Response by Nicholas Lewis to a letter dated 24 September from Julia Slingo, chief scientist at the UK Met Office, responding to my comments on the Met Office July report "The recent pause in global warming (3): What are the implications for projections of future warming?"

The tone of Julia Slingo's letter is restrained and it contains reasoned discussion of scientific points. That is very welcome. However, the arguments made in the letter do not refute, as they set out to do, the main point I made in relation to the Harris et al (2013) study, namely that it is unsatisfactory since it effectively rules out from the start the possibility that both aerosol forcing and climate sensitivity are modest, the combination that – according to a number of peer-reviewed studies – recent observations are supporting.

Before commenting on what Dr Slingo's letter says, I will point out what it doesn't say. In my commentary on the 3rd Met Office July report, I made a number of allegations that it contained a variety of misrepresentations and erroneous and misleading statements. Dr Slingo's letter barely touches any of those points, save that concerning methods of estimating transient climate response (TCR). In effect, the Met Office has implicitly accepted those criticisms. Yet I have seen no sign that it proposes to correct or withdraw that July report.

I have some detailed comments on particular statements (italicised) in Dr Slingo's response letter:

1. *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.*

   My complaint was not the inclusion in Figure 1 of the report (estimates of TCR) of the Otto et al (2013) results for the 1970-2009 period, but rather that its primary results, those for the 2000-2009 period, were not also included. I did not, as Dr Slingo's letter hints I did, suggest that the report should *just* have given the result for the 2000-2009 period. And while internal decadal variability is obviously a larger factor when considering results from a single decade, the Otto et al uncertainty ranges allowed for that fact.

2. *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.

Whilst I do not disagree with these points, it was the Met Office in their July report who said: "TCR can be estimated in a variety of ways. These include estimates from simulations made with climate models, estimates made from observations, and estimates made by combining climate model and observationally-derived values." I was following their use of language.

3. *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.*

   I did not mention the use of results from the CMIP3 ensemble because that is not relevant to the main points I made. Use of the CMIP3 ensemble results does not enable the HadCM3 model to simulate a climate system with modest aerosol forcing and modest sensitivity. The effects of the CMIP3 ensemble results, reflected in discrepancy terms, appear to be minor in any event: doubling the standard discrepancy factor changes the TCR best estimate and bounds only by about 1%.

   Incidentally, this paragraph reveals some confusion. Although parameters are perturbed in a PPE, the acronym actually stands for perturbed physics ensemble.

4. *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.*

   This is a key paragraph, which in effect concedes that my main criticism is valid.

   I don't dispute the point that in HadCM3 – and very possibly other models – the relationship between aerosol forcing and ECS is an emergent property. That is precisely why HadCM3 is not suitable for a PPE study in which, supposedly, "uncertainty in the response of the climate system to CO2 forcing is comprehensively sampled". To achieve that, it must be possible, by varying the model's input parameters, to obtain all combinations of aerosol forcing and ECS

that are to any extent plausible. That is the case, for instance, for the quite complex 2D MIT climate model used in Lewis 2013. But it is not possible with HadCM3 and at least some other complex 3D climate models. One senior climate scientist who used to work on PPE studies told me that he basically agreed with me: his group had soon worked out that they were of limited use. With the model he was using, it proved impossible even to get ECS down to 3°C by perturbing its parameters. He thought the Met Office were a bit stuck in the mud about their ensembles, due to some combination of corporate image and sunk costs of ensemble creation.

5. *It is not surprising that such a relationship [between aerosol forcing and ECS] might be found, given, for example, the key role played by clouds in simulations of both climate sensitivity and aerosol forcing.*

   True, but irrelevant since the mere existence of such a relationship would say nothing about the form of the equation between aerosol forcing and ECS representing that relationship or the values of the coefficients it contained. There is no reason to think that the effective relationship between aerosol forcing and ECS in HadCM3 (as represented in Figure S2 and equation S9 of Harris et al 2013) reflects the real climate system. A study involving HadCM3 and another model[1], co-authored by one of the Met Office's own scientists, found substantial differences in the nature of the cloud-ECS relationship between the two models, so there is clearly not an accepted standard basic-physics relationship between them. The key point I made remains untouched: if there are no parameter combinations that achieve low aerosol forcing and low climate sensitivity in HadCM3, the PPE study cannot sample that region.

6. *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.*

   That strength cannot overcome the basic problem that HadCM3 cannot sample low aerosol forcing, low climate sensitivity combinations and is therefore an unsuitable model for this PPE study.

7. *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.*

[1] Yokohata, Tokuta, Mark J. Webb, Matthew Collins, Keith D. Williams, Masakazu Yoshimori, Julia C. Hargreaves, James D. Annan, 2010: Structural Similarities and Differences in Climate Responses to CO2 Increase between Two Perturbed Physics Ensembles. *J. Climate*, **23**, 1392–1410. http://journals.ametsoc.org/doi/pdf/10.1175/2009JCLI2917.1

I never claimed that the HadCM3 PPE study did not explore low ECS values, but it only does so by using its statistical emulator to extrapolate beyond the lowest ECS value achieved by HadCM3 itself, and only in combination with highly negative aerosol forcing. My point was that the HadCM3 PPE did not explore combinations of low aerosol forcing and low ECS values. That is evidently because even by perturbing 31 key main atmospheric parameters, and subsidiary aerosol module parameters, HadCM3 was unable to do so. I give Glen Harris and his colleagues credit for trying hard to explore uncertainties, but in this respect they were unable to overcome structural rigidities in HadCM3.

The statements also show a slightly unusual view of what constitutes low climate sensitivity. The reference is presumably to Figure S2 in Harris et al, since there is no Figure 2S. That figure shows that the lowest ECS value directly achieved by HadCM3 is circa 2.1°C; Figure 1a shows the same. An ECS of 2.1°C is not particularly low; all four of the recent observationally-based studies I cited gave best estimates for ECS of below 2.1°C.

8. *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.*

These two statements are respectively invalid and irrelevant, and reveal a worrying ignorance about statistical inference. In observationally-constrained studies, what is important is that all values of the climate system properties being estimated *that are not effectively ruled out by the observations used* should be sampled. Since the Harris et al PPE study did not explore low aerosol forcing, low ECS combinations in the region which several fully observationally-constrained studies found to be quite probable, it clearly did not fully explore the appropriate range of combinations of aerosol forcing and ECS.

For all the studies I highlighted that estimated aerosol forcing from observational constraints, it appears that the inferred values were <u>not</u> affected by any lack of sampling of large negative aerosol forcing values: they were anyway ruled out by the observations used. The observations concerned were not of aerosols, but rather of temperature – particularly of differences in surface temperature evolution between the hemispheres or latitudinal zones. It is standard in studies like these to derive (inverse) estimates of aerosol forcing from such data. Since in Harris et al the PPE was constrained by similar surface temperature observations, it is highly likely that those observational constraints would also strongly contraindicate possible highly negative aerosol forcing values that were not ruled out by the model and reanalysis studies cited.

Moreover, in order for inferred probabilistic estimates for those climate system properties to be realistic, it is also important that the weighting given at each possible combination of their sampled values should reflect how informative the observations are about the climate system properties at those values. That is to say, an objective Bayesian approach (as in Lewis 2013), rather than the subjective Bayesian approach used in the Harris et al study, should be used. This point was made in a presentation given at the recent European Meteorological Society Conference in Reading, and the use of subjective Bayesian methods in UKCP09 adversely commented on.

9. *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.*

It is certainly the case that Harris et al only reflects the 'joint space' of model (almost purely HadCM3) and observational plausible outcomes. However, because the HadCM3 PPE has such a strong bias (100 times or more) against the low aerosol forcing, low ECS region, and the observations do not have a high signal-to-noise ratio, the region of outcomes considered least implausible is dominated by the characteristics of HadCM3, not by the observations. Since the modelling of detailed climate system physical processes is a young science, with many processes poorly understood and/or impracticable to model from basic physics, and others perhaps overlooked, it seems totally inappropriate to let model characteristics dominate over observational evidence. To my mind, doing so represents a major failing in climate science in general, and of the Met Office in particular.

Moreover, although Harris et al was able to use a large variety of climate variables, it in fact reduced these down to just six composite variables when comparing the model output with recent mean climate at the key first stage. None of the thousands of observational variables going to make up the six composite variables represented sub-surface ocean layer temperatures – a surprising omission given that observational studies tend to find such ocean observations to be of critical importance in constraining ECS estimates. In addition, at a later stage the study used four variables relating to historical changes in climate, all of which actually represented surface temperatures.

10. *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.*

    This shows a misunderstanding of Otto et al. That study did not treat aerosol forcing as an independent parameter to be varied in seeking to fit historical surface temperature changes. Otto et al in fact used aerosol forcing as represented in an ensemble of CMIP5 models, although it made an adjustment to the mean aerosol forcing used, in order to reflect the estimated excess of CMIP5 model aerosol forcing over satellite-observation derived estimates. Dr Slingo's description is more accurate in relation to the other three observationally-based studies that I cited (Aldrin et al 2012, Ring et al 2012 and Lewis 2013), although those studies estimated effective ocean vertical diffusivity (and other variables in the case of Aldrin et al) as well as aerosol forcing and ECS, and their estimates were based on historical observations of ocean layer as well as surface temperature changes.

11. *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.*

    The distribution of the varying central aerosol forcing and ECS estimates along the sloping curve shown in the figure in my response to the Met Office Report, and the elongation of their contour regions along that curve, simply reflects the fact that relative uncertainty in aerosol forcing changes dominates relative uncertainty in global surface temperature change, combined with the forms of the basic physical and mathematical relationships involved.

    It is unclear that these four studies used less important information than Harris et al 2013, even though they used a lesser number of climatic variables. The inclusion in these four studies of a variable for ocean layer temperature is arguably of more importance than the fact that thousands of other climate variables contributed towards the six mean climate (not change in climate) variables used in Harris et al.

    The difference in the results of these observational studies from Harris et al's results almost certainly reflects much more the failure of the HadCM3 PPE to explore the region of low aerosol forcing, low ECS combinations that the four observationally-based studies show to have substantial probability, than differences in the observations used to constrain the estimates.

12. *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.*

    What this "even-handed way" means in practice is that estimates from unconstrained simulations by complex climate models are given similar, perhaps more, weight than observational evidence in the assessments of ECS and TCR by the IPCC. That may have been unavoidable prior to the last decade, as the signal of anthropogenic influence was only just emerging from the noise of internal climate variability. But that signal is now strong enough, and observational methods have advanced enough (due in particular to satellite-derived estimates and, recently, the deployment of the Argo network of deep ocean buoys) for a much higher weight to be put on estimates that are fully constrained by observations, even though still involving some use of models. Indeed, it is unclear to what extent estimates by complex models that are not constrained by observations, or only very loosely so constrained, should be considered valid scientific evidence at all.

13. *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.*

    As explained in point 3, the inclusion – via discrepancy terms – of information from other CMIP3 models in the Harris et al study, and hence in the UKCP09 projections, provides no evidence whatsoever that important areas of the aerosol forcing/ECS space have not been inappropriately ruled out. Hence it provides no evidence that the UKCP09 experimental setup gives a credible estimate of future climate change.

    All the inclusion of information from other CMIP3 models, and the resulting minor discrepancy terms, shows is that by suitable variation of its input parameters HadCM3 is able to produce similar simulations of climate change to that produced by the other CMIP3 models in their standard configurations. As combinations of low aerosol forcing and low ECS supported by observations are not represented by those configurations, that does not address the fundamental shortcoming in the Harris et al study, and hence in UKCP09, that I have identified.

14. *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.*

> This appears to represent an attempt to stifle reasoned scientific debate. I am perfectly open to review by my peers. Anyone with expertise in climate science is welcome to try to pick holes in my critique of the Met Office July report (3) and, in particular, in what I say about the Harris et al study. The Met Office has singularly failed to do so, and in the course of its attempts has displayed worrying misunderstandings by its most senior scientist on several important topics. The limitations of peer review are shown by the fact that the fundamental weakness of the Harris et al study – and its sister study Sexton et al 2012 – was evidently not focussed on by the peer reviewers. And as has been commented on elsewhere, persuading a journal to publish a critical comment about a paper it has published is not easy – maybe particularly so when the paper is by scientists at the Met Office, a major source of journal papers.
>
> Peer review certainly has its place. However, often peer review gives a veneer of respectability to work that conforms with a consensus, but is in fact deeply flawed. Conversely, good work that contradicts the consensus may be kept out of the peer-reviewed literature, or at a minimum delayed, by gatekeepers defending consensus positions. The profound potential implications of major climate change, the huge costs involved in mitigation attempts, the unsettled state of scientific understanding of many of the key climate processes involved and the need for members of the public – particularly those with scientific or technical expertise – to place trust in the climate science involved surely all point to the need for an open scientific debate alongside the publication of peer reviewed studies. When protagonists refuse to provide reasoned and convincing defences to non-peer reviewed technical criticisms of their work by other researchers with established expertise in the area involved, that is a signal not to trust their work, irrespective of it having been peer reviewed.

In conclusion, Dr Slingo's letter effectively concedes my main criticism of the Harris et al (2013) HadCM3-based PPE study, that it fails to sample low aerosol forcing, low ECS combinations that several recent observational studies indicate have a substantial probability of representing the real climate system. In effect, her view seems to be that it doesn't matter what the observations imply, because the models rule out the possibility of low aerosol forcing, low ECS combinations. I am reminded of a famous line by Bertold Brecht to the effect of: "The people have failed the government. The government must elect the new people." But the Met Office can no more replace the real climate system with one that agrees with the models than a communist government could replace the people with one that satisfied its ideology.

25 September 2013