The paper by Masters concerns the issue of short-term variations in the rate of global temperature rise. This issue has met considerable public interest recently, even though scientists tend to regard short-term trends with great caution due to their lack of robustness.

The paper consists of three main parts (corresponding to sections 3, 4 and 5) that I will discuss in turn.

Part 1. The first part uses a simple 1-box energy balance model to test whether a mixed climate signal (consisting of ENSO, solar cycles, two volcanic eruptions and a trend) in this model can be decomposed with linear regression methods.

The first finding is that *if* the climate response is described by a long e-folding time scale, then the recovery of the signal components by linear regression is poor, while it is much better using an approach that explicitly includes such an e-folding time scale. This is obvious even without performing the test, but it hinges on a huge '*if*'. The key question is whether the real climate system is well described by such a single e-folding time scale, and as the author himself shows in the next section with proper climate models, the answer is *no*. That is why we decided against including a "memory" in form of an e-folding time scale in Foster & Rahmstorf (2011).

The second finding of Masters is that the regression performs poorly in that it overestimates the response to solar cycles in the 1-box model. The author tacitly assumes that it overestimates this also for the real world and puts this down to there being "only a few solar cycles" to fit. However, that is very unlikely, because almost the same amplitude in the response to solar cycles was also found by Lean and Rind (*GRL* 2008) with data starting in the 1880s, thus including 11 solar cycles. Recovering a quasi-periodic signal from noisy data when 11 full cycles are available is quite robust, and Lean and Rind found that global temperature lags solar forcing by just 1 month, in full agreement with our regression analysis starting in 1979. This rapid response time scale found in the data is incompatible with the model of Masters, where he considered response time scales of 14, 57 and 194 months. By design his model excludes the important rapid response component, which is the one captured by linear regression.

The fallacy of using a single time scale can be easily understood by looking at a slightly more realistic energy balance model, one that consists of one third land (including ice-covered ocean) area with essentially zero heat capacity, and two thirds ocean area with a mixed layer heat capacity. In that case on third of the temperature response would occur immediately and two thirds with the e-folding time scale considered by Masters. For his standard case of a 57-months mixed layer time scale, his model would get 19% of the response in the first year. The model that includes land surface would get 46% of the response in the first year. The one-box model greatly underestimates the initial rapid response of the climate system not only by ignoring the land-covered part of our planet but in addition by ignoring the fraction of radiative forcing absorbed straight in the atmosphere. Of the absorbed incoming solar radiation ~30% is absorbed in the atmosphere before reaching the surface.

The need for including multiple response time scales has often been discussed in the literature, recently e.g. by Caldeira and Myhrvold (*ERL* 2013) who fitted exponentials to GCM output and found that two or three time scales were needed – even though they ignored the sub-annual response by using annual-mean data in the analysis. They write that "substantial temperature changes over land

are observed to occur within days in climate model simulations of step-function changes in radiative forcing. The median value of the shortest time constant in the 3-exp fits was 0.6 years, which is less than the annual resolution used in this analysis."

Part 2. In the second part, a similar exercise as with the simple box model is repeated with an ensemble of global climate models. The results are strikingly different: as can be seen in Figs. 5 and 6, the volcanic response is captured rather well by linear regression. This result invalidates the conclusion of the first part, namely that the box model performs much better than linear regression. This is despite of the evidence that global climate models significantly overestimate the response to volcanic eruptions both in amplitude and duration (see section below). With this in mind, the linear regression model very likely would perform even better for the real world than for the global climate models.

Remarkably, Masters does not say what e-folding time k was found for the fit shown in Figs. 5 and 6 and how this compares to the values assumed in the previous section, but the proximity of the curves for Method 1 and Method 2 shows it must be much shorter than any of the values considered before.

For ENSO, Masters found in the first part that the regression analysis performs very well, but here again he comes to the opposite conclusion with the GCMs, finding a poor performance (even the wrong sign for post-2000 temperatures) and concluding that "the proxy for ENSO (the Niño3.4 SST in this case) diverges from the actual influence of ENSO". This problem is not relevant for Foster and Rahmstorf (2011), since we did not use the Niño3.4 SST but the multivariate ENSO index as a more comprehensive index measuring the amplitude of El Niño and La Niña events. Also, it needs to be tested whether the linkage between ENSO and global-mean temperature might differ between models and the real world.

Part 3. In the third part of the paper Masters proposes a new way of adjusting the global temperature time series. It relies on using the output of climate model simulations in order to subtract the solar and volcanic signals from the data.

There are two problems with this: (a) such climate model ensembles are typically only available with at least five years delay, so this approach is not useful for operational adjustments of global temperatures on an annual basis or so, and (b) this approach assumes that the models are reliably capturing the temperature response to volcanoes and solar cycles (as the author writes himself). In both cases this is not obvious. Evaluating the performance of this proposed adjustment procedure would require that the author makes an attempt at estimating error bars on the simulated volcanic and solar response, which one would expect from a paper that proposes such a method. Below I give some reasons for caution about the models in this regard.

Then in a second step Masters corrects the global temperature series for ENSO using the Niño3.4 SST-based method which he found to give the wrong sign of response in part 2. This is rather surprising, as one might expect that the method passing the test would have been the basis for its application for a proposed new temperature adjustment. Masters gives no scientific reason for using Niño3.4 instead of the more comprehensive multivariate ENSO index, but only motivates it with the "ease of calculating".

Based on these questionable adjustments, Masters then claims a "significant deceleration" of global warming. However, this claim of significance is not based on any systematic uncertainty analysis, and in my view none of the sweeping claims made in the conclusions section are properly supported by sound methods in this paper.

Do climate models properly capture the fast time scale of response to volcanic eruptions?

It is well-known that climate models tend to smear out rapid time scales because of their limited grid resolution, leading to numerical diffusion and generally "over-mixed" models both in the horizontal and vertical. Hansen et al. (*Atmos. Chem. Phys.* 2011) extensively looked at the response functions of climate models and conclude: "We believe, for several reasons, that the GISS modelE-R [also used by Masters] response function is slower than the climate response function of the real world" and that "the slow response ... is typical of most IPCC models." This conclusion is directly relevant for Masters' statement that his "method relies to some degree on the multi-model-mean approximately matching the actual response time of the real world".

This too-slow model response can be seen clearly for the example of the Pinatubo eruption in IPCC models, see the following figure.
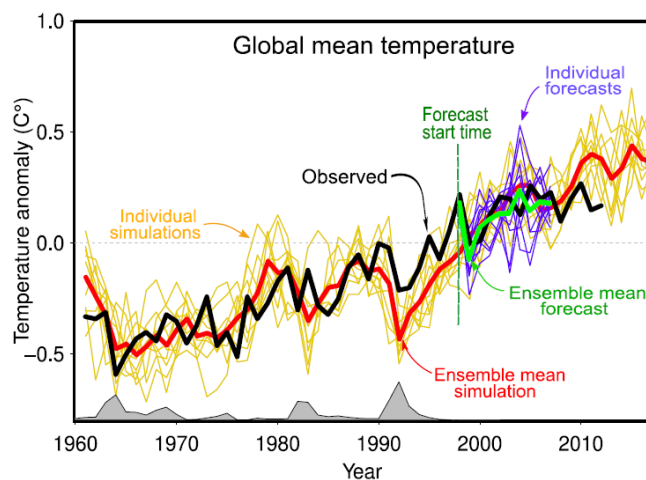
Figure: Observed vs. modelled temperature, from Box 11.1 of the IPCC AR5 WG1 report. Note that all models are too cold starting from the Pinatubo eruption in 1991 for at least five years.

More systematic comparisons for volcanic eruptions over the last 400 years, using coral data as reliable high-resolution recorders of sea surface temperature, show that climate models systematically overestimate both the amplitude and duration of the response to the eruptions (Tierney 2013).

Masters already published the basic arguments of the current manuscript a year ago in his blog where he concluded:

"From what I can tell, in this case, the multiple regression approach used by Foster and Rahmstorf (2011)... can produce misleading results, even failing to recognize a pause in the underlying signal. [...] Of course, in this particular case, much of it seems to stem from a lingering effect of the Pinatubo recovery into the 21st century, which I am currently skeptical exists in the real world."

That this skepticism is highly justified is shown by the GCM results shown in the manuscript, combined with the fact that even those GCMs very likely overestimate the response time scale.

Conclusion

In Foster and Rahmstorf (2011) we deliberately used a simple, straightforward regression analysis to obtain the contributions of ENSO, solar cycles and volcanoes to the global temperature evolution. The idea was to derive those straight from data without the use of models, since numerous modelling studies to interpret the temperature evolution already exist, but the ability of models to correctly reproduce these effects remains disputed. Thus a complementary, purely data-based approach seemed useful. Our regression analysis relies on the assumption that a rapid temperature response without long memory is a reasonable approximation.

Masters calls this into question in his manuscript, which is a comment on our paper that is based on an earlier blog post on it. His argument hinges on the existence of "a lingering effect of the Pinatubo recovery into the 21st century", but he provides no evidence for its reality. In the first section he simply uses an inadequate model where such a lingering effect is basically built in from the start due to the inclusion of only one slow response time scale. In the second section he disproves this himself when his e-folding model is fitted to GCM results and can only fit those when a much shorter time scale is used, so that his energy balance model results become almost identical to our regression results – and this despite the fact that also the GCMs likely overestimate the response time scale. Finally he proposes a way to adjust observational data with the help of model simulations and an ENSO adjustment which earlier he showed to perform poorly, but he does so without proper uncertainty analysis and the whole approach is questionable in the way it conflates models and observational data. A better approach is to use climate models to interpret rather than adjust the data, by careful comparison of model results with data, as is done e.g. in the IPCC reports.