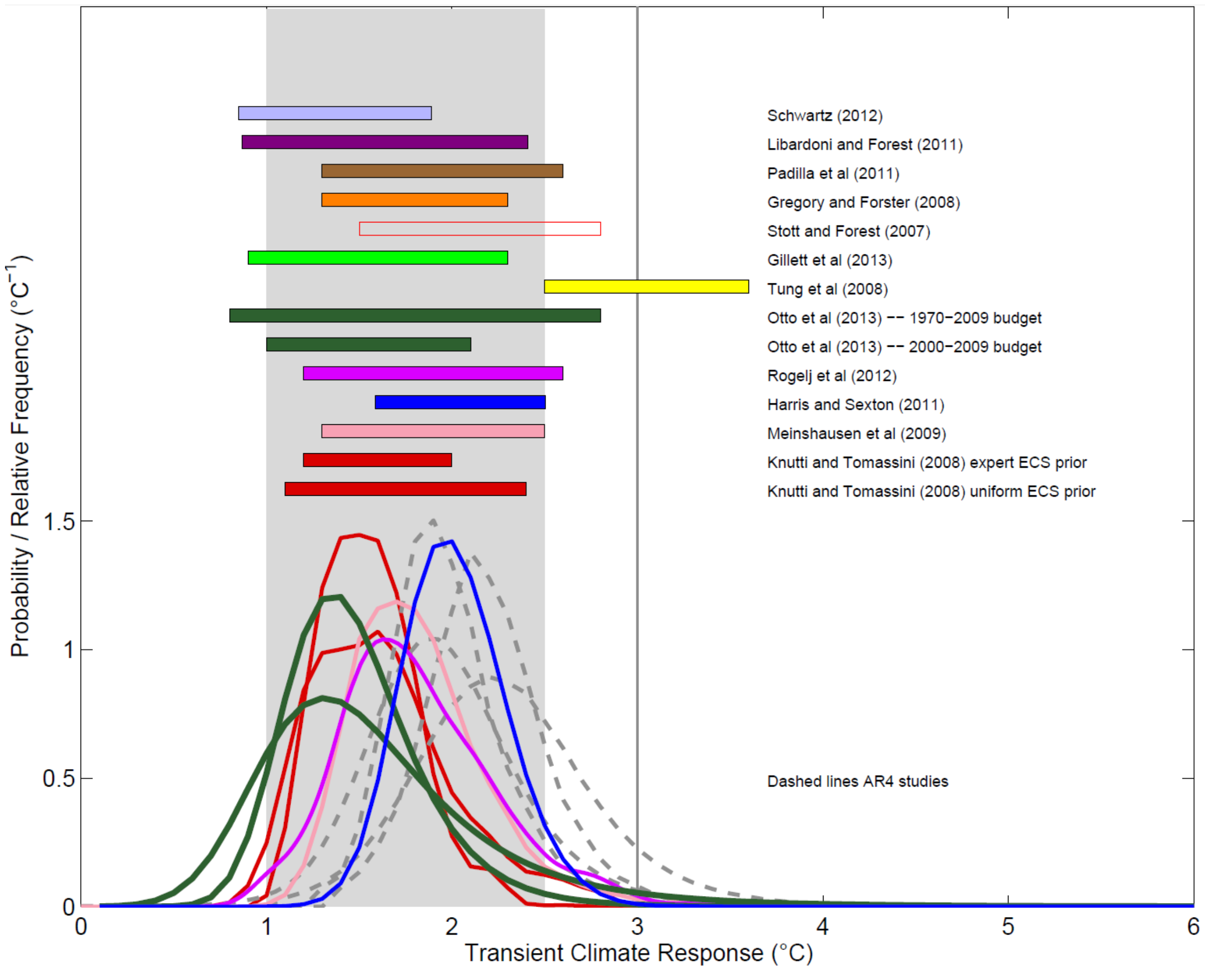
Brief critiques of observationally-based TCR estimates included in AR5 WG1 Figure 10.20.a)

This document provides critical assessments of all the probabilistic transient climate response (TCR) estimates included in Figure 10.20.a) of the accepted version of the IPCC fifth assessment Working Group 1 report (AR5 WG1), published on 30 September 2013. Figure 10.20.a) is shown below.



Reproduction of Figure 10.20.a) from the accepted version of AR5 WG1. Bars show 5–95% uncertainty ranges for TCR. Note that the PDFs and ranges given for Otto et al (2013) are slightly too high in this version of Figure 10.20.a). It is understood those in the final version of AR5 WG1will agree to the ranges in the published study.

a) **Gillett et al (2013)**

This study uses temperature observations over 1851–2010 and a detection-and-attribution regression method to scale AOGCM TCR values. That appears to be a reasonable approach, although scaling results obtained from running models at parameter settings that produce simulations not matching the observations does rest on certain assumptions. The range of 0.9–2.3°C given in Figure 10.20.a) is based on multiplying the mean model TCR by a common scaling factor derived from a single regression using all the models, and then allowing for the spread of the individual model scaling factors and for observational uncertainty. It is unclear whether this is as suitable a method as the alternative of carrying out individual regressions for each model, multiplying the resulting separate scaling factors by the related model TCRs and estimating the range from the spread of the resulting observationally-constrained TCRs. Figure 4(b) of Gillett et al (2013) shows the results of so calculating observationally-constrained TCRs on a model-by-model basis, enabling use of that alternative method. Doing so provides an observationally-constrained TCR range of 0.9–1.9°C, with a median of 1.4°C. This range is from fitting a Normal distribution to the best estimates, after excluding one outlier at each end of the TCR range; both outliers also have poorly-constrained scaling factors. Making allowance for uncertainties in the observations and the individual scaling factors is estimated to widen the range to 0.8–2.0°C, although this still excludes the two outliers.

b) **Gregory and Forster (2008)**

This study based its TCR range on forcing estimates that increased, in aggregate, by only 1.4 W/m^2 over the 1850–2006 period considered. That will almost certainly have biased upwards its estimation of TCR: the updated forcing best estimates in AR5 have a corresponding rise of 1.9 W/m^2. A crude adjustment of the study's 1.3–2.3°C TCR range for the implied 35% over-estimation of TCR would reduce it to 0.95–1.7°C. However, the largest part of the adjustment relates to aerosol forcing, which the study attempted to minimise sensitivity to. Unfortunately, since global aerosol and greenhouse gas forcing time series are almost perfectly negatively correlated, their effects can, at best, only be partially disentangled when using, as this study did, only global mean surface temperature data. Therefore, the study's TCR range is very likely too high, but the extent of its over-estimation cannot be reliably determined.

c) **Harris et al (2013)**

This perturbed physics/parameter ensemble (PPE) study's TCR range, like its ECS range, almost entirely reflects the characteristics of the Met Office HadCM3 model (including its HadSM3 slab-ocean variant). Despite the HadCM3 PPE (as extended by emulation) sampling a wide range of combinations of values for 31 important model atmospheric parameters, none of the actual or emulated cases is able to produce a combination of low-to-moderate climate sensitivity and low-to-moderate aerosol forcing – nor could perturbing aerosol model parameters achieve this. Parameter combinations that produce relatively low ECS and TCR values give rise to extremely negative aerosol forcing; those that produce very high ECS and TCR values give rise to only modestly negative aerosol forcing. Both combinations are inconsistent with observational data. So all the comparison with the observational data does is concentrate the probability distribution in the only area of the model PPE's fairly narrow prior sensitivity and aerosol forcing region that is consistent with the observations, being ECS values in the range 2.4–4.3°C (corresponding to TCR values of

1.6–2.5°C) and total aerosol forcing values in the range -0.6 to -2.0 W/m^2 – on average considerably more negative than the AR5 best estimate of -0.9 W/m^2. Most of the concentration of the probability distribution takes place when the model PPE is constrained by observations of mean climate in stage one of the study. There is almost no change in the posterior PDF for TCR when the recent climate change observations are incorporated in stage two, in which uncertainty in aerosol and ocean model parameters is also sampled. Since the study's ECS and TCR best estimates very largely reflect the characteristics of the HadCM3 model and are only to a minor extent constrained by recent observed climate change, they cannot realistically be regarded as being observationally-based estimates. See http://niclewis.files.wordpress.com/2013/09/metoffice_response2g.pdf, Box 1 for further details.

d) **Knutti and Tomassini (2008)**

This study used estimates of direct and indirect (cloud interactions) aerosol forcing totalling -1.3 W/m^2 in 2000. That is much higher than the best estimate in AR5 of -0.9 W/m^2, which itself exceeds the -0.78 W/m^2 mean per the satellite-observation based studies taken into account when the AR5 aerosol forcing range was formed. Although the method used encompasses inverse estimation of aerosol forcing via scaling factors, only global mean observational temperature data is used, so the inverse estimate arrived at will be unreliable. The very high (negative) correlation between the time evolution of greenhouse gas and aerosol forcings on a global scale makes it impossible robustly to distinguish between different combinations of ECS and aerosol forcing values that each satisfy the energy budget constraint. In addition, the prior distribution used for the indirect aerosol forcing scaling factor rules out values as low as those implied by satellite-observation studies. It appears that the study's resulting recent estimated total aerosol forcing has a mean of about -1.1 W/m^2, well above that per the two inverse estimates of aerosol forcing cited in section 10.8.3 of AR5 WG1 that used latitudinally-resolved surface temperature data. That is unsurprising given the high initial forcing estimate, the use of only global mean temperature and the restrictive priors in this study. The high estimated aerosol forcing will have significantly biased upwards estimation of TCR. Moreover, one of the study's TCR PDFs used a uniform prior on ECS, which will almost certainly have made its TCR estimate 95% bound too high. That upper bound is moreover stated on page 10-60 of AR5 WG1 to be 2.3°C, but shown in Figure 10.20.a) as 2.4°C. Accordingly, both TCR PDFs and associated ranges from this study will be biased high, with the upper bound based on a uniform prior for ECS particularly affected.

e) **Libardoni and Forest (2011)**

This model-multiple-simulations – observation comparison Bayesian study (actually a corrigendum to a study originally published in 2011) uses an informative "expert" prior distribution for ECS, and an inappropriate uniform prior distribution for ocean heat uptake efficiency (the square root of ocean effective diffusivity, K_v: the study states that the prior is uniform in K_v itself, but its computer code shows that to be incorrect). Use of such prior distributions will have biased the study's ECS and TCR estimation. Using one surface temperature dataset, Libardoni and Forest find ECS to be lower, K_v to be completely unconstrained, and aerosol forcing to be more negative, than when using the other two datasets. Yet with greenhouse gas forcing being offset to a greater extent by negative aerosol cooling and more heat being absorbed by the ocean, energy conservation implies that ECS would need to be significantly higher to match the 20th century rise in global temperatures, not lower. Although various errors pointed out in Lewis (2013) were addressed in this corrigendum, at least one was incorrectly dealt with, and the unsatisfactory way surface temperature data was used (see Lewis, 2013, doi: http://dx.doi.org/10.1175/JCLI-D-12-00473.1) was not altered, which may account for these problems. Since there are question marks over the Libardoni & Forest implementation of its methodology and its results defy conservation of energy, they must be considered unreliable. Moreover, the 95% limit of the 0.9–2.4°C TCR range arises from the case where the outdated HadCRUT2 surface temperature dataset is used; the highest upper limit using the other datasets, including HadCRUT3, is 2.25°C.

f) **Meinshausen et al (2009)**

This study does not set out to provide an observationally based estimate of TCR, nor does it do so. Its only PDF for TCR is based on a prior PDF for ECS that was explicitly chosen, for illustrative purposes, so that the posterior PDF for ECS exactly matches the Frame et al (2006) PDF for ECS – selected because it provided the best match to the AR4 likely range and best estimate for ECS. The study's PDF for TCR represents a conversion of the Frame et al (2006) PDF for ECS into a PDF for TCR using the ocean heat uptake characteristics implied by the observations, which since the PDF for ECS has already been chosen have a limited influence on the TCR PDF. Accordingly, the study's TCR range should be seen as representative of the poorer state of knowledge in the years leading up to AR4, not current knowledge (as represented in AR5).

Note that although the paper claims to obtain from the observations, using frequentist statistical methods, broadly similar confidence intervals for ECS to those from Frame et al (2006), it uses aerosol forcing distributions based on those in AR4 (which has a significantly more negative total aerosol forcing distribution than that in AR5). With 82 parameters being estimated using observational data with only 9 degrees of freedom, it seems most improbable that (profile) likelihood frequentist confidence interval estimation will be reliable. Indeed, the AR4-matching total aerosol forcing distribution used appears to be hardly altered by the comparison of model simulations with the observational data, and so will almost certainly have significantly biased upwards estimation of ECS. Therefore, the support given from such confidence intervals to the Frame et al (2006) based PDF for ECS must be regarded as highly suspect.

g) **Otto et al (2013)**

This is an energy budget study, and so should provide robust estimates. The TCR estimates given in Figure 10.20.a) are based on changes, between 1860–1879 and either 2000–2009 or 1970–2009, in observed global surface temperature and in mean total forcing as simulated by the sample of CMIP5 models analysed in Forster et al (2013). The post mid-twentieth century increase in forcing is shown to approximate to a linear ramp, so the ratio of those changes should provide a valid estimate of TCR provided that the forcing estimate used is realistic. The multi-model mean forcing changes were adjusted to reflect an assessment that estimates based on satellite observations indicate recent aerosol forcing to be some 0.3 W/m^2 less negative than in CMIP5 models. Allowing for the different F_{2xCO2} values involved, the resulting change in mean forcing between 1860–1879 and 2000–2009 used in Otto et al is closely in line with the AR5 best estimate of the forcing change between the same periods after adjustment to reflect the mean of the recent satellite-observation derived estimates used to inform estimation of total aerosol forcing in AR5. Since mean total forcing was higher over 2000–2009 than over 1970–2009, and was also less affected by volcanic activity, the TCR estimate based on 2000–2009 data is less uncertain, and arguably more reliable, than that based on 1970–2009 data.

Note: the writer was a co-author of Otto et al (2013).

h) **Padilla et al (2011)**

This paper estimates what it calls TCS, which should exceed TCR on account of heat absorbed by the top 50 to 100 m of the ocean – the mixed layer (MLD) – and should depend on the assumed depth of the MLD. But the estimate of TCS is little changed over an MLD range of 1 m to 400 m. That suggests it is questionable what exactly TCS represents, and that it should not necessarily be considered – as the study assumes – to be almost equivalent to TCR. Moreover, methods like this that seek to estimate the time evolution of aerosol forcing separately from that of greenhouse gas etc. forcing using only global temperature data are suspect, because of the very high co-linearity between the global aerosol and greenhouse gas forcing series. The TCS estimate is sensitive to the particular forcing dataset used. It is also sensitive to the period used. The TCS range based on using post 1970 data only, which the study's authors suggest avoids the period of greatest uncertainty about aerosol forcing, lowers the 90% range from 1.3–2.6°C to 1.1–1.9°C with the median falling from 1.7°C to slightly under 1.4°C. In view of the uncertain relationship between TCS and TCR and the severe difficulty studies that only consider global temperature data have in distinguishing between different combinations of aerosol forcing and climate sensitivity – reflected here in significantly different TCR ranges using data from different periods – the 1.3–2.6°C TCR range preferred by the authors should be regarded as unreliable.

i) **Rogelj et al (2012)**

This study explicitly sets out to generate a PDF for ECS that simply reflects the AR4 likely range and best estimate; in fact it reflects an slightly higher range. The study does not in any sense provide an observational estimate for either ECS or TCR. Moreover, the paper and its Supplementary Information do not even mention estimation of TCR or to provide any estimated PDF for TCR. The TCR PDF and range were presumably derived by the authors from the study's ECS range after its publication. In the circumstances the Rogelj et al TCR PDF and range should be disregarded entirely.

j) **Schwartz (2012)**

This study derived TCR by a zero-intercept regressions of changes from 1896–1901 in observed global surface temperature on corresponding changes in forcing, up to 2009, based on forcing histories used in historical model simulations. The regression slope, which is based on 1965 to 2009 data, will be dominated by the larger changes in the later part of the period, and thus largely reflect the surface temperature change towards the recent part of the post mid-twentieth century period of strongly increasing forcing, thereby approximating to the 70 year linear forcing ramp specified in the TCR definition. The approach therefore appears reasonable subject to the forcing history estimates being realistic and noise not being so high as to prevent strong regression relationships. The change in forcing between 1896–1901 and 1990 (pre the Mount Pinatubu eruption) per the five datasets used to derive the TCR range varied from 1.1 to 2.1 W/m^2. The mean of these estimates is close to the best estimate of the forcing change per AR5 of 1.55 W/m^2. The mean regression R^2 value was reasonably high at 0.6, although it was only 0.3 for the forcing dataset that gave the highest TCR estimate.

k) **Stott and Forest (2007) – range shown by outline bar only: based on an AR4 study**

This TCR estimate is based on the analysis in Stott et al (2006), discussed below under AR4 studies and found to have been biased upwards by almost 40% through use only of 20th century temperature observations. Indeed, it appears that the Stott and Forest (2007) range is a virtual duplication of the combined-models range estimated in Stott et al (2006), which is represented by one of the unlabelled dashed line AR4 PDFs in Figure 10.20.a). Stott and Forest (2007) uses a slightly different way of deriving a TCR range from the Stott et al (2006) greenhouse gas (GHG) regression coefficients for the three CMIP3 models it used. Each model's regression coefficient is used to scale its simulated GHG-caused 20th century temperature rise rather than its TCR value, thereby deriving estimates for 20th century warming to be attributable to greenhouse gas forcing. Stott et al (2006) did this and produced a combined model range of 0.7–1.3°C for GHG-attributable warming. Stott and Forest (2007) derived its TCR range of 1.5–2.8°C therefrom (presumably using unrounded figures) by multiplying by the ratio of F_{2xCO2} (taken as 3.74 W/m^2) to their estimate of the 20th century GHG forcing increase (1.66 W/m^2 per century). That TCR range is, not very surprisingly, identical to the one in Stott et al (2006). Note that the estimate of the 20th century GHG forcing increase used is 11% less than the AR5 best estimate. That would bias the TCR 5% and 95% bounds up by 12% even without the bias from using only 20th century temperature data. In conclusion, this study's range adds virtually nothing to knowledge about TCR beyond that conveyed by the AR4 study Stott et al (2006), which as discussed below appears to give a greatly excessive estimate of TCR.

l) **Tung et al (2008)**

This estimate depends on the response to the 11 year solar cycle. It is discounted in section 10.8.1 of AR5 WG1 on, inter alia, the basis that it may be affected by different mechanisms by which solar forcing affects climate. Its very high TCR range appears inconsistent with robust energy balance constraints (Otto et al, 2013).

m) **AR4 studies – unlabelled dashed line PDFs; no ranges**

All four of the dashed grey line AR4 PDFs come from Stott et al (2006). That study uses twentieth century temperature observations and a detection-and-attribution regression method to scale TCR values from the HadCM3, PCM and GFDL R30 AOGCMs. The PDFs represent observationally-constrained TCRs for each individual model as the product of each model's greenhouse gas scaling factor (regression coefficient) and its TCR, with an allowance added for internal variability. The fourth PDF is a combination, by averaging, of the individual model PDFs. The median combination TCR estimate of 2.0°C is 40% higher than the comparable figure of 1.4°C obtained from the Gillett et al (2013) study (see the assessment of that study, above) despite almost identical methodology. The likely principal reason is that use in Stott et al (2006) of purely twentieth century observational data biases TCR estimation upwards. Gillett et al (2012) showed that using just twentieth century observations to constrain TCR results in a best estimate for TCR that is almost 40% higher than when using data spanning 1851–2010. Gillett et al (2012) attributed this discrepancy partly to the first two decades of the 20th century being anomalously cool and partly to the decade 2001–2010, although warmer than the 1990s, exhibiting a slower temperature rise. From the discussion of Stott and Forest (2007), it also seems likely that the three CMIP3 AOGCMs involved underestimate the 20th century increase in greenhouse gas forcing, which will have further biased upwards their TCR estimation. Given these facts, the Stott et al (2006) PDFs for TCR must be seen as unreliable, almost certainly biased substantially high, and superseded by the results of Gillett et al (2013). That study not only uses a much longer temperature record but also a set of AOGCMs that is both much larger and from a later generation than those used in Stott et al (2006).