Met Office

24 September 2013

Dear Mr Lewis,

I refer to your comments on the third report in our series on ‘The recent pause in global warming’ in which we examine, based on the peer reviewed literature, implications of the pause for projections of future warming.

I am aware of your contribution to the literature on estimating climate sensitivity and I respect your efforts. This letter is intended to be a reasoned discussion of the evidence and in particular a response to your main points. I very much welcome your careful reading of the report.

Firstly, one of your main concerns appears to be that we chose to focus our discussion on the results from Otto et al (2013) for the full period, 1970-2009, rather than the most recent decade of 2000-2009. In fact the report does show the results for each of the periods in Table 1; it would be misleading to just give the result for only one of the decade of analyses, particularly one which for which there is increasing evidence that natural decadal variability has been the major determinant for recent global mean surface temperatures, as we discuss in the second report in our series. It is widely accepted that there are advantages to using a longer period of observations to minimise the impact of natural variability and that there are risks in focusing on too short a period to draw fundamental conclusions about trends in global mean temperatures; indeed the original Otto et al. (2013) paper makes these points.

Secondly, you make some strong statements about different methods for estimating climate sensitivity. Language is getting in the way here. You refer to ‘observationally based assessments’, but these methods employ a model for Earth’s energy balance and make certain assumptions, such as there being a linear radiative feedback.

Furthermore, in order to determine the radiative forcing from carbon dioxide alone, these assessments have to calculate the forcing from components other than carbon dioxide again using results from models. Observations alone cannot provide that information. I will return to this point later.

I agree with your comment that ‘*In science, it is standard to test the validity of theoretical models by comparing their predictions to observational data*’. This is fundamental to the way in which we conduct all our research. Climate models are rooted in physical theory and draw heavily on observations from field experiments, satellites and other observing platforms to understand and represent the myriad of processes that make up the climate system. And the model’s performance is comprehensively tested against the real world. Regarding HadGEM2 and its projections of future warming, it is true that it is one of the most sensitive models within the IPCC AR5, but we also know that it is also one of the most skilful in terms of simulating many aspects of the mean climate and its variability.

What you omit to say is that observations of the real world, including those you need to compute TCR and ECS are themselves seriously incomplete and therefore inherently uncertain. Indeed one could argue that models provide a more physically consistent representation of the real world than spatially sparse and poorly sampled observational data. This is why we should look across all estimates and not claim that one method is superior to another. IPCC has correctly used all the evidence to set the range.

I do need to comment on your third key point, the interpretation of the results in Harris et al. (2013). You rightly point out that the results are based on the perturbed parameter ensemble (PPE) approach using the HadCM3 model, though you omit to mention that results from alternative (CMIP3-generation) climate models are also used with the ensemble to form the probabilistic projections that underpin UKCP09. This is a key component that adds sampling of structural uncertainties in model formulation to the methodology.

Having said that, it is true that the relationship between historical aerosol forcing and equilibrium climate sensitivity (ECS) depicted in your Figure B1 is based only on the PPE. But we disagree with your assertion that the results from HadCM3 are fundamentally biased. It is certainly the case that versions of HadCM3 with low climate sensitivity and strongly negative aerosol forcing are incompatible with the broad range of observational constraints. But the key point is that the relationship between aerosol forcing and ECS is an emergent property of the detailed physical processes sampled in the PPE simulations. It is not surprising that such a relationship might be found, given, for example, the key role played by clouds in simulations of both climate sensitivity and aerosol forcing.

A key strength of the Harris et al. approach is the application of multiple observational constraints designed to measure the detailed physical credibility of the simulations. This enables the performance of different model variants to be tested in a more physically comprehensive manner than could be achieved by relying exclusively on a few criteria derived only from historical changes in global or continental-scale surface temperature. This reduces the risk that certain model variants could erroneously receive a high weight because they happen to match historical changes due to a fortuitous cancellation of errors in the effects of different physical processes. One of the important results of the paper is the demonstration that the greater the range of observational constraints, the more uncertainty in the ECS and future projections can be reduced.

You have questioned the correlation between aerosol forcing and ECS in the PPE through a comparison with several other studies based on simple models and observations. From this you infer that HadCM3 cannot support low values of ECS. This is not correct. Firstly, we do explore a wide range of ECS values, as Figure 2S in Harris et al. shows. Secondly there is good evidence that we explore a more appropriate range and distribution of aerosol forcing than the simple model and observational constraint studies you highlight. Indeed, there are numerous recent aerosol forcing estimates that suggest the possibility of large negative aerosol forcing which you do not show.

For example, Shindell et al (2013) report a 5-95% confidence interval for total aerosol forcing (for 2000 relative to 1850) of ~-0.7 Wm^{-2} to ~ -1.7 Wm^{-2} (best estimate ~ -1.2 Wm^{-2}), based on an ensemble of ten contemporary climate models, and using aerosol optical depth (AOD) measurements derived from satellites and sun photometers to screen out less reliable model results. Bellouin et al. (2013) obtained a range of -0.3Wm^{-2} to -3.0Wm^{-2} for the short-wave component of total aerosol forcing, based on a reanalysis approach in which satellite AOD observations are assimilated into global model. Both of these results show support for a best estimate value near or outside the “likely” ranges indicated by any of the studies in the bottom right quadrant of your Figure B1, and for the exploration of negative values approaching or exceeding -2.0Wm^{-2}.

The fact is that the differing relationships between aerosol forcing and ECS found by Harris et al. and (say) Otto et al. reflect fundamental differences in methodological approach: Harris et al. seek emergence of behaviour from detailed physical processes, and *then* explore the consequences of the historical climate record, going way beyond just surface temperature, in shaping the joint space of plausible outcomes. In contrast Otto et al. assume no prior physical understanding of any relationship, and treat aerosol forcing and ECS as independent parameters to be varied in seeking to fit historical surface temperature changes as well as possible. The slope found in Otto et al. and related studies, is an unsurprising consequence of these chosen prior assumptions and observational constraint. All use a paradigm based on much simpler representations of climate system physics, and using less information (based only on gross characteristics of post-industrial climate change) to provide an observational constraint. So it is not surprising that the results are different.

Your comments imply a judgement that one class of climate sensitivity estimates is inherently more reliable than the other. Rather, such differences reflect different experimental design choices, the degree of sophistication in the chosen modelling approach, and the types and extent of observational information used. We support the approach of synthesising the different lines of evidence in an even-handed way, as in the assessments of ECS and TCR provided by the IPCC. In addition, the fact that the final projection ranges used in UKCP09 also include information from the other available CMIP3 models provide the evidence that we have not inappropriately ruled out important areas of the forcing/ECS space, and, more importantly, that our experimental set up gives a credible estimate of future climate change.

I trust that you will take these comments in the spirit in which they are offered – as part of a constructive scientific debate. As I said we appreciate your contributions to the literature on these topics; but the implications of climate change are so profound that it is essential that scientific debate takes place in the appropriate forum. With this in mind I think it is appropriate that further discussion be subject to proper peer review, through the scientific journals.

Yours sincerely,

Julia Slingo.

Professor Julia Slingo OBE DSc