Dear Doug Keenan,

Thank you for your comments. It seems that we agree that there should be a new way forward in summarising the uncertainty in the observations chapter of the forthcoming IPCC report. A linear trend, while conveniently easy to understand and apply, is simply inadequate to capture all of the timescales that are apparent in the Earth system. To be fair to the authors of the chapter, they have tried to express the uncertainty in the global mean temperature in other ways, for example by comparing a periods at the beginning and end of the timeseries. Further, the statistical assumptions about trends made in this chapter are not crucial to the detection of global warming, and its attribution to largely anthropogenic influences. I think, however, that the "observations" chapter of the forthcoming IPCC summary would ideally just report on the state of the climate, and leave any assessment of detection and attribution to other chapters.

Aside from that, a colleague reminds me that the first "uncertainty issue" that had to be dealt with by chapter 3 was that of the observations themselves. The first question to ask upon seeing an apparent trend was "could this be an artifact of an inadequate observational network?". It is easy to see how this question might get conflated with "is the trend unexpected", and "what are the causes of the trend?". Given the estimated observational uncertainty, it would be hard to imagine arguing that the Earth's surface has not warmed, regardless of the driving processes.

Thanks for clarifying that you think that GCMs are the best way to understand the climate system. While a focus on observations is good, this discussion only highlights how difficult it is to examine observational evidence in isolation from our physical understanding, which often finds its expression in models of various complexity.

I must take issue with one of your comments however, you say that:

"The full situation is even worse: *there does not seem to be any statistically-valid observational evidence for global warming"*

and go on to say

"What would happen if policy makers were officially told that there is no observational evidence for global warming?"

As highlighted in my previous email, there is a great deal of observational evidence that agrees well with our physical understanding of the system - much "physically valid" observational evidence, if you like.

However, I think that it is important to be very clear what we mean by "statistically valid", before we decide whether the evidence that we have conforms to that or not. What would statistically valid evidence for global warming look like? Perhaps more usefully, what would statistically valid evidence *against* the theory of anthropogenic induced warming look like?

To say that observational evidence is not "statistically valid" is probably more a comment on our statistical framework, than our knowledge of the climate. I think it is unfair to make the argument that there is no "statistically valid" evidence, without stating what statistical framework we are working under, and then going on and showing that the evidence for warming is not statistically valid. To further extend this and say that there is "no observational evidence" for global warming is stretching the point even further.

With that in mind, I think your instinct to question the statistical models used in climate science is a good one. We should certainly test a number of timeseries (including those from models) against a number of statistical models and assumptions. You have, however, made a good point about the inappropriateness of a linear trend assumption for global mean temperatures. It would seem unfair then, to test our best statistical models against this statistical model? The appropriate test is surely against our physical models, worked out from first principles, or against a physical-statistical model?

A good example of this approach is used in a paper that I was only made aware of since last writing to you. The paper is Mann (2011), where the author looks at long range dependence in global temperature series. The author claims that the value of the Hurst coefficient seen in the observational temperature record, can best be reproduced in a very simple climate model, with a linear response to natural and anthropogenic forcing, combined with stochastic noise. It also finds that purely stochastic forcing can produce the observed effect, and that the observational record is perhaps too short to distinguish between the two. The structure of the paper seems to me to be a good template for a comparison of physical and statistical models. As a bonus, the code for the experiments is available online. The paper is an editorial comment of Rea et al. (2011), which also looks relevant to our discussions, but which I haven't had time to digest yet.

Later in your email, you say:

"The central question, though, is this: what statistical models should analyses be based on?. Any statistical model should have both physical realism and a good statistical fit to the data. The only statistical model of which I am aware that has been based on the underlying physics is fGn."

I think the first sentence is indeed a central question, and I agree with the second sentence. I wonder though, if fGn is really the only statistical model that takes into account the underlying physics? I doubt that this is the case.

I notice that there is a package FGN for R, which looks promising. Regarding your suggestions for fitting fGN to models or observations, I have a number of questions:

1. How we could test that fGn was an appropriate statistical model for climate model or observational records?
2. If fGn does fit either (or both) well, what physical interpretation does this suggest?
3. If we were able to fit fGn models to both GCMs (or simpler physical models) and the observations, what would that tell us about the Earth system that was new?

I think there could be an interesting study in comparing an fGn model against an appropriate physical-statistical model, but we would have to be sure to set up a fair test beforehand, and be careful in interpreting the results.

Kind regards,

Doug McNeall

M. E. Mann (2011) On long range dependence in global surface temperature series, an editorial comment, *Climatic Change*, Volume 107, Numbers 3-4, 267-276, DOI: 10.1007/s10584-010-9998-z

W. Rea, M. Reale and J. Brown (2011) Long memory in temperature reconstructions, *Climatic Change*, Volume 107, Numbers 3-4, 247-265, DOI: 10.1007/s10584-011-0068-y