Marvel et al.'s new paper on estimating climate sensitivity from observations

Introduction and summary

Recently a new model-based paper on climate sensitivity was published by Kate Marvel, Gavin Schmidt (the head of NASA GISS) and others, titled 'Internal variability and disequilibrium confound estimates of climate sensitivity from observations'.[1] It appears to me that the novel part of its analysis is faulty, and that the part which isn't faulty isn't novel.

As some readers may recall, I found six serious errors in a well-publicised 2016 paper by Kate Marvel and other GISS climate scientists on the topic of climate sensitivity.[2] Two of the six errors were subsequently corrected.

With regards to the new Marvel et al paper, I find that:

- the low ECS estimates Marvel et al. obtain when using current (CMIP5) climate models' historical simulation data arise from using a period with unbalanced volcanic forcing, with the low bias disappearing when that problem is addressed; and

- the low ECS estimates they obtain when using data from AMIP simulations (those where models are driven by observed evolving sea-surface temperature patterns as well evolving forcing) are not news. They more likely indicate problems with CMIP5 models' ocean modules, than (as Marvel et al. suggest) that internal variability in recent decades was particularly unusual.

The background to the claims made by Marvel et al.

The paper's abstract commences by saying:

> "An emerging literature suggests that estimates of equilibrium climate sensitivity (ECS) derived from recent observations and energy balance models are biased low because models project more positive climate feedbacks in the far future."

While this statement is technically correct in that there have been several recent papers to this effect, these papers are based on flawed arguments. First, the fact that models project more positive climate feedbacks in the future does not in any way prove that the models are correct in doing so. Secondly, the more detailed explanation in the paper itself supports the statement with several different, mainly invalid, arguments:

(a) that tropospheric aerosols and land use change have a high efficacy (a strong effect on surface temperature relative to the effective radiative forcing (ERF) they exert, compared with that for CO_2);

(b) that the energy balance framework used by the studies that they are implicitly criticising,[3] and the forcing-adjustment-feedback paradigm on which it is based, assumes that perturbations to the climate system are small enough that feedbacks can be considered constant, but that recent work "shows that this assumption rarely holds even for the quadrupled-CO_2 state from which ECS is frequently inferred"; and

(c) that current climate models show a lower sensitivity when their atmospheric modules are driven by the observed historical evolution of sea surface temperature (SST) patterns; they also mention briefly related arguments about the effects of ocean heat uptake patterns.

The evidence for argument (a) is very weak. Marvel's own 2016 paper showed that the efficacy of aerosol ERF was almost exactly one – the same as that for CO_2. While it did show a high efficacy for the minor land use change forcing, to a substantial extent because of an outlier run,[4] Hansen's seminal 2005 forcing efficacy study estimated land use change efficacy to be close to one,[5] and a subsequent study found it to be very low.[6]

Marvel et al. cite two studies in support of argument (b).[7] The first paper cited has nothing to do with what Marvel et al. assert. The second is relevant to increases in CO_2 concentration from a doubling to a quadrupling, but its findings are fully explicable by the fact that CO_2 forcing increases very slightly faster than logarithmically with concentration.[8] In any event, observational climate sensitivity studies involve extrapolating only from ~$1.4\times$ to $2\times$ CO_2, over which the departure from a logarithmic forcing-concentration relationship is minute.[9]

I will leave argument (c) for now and come back to it later.

Marvel et al. do not go into the main explanation for most CMIP5 models projecting more positive feedbacks in future. In these models the pattern of SST warming changes over time after forcing is applied, and on average the feedbacks applying to the later warming pattern are more positive. However, across CMIP5 models the median estimated downwards bias this would induce in estimates of ECS derived from data over the historical period is only ~10%.[10]

What Marvel et al. did

I turn now to the meat of the paper. This is what the abstract says about the model-based analysis they carried out:

> "Here, we use simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5) to show that across models, ECS inferred from the recent historical period (1979-2005) is indeed almost uniformly lower than that inferred from simulations subject to abrupt increases in CO_2 radiative forcing. However, ECS inferred from simulations in which sea surface temperatures are prescribed according to observations is lower still."

Marvel et al. go on to say that "One interpretation is that observations of recent climate changes constitute a poor direct proxy for long term sensitivity." Indeed so. But, as I will show, a better interpretation is that estimating ECS by using changes over a twenty-six year period is unwise. Climate scientists who make serious attempts to estimate ECS from observed changes in the Earth's temperature and energy balance normally use much longer periods.

Marvel et al. estimated ECS in models using changes over 1979-2005 in global temperature ΔT, ERF ΔF and top-of-atmosphere radiation imbalance (their ΔQ, but usually ΔN) simulated in two CMIP5 "experiments": historical and AMIP, which ran to respectively 2005 and 2008. They used the well-known energy-balance estimation formula:

$$\text{ECS} = F_{2\times CO2}\ \Delta T\ /\ (\Delta F - \Delta N) \qquad\qquad (1)$$

where $F_{2\times CO2}$ is the ERF for a doubling of atmospheric CO_2 concentration. Marvel et al. actually inferred ECS by regressing annual mean $(\Delta F - \Delta N)$ on ΔT to estimate the climate feedback parameter λ, and then calculated $\text{ECS} = F_{2\times CO2}\ /\ \lambda$. They reported that simply subtracting the first decade from the last yielded similar results.[11]

Both the historical and AMIP experiments involved changing a model's atmospheric composition and/or emissions that affected its composition, and land use, in a way intended to imitate real-world conditions in each corresponding year. In the AMIP experiments, instead of the model's ocean module responding to the imposed forcing, prescribed SST patterns evolving in line with observations are used to drive an atmosphere-only model.

Unfortunately, it is generally not known what total ERF the changing atmospheric composition and/or emissions in these experiments produced in each model. Marvel et al. therefore estimated ΔF, for all models, from the IPCC AR5 time-series for total ERF, and used the corresponding AR5 value of 3.7 Wm^{-2} for $F_{2\times CO2}$. Given the wide spread between CMIP5 models in, *inter alia*, the level of aerosol forcing, and in estimated ERF from CO_2, this will likely cause considerable inaccuracy when using equation (1) to estimate ECS for individual models. Averaged over all models, the inaccuracy will be smaller. In general the method would be likely to produce a downwards bias in ECS estimates due to aerosol ERF being on average more negative in CMIP5 models than per the AR5 time-series. However, post-1979 the changes in aerosol ERF are relatively small, so there may be little downwards bias.

Figure 1 shows the resulting ECS estimates Marvel et al. obtained for each simulation run by the 22 models they studied.
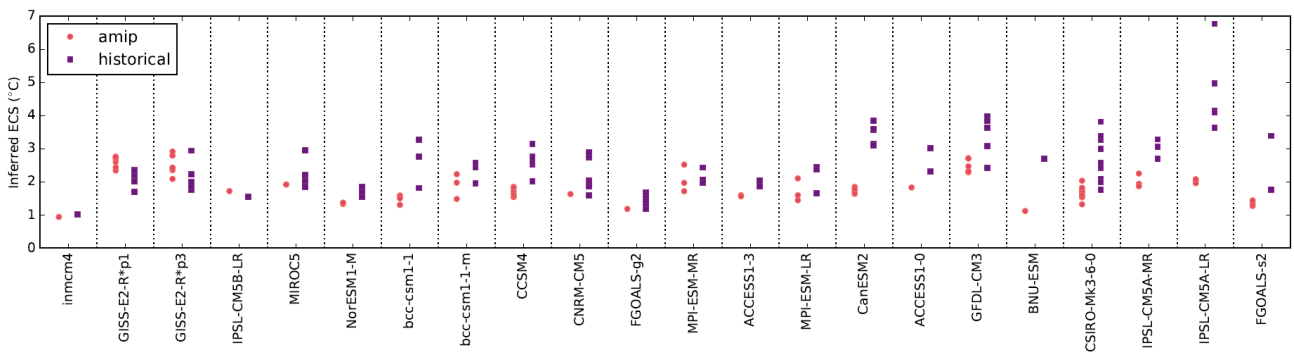

Figure 1. ECS estimated from recent (1979-2005) AMIP and historical simulations for each model's ensemble of runs. Models are ordered by increasing estimated long-term ECS. Reproduced from Figure 1 of the Supporting Information for Marvel et al. (2018).

ECS estimates from historical simulations

The median ECS that Marvel et al. infer from1979-2005 historical simulation data is 2.3°C, significantly lower than the median long-term ECS estimate of 3.1°C.[12] However, there is an obvious possible explanation for these low ECS estimates from historical simulation data.

The 1979-2005 period is particularly unsuitable for ECS estimation since strong negative volcanic forcing arose during its first half, but not thereafter. There is evidence (including from Marvel et al.'s 2016 paper) that volcanic forcing has a low efficacy – it has much less effect on global temperature than the same CO_2 forcing.[2][13] Accordingly, over the 1979-2005 period one would expect volcanism to increase the trend in F by a greater percentage than the trend in T, hence increasing the estimate of λ and depressing that of ECS.

It is simple enough to investigate the effect on short-period ECS estimation of avoiding significant influence from volcanism. I do so by using historical simulation data from the almost identical 1977-2005 period and Marvel et al.'s alternative decadal changes ECS estimation method. I made up the base ten years by combining the volcanic-free 1977-1981 and 1986-1990 periods. I took average changes from the base ten years to the final decade, 1996-2005, which is also free of eruptions. Doing so avoided the 1982 El Chichon and 1991 Mount Pinatubo eruptions and the main parts of the recoveries from each of them.

Figure 2 shows the resulting ECS estimates, upon applying equation (1).[14] The ECS estimates from individual simulation runs (red circles) are all over the place, as one would expect when estimating ECS from changes taking place over an average period of under twenty years. The change ΔF in average ERF is only 0.7 Wm^{-2}, so in the odd run where a model exhibits large positive internal variability in ΔN between the split base period and 1996-2005 the denominator in (1) will be small, and thus the ECS estimate very high. In a modern observationally-based ECS estimate the ΔF value would typically be three times as large.
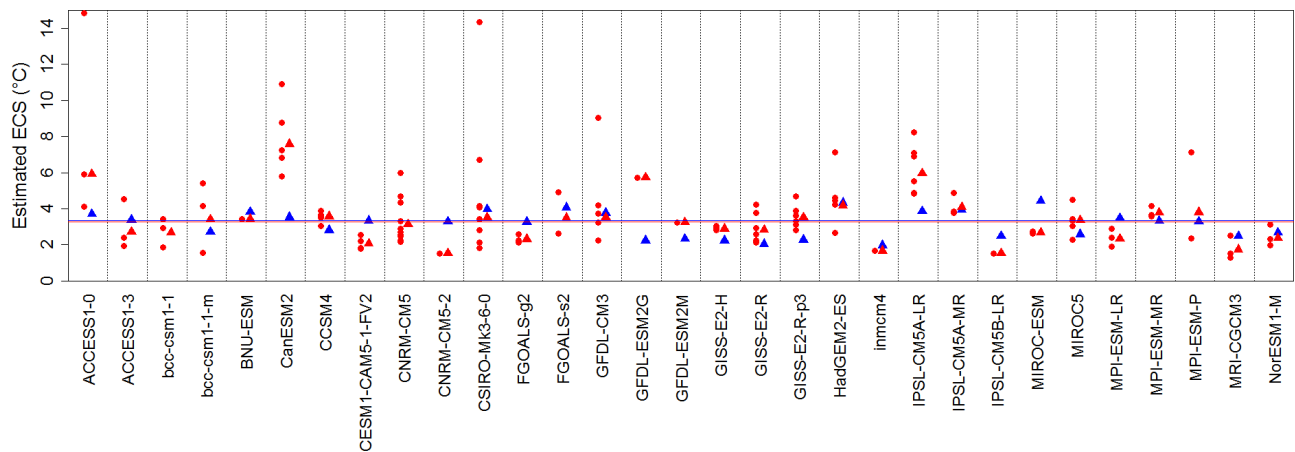

Figure 2. ECS estimated from non-volcanic periods in recent (1977-2005) historical simulations. Red triangles and circles show ECS estimated respectively from each model's ensemble-mean values and from individual runs. Blue triangles show estimated long-term ECS. The red and blue lines (which overlap) show the multimodel-ensemble medians of respectively ensemble-mean ECS estimates and long-term ECS estimates. Long-term ECS was estimated using the same method as Marvel et al.

Where several historical simulation runs were carried out by a model, the ECS estimates using mean values from its ensemble of runs (red triangles) are less wild. But the interesting point shown in Figure 2 is that, across all models, the median of the long-term ECS estimates (blue line: 3.29°C) is

almost identical to the median of the model-ensemble means based ECS estimates (red line: 3.37°C).[15] So, when care is take to avoid volcanism distorting the estimates, it is not true that ECS inferred from the recent historical period is "almost uniformly lower than that inferred from simulations subject to abrupt increases in CO_2 radiative forcing", as claimed by Marvel at el.

It is not possible to find a long period in historical simulations that avoids both significant volcanic activity and a large change in aerosol forcing. However, it is possible to improve the estimation of CMIP5 model ECS values by extending the period forward to 2016, splicing on data from RCP8.5 simulation runs that continue historical simulation runs after 2005, so as to use a final period of 2007-2016, as before taking changes relative to the combined 1977-81 and 1986-1990 periods.[16] The median within-model standard deviation of the resulting ECS estimates based on single simulation runs is then 13% of the median ensemble-mean ECS estimate. If that is taken as a proxy for the effect of internal variability on ECS estimation, it is not too bad given that this estimate is based only on data spanning a thirty year period, and on averaging over single decades.

For observationally-based energy-balance climate sensitivity estimation, where concern about model aerosol ERF strength is not a concern, one would normally use a much earlier (and typically rather longer) base period, thereby achieving a higher signal-to-noise ratio. If the full historical period to date is used to estimate model ECS values from simulation data, better precision is achievable. When using changes between the means for 1859-1882 and 1995-2016, two volcanism free periods, the median single-run ECS estimate standard deviation is only 8% of the median ensemble-mean ECS estimate. On that basis, uncertainty in observationally-based ECS estimation arising from internal variability is minor compared with other uncertainties.

ECS estimates from AMIP simulations

Marvel et al.'s median ECS estimate from CMIP5 AMIP simulations (1.8°C) was lower than that from historical simulations. A similar finding was shown (with volcanic years excluded) in Tim Andrews' Ringberg talk in March 2015, and Gregory and Andrews (2016) gave sensitivity estimates for all models with AMIP simulations, albeit without identifying them, as well as their average.[17] It appears that the observed evolution of SST gave rise to enhanced tropical low-cloud cover compared to that in CMIP5 models' historical simulations. The AMIP runs, which generally span 1979-2008, are too short to tell one much about the underlying cause, but in this case I think the lower ECS estimates for models are probably primarily genuine, rather than artefacts arising from use of a period with unbalanced volcanism. This is a reflection of Marvel et al.'s argument (c), which I put to one side earlier.

Marvel et al. claim that the low ECS values when models are driven by the observed evolution of SST patterns suggests that the "specific realization of internal variability experienced in recent decades provides an unusually low estimate of ECS." However, as they admit, this is based on the perfect-model framework, which assumes "that the models as a group provide realistic descriptions of the mechanisms underlying observed climate variability".

An alternative explanation for the models as a group misestimating the actual temporal evolution of SST change patterns is that the models as a group are imperfect. To my mind that should be the null

hypothesis, rather than that internal variability over the last few decades results in an unusually low estimate of ECS. Indeed, the fact that internal variability linked to the Atlantic multidecadal oscillation is thought to have boosted warming over 1979-2005[18] makes it seem even less likely that in the real climate system ECS estimates based on this period would be biased low. Moreover, internal variability sufficient to produce a 20-year excursion of the magnitude required to account for the CMIP5 model average difference in N between AMIP and historical simulations does not appear to occurred in any of the 13,000 odd overlapping 20 year segments of their preindustrial control simulations.

Even if CMIP5 models don't do too bad a job of simulating atmospheric behaviour, it is entirely possible that the real ocean is better able to move heat around the Earth's climate system, in a way that reduces average surface temperature, than CMIP5 model oceans are able to do in their simulated climate systems. Marvel et al. recognize this, saying that the low ECS estimates derived from AMIP simulations "could also arise from the failure of the coupled models to reproduce aspects of the forced response". Moreover, it is not the case that low model ECS estimates when driven by observed evolving SST patterns are limited to the last few decades. For now I will refrain from further discussion of this interesting area, which is a focus of research activity, as this article is already overlong.

Nic Lewis   5 February 2018

---

[1] The paper itself is pay-walled, but the Supporting Information is not.

[2] Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. S. (2016). Implications for climate sensitivity from the response to individual forcings. Nature Climate Change, 6(4), 386.

[3] They mention, as examples:
Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner (2002), An Observationally Based Estimate of the Climate Sensitivity, J. Climate, 15 (22), 3117-3121;
Otto, A., F. E. Otto, O. Boucher, J. Church, G. Hegerl, P. M. Forster, N. P. Gillett, J. Gregory, G. C. Johnson, R. Knutti, et al. (2013), Energy budget constraints on climate response, Nature Geoscience, 6 (6), 415-416;
Lewis, N., and J. A. Curry (2015), The implications for climate sensitivity of AR5 forcing and heat uptake estimates, Climate Dynamics, 45, 1009-1023.

[4] In the outlier land use change forcing run by the GISS-E2-R model that they used, ocean convection appears to have partly collapsed in the North Atlantic, as it does in some of that model's main CMIP5 simulations.

[5] Hansen, J. E. et al. Efficacy of climate forcings. J. Geophys. Res. 110, D18104 (2005).

[6] E. L. Davin, N. de Noblet-Ducoudre, and P. Friedlingstein (2007), Impact of land cover change on surface climate: Relevance of the radiative forcing concept. Geophys. Res Lett, 34, L13702.

[7] Armour, K. C., C. M. Bitz, and G. H. Roe (2013), Time-varying climate sensitivity from regional feedbacks, Journal of Climate, 26 (13), 4518-4534; Gregory, J. M., T. Andrews, and P. Good (2015), The inconstancy of the transient climate response parameter under increasing CO_2, Philos. Trans. R. Soc. London. (Described by Marvel et al. as "in press" but in fact published in October 2015.)

[8] Byrne, B., and C. Goldblatt (2014): Radiative forcing at high concentrations of well-mixed greenhouse gases. Geophys. Res. Lett., 41, 152–160, doi:10.1002/2013gl058456; and
Etminan, M., G. Myhre, E. J. Highwood, and K. P. Shine (2016): Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. Geophys. Res. Lett. 43(24) doi:10.1002/2016GL071930.

[9] Since ECS is defined as the eventual temperature rise going from 1× to 2× (preindustrial) CO2 levels, and recent levels are approximately 1.4× preindustrial. If feedbacks change with a perturbation of 4× CO2, that would be a problem when using climate model simulations involving 4× CO2 to estimate their ECS, as is typically done, but there is little model evidence of that being the case.

[10] See my analyses here and here. The best estimates of ECS for CMIP5 models are now generally obtained by scaling the x-intercept of a regression fit to years 21-150 of ΔT and ΔN data from a simulation in which a model's CO_2 concentration is abruptly quadrupled ('abrupt4xCO2'), thus omitting the early decades in which higher feedback strength (lower sensitivity) is exhibited.

[11] They presumably estimated λ as the ratio of the inter-decade change in $(\Delta F - \Delta N)$ to that in ΔT. This method is arguably more robust than using regression.

[12] Derived from scaling the x-intercept of a regression fit to years 1-150 of ΔT and ΔN simulation data after a model's CO_2 concentration is abruptly quadrupled. On average, this method appears to underestimate CMIP5 models' ECS values, but only by 5-10% compared to estimates derived from the now generally preferred method of regressing over years 21-150.

[13] E.g., Gregory, J. M., Andrews, T., Good, P., Mauritsen, T., & Forster, P. M. (2016). Small global-mean cooling due to volcanic radiative forcing. *Climate Dynamics*, *47*(12), 3979-3991.

[14] I derived ECS estimates for all models for which I could obtain data for their historical, preindustrial control and abrupt CO_2 quadrupling experiments, using data from the latter two experiments to estimate a model's long-term ECS.

[15] If 1977 and 1978 are excluded from the initial years, there is little change in the average ensemble-mean ECS estimate: the mean increases slightly and the median is marginally lower.

[16] I extended the AR5 forcing series from 2011 to 2016 using primarily observationally-based estimates. The resulting increase in anthropogenic ERF over that period was 0.23 Wm^{-2}, the same as per the RCP8.5 forcings dataset.

[17] Gregory, J. M., and T. Andrews (2016), Variation in climate sensitivity and feedback parameters during the historical period, Geophys. Res. Lett, 43 (8), 3911-3920.

[18] E.g., DelSole, T., Tippett, M. K., & Shukla, J. (2011). A significant component of unforced multidecadal variability in the recent acceleration of global warming. Journal of Climate, 24(3), 909-926.