a "yes" or "no" answer depending on the hypothesis being tested. Is the GHG-forced hot spot a fingerprint in models? Yes. Is the hot spot a GHG fingerprint in the real world? Evidently not.

In a comment Bart Verheggen then rephrased the question/statement, in order to avoid confusion about different interpretations of the word "fingerprint" and noted that all three participants agreed that:

> No, the hot spot in the tropics is not specific to a greenhouse mechanism.

It appears that on the one hand Christy acknowledges that the tropospheric amplification is not specific to a GHG mechanism, but on the other hand he maintains that the magnitude of the tropospheric temperature trend indicates that the GHG influence on warming is less than expected. This is related to slightly different interpretations of the terms “hot spot” and “fingerprint”.

Christy wrote in his guest blog:

> So, what has the extra CO2 and other greenhouse gases done to the climate as of today? Climate model simulations indicate that a prominent and robust response to extra greenhouse gases is the warming of the tropical troposphere (…)

So for him a lack of warming in the tropical troposphere would falsify the models and could indicate that the effect of greenhouse gases is not so large.

Mears and Sherwood used a stricter interpretation of the term fingerprint (i.e. implying specificity) in their guest blog and comments along with an interpretation of the hot spot as referring to the amplification of surface warming in the tropical troposphere. Since this amplification is expected whatever the cause of warming is, they don’t see why this should be such a controversial topic and why people use it to imply something about the relative role of greenhouse gases specifically.

In his introduction *Sherwood* for example indicates the (non-)existence of a hot spot would have no clear implications for global warming because the regulation of lapse rate changes by atmospheric convection is expected to work exactly in the same way whether global temperature changes are natural or caused by anthropogenic greenhouse gases. He also wrote in his guest blog:

> Perhaps the most remarkable and puzzling thing about the “hot spot” question is the tenacity with which climate contrarians have promoted it as evidence against climate models, and against global warming in general.

Mears in a comment agreed with this statement of Sherwood adding:

> Like Steve, I am somewhat mystified about all the attention given to the tropical hotspot, as I don’t think it is very important for global warming theory, and it is relatively poorly observed.

Mears added that the tropospheric hot spot is not some sort of lynchpin of global warming theory:

> Surface warming due to any cause would show a tropospheric hotspot in the absence of other changes to the heating and cooling of the atmosphere. Nevertheless, the tropospheric hotspot is often presented as some sort of lynchpin of global warming theory. It is not. It is just a feature of a close-to-unstable moist atmosphere.

Summary

The use of different definitions for both the “hot spot” and “fingerprint” make it possible to either present the hot spot controversy as a relatively minor issue (Sherwood and Mears) or to present the hot spot as a pronounced human feature in the models which, if missing in the observations, would have serious consequences for our understanding of the climate system (Christy). This explains why it has been a controversial topic for years.

Table 2

	Sherwood	**Christy**	**Mears**
Can the hot spot in the tropics be regarded as a specific fingerprint of greenhouse warming?	No	No	No
Do climate model simulations indicate that extra greenhouse gases give a prominent warming of the tropical troposphere?	Yes	Yes	Yes
Would a lack of warming in the tropical troposphere have serious implications for attribution of global warming to GHG?	No	Yes	No

Is there a significant difference between modelled and observed amplification of surface trends in the tropical troposphere (as diagnosed by e.g. the scaling ratio)?

The key word again in this question is “amplification” (or “scaling ratio”). Mears and Christy both provided graphs to explain the issue. Here is first Christy’s figure 4:

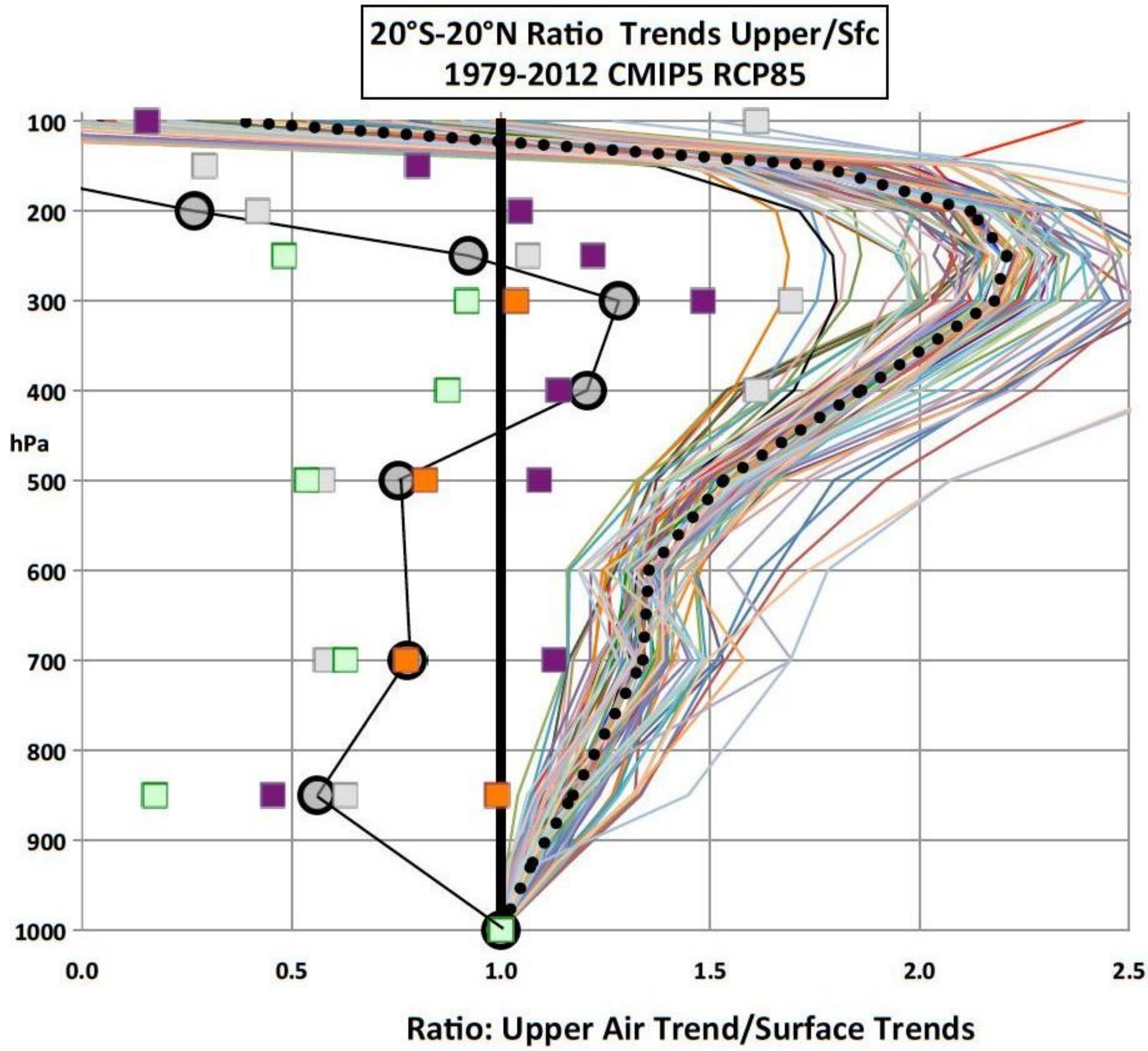

Figure 3, a repeat of figure 4 in Christy’s guest blog. Value of the 1979-2012 temperature trend at various upper levels divided by the magnitude of the respective surface trend, i.e. the ratio of upper air trends to surface trends. Model simulations are lines with the average of the models as the dotted line. Squares are individual balloon observations (green – RATPAC, grey RAOBCORE, purple – RICH and orange – HadAT2) with the averages of observations the grey circles.

The figure shows the ratio between values higher up in the tropical troposphere divided by the trend at the surface. Values lower than one means the trend at some height is lower than the trend at the surface. Values higher than one mean amplification of the surface trend. The figure shows that models on average show more amplification than the observations. Christy concluded (our bold):

What this figure clearly indicates is that the second aspect of this discussion, i.e. namely the rising temperatures with increasing altitude, is also over-done in the climate models. **The differences of the means between observations and models are significant**.

Mears also showed a figure that is relevant to this discussion:

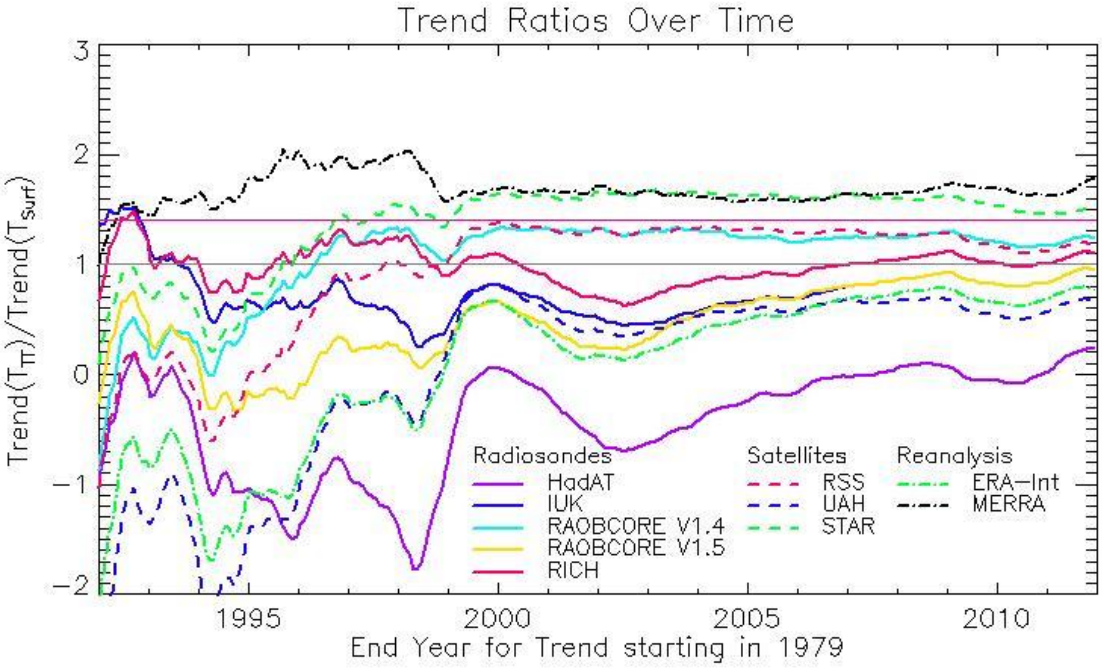

Figure 4, a repeat of Mears' figure 1. Ratio of trends in T_{TT} to trends in T_{Surf} as a function of the ending year of the trend analysis. The starting point is January 1979. The surface dataset used is HADCRUT4. The pink horizontal line is at a value of 1.4, the amplification factor for T_{TT} in reference 1.

So here the amplification factor is shown on the vertical axis and on the horizontal axis the length of the trend period is shown. T_{TT} stands for the temperature of the tropical troposphere.[vii] The pink horizontal line represents the average amplification factor of the models. Now contrary to the figure of Christy, two datasets actually show more amplification than the models.

Mears therefore concluded (our bold):

> [T]he values that are ultimately reached strongly depend on which upper air dataset is used. For some datasets (HadAT, UAH, IUK, RAOBCORE 1.5, ERA-Interim), the trend ratio is less than 1.0, indicating lack of a tropospheric hotspot. For other datasets (RICH, RAOBCORE 1.4, RSS, MERRA, and STAR), the ratio is greater than one, indicating tropospheric amplification and the presence of a hotspot. (…)
>
> Often one or more of these datasets is used to argue that a tropical hotspot exists or does not exist. **A more balanced analysis shows that it is difficult to prove or disprove the presence of the tropospheric hotspot given the current state of the data**.

Both *Sherwood* and *Mears* **agree** this question (whether there is a significant difference between modelled and observed amplification of surface trends in the tropical troposphere) cannot be answered because the observations are not stable enough over time to determine whether a hot spot exists or not, or is as prominent as we would expect. This is in part due to the added noise that one gets when calculating the ratio of two small, relatively similar, uncertain numbers.

Sherwood in his guest blog wrote the following:

> Although there has been more to-ing and fro-ing in the literature since then, as described in the opening article for this exchange, I still remain unconvinced that we can observe the small changes in temperature structure that are being discussed. Tests of radiosonde homogenisation methods (e.g., Thorne et al. 2011) show that they are often unreliable. MSU is not well calibrated and its homogenisation issues are also serious, as shown by the range of results previously obtained from this instrument series. To obtain upper-tropospheric trends from Channel 2 of MSU requires subtracting out a large contribution to trends in this channel coming from lower-stratospheric cooling. The latter remains highly uncertain due to a discrepancy between cooling rates in radiosondes and MSU.

Sherwood even thinks that there is a chance that all the satellite datasets (RSS, UAH and STAR) *underestimate* the real warming in the troposphere (our bold):

> Tropical ozone trends are sufficiently uncertain so as to render either of these physically plausible (Solomon et al. 2012). I used to think (as do most others) that the radiosondes were wrong, but in Sherwood et al. 2008 we found (to my surprise) that when we homogenised the global radiosonde data they began to show cooling in the lower stratosphere that was very similar to that of MSU Channel 4 at each latitude, except for a large offset that varied smoothly with latitude. Such a smoothly varying and relatively uniform offset is very different from what we'd expect from radiosonde trend biases (which tend to vary at lot from one station to the next) but is consistent with an uncorrected calibration error in MSU Channel 4. If that were indeed responsible, it would imply that there has been more cooling in the stratosphere than anyone has reckoned on, and that **the true upper-tropospheric warming is therefore stronger than what any group now infers from MSU data**. By the way, our tropospheric data also came out very close to those published at the time by RSS, both in global mean and in the latitudinal variation (Sherwood et al., 2008).

An interesting point that unfortunately wasn't discussed in more detail. Later in the dialogue, discussions about datasets mainly took place between Mears and Christy who are closely involved in the RSS and UAH satellite datasets respectively.

Mears believes part of the controversy surrounding this topic is that it was discussed before enough data was stable over time. The presence (or not) of the tropospheric hotspot depends on which pair of datasets you use. Ross McKitrick, who participated in the public comments, agreed with Mears and Sherwood. He wrote (our bold):

> Mears' Figures 1 and 3 (+4) nonetheless show that the amplification rate in models is high relative to the distribution in the observations. Current data sets are too short to say whether the difference is statistically significant or not. I doubt they will ever be long enough. **The statistical issues involved in figuring out the distributions of ratios of random numbers get complicated quickly, and I wouldn't be surprised if the problem is intractable**.

So Christy is rather alone in claiming that the amplification factors differ significantly between models and observations. His conclusion is based on a different selection of datasets than Mears used. Also, Christy averaged different observational datasets and compared these averages with the ensemble mean of the models, while Mears compared all the datasets individually. Sherwood criticized the averaging procedure suggested by Christy because the subset of numbers they are averaging exhibit a very wide spread. The average would thus have a very wide uncertainty range and discarding other data, based on just the distance to this average (without taking uncertainties into account) is misleading.

In general Christy seems to favour "colder" datasets while Mears is in favour of "warmer" datasets. In Mears' figure 1 for example two datasets show an amplification factor that is higher than the amplification factor of the models: MERRA and STAR. Christy doesn't use either of them. So to understand the positions on this issue a rather technical discussion about the pros and cons of different datasets is unavoidable. This is what has happened in several peer reviewed papers and also in this Climate Dialogue in several long comments. This discussion mainly took place between Mears and Christy.

Several data issues were discussed, which can be summarized by the following questions:

1) What observations or data sources (i.e. satellites, radiosondes/balloons and re-analysis) should be considered?
2) What satellite metric is the best indicator for the tropical troposphere, the Temperature Tropical Troposphere (TTT) favoured by Mears or the Temperature of the tropical Mid-Troposphere (TMT) used by Christy?
3) What could explain the relatively large difference in tropical trends between the UAH and the RSS dataset? (which was question 4 in our introductory article)

For the reader who is not interested in all the technical details here is first a short summary of the differences between Mears and Christy. In general the disagreement about what datasets to use remained. Christy dismisses the STAR satellite dataset, the one showing the largest trend of the three satellite products (UAH, RSS and STAR). Mears includes it.

Something intriguing is going on with the UAH and RSS satellite datasets. Note that Christy is in charge of the UAH dataset, while Mears is in charge of the RSS dataset. Globally the RSS and UAH temperature trends of the lower troposphere are pretty similar with UAH showing a slightly larger trend of 0.14°C/decade since 1979 than RSS (0.12°C/decade). However in the mid-troposphere, the datasets start to differ with RSS globally showing considerably more warming (0.078°C/decade) than UAH (0.046°C/decade). This also applies to the tropical troposphere, where the RSS trend is 0.09°C/decade and the UAH trend is 0.03°C/decade. Mears and Christy agree the difference must have something to do with diurnal cycle adjustments but both – not surprisingly - see their own dataset as the more reliable one.

They also disagree about which satellite metric (TTT or TMT) is the most suitable. The only thing on which they agreed after the discussion is that reanalysis datasets (MERRA and ERA) should not be used yet. Below follows a much more technical summary of the discussions about datasets. If you want to skip the technical discussion read on at page 16 under the subsection “Are models warming significantly faster than the observations?”.

What observations or data sources (i.e. satellites, radiosondes/balloons and re-analysis) should be considered?

Let's first give an overview of the datasets that were mentioned in the discussion. In bold the datasets that were 'rejected' by either Mears or Christy.

Table 3

Data source	Source Type	Used by Christy	Used by Mears
RATPAC	Radiosonde	Y	**N**
RAOBCORE	Radiosonde	Y	Y
RICH	Radiosonde	Y	Y
HadAT2	Radiosonde	Y	Y
RSS	Satellite	Y	Y
UAH	Satellite	Y	Y
STAR2.0	Satellite	**N**[a]	Y
ERA	Re-analysis	**N**	**N**
MERRA	Re-analysis	**N**	**N**

[a] This dataset is less reliable than UAH and RSS according to Christy; the newer version STAR3.0 though is more reliable and its results are closer to RSS.

Table 3 *Overview of the data sources discussed. It is indicated which datasets are used by Christy and Mears, the text explains the reasons why. RATPAC (Radiosonde Atmospheric Temperature Products for Assessing Climate) created by scientists from NOAA, NCDC and GFDL. RAOBCORE (RAdiosone OBservation COrrection using REanalyses) and RICH (Radiosonde Innovation Composite Homogenization) from the university of Vienna. HadAT2 developed by the United Kingdom Met Office Hadley Centre. RSS prepared by Remote Sensing Systems, a private company in California. UAH prepared by the University of Alabama in Huntsville. ERA is the reanalysis data archive developed by ECMWF. MERRA (Modern Era Retrospective-analysis for Research and Applications) developed by NASA. STAR2.0 (Satellite Applications and Research) developed by NOAA the Satellite and Information Service (NESDIS).*

At first sight the differences don't seem very large. Mears did show the MERRA and ERA reanalysis trends in his figure 1. However, after some discussion he agreed that these datasets are not very reliable yet:

> I tend to de-emphasize reanalysis output, because I think reanalysis is even less ready than the satellite data for use in global temperature trend assessment. In general, the reanalysis projects ingest uncorrected satellite data, and hope that their analysis system can make the needed adjustments. This has certainly not been proven to be the case, and there are many examples of it not working out — e.g. problems with vapour and clouds in the MERRA reanalysis caused by the advent of AMSU brightness temperatures.

The 'only' apparent difference that's left between Mears and Christy is that contrary to Christy Mears accepts STAR as a dataset while he dismisses the RATPAC radiosonde dataset. RATPAC is left out by *Mears* because he argues the individual station data are adjusted before 2005, and then not adjusted after 2005.

STARv2.0 is left out by *Christy* but included by Mears. However, at the time that this summary was written (a year after the dialogue) STAR2.0 was replaced by STAR3.0 which is also accepted by Christy.

In the table below we summarize several different temperature trends in the tropical troposphere:

Table 4 Tropical tropospheric temperature trends since 1979

Data source	Temp Type	^{0}C/dec Christy	^{0}C/dec Mears
RSSv3.3	TMT	0.088	0.086 ± 0.04
UAHv5.6	TMT	0.031 ± 0.05	0.033
RSS+UAH	TMT	0.060 ± 0.03	0.060 ± 0.03
STAR3.0[a]	TMT	0.106	0.102
All satellites[b]	TMT	0.075	0.074
HadAT2	TMT	Not Updated through 2013	
Raobcore	TMT	0.055	0.058
RICH	TMT	0.087	0.100
RATPAC	TMT	0.016	Not Adjusted After 2005
Radiosondes	TMT	0.049 ± 0.035[c]	0.079[d]
74 models	TMT	0.26	0.278[e]
RSS	TTT[f]	0.123	0.121
UAH	TTT	0.068	0.067
STAR3.0	TTT	0.145	0.144
All Satellites	TTT	0.112	0.111
HadAT2	TTT	Not Updated through 2013	
Raobcore	TTT	0.081	0.085
RICH	TTT	0.128	0.135
RATPAC	TTT	0.071	Not Adjusted After 2005
102 models	TTT	0.316	0.330[g]

[a] During the dialogue there was much discussion about the reliability of STAR2.0; STAR3.0 though is accepted by both Christy and Mears.
[b] Including STAR3.0.
[c] Based on Raobcore, RICH and RATPAC
[d] Based on Raobcore and RICH
[e] Based on 33 model runs.
[f] TTT = 1.1*TMT - 0.1*TLS where TLS is Temperature of the Lower Stratosphere.
[g] Based on 33 model runs.

Table 4 *Tropical tropospheric temperature trends based on different radiosonde and satellite datasets for the period 1979-2013 and the area 20S-20N. TMT=Temperature of the tropical Mid Troposphere, TTT=Temperature of the Tropical Troposphere. Note that this table was made a year after the actual dialogue together with active input from Christy and Mears.*

Note that in this table we no longer compare amplification factors between surface and troposphere but absolute trends in the tropical troposphere. The table is so long because Christy favours TMT while Mears favours TTT. As one can see TTT trends are generally higher than TMT trends, but the model trends for TTT are also higher (0.32°C/decade) than the TMT model trends (0.26°C/decade).

This table was prepared a year after the actual dialogue when we were working on the summary and tried to understand why sometimes Mears and Christy mentioned *different* trends for the *same* dataset. During this process Christy and Mears converged almost completely as can be seen from the table. Also both agree that the absolute trends of models and observations differ significantly. As Mears wrote in an email: "I don't think any of these details affect the main story. The observed tropical trends are outside the range predicted by almost all models."

Warmest vs coldest

Christy argued during the dialogue that *Mears* tends to discuss those datasets which are the "warmest", i.e. RSS, MERRA and STAR. Also, none of the datasets can claim to be perfect, but many of their differences can be explained (see Christy et al 2010 and 2011).

Mears in his turn emphasized that *Christy* likes to use arguments based on short term trend differences and jumps to throw out datasets — usually those with *warmer* than average trends. In Mears et al (2012) the entire time series is assessed, as opposed to only analysing one or more segments that are under suspicion. When doing so, the STARv2.0 dataset had short term trends that were closest to those in the various adjusted radiosonde datasets.

Christy said about STAR:

> STARv2.0 contains a spurious warming shift on 1 Jan 2001 which will be corrected in the new v3.0 to be released later this year [2013]. So, STAR's results in Mears's contribution overstate the warming. (…)
>
> I did not include STAR due to the known shift in its temperature and the fact it uses the identical diurnal corrections as RSS – thus it is very similar to RSS but with a known spurious shift.

Mears explained STAR:

> In Christy's last post, he made an argument for excluding STAR V2.0 based on a small positive jump in temperature in 2001, and that it is the same as RSS, since it uses the same diurnal correction. First, the jump in 2001 is fairly small, and does not change the 34 year trend very much when removed in V3.0 (the global trend in STAR V3.0 TMT will be about 0.015 K/decade lower — but still warmer than RSS). Second, the STAR analysis uses a completely different calibration scheme based on simultaneous nadir overpasses. In the STAR scheme, the satellite calibration is not polluted by errors in the diurnal correction, because it occurs before the diurnal correction in processing. So I really think STAR 2.0 is an independent dataset, and cannot be excluded based on dependence on RSS. Also note that STAR 3.0 TMT will no longer use the RSS diurnal correction, but still shows more warming than RSS TMT.

Christy replied:

> Since the STAR shift is known to its authors, I calculated the value relative to RSS (since both STAR and RSS use the same diurnal correction and the AMSUs were in use in 2001, this focuses on the shift apart from other adjustments) as +0.056 °C. Subtracting this shift at 1 Jan 2001, now produces a time series almost identical with RSS with a difference in trend of only +0.004 °C/decade.

So no agreement was reached on this issue. Note that this discussion took place in September 2013. The summary was written almost a year later when STAR 3.0 was available. This dataset is accepted by Christy too. Table 4 mentions up to date data.

Christy emphasized that even if STAR were included as an "independent" dataset, the significance of the results would not change. The same trends calculated from observations, i.e. the mean of four balloon and mean of two satellite datasets, are slightly less than 0.05 and 0.06 °C/decade respectively. However this might only be the case if STAR is averaged with the other datasets.

Closely related to the discussion about STAR was our question 4: **What could explain the relatively large difference in tropical trends between the UAH and the RSS dataset?**

As can be seen in table 4 the trend differences between the UAH and RSS datasets are quite large in the tropical troposphere. Globally UAH and RSS trends are quite similar for the lower troposphere (TLT). The UAH trend nowadays is even slightly higher than the RSS trend. However for TMT the agreement is much less, both in the tropics and globally, as was noted in a public comment by Paul S:

> That's pretty much true for TLT but there is a reasonable discrepancy for global TMT: 0.078 against 0.046ºC/Dec. That is intriguing though, given that TLT is produced from the same base data as TMT. It suggests the similarity in global TLT is largely due to compensating errors rather than agreement.

Mears in a comment agreed with this last remark of Paul S about compensating errors. Ross McKitrick in a comment wrote about the TMT trends of UAH and RSS[viii]:

> For [T]MT:
>
> (a) The UAH trend (0.040 C/decade) is insignificant, the RSS trend (0.111 C/decade) is significantly different from zero at 5%, the HadAT trend (0.018 C/decade) is insignificant and the RICH trend (0.025 C/decade) is insignificant. The 4 series averaged together have a trend (0.025 C/decade) that is statistically insignificant (p=0.53). So only RSS exhibits a trend at the MT layer. (...)
>
> (c) UAH and RSS are significantly different from each other at 5%. The MSU series averaged together is not significantly different from the balloon series averaged together (p>0.4). The disagreement within the basic data types is stronger than that across the data types, and is not large compared to the difference between observations and models.

McKitrick's remark that "only RSS exhibits a trend at the MT layer" shows the relevance of a discussion about the differences between the RSS and UAH TMT trends. Note that McKitrick didn't discuss STAR.

Mears*:* The problem is that the uncertainty in UAH has not been documented well enough for me to feel comfortable doing analysis with it. Using the uncertainty analysis for RSS (using a Monte-Carlo approach), in general for TLT, UAH is within the 95% confidence interval of RSS, while for TMT, it is not. This suggest that for TMT, the RSS/UAH differences may be significant.

Two main possibilities are differences in 1) the non-linearity correction or the "target factor"[ix] and 2) in the diurnal adjustment applied to account for changes in measurement time. The target factor differences are largest for the NOAA-09 satellite (Po-Chedley and Fu, 2012). The diurnal adjustment is most important for NOAA-11, due to its long life and large measurement time drift.[x] In 2005, UAH made changes to their diurnal adjustment to TLT that brought their results into closer agreement with RSS. At any rate, the geographic distribution of RSS/UAH TMT differences points to the diurnal cycle as the most likely culprit. The best agreement is in the Southern Hemisphere Extratropics, where the diurnal adjustment is small due to the prevalence of ocean. The largest disagreement is in the tropics, where the diurnal cycle tends to be large. If the main culprit were the target factors, the differences would be more similar in the different regions. There is a **tool** that allows to look at trend difference between RSS, UAH, and STAR, as well as the radiosonde datasets. The error analysis is documented in Mears et. al 2011.

Christy agrees that in a statistical sense TMT for RSS and UAH are significantly different from each other in the tropics and that the divergence in the mid-1990s is the key and likely relates to the diurnal adjustments.[xi] UAH uses an empirical technique drawn from (admittedly noisy) observations and RSS relies on a climate model simulation. Neither will be perfect, and hence an average of the two is the best way to deal with the differences, since UAH and RSS are within error ranges of their common mean value (see table 4). Whatever spurious warming or cooling there might be in the separate constructions it will likely be minimized in the average.

Christy disagrees that the uncertainty in UAH has not been documented well enough. According to published results the same could be said about RSS. In the UAH dataset a collection of “adjusted” individual multi-country radiosondes are used to perform the corrections which are, however, plagued with unknown instrumentation and other changes (Mears et al. 2011). Many require “adjustments” for undocumented changes whose magnitudes impact the trend to an extent greater than the true trend-signal itself (see Christy and Norris 2004). The tropical average of these radiosonde datasets is a way to minimize their individual errors. The question of “which is better?” ends up being a dilemma because for both datasets strong claims can be made to back up the decisions as to why a particular method of testing was chosen. It is entirely understandable that an outsider would be suspicious of any dataset where the methodology of evaluation supports the results of that dataset. To by-pass the question “Which dataset is better?” one can simply utilize the average of UAH and RSS. By so doing, it is essentially assumed that UAH and RSS contain an equal amount of error on either side of the truth.

By averaging, the independent errors can be reduced. The fact that TMT trends from the average of two very different and independent set of monitoring systems, i.e. the average of the balloons and the average of UAH plus RSS (see Table 2), are within 0.01 °C/decade of each other (i.e. 0.047 vs. 0.059, see Table 3), lends confidence to the result.

Sherwood disagrees because according to that reasoning there is no longer any doubt about equilibrium climate sensitivity, because the average of the models and of various estimates based on past data are each around 3^0C (e.g., IPCC 2007).

TTT versus TMT

The third dataset discussion was about TTT versus TMT. Mears prefers the Temperature Tropical Troposphere (TTT) while Christy favours the Temperature of the tropical Mid-Troposphere (TMT).

Mears indicates that TMT, as directly observed by the Advanced Microwave Sounding Units (AMSU) on board of satellites, is not really the mid-tropospheric temperature because it also includes part of the lower stratosphere. Since the stratosphere is cooling it tends to cancel some of the tropospheric warming. Therefore, TTT should be used which adjusts for this cooling effect:

$$TTT=1.1*TMT-0.1*TLS \qquad (1)$$

where TLS is the observed AMSU Temperature of the Lower (tropical) Stratosphere. In this way TTT is centered in the mid to upper tropical troposphere, where we expect the hot spot to be most pronounced.

Christy argues that the observations of the lower stratosphere (TLS) have greater uncertainty than TMT, and therefore contribute to the spread of TTT results among the various datasets. Since TLS only contributes about 7% to the TMT-signal, a more direct method that reduces observational error is simply to use TMT, which is directly measured and captures the bulk of mid-to-upper troposphere temperature and avoids the compounding of errors that TTT introduces (see Christy et al. 2010 and 2011 for details).

Mears prefers TTT because it separates the trends in the troposphere and stratosphere more than TMT does. Using the argument that TTT is not directly measured, many satellite retrievals would not be acceptable also, including, to cite a RSS example, all the wind speed and total column water vapour retrievals from microwave imaging instruments, which are derived from measured radiances. Even the MSU/AMSU measurement are derived from radiances, which are derived from small currents crossing the PN junction in a detector diode. So *none* are directly measured. If you don't believe in the ideas behind calculating TTT, then you don't believe in the possibility of atmospheric sounding with microwaves.

Furthermore, TTT in the worst case increases the uncertainty from 0.038 K/decade (i.e. the uncertainty of TMT) to 0.0478 by adding the errors of TMT and *twice* the error of TLS as in formula (1), which is a factor of about 1.25. However, TTT *also* increases the signal we want to see by a factor of 1.35 for both RSS (i.e. from 0.117 to 0.158 K/decade) and STAR (i.e. from 0.144 to 0.194). For UAH the factor is even larger, about 1.85 (from 0.05 to 0.091 K/decade). So, by using TTT instead of TMT, the signal to noise ratio has increased! Also, when comparing with radiosonde data (i.e. measurements from weather balloons) it is better to use TTT instead of TMT because it is fairly well established that the problems with the radiosonde increase at higher altitude, with most indications being that even the homogenized records show spurious cooling at high altitude. By using TTT instead of TMT, the contribution for the radiosonde levels at high altitude (i.e. above 100 hPa) are very much reduced, reducing the contribution to the error from these troublesome levels.

Christy disagrees because he thinks the errors in TTT are larger than estimated by *Mears*. As far as the models go, the average ratio of TTT to TMT is 1.18, meaning very little new information is provided. The satellite ratios are much larger (1.35 to 1.82) than the models simply because the denominators (i.e. observed TMT trends) are so much smaller. But irrespective of the disagreement on TTT, *Christy* argues that one still must come to the conclusion that the trend of TMT is a problem for models to replicate, which in his view seems to be accepted by Mears and Sherwood too.

Table 5

	Sherwood	**Christy**	**Mears**
Is there a significant difference between modelled and observed amplification of surface trends in the tropical troposphere?	Data too uncertain to answer this question	Yes	Data too uncertain to answer this question
What satellite metric should be used?	x	TMT[a]	TTT[b]
Is STAR a reliable additional dataset?	x	STAR3.0 is more reliable than STAR2.0	Yes
Are UAH and RSS tropical troposphere trends significantly different from each other?	x	Yes	Yes
What is likely the main reason for the difference between the UAH and RSS TMT trends?	x	Diurnal adjustments	Diurnal adjustments

[a] TMT=Temperature of the tropical Mid Troposphere
[b] TTT=Temperature Tropical Troposphere

Are models warming significantly faster than the observations?

This brings us to what is the key issue for Christy: the fact that climate models show significantly more warming in the tropical troposphere than the observations.

For Christy the main issue is that models fed with greenhouse forcing produce large warming trends in the tropical mid-troposphere. All the observations including even STAR show much less warming in this region than the models.

From Christy's first comment:

> The simple numbers tell the story and can't be overlooked. From 73 CMIP-5[xii] model runs, the 1979-2012 mean tropical TMT trend is +0.26 °C/decade. The same trends calculated from observations, i.e. the mean of four balloon and mean of two satellite datasets, are slightly less than +0.06 °C/decade. Tropical TMT is a quantity explicitly tied to the response of models to the enhanced greenhouse effect (or any applied forcing). Because the sample of climate model runs is relatively large (N = 73) we have a very confident assessment of the model-mean value and its error range does not encompass the observations. In addition, the agreement of the means of two independent observational systems further indicates that we have a very good idea of the actual TMT trend. The mean of the models (often used as the "best estimate" in IPCC assessments) and observations differ by +0.20 °C/decade which is highly significant. And, we are not talking about 10 or 15-year trends – this is a 34-year period over which this discrepancy has grown. Regarding the highly significant nature of the differences in my initial posting, I failed to mention the many papers led by Ross McKitrick (e.g. McKitrick et al. 2010, McKitrick et al. 2011 and others) in which they demonstrate with more advanced statistical tools that the models and observations are indeed significantly different regarding tropical tropospheric temperature trends.

In his guest blog Christy posted a figure showing this discrepancy between models and observations:

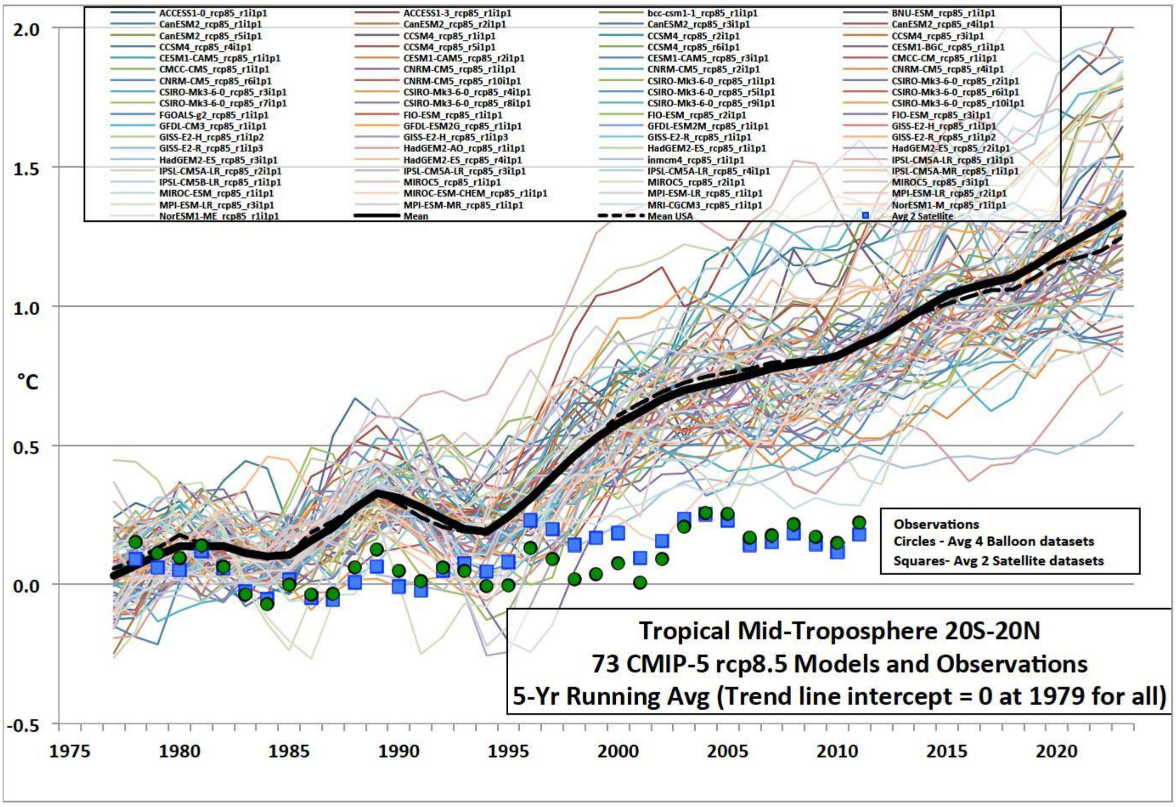

Figure 5, a repeat of Christy's figure 1. Time series of the tropical mid-tropospheric temperature (TMT) of 73 CMIP-5 climate models (RCP8.5) compared with observations (circles are averages of the four balloon datasets and squares are averages of the two satellite datasets.) Values are running 5-year averages for all quantities. [There are four basic RCP emission scenarios applied to CMIP-5 models, but their divergence occurs after 2030. Thus, for our comparison which ends in 2012, there are essentially no differences among the RCP scenarios.] The model output for all figures was made available by the KNMI Climate Explorer.

McKitrick in a public comment confirmed these conclusions of Christy.

> UAH and RSS are individually and jointly significantly different from (i.e. below) the models at the 1% and 5% levels respectively. HadAT and RICH are jointly significantly different from (i.e. below) the models at the 1% level. (We didn't test them individually.) All 4 series averaged together have a trend significantly different from models at 5%. So I conclude the data are significantly below MT [Mid Troposphere] model trends.

Although both Mears and Sherwood focused on the amplification aspect of the hot spot they actually agreed in their guest blogs and comments that the differences between the models and observations are large or even significant.

For example in his first comment Mears said :

> I think all three of us agree that the observed temperature changes in the tropics (and globally) are less than predicted over the last 35 years. John uses this fact to argue that there are fundamental flaws in all climate models, and that there results should be excluded from influencing policy decisions. This goes much too far.

Sherwood in a comment wrote:

> I think we all agree that recent warming in the Tropics has been less than we would have expected no matter how it is measured, and I agree this merits further research (and indeed has spurred a flurry of efforts in the last year or two, so rest assured more papers will be coming out looking at this).

In his guest post though Sherwood emphasized there are other model discrepancies that are more interesting, e.g. the decrease in Arctic sea ice that is larger in the observations than in the models. Sherwood assumes sceptics are so focused on the "missing" hot spot because it can be spun into a tale of model exaggeration:

> If I were looking for climate model defects, there are far more interesting and more damning ones around. For example, no climate model run for the IPCC AR4 (c. 2006) was able to reproduce the losses of Arctic sea ice that had been observed in recent decades (and which have continued accelerating since). No model, to my knowledge, produces the large asymmetry in warming between the north and south poles observed since 1980. Models underpredict the observed poleward shifts of the atmospheric circulation and climate zones by about a factor of three over this same period (Allen et al. 2012); cannot explain the warmings at high latitudes indicated by paleaoclimate data in past warm climates such as the Pliocene (Fedorov et al. 2013); appear to underpredict observed trends in the hydrological cycle (Wentz et al. 2007, Min et al. 2011) and in their simulated climatologies tend to produce rain that is too frequent, too light, and on land falls at the wrong time of day (Stephens et al. 2010). Finally, the tropical oceans are not warming as much as the land areas, or as much as predicted by most models, and this may be the root cause of why the recent warming of the tropical atmosphere is slower than predicted by most models (there is a nice series of posts about this on Isaac Held's blog). What makes the "hot spot" more important than these other discrepancies which, in many cases, are supported by more convincing evidence? Is it because the "missing hot spot" can be spun into a tale of model exaggeration, whereas all the other problems suggest the opposite problem?

Christy reacted to this argument by saying that the "tropical atmospheric temperature is key to model fidelity":

> When pointing out other model problems Sherwood notes, as an example, that no climate model has replicated the rapid north polar ice loss and that this is an interesting problem. However, none of the models have shown an increasing extent of sea ice in the southern hemisphere either – so we have a problem at both poles for which models have opposing answers and thus opposing issues to solve. There are other such examples of models overwarming the climate. However, as stated in my original post, the importance of the tropical atmospheric temperature is key to model fidelity to the real world because it involves the complicated and ubiquitous interrelationships among the various water components of the climate system.

And in another comment Christy wrote:

> The "hot spot", as I stated earlier, represents an integration of much of our understanding of the energy cycle of the climate system. It is the energy cycle that must be well-characterized before attempting to forecast the climate response to a very slight increase in total energy forcing due to the enhanced greenhouse effect. The tropical atmosphere represents about 30% of the global atmospheric mass, holds a significant role of the planetary hydrologic cycle, and is the entry point for about half of the Earth's solar energy. If the processes that combine to create the observed tropical structure, variations and change are not understood and replicated well, then we cannot claim we know enough about the system to make confident

predictions. Thus, I agree with the instigators of this blogpost, by saying "…DON'T move along now, because there IS something to see here."

In a later comment Mears wrote (our bold):

> None of this changes my overall conclusions:
>
> 1. The presence (or not) of the tropospheric hotspot depends on which pair of datasets you use. Thus the result is not statistically significant in the grossest sense.
>
> 2. Measured trends in the tropical troposphere are less than all of the modelled trends (or almost all in the case of STAR 2.0). This is an important, **statistically significant**, and substantial difference that needs to be understood. I addressed this in my last post.

Here his conclusion 1 refers to the amplification factor which we discussed before. With the second conclusion he agrees with Christy and McKitrick. In our introductory article we mentioned the debate that lasted for years and was subject of two papers by Douglass et al (on which Christy was a co-author) and Santer et al (on which Mears and Sherwood were co-authors) in 2008.[xiii] At the time Douglass et al already claimed a statistically significant difference between models and observations. Santer et al. disagreed and said Douglass et al underestimated the uncertainties, that the differences were not significant.

This point was repeated in a public comment by GavinCawley :

> It is worth noting that the statistical test used in Douglass et al. (2008) is obviously inappropriate as a perfect climate model is almost guaranteed to fail it! This is because the uncertainty is measured by the standard error of the mean, rather than the standard deviation, which falls to zero as the number of models in the ensemble goes to infinity. (…)
>
> Had we used +/- twice the standard deviation, on the other hand, the perfect model would be very likely to pass the test. Having a test that becomes more and more difficult to pass as the size of the ensemble grows is clearly unreasonable. The spread of the ensemble is essentially an indication of the outcomes that are consistent with the forcings, given our ignorance of the initial conditions and our best understanding of the physics. Adding members to the ensemble does not reduce this uncertainty, but it does help to characterise it.
>
> The thing that really concerns me though, is that Douglass and Christy (2013) discuss their earlier paper quite uncritically, despite the statistical shortcomings of that paper having been widely discussed, both on-line and in the peer-reviewed literature.

Although Christy replied to this comment he didn't specifically address the criticism of using the standard error of the mean instead of the standard deviation. Mears and Sherwood didn't join this specific discussion about the statistics in Douglass (2008).

However it seems that the supposedly statistical errors in Douglass (2008) have become irrelevant for the discussion as more recent papers like those of McKitrick[xiv] (using more advanced statistical tools) confirm that there is a statistically significant difference between the models and the observations in the tropical troposphere and Mears and Sherwood also accept that. This is a step forward in the debate and we can now focus on the possible reasons for this discrepancy between models and observations.

Sherwood (when asked later by email) agreed with this conclusion (that model trends differ significantly from the observations) for the satellite period (since 1979) but not for the full period since 1958.

Table 6

	Sherwood	Christy	Mears
Are models showing significantly more tropical tropospheric warming than observations?	Yes (since 1979) No (since 1958)	Yes	Yes
Is this difference between models and observations an important issue for understanding anthropogenic global warming (e.g. attribution or sensitivity)?	Somewhat, but other model-observation discrepancies are less uncertain and more interesting	Yes, tropical atmospheric temperature is key to model fidelity	Yes, but no reason yet to dismiss the models

What explanation(s) do you favour regarding the apparent discrepancy surrounding the tropical hot spot?

Given the confusion that existed about the definition of the hot spot, answers to this question sometimes referred to the narrow definition of the hot spot (amplification of surface trends higher up in the tropical troposphere) and sometimes to the broader definition (magnitude of warming expected in the tropical mid-troposphere). In our introduction we gave several suggestions of possible reasons which we will discuss below.

Suggestion1: Satellite data show too little warming
As we mentioned earlier in the discussion of the datasets, Sherwood regards this as a serious option. Sherwood et al. 2008 suggests there has been more cooling in the stratosphere than anyone has reckoned and thus the true upper-tropospheric warming could be stronger than what any group now infers from the satellite data. Mears and Christy did not react on this possibility.

Suggestion 2: Surface data show too much warming
Both Mears and Sherwood find this highly unlikely. In a comment to McKitrick Sherwood said (our bold):

> Ross McKitrick seems to be implying that we should not trust the surface warming record, and should regard the atmospheric temperature record as a separate measure of climate change which somehow discounts global warming itself. He and others seem unwilling to accept the clear evidence that **the surface and near-surface warming records are**, collectively (that is including independent ocean surface, near-surface maritime, and 2-meter terrestrial records which at least over the 20th century are all quite consistent) **far more trustworthy and solid than the dodgy free-atmosphere trends**.

Mears in his guest blog wrote:

> It has been suggested that the lack of a tropospheric hotspot (if there is such a lack) is mostly due to errors in the surface temperature datasets, which are (in this story line) suspected of being biased in the direction of too much warming. This seems unlikely. Clearly, the above spread in results for different upper air datasets reveals considerable structural uncertainty (Thorne et al, 2005) for the upper air data, and the error bar on the RSS trend values is much larger than the error bar for the HadCRUT4 value. Also, the various surface datasets are much more similar to each other.

Christy on the other hand indeed claims there is evidence that the surface temperatures have a warm bias:

> Sherwood indicates that the surface temperature record is robust for climate purposes. I have three comments to make here. First, we and others have shown that the land surface record, as represented by daily mean temperature, is likely contaminated by a warming night-time trend due to surface development around the world (e.g. Christy et al. 2006, McKitrick and Michaels 2007, Christy et al. 2009, McKitrick 2010, McKitrick and Nierenberg 2010, McNider 2012, Christy 2013). Secondly, if the surface temperature is the most robust and important metric, we find that the 73 models shown in the earlier post, on average, produce a surface trend in the tropics that is almost twice too warm since 1979 even with the contaminated observational data (+0.19 vs. +0.11°C/decade).[xv] Thus, even with the surface temperature metric, there are problems for models.

So Christy in fact is saying that probably at least the land surface record (in the tropics) has a warm bias. If corrected this could then mean that after all surface warming is amplified aloft. However it doesn't change the fact that models overestimate the warming both at the surface and in the troposphere. This discrepancy between models and observations would then become even bigger.

In his guest blog Sherwood also hinted that surface trends in the oceans are lower than expected and this could play a role in the missing hot spot discussion:

> [T]he tropical oceans are not warming as much as the land areas, or as much as predicted by most models, and this may be the root cause of why the recent warming of the tropical atmosphere is slower than predicted by most models

Suggestion 3: The theory (of moist convection leading to more tropospheric than surface warming) overestimates the magnitude of the hotspot.
Sherwood tends to **agree** because he indicates that convective schemes in global atmospheric models need improving. Current schemes enforce the theoretical moist-adiabatic lapse rate quite strongly and it might turn out that they are too heavy-handed in this respect, and that a better model would anchor the upper tropospheric temperature less firmly to the surface temperature. Other problems with models, such as difficulties in generating proper hurricanes or a tropical phenomenon known as the Madden-Julian oscillation, may also derive from the schemes triggering convection too easily and enforcing these lapse rates too vigorously.

Mears indicates that temperature changes in the upper troposphere and lower stratosphere have been shown to be very sensitive to the stratospheric ozone concentrations used (Solomon et al, 2012). The ozone dataset used in the CMIP-5 simulations is the one with the most conservative trends in ozone. If one of the other datasets had been used, the models would have shown less upper tropospheric warming and the hot spot would have been overestimated less.

Christy **agrees** with respect to the likely tendency of models to tie the surface trends too tightly to upper tropospheric temperature. *Christy* adds that the fact the average model is accumulating heat in the upper atmosphere at a rate three times faster than the observations has serious implications for representing the entire climate system.

Table 7

What explanation(s) do you favour regarding the apparent discrepancy surrounding the tropical hot spot?	Sherwood	Christy	Mears
Satellite data show too little warming	Likely	Unlikely	As likely as not
Surface data show too much warming	Very unlikely	Very likely	Unlikely
The theory is not perfect yet due to issues with fundamental model physics (e.g. water vapour and cloud feedback)	Unlikely	Very likely	Unlikely
The solar, ozone and stratospheric aerosol forcings in the CMIP5 models may be wrong	As likely as not	Likely (to some extent)	Likely

Options: don't know, very unlikely, unlikely, as likely as not, likely, very likely)

What consequences, if any, would your explanation have for our estimate of the lapse rate feedback, water vapour feedback and climate sensitivity?

Sherwood was quite clear in his guest blog that even if there turns out to be no hot spot the broader implications for global warming would be nil. As he wrote:

> Let us suppose for the moment that the "hot spot" really has been missing while the surface has warmed. What would the implications be?
>
> The implications for attribution of observed global warming are nil, as far as I can see. The regulation of lapse rate changes by atmospheric convection is expected to work exactly the same way whether global temperature changes are natural or forced (say, by greenhouse gases from fossil fuel burning).
>
> **The implications for climate sensitivity are also roughly nil**. The total feedback from water vapour and lapse-rate changes depends only on the changes in relative humidity in the upper troposphere, not on the lapse rate itself (see Ingram, 2013). In fact, in climate models where the lapse rate becomes relatively steeper as climate warms (as would be the case with a missing hot spot), the total warming feedback is very slightly stronger because the increased lapse rate increases the greenhouse effect of carbon dioxide and other well-mixed greenhouse gases. So a missing hot spot would not mean less surface warming, at least according to our current understanding.
>
> Moreover, the discrepancy with models was opposite from 1958-1979 (Gaffen et al. 2000)—that is to say, the observed tropical upper-tropospheric warming was evidently stronger than expected. But the world was warming then too. So if this interesting phenomenon is real, it probably is not connected to global warming.

Note that the lapse rate change is a negative feedback (it decreases surface warming) while the water vapour feedback is a positive feedback (it increases the surface warming). Note also that Sherwood in this comment argues based on the narrow definition of the hot spot (i.e. the amplification). So here he is saying "if it turns out there is no amplification of the surface trend, it doesn't mean climate sensitivity is smaller, it could be even larger". This is because the hot spot is directly linked to the (negative) lapse rate feedback.

Elsewhere in his guest blog and comments Sherwood did also address the broader aspect of the hot spot, namely models expecting more warming in the tropical troposphere than the observations indicate (our bold):

> Currently none of the explanations I can see for the "missing hot spot" would change our estimate of future warming from human activities, except one: that **the overall warming of the tropics is simply slower than expected**. **It does seem that global-mean surface warming is starting to fall behind predictions**, and this is particularly so in the tropical oceans (though not, curiously, on land). Possible causes are (a) aerosols, solar or other forcings have recently exerted a stronger (temporary) cooling influence than we think; (b) negative feedbacks from clouds have kicked in; or (c) the oceans are burying the heat faster than we expected. If (b) were true, we would revise our estimates of climate sensitivity downward. There are observations supporting options (c) and to a small extent (a), but there is plenty of room for new surprises. If it is (c) (which appears most likely), we then have to decide whether this is a natural variation or if it is a feature of global warming. In the former case the heat will soon come back; in the latter, the oceans will delay climate change more effectively than we thought. Another decade or so of observations should reveal the answer.

So Sherwood's favourite reason for the lack of warming in the tropics is that oceans are burying heat faster than expected.

Mears in his first comment gave several reasons for the discrepancy between models and observations. He summarized them in three categories: bad luck, bad forcings and bad model physics. With bad luck he meant that decadal variability which is not well represented by a model mean in reality went the other way. With bad forcings he meant that some forcings are maybe different than we expected. He gave several examples: stratospheric aerosols from volcanic eruptions, solar output, stratospheric ozone and black carbon aerosols. With bad model physics he meant that models are known to have difficulty with such things as clouds and aerosols. And like Sherwood Mears said there "is some evidence that heat is being subducted into the ocean at a rate higher than the models expect, though exactly where it is going is less clear (Balmaseda et al,, 2013, Levitus et al., 2012)." Mears thinks the discrepancy has several reasons:

> In summary, there are a large number of possible explanations for the model/measurement discrepancy in recent temperature rise. Only a few of these, such as errors in cloud feedback, affect the long-term predictions, while others, such as errors in the natural forcings used as model input, or simulated ocean heat uptake do not. At this time, we simply do not know the exact cause or causes, but I strongly suspect that it is due to a combination of causes rather than one dominant cause.

In his guest blog Christy writes there is no clear answer yet why models overwarm the troposphere compared to observations by such large amounts:

> While there is much that can be discussed from these results, we wonder simply why the models overwarm the troposphere compared with observations by such large amounts (on average) during a period when we have the best understanding of the processes that cause the temperature to change. During a period when the mid-troposphere warmed by +0.06 °C/decade, why does the model average simulate a warming of +0.26 °C/decade?
>
> Unfortunately, a complete or even satisfactory answer cannot be provided. Each model is constrained by its own sets of equations and assumptions that prevent simple answers, especially when all of the individual processes are tangled together through their unique complex of interactions. The real world also presents some baffling characteristics since it is constrained by the laws of physics which are not fully and accurately known for this wickedly complex system.

Christy then sums up several possible reasons. Like Sherwood and Mears he mentions the heat going in the deep ocean. But unlike Sherwood Christy finds this possibility unlikely:

> This requires extremely accurate measurements of the deep ocean (better than 0.01 °C precision) which are not now available comprehensively in space and time. Current studies based only on observations suggest this enhanced sequestration of heat is not happening.

Christy's favourite option is that cloud and water vapour feedback (which are positive in the models) in reality are less positive or even negative:

> We have actually measured large temperature swings that were preceded by changes in cloudiness in our global temperature measurements. So a response to the extra CO2 forcing by clouds and water vapour, which have a massive impact on temperature, could be the reason for the rather modest temperature rise we've experienced (Spencer and Braswell, 2010).

But the bottom line according to Christy is we don't know yet:

> The bottom line is that, while I have some ideas based on some evidence, I don't know why models are so aggressive at warming the atmosphere over the last 34 years relative to the real world. The complete answer is probably different for each model. To answer that question would take a tremendous model evaluation program run by independent organizations that has yet to be formulated and funded.

Christy does think though that the lack of warming in the tropical troposphere suggests the climate is relatively insensitive to CO2 forcing:

> With so much more to learn, and the apparent relative insensitivity of the climate system to CO2 forcing as demonstrated by very modest temperature trends, I believe we are in a situation to question the presumed outcomes of specific carbon-control proposals which will also have tremendous economic impacts. These outcomes are based on model projections which to this point have low credibility in my view.

Mears replied that many imperfect models are used to inform policy makers in many areas, including models of the economy, population growth, environmental toxins, new medicines, traffic flow, etc. etc. Policy makers are used to dealing with uncertain predictions. If we throw out all imperfect models, we will be reduced to consulting the pattern of tea leaves on the bottom of our cups to make decisions about the future.

Table 8

	What do you consider the most likely cause for explaining the lack of warming in the tropical troposphere?
Sherwood	Radiosonde and satellite data likely underestimate atmospheric warming and/or negative forcings
Christy	Water vapour and cloud feedback and therefore climate sensitivity smaller than we thought
Mears	A combination of internal variability, heat subduction into the deep ocean, and solar, volcanic aerosol, and ozone forcing

[i] University of Alabama in Huntsville (UAH) satellite temperature dataset

[ii] Douglass DH, Christy JR, Pearson BD, Singer SF. A comparison of tropical temperature trends with model predictions. Int J Climatol 2008, 27:1693–1701

[iii] Santer, B.D.; Thorne, P.W.; Haimberger, L.; Taylor, K.E.; Wigley, T.M.L.; Lanzante, J.R.; Solomon, S.; Free, M.; Gleckler, P.J.; Jones, P.D.; Karl, T.R.; Klein, S.A.; Mears, C.; Nychka, D.; Schmidt, G.A.; Sherwood, S.C.; Wentz, F.J. Consistency of modelled and observed temperature trends in the tropical troposphere. Int. J. Climatol. 2008, doi:1002/joc.1756

[iv] The laps-rate is the rate of decrease of atmospheric temperature with increase in altitude. Unsaturated air loses about 1°C per 100 m which is the dry adiabatic lapse rate, whereas saturated air loses an average of 0.5°C per 100 m, i.e. the saturated or Moist Adiabatic Lapse Rate (MALR).

[v] Douglass DH, Christy JR, Pearson BD, Singer SF. A comparison of tropical temperature trends with model predictions. *Int J Climatol* 2008, 27:1693–1701

[vi] Douglass, D.H. and J.R. Christy, 2013. Reconciling observations of global temperature change: 2013. Energy and Env., 24, 415-419.

[vii] For MSU/AMSU, T_{TT} is equal to 1.1*TMT – 0.1*TLS where TLS stands for the Temperature of the Lower Stratosphere.

[viii] These conclusions were at the time under review and meanwhile published in McKitrick, Ross R. and Timothy Vogelsang (2014) HAC-Robust Trend Comparisons Among Climate Series with Possible Level Shifts, Environmetrics DOI: 10.1002/env.2294.

[ix] The target factor accounts for errors due to changes in temperature in the hot calibration target.

[x] Christy: "NOAA-11 crossed the equator around 1:30 local time but then drifted to at least 5:00 over the next several years. UAH cuts off NOAA-11 and NOAA-14 earlier than RSS to avoid the extreme corrections (and their associated error) required for the longer drifting."

[xi] According to Christy adjustments for NOAA-14 are even more important than those for NOAA-11.

[xii] CMIP-5 is the Coupled Model Intercomparison Project phase 5. C*MIP-5* is meant to provide a framework for coordinated climate change experiments and thus includes simulations for assessment in the Fifth Assessment Report of IPCC (AR5) as well as others that extend beyond the AR5.

[xiii] Douglass DH, Christy JR, Pearson BD, Singer SF. A comparison of tropical temperature trends with model predictions. Int J Climatol 2008, 27:1693–1701; Santer, B.D.; Thorne, P.W.; Haimberger, L.; Taylor, K.E.; Wigley, T.M.L.; Lanzante, J.R.; Solomon, S.; Free, M.; Gleckler, P.J.; Jones, P.D.; Karl, T.R.; Klein, S.A.; Mears, C.; Nychka, D.; Schmidt, G.A.; Sherwood, S.C.; Wentz, F.J. Consistency of modelled and observed temperature trends in the tropical troposphere. Int. J. Climatol. 2008, doi:1002/joc.1756

[xiv] These conclusions were at the time under review and meanwhile published in McKitrick, Ross R. and Timothy Vogelsang (2014) HAC-Robust Trend Comparisons Among Climate Series with Possible Level Shifts, Environmetrics DOI: 10.1002/env.2294.

[xv] Christy used the average of the NCDC and HadCRUT3 datatsets.