Marvel et al. – Gavin Schmidt admits key error but disputes everything else

Nicholas Lewis

Introduction

Gavin Schmidt has finally provided, at the GISS website, the iRF and ERF forcing values for a doubling of CO_2 (F_{2xCO2}) in GISS-E2-R, and related to this has made wholesale corrections to the results of Marvel et al. 2015 (MEA15). He has coupled this with a criticism at RealClimate of my appraisal of MEA15, writing about it "As is usual when people try too hard to delegitimise an approach or a paper, the criticisms tend to be a rag-bag of conceptual points, trivialities and, often, confused mis-readings – so it proves in this case". Personally, I think this fits better as a description of Gavin Schmidt's article. It contains multiple mistakes and misconceptions, which I think it worth setting the record straight on.

Corrected values for the forcing from a doubling of CO_2 concentration (F_{2xCO2})

I will start with the one fundamental problem in MEA15 that I identified in my original article about which Gavin Schmidt admits I was right. All the efficacy, TCR and ECS results in MEA15 scale with value of F_{2xCO2} used. That value varies between the three measures of radiative forcing involved: instantaneous radiative forcing at the tropopause (iRF, or Fi per Hansen et al. 2005[i]); stratospherically-adjusted forcing (Fa per Hansen, RF in IPCC AR5); and effective radiative forcing (Hansen's Fs). For results involving efficacy to be valid, they must use the same forcing measure when comparing the response to CO_2 forcing with that to other forcing agents. MEA15 did not do so. It used the RF value for F_{2xCO2}, 4.1 W/m^2, when calculating efficacies, TCR and ECS values for non-CO2 forcings measured in terms of iRF and ERF, the two alternative measures used in MEA15. As it was obvious to me that this was fundamentally wrong, around the turn of the year I emailed GISS asking for the iRF and ERF F_{2xCO2} values. GISS have now finally revealed them, as 4.5 W/m^2 for iRF and 4.35 W/m^2 for ERF. Correcting the erroneous F_{2xCO2} values used in the originally-published version of the paper increases all the MEA15 efficacy, TCR and ECS estimates for non-CO2 forcings by 10% for iRF, and by 6% for ERF. Since the paper was all about the divergence of the calculated values of these estimates from those applying to CO_2, changes of 10%, and even 6%, are quite significant.

The GISS website says: "There was an error in the Early-Online version of the paper (which will be fixed in the final version) in the definition of the F_{2xCO2} which was given as F_a (4.1 W/m^2) instead of F_i (4.5 W/m^2) and F_s (4.3 W/m^2)." It will be interesting to see whether Nature Climate Change takes the same stance as Nature Geoscience did with Otto et al. (2013), where as the incorrect Supplementary Information had already been published online (as here), it has been kept available alongside the corrected version.

Gavin Schmidt's comments on the other five fundamental problems I identified

I will now deal with Gavin's responses to the remaining five of my six points, using the same numbering.

1. Use of an inappropriate climate state to measure forcings that are sensitive to climate state

For some reason Gavin Schmidt paraphrases this, completely wrongly, as "MEA15 is working with the wrong definition of climate sensitivity". He writes "Point 1 is a misunderstanding of the concept of climate sensitivity and in any case would apply to every single paper being discussed including Lewis and Curry and Otto et al. It has nothing to do with whether those papers give reliable results."

I can only think that there is some "confused misreading" involved. [Equilibrium] climate sensitivity is defined as the increase in global mean surface temperature (GMST), once the ocean has reached equilibrium, resulting from a doubling of the equivalent atmospheric CO_2 concentration, being the concentration of CO_2 that would cause the same radiative forcing as the given mixture of CO_2 and other forcing components.[ii] It is usual to assume that forcings from different agents are linearly additive; results from GCMs generally support this assumption. Nothing I wrote under my point 1. conflicts with any of this.

My point was that the forcing produced by certain agents (anthropogenic aerosols and ozone) appeared to be very different in the 1850 climate state (in which MEA15 measured forcings, both iRF and ERF) to that produced in the recent, warmer, climate state. Lewis and Curry and other observationally-based papers use estimates for recent values of these forcings that reflect the contemporary climate state, not the 1850 climate state. That is appropriate since it is the radiative forcing produced by aerosols, ozone etc. in the recent climate state, not in the 1850 climate state, that determines their effect on recent temperatures.

I pointed out that in GISS-E2-R the 2000 level of anthropogenic aerosol loading produces direct aerosol TOA radiative forcing of –0.40 W/m^2 in the 2000 climate, but zero forcing in the 1850 climate; and that when the climate state is allowed to evolve as in the all-forcings simulation ozone iRF forcing in GISS-E2-R is 0.28 W/m2 in 2000 versus 0.45 W/m2 per MEA15. In both cases, using the forcing values calculated in the 1850 climate state would appear to lead to a downward bias in estimation of efficacies, TCR and ECS in the Historical, all-forcings combined, case.

Gavin Schmidt complains, I presume in relation to this point, that: "He conflates different model versions (fully interactive simulations in Shindell et al (the p3 runs in CMIP5), with the non-interactive runs used in MEA15 (p1 runs)), and different forcing definitions (Fi and Fa)".

Well, for aerosols I took my comparison from Miller et al (2014)[iii] where it states in relation to the basic, non-interactive, NINT model version: "Koch et al. [2011] similarly found that NINT aerosols in the year 2000 result in TOA direct forcing of 0.40 W/m^2 when using the double-call method (compared to our value of 0.00 W/m^2 based upon the 1850 climate)." Gavin Schmidt is the second author of that paper: even if that statement conflates different model versions or forcing measures I don't think I can be blamed for relying on it.

For ozone, I used iRF values in both cases, but the 0.28 W/m^2 value was for the fully-interactive TCADI version of GISS-E2-R, for which the iRF value in the 1850 climate state is 0.39 W/m^2 not 0.45 W/m^2. My mistake, but the impact of using the correct 1850 climate state forcing value is minor. The year 2000 ozone concentration still produces a 39% higher iRF when imposed in the 1850 climate state than in the climate state produced by the Historical, all-forcings simulation (and 77% higher than when constant present day conditions are imposed). However, these are the effects in the TCADI version. Since no values based on the recent climate state appear to have been computed for ozone forcing in the NINT version, it is impossible to be sure what values should be used for its recent ERF and iRF ozone forcing.

2. All previous papers using the historical records to estimate ECS actually estimate 'effective' climate sensitivity, which is smaller than ECS.

Gavin Schmidt's paraphrasing states that effective climate sensitivity is smaller than ECS. But what I actually wrote was that *in GISS-E2-R* effective climate sensitivity increases with time since the forcing was applied, as it does in many GCMs. While effective climate sensitivity is smaller than ECS in many GCMs, it is not known whether that is the case in the real climate system, and MEA15 has nothing to contribute on this question.

The point I was making was that even if all forcing agents had an efficacy of one, as for CO_2, estimating the ECS of GISS-E2-R from simulated changes over the historical period would be expected to give too low a value, since its effective sensitivity over such a period is lower than its ECS.

Gavin Schmidt claims that my point "begs the question entirely (why do analyses of transient simulations under-estimate ECS?". On the contrary, I want to separate out the effects on climate sensitivity estimation of varying GMST responses to different forcing agents, which is what MEA15 is about, from the effects of time-varying climate sensitivity in GISS-E2-R. Conflating these two completely different issues makes no sense.

4. *MEA15 shouldn't have used ocean heat content data (or should have done so differently)*

Gavin Schmidt says that the point "misunderstands that MEA15 were trying to assess whether real world analyses give the right result. Using TOA radiative imbalances instead of ocean heat uptake (which cannot be directly observed with sufficient precision) would be pointless". He goes on to write: " What if you account for the additional energy storage (apart from the ocean) in the system? … The bottom line is that … assuming that ocean heat uptake is only 94% of the energy imbalance makes no qualitative difference at all."

This is misleading. None of three 'real world' studies analysed in MEA15 used estimates of ocean heat uptake (OHU) only. One of them[iv] did not even estimate ECS, and therefore used no estimate of heat uptake; it is not clear what the value for OHU attributed to that study in MEA15 represents. But Otto et al and Lewis and Curry both used increases in the Earth's energy inventory (the integral of its total radiative imbalance), as estimated in IPCC AR5. As well as ocean heat uptake, these estimates included energy change in the atmosphere, land and from ice melt.

Moreover, although OHU represented ~93% of the total energy change for real world estimates, that is not the case for the MEA15 values for OHU in GISS-E2-R. For the Historical (all-forcings) case OHU represents, as I wrote, 86% of the total radiative imbalance, looking at the whole period. The ratio appears to be lower, only 83%, over recent decades, which are more relevant to the estimation in MEA15 of equilibrium efficacy and ECS. Furthermore, the ratio varies between forcing agents.[v]

The claim that assuming that ocean heat uptake is only 94% of the energy imbalance makes no qualitative difference is irrelevant, even if true (the link given does not appear to show the effects of such an assumption). I am interested in quantitative results. Gavin Schmidt has made no attempt to counter my estimate that allowing for non-ocean energy absorption would increase most of the equilibrium efficacy and ECS estimates, typically by 5–10%.[vi]

5. The regressions in MEA15 in the iRF case should have been forced to go through zero.

Gavin Schmidt says this is "easily tested and found not to matter in the slightest (as could easily be inferred from the graphs)". This claim that is self-evidently wrong, other than as regards ease of testing. As I wrote originally, when the regression best-fit lines are required to pass through the origin, substantially different iRF efficacy estimates are obtained for land-use change (LU), ozone (Oz), solar (SI) and volcanoes (VI) forcings. Based on the corrected F_{2xCO2} value of 4.5 W/m^2, iRF transient efficacy changes from 4.27 to 1.18 for LU; from 0.66 to 0.77 for Oz; from 1.68 to 1.47 for SI; and from 0.61 to 0.54 for VI.[vii] That the regression slopes involved will change, radically for LU, is obvious from the graph in Gavin Schmidt's Ringberg15 presentation. It is less obvious from the equivalent graph in MEA15, as there the area around the origin is obscured by large decadal-mean blobs.

I have some other objections to the regressions. MEA15 states, in the SI:

"TCR and ECS are calculated by regressing ensemble-average decadal mean forcing or forcing minus ocean heat content change rate against ensemble-average temperature change."

However:

- MEA15 actually regressed the opposite way round. They decadally regressed temperature change (as the *y* variable) against (as the *x* variable) forcing or forcing minus ocean heat content change rate. In some cases, this makes a significant difference to the results.
- MEA15 didn't regress ensemble-average values. They actually regressed individual run values and then took the ensemble mean of the regression slopes. This makes no difference for TCR and transient efficacy estimates, but it does for equilibrium efficacy and ECS estimates.
- MEA15 seriously miscalculated their *t*-distribution based uncertainty ranges. They are all double the correct value, except for the Historical All-forcings case, where they are more than double. In that case, they seem also to have overlooked that there is one more simulation run than in the other cases.

6. The linearity of the different forcings is only approximate.

As I wrote originally, the differences between the sum of (ensemble mean) values for the individual forcing simulations and the Historical (All forcings) simulations are ~10% for ΔT and iRF ΔF values. Such a difference is not insignificant in the context of a shortfall in efficacy (averaging the transient and equilibrium estimates) that is only slightly larger, at 13%.

I also wrote that "For unknown reasons, both plotted iRF ΔF values are shifted by approaching 10% relative to the data." This was not an important point: as the iRF regressions use an intercept term, shifting the ΔF values does not affect the slope and hence has no effect on MEA15's results. I was just noting it as another unexplained oddity in MEA15, which it was. However, Gavin Schmidt has gone to town on this, writing:

"His calculations didn't use the decadal mean forcings/responses that were used in MEA15 and thus he 'found' a -0.29 W/m^2 'error' in our graphs. [Despite having been told of this error weeks ago, no acknowledgement of this mistake has been made on any of the original posts]."

I think he is referring to his response at RealClimate to a comment of mine, in answer to a query by another reader, reiterating that the iRF for volcanoes appears to have been shifted by ~+0.29 W/m^2 from its data values. Gavin Schmidt responded:

"You are confused because you are using a single year baseline, when the data are being processed in decadal means. Thus the 19th C baseline is 1850-1859, not 1850. We could have been clearer in the paper that this was the case, but the jumping to conclusions you are doing does not seem justified. – gavin".

Indeed, I used forcings in the year 1850, when they were zero, as the baseline. Since MEA stated (in Figure 1 of the SI) that ensemble-average temperature response anomalies were relative to 1850, and nowhere did the paper suggest that forcings were treated differently, as anomalies relative to 1850-59 or any other period, it seemed to me to be natural to use the forcing values as they were.

Moreover, it seems that Gavin Schmidt is himself confused. Although the corrected Supplementary Information repeats the statement (in Figure S1) "(h): Ensemble-average temperature anomalies (relative to 1850) for each single-forcing simulation", the current version of Figure S1 at the relevant GISS webpage states "(h): Ensemble-average temperature anomalies (relative to 1850–59) for each single-forcing simulation". Furthermore, both these contradictory statements appear to conflict with the statement, in the text of both versions of the SI, that temperature anomalies are calculated with respect to pre-industrial control averages, with any temperature drift removed.

I also raised another point under this heading – that there were hints that land use forcing might have been omitted from the calculated values for Historical forcing. As I wrote, if that were the case, it would incorrectly depress the efficacy estimates relating to Historical iRF.

Additional issues

Gavin Schmidt also writes:

"Lewis in subsequent comments has claimed without evidence that land use was not properly included [viii] in our historical runs, and that there must be an error [ix] in the model radiative transfer. He has also suggested that it is statistically permissible to eliminate outliers in the ensembles because he doesn't like the result (it is not). These are simply post hoc justifications for not wanting to accept the results."

Let's see what I actually wrote, and my justification for doing so.

a) *Inclusion of land use forcing*

The link given is to my update article at Climate Audit; it contains a section headed "Possible omission of land use change forcing from Historical forcing data values". In it, I showed regression results that provided evidence strongly suggesting that LU forcing had indeed been omitted from the reported Historical forcing data values. But, despite this evidence, I didn't state that it had definitely been omitted. Rather, I concluded:

" I really don't know what the explanation is for the apparently missing Land use forcing. Hopefully GISS, who alone have all the necessary information, may be able to provide enlightenment."

I have in fact had some perfectly friendly correspondence with Ron Miller of GISS, another author of MEA15, about this issue. He has looked into it and can see no evidence of LU forcing having

been omitted for the Historical simulations themselves. I accept this, having no strong evidence to the contrary. That leaves the possibility that whilst LU forcing was included in the forcings applied during the Historical simulations, it somehow wasn't included when computing the value of the total forcing applied in that simulation. That had originally seemed to me unlikely, but it has been pointed out[x] that the forcing value is calculated separately. Certainly, it seems to me that LU forcing could have been omitted from the calculation of total forcing if there was some bug in the code used to perform the calculations (or possibly if there were an error in the settings used).

In addition to the regression results based on global data, there is almost no trace of LU forcing in the spatial pattern for Historical, All forcings together. Compare Figures 1 and 2 below. Other forcings are fairly uniform in the regions where patches of extremely negative LU forcing are located. If LU forcing was included in the calculation of All forcings together, then why is there no trace of its spatial pattern? Is there something very singular about the workings of GISS ModelE2?

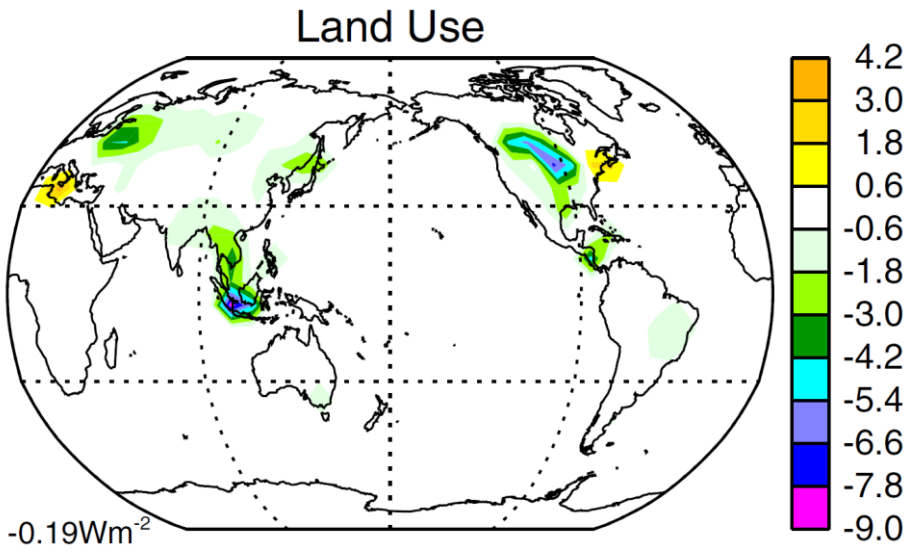

Figure 1. Reproduction of Figure 4c from Miller et al 2014: LU forcing in 2000 (vs 1850)

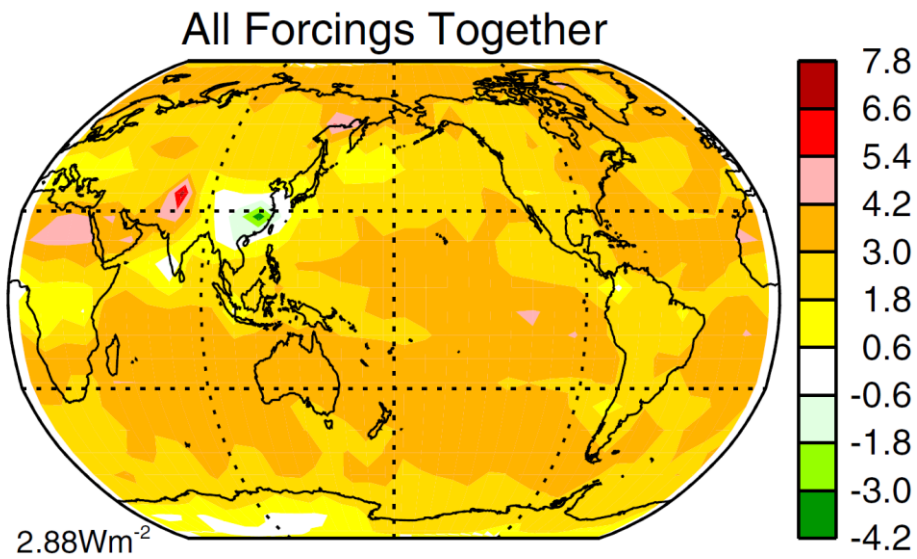

Figure 2. Reproduction of Figure 4d from Miller et al 2014: Historical forcing in 2000 (vs 1850)

b) *There must be an error in the model radiative transfer code*

The link given for this is to a comment of mine at Climate Audit, where I said:

""The GISS-E2-R increase in GHG ERF is 3.39 W/m^2. The 1850-2000 increase in GHG RF and ERF per AR5 Table AII.1.2 is 2.25 W/m^2, but I use the higher 1842–2000 increase of 2.30 W/m^2 since the 1850 CO_2 concentration in GISS ModelE2 was first reached in ~1842". If one strips out the CO_2 contributions, of 1.38 W/m^2 for AR5 (based on an F_{2xCO2} of 3.71 W/m^2) and of ~1.53 W/m^2 for GISS-E2-R (based on an ERF F_{2xCO2} of 4.1 W/m^2) the contribution of the other long lived GHG is 0.92 W/m^2 per AR5 and ~1.86 W/m^2 for GISS-E2-R.

That is, methane, nitrous oxide, CFCs and minor GHGs add TWICE as much forcing in GISS-E2-R as per the AR5 best estimate.

As I wrote, it looks as if GISS-E2-R radiative transfer computation in GISS-E2 may be inaccurate."

When I write "may be inaccurate", I mean just that. I do not mean, as Gavin Schmidt implies I do, that "there must be an error". Moreover, far from making a claim based on no evidence, I set out in detail the evidence that it was based on.

The divergence is slightly smaller using the corrected ERF F_{2xCO2} value of 4.35 W/m^2: ERF attributable to about non-CO_2 greenhouse gases is then about 190% higher in GISS-E2-R than it is according to AR5. Gavin Schmidt has not attempted to justify the large difference, or to refute my calculation.

c) *My alleged suggestion that it is statistically permissible to eliminate outliers in the ensembles because I didn't like the result*

This is quite wrong. I made no such suggestion, and Schmidt cites no evidence that I did so. What I wrote about the extreme outlier LU run 1 in my original article was this:

"It appears that the very high (although not statistically significant) best estimates for LU efficacy are affected by an outlier, possibly rogue, simulation run…. The difference from the ensemble mean is over four times as large as for any of the other 35 simulation runs. The LU efficacies estimates are greatly reduced if run 1 is excluded."

In my second article I expanded on the issue as follows:

"Whatever the exact cause of the massive oceanic cold anomaly developing in the GISS model during run 1, I find it very difficult to see that is has anything to do with land use change forcing. And whether or not internal variability in the real climate system might be able to cause similar effects, it seems clear that no massive ocean temperature anomaly did in fact develop during the historical period. Therefore, any theoretical possibility of changes like those in LU run 1 occurring in the real world seems irrelevant when estimating the effects of land use change on deriving TCR and ECS values from recorded warming over the historical period."

Maybe Gavin Schmidt doesn't understand this point. My case for considering excluding LU run 1 has nothing to do with whether I like the result of the run or not.

Schmidt's reworking of the Otto et al. results

A final point. Gavin Schmidt writes:

"If one was to redo those papers, you would choose the efficacies most relevant to their calculations (i.e. the ERF derived values for Otto et al) along with their adjustment for the ocean heat uptake (in our sensitivity test), and conclude that instead of an ECS of 2.0ºC [likely range 1.4-3.2], you'd get 3.0ºC [likely range 1.8-6.2]."

This is wrong. It appears he doesn't understand that the underlying forcing estimates used in Otto et al. are not simple ERF values. Rather, they were calculated from the GMST response of CMIP5 models, their effective climate sensitivity parameters and their radiative imbalances. Since they reflect the actual model responses to the applied forcings, *they already incorporate efficacies*. If volcanic forcing produces only half the GMST and radiative imbalance response in CMIP5 models as does the same forcing by CO_2, for instance – implying that volcanic forcing has an efficacy of 0.5 for the measure of forcing used – then the calculation of total forcing involved will automatically down weight by 50% the contribution from volcanic forcing.

11 February 2016

---

[i] Hansen J et al (2005) Efficacy of climate forcings. J Geophys Res, 110: D18104, doi:101029/2005JD005776

[ii] Paraphrased from IPCC AR5 Annex III

[iii] Miller, R. L. et al. CMIP5 historical simulations (1850_2012) with GISS ModelE2. J. Adv. Model. Earth Syst. 6, 441_477 (2014). Open access.

[iv] Shindell, DT (2014) Inhomogeneous forcing and transient climate sensitivity. Nature Clim Chg: DOI: 10.1038/NCLIMATE2136

[v] It is also 86% for well-mixed greenhouse gases (GHG), and similar at 85% for aerosols. But it is only 80% for land use, 82–83% for solar and volcanoes, and it is 90% for ozone forcing.

[vi] The exception being for land use forcing, where the heat uptake has the opposite sign to the GMST change.

[vii] Transient efficacy for Historical (All forcings) would decline, from 0.96 to 0.85, but this may be affected by the issue of whether LU forcing was included in the measure of Historical forcing.

[viii] http://climateaudit.org/2016/01/21/marvel-et-al-implications-of-forcing-efficacies-for-climate-sensitivity-estimates-an-update/

[ix] niclewis I'm unsure why this reference given at RealClimate was to my web pages.

[x] See http://climateaudit.org/2016/01/21/marvel-et-al-implications-of-forcing-efficacies-for-climate-sensitivity-estimates-an-update/#comment-766295