Patrick Frank
Earth and Space Science Manuscript 2017EA000256
Response to Review #6

Summary response:

1. [*John and Soden*, 2007] does not physically validate simulation anomalies; item 1.1.
2. Differencing from a base state does not subtract away simulation errors; items 1.2 and 1.3.
3. [*Dessler*, 2013] does not validate model simulations against observations; item 1.4.
4. Simulation anomalies have unrecognized uncertainty; item 1.5.
5. The reviewer has misconstrued statistical uncertainties to be physical magnitudes; item 2.2.1.
6. The reviewer has confused model precision with model accuracy; items 1.1, 3.2 and 3.3.1.
7. The reviewer has inadvertently validated the manuscript error propagation; item 3.3.1.
8. The reviewer has misperceived manuscript eqn. 6 as representing climate physics; item 4.1.
9. The reviewer has misunderstood the origin of the 0.42 sensitivity fraction and of eqn. 6; items 4.2, 4.3 and 4.4.

The reviewer is quoted in italics, and the indented author response follows.

1. The paper makes an elementary but fundamental error: it confuses errors in the models' base state with errors in the models' predictions of how the climate will change. The fact that models can have large biases in their base state is well documented; e.g., previously published work has shown biases in their water vapor and temperature fields (e.g., John and Soden (2007), Temperature and humidity biases in global climate models and their impact on climate feedbacks, Geophys. Res. Lett., 34, L18704, doi: 10.1029/2007GL030429), and I have no doubt that some GCMs have large biases in their cloud fields (as this paper argues).

*However, this does not mean that the *change* in these fields as the climate warms in the models is wrong. John and Soden showed that, despite the biases in the water vapor fields, the **change** in water vapor in response to warming is nearly identical among the GCMs, meaning that the water vapor feedback is nearly identical. Comparisons of the cloud feedbacks in the GCMs shows good agreement among the GCMs, and with observations (e.g., Dessler, A. E. (2013), Observations of climate feedbacks over 2000-10 and comparisons to climate models, J. Climate, 26, 333-342, doi: 10.1175/jcli-d-11-00640.1). This means that, despite large differences in the cloud fields, the change in clouds as the climate warms is basically the same.*

> 1. To summarize: the reviewer makes two arguments here. The first is that errors in states ("*base states*") do not translate into uncertainties in anomalies. Second, that agreement among models is a demonstration of accuracy. These ideas are taken in turn.
>
> Response item 1, below, shows that [*Covey et al.*, 2003; *John and Soden*, 2007] do not validate simulated climate change, that model error is not known to be constant, that climate models are not known to follow linear response theory, that [*Dessler*, 2013] does not does not convey observational verification of climate models, and that uncertainty, when present, only increases when taking anomalies.

These points are taken up in sequence below, in numbered subsections.

1.1 [*John and Soden*, 2007] does not validate that the change in climate fields is accurately modeled.

It is first is noted that the reviewer presents inter-model biases as though they are physical errors. They are not. Physical errors are obtained relative to observations. They are a measure of accuracy. Inter-model biases represent precision. This distinction was emphasized in the Introduction, but the reviewer has evidently not encompassed it.

The CMIP3 models in [*John and Soden*, 2007] are listed below with error in Total Cloud Amount (CA) and TOA Longwave Cloud Forcing (LWCF). Table R1 shows these CMIP3 errors as read off the corresponding Taylor diagrams in Figure 3 of [*Lauer and Hamilton*, 2013].

Table R1: Selected Errors for the CMIP3 Models in [*John and Soden*, 2007].

Models	RMS Error Total CA (%); (Correlation)	RMS Error TOA LWCF (Wm^{-2}); (Correlation)
BCCR_CM2_0	0.9±1.2; (0.65)	---
CNRM_CM3	0.75±0.8; (0.70)	0.75±1.3; (0.82)
CSIRO_MK3_0	1.1±0.95; (0.71)	0.8±1.5; (0.82)
GFDL_CM2_0	1.1±1.2; (0.55)	0.49±1.1; (0.9)
GFDL_CM2_1	1.1±1.2; (0.52)	0.5±0.9; (0.9)
GISS_MODEL_E_H	1.2±1.2; (0.45)	0.8±0.9; (0.66)
GISS_MODEL_E_R	1.2±1.2; (0.49)	0.6±0.95; (0.78)
IAP_FGOALS1_0_G	1.25±0.9; (0.11)	0.6±1.1; (0.84)
INMCM3_0	0.8±0.95; (0.61)	0.8±1.3; (0.79)
MIROC3_2_MEDRES	0.7±1.3; (0.85)	0.55±1.2; (0.9)
MPI_ECHAM5	0.9±1.1; (0.63)	0.52±1.2; (0.9)
MRI_CGCM2_3_2A	0.8±1.2; (0.73)	0.5±1; (0.9)
NCAR_CCSM3_0	0.75±0.8; (0.70)	0.6±1.3; (0.88)
NCAR_PCM1	1.25±1.2; (0.33)	0.8±1.25; (0.79)
UKMO_HADCM3	0.8±1; (0.61)	0.5±1; (0.88)
UKMO_HADGEM1	0.7±1.2; (0.82)	0.6±1.1; (0.84)
Observations	0±1; (1.00)	0±1; (1.00)

CA is total cloud amount, LWCF is longwave cloud forcing. Correlations are between modeled and observed spatial distributions.

The correlation of simulated with observed percent CA ranges from 0.11 to 0.85. In HADCM3, for example, CA is a reasonable 0.8±1, but the correlation with observed CA is 0.61. Thus, even though the amount and variation of HADCM3 CA is good, the cloud distribution is poor. Similar concerns accrue to LWCF. Thus, the simulated tropospheric energy flux distributions are not physically correct and vary from model to model.

These errors indicate the CMIP3 models produce climate states different from the correct terrestrial state. Additionally however, they also do not produce climates states similar even among themselves. None of the modeled climate states include physically correct atmospheric energy flux distributions.

Figure R1 shows Figure 2 of [*John and Soden*, 2007], with author additions.

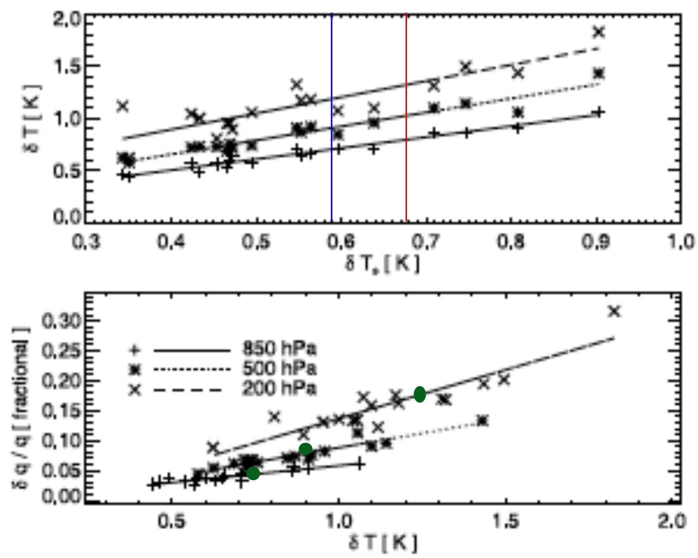

Figure R1, original Figure 2 legend: "*(top) Response of T at three different atmospheric levels (850, 500, and 200 hPa) to change in surface temperature (T_s). (bottom) Fractional response of q at three different atmospheric levels (850, 500, and 200 hPa) to change in T at those levels. Different symbols represent different coupled GCMs used in this study. Tropical means are used. ∂T, ∂T_s, and ∂q/q are the difference between the first 10 year and the last 10 year means of 20th century of each variable.*"

The data in Figure R1 (Figure 2 of [*John and Soden*, 2007]) display simulated anomalies. The fitted black lines in the original Figure illustrate correlation among models, not correspondence with physical observables. That is, original Figure 2 is again about model precision, not model physical accuracy.

In Figure R1 top, the colored vertical lines are the 20th century ∂T_s ten-year means of the GISS (red, 2007/01/08 version) and CRU (blue; May 2005 version) surface temperature records. The physically correct values of the ordinate ∂T(K) for 850, 500 and 200 hPa presumably lay somewhere on these lines.

The three horizontal intercepts through these lines then represent the best-estimate model ∂T(K) values. These are (pressure level, best ∂T(K), (model ∂T(K) range)): 200 hPa, 1.3, (0.7-1.7); 500 hPa, 1, (0.6-1.3); 800 hPa, 0.75, (0.45-1.0).

The simulation anomalies distant from these intercepts are clearly incorrect. Greater distance represents greater physical error. However, the physically correct values need not lie on the fitted lines, but may be vertically displaced. Therefore, even the best-estimate points are not known to be physically correct.

This analysis clearly shows that the mere fact the inter-model simulation anomalies are correlated does not mean their physical error subtracts away. None of the simulation anomaly lines or points in Figure R1 is known to be correct.

Further, calibration experiments have shown that T_s is contaminated with non-normal systematic sensor measurement error [*Brooks*, 1926; *Hubbard and Lin*, 2002; *Huwald et al.*, 2009; *Lin et al.*, 2005; *Saur*, 1963]. Non-normal error does not average away, and is responsible for a lower limit uncertainty of ±0.5 K in T_s, which has been generally ignored [*Patrick Frank*, 2010; 2011; 2015]. Uncertainty combines in quadrature when taking anomalies [*Bevington and Robinson*, 2003]. Therefore a mean of the vertical CRU and GISS anomaly ∂T_s lines includes a horizontal ±0.7 K uncertainty.

This level of uncertainty causes the GISS and CRU anomalies uncertainty ranges to overlap at the 1 σ level, and the horizontal uncertainty bars extend right across the entire width of Figure R1 top. This wide uncertainty in ∂T_s makes it impossible to discover the physically correct value of ∂T at any hPa. It is therefore impossible to know which simulation value is the more correct, or indeed whether any of them are correct. Thus, the uncertainty in target magnitude necessarily imposes wide confidence intervals about the simulation. None of Figure R1 has any clear physical meaning.

In Figure R1, bottom, the green points plot the mean best model ∂T(K) intercept values on the fitted $\partial q/q$ lines. The horizontal intercepts are again the best-estimate model $\partial q/q$ values. These are (pressure level, mean %, (model range)): 200 hPa, 0.18, (0.075-0.27); 500 hPa, 0.09, (0.05-0.17), and; 850 hPa, 0.045, (0.03-0.06).

Once again, the simulation anomalies more distant from these points represent greater apparent physical error. Once again, however, even the best-estimate points are not known to be physically correct.

Once again, the fact that the inter-model simulation anomalies are correlated clearly does not mean their physical error subtracts away. Indeed, the physical uncertainties are so large that error is impossible to evaluate. How can errors be said to subtract away, when the errors themselves are unknowable?

Additionally, [*John and Soden*, 2007] Figure 2 illustrates the conclusion following from Table R1. The different CMIP3 models simulate <u>disparate</u> temperature and heat flux changes despite the <u>identical</u> 20th century forcings.

All of the simulated anomalies in [*John and Soden*, 2007] Figure 2 include large physical uncertainties, but the true magnitudes of the errors remain unknown because the physically true magnitudes are not known.

These are all tuned models. They deploy alternative suites of parameters. Each parameter has a significant range of uncertainty. Disparate sets of parameter magnitudes are used among the different models. These parameter sets encode disparate physical relationships among the respective climate variables. The linear fits merely show the effects of parameter tuning, namely correlated model expectation values.

The disparate physical relationships of the variables deployed within the models necessitate that none of the model expectation values represent unique solutions to the problem of the climate energy-state. Nevertheless, [*John and Soden*, 2007] Figure 2 included no uncertainty bars.

In short, [*John and Soden*, 2007] illustrates exactly the problem addressed by the present study. This is the problem of physical accuracy versus model precision. [*John and Soden*, 2007] is about model precision. It establishes nothing about physical accuracy. Model tuning hides the simulation uncertainties. Physical error is ignored, and is not propagated through the simulations. The results in [*John and Soden*, 2007] are misleading because they reveal only model precision while taken to imply model accuracy, and they completely lack valid physical uncertainty estimates.

1.2 Simulation errors do not subtract away.

Taken up here is the reviewer's suggestion that changes in climate fields are correct, even when the fields themselves are wrong. Also implied is that climate models cohere with linear response theory. This further implication is taken up in 1.3 below.

The problem of constant model error was thoroughly addressed in manuscript Section 2.4.3, lines 639-676, with reference to an even more detailed discussion in SI Section 7. However, the reviewer apparently did not consult this material.

The reviewer proposes error-free anomalies, and has not limited this freedom from error to only those anomalies calculated from adjoining step-wise realizations. That is, the reviewer asserts anomalies taken across distant points such as, e.g., the twentieth century anomalies of [*John and Soden*, 2007], are also correct. This reviewer-imposed condition requires simulation errors of constant magnitude.

Suppose a climate projection, starting from a zeroth climate state, projects a series of successive climate states. These states will have air temperatures with physically true magnitudes, $T_0, T_1, T_2, ..., T_n$.

The projected temperatures are $T_0+\boldsymbol{\varepsilon}_0, T_1+\boldsymbol{\varepsilon}_1, T_2+\boldsymbol{\varepsilon}_2, ..., T_n+\boldsymbol{\varepsilon}_n$, where $\boldsymbol{\varepsilon}$ is the error in the simulated temperature. The zeroth state is a simulated state, and thus does have an associated error.

According to the reviewer, the anomaly $\Delta T_{1,0} = (T_1+\boldsymbol{\varepsilon}_1) - (T_0+\boldsymbol{\varepsilon}_0) = (T_1-T_0) + (\varepsilon_1-\varepsilon_0)$, and $\varepsilon_1-\varepsilon_0 = 0$. The reviewer's condition requires this be also true for an anomaly of any span, e.g., $\Delta T_{m,i} = (T_m+\boldsymbol{\varepsilon}_m) - (T_i+\boldsymbol{\varepsilon}_i) = (T_m-T_i) + (\varepsilon_m-\varepsilon_i)$ and $\varepsilon_m-\varepsilon_i = 0$ for any distant pair of time-indexed magnitudes.

That is, the reviewer's condition requires that the errors are constant offsets, such that $\varepsilon_0 = \varepsilon_1 = \varepsilon_2 = ... = \varepsilon_n$, and every $\varepsilon_m - \varepsilon_i = 0$ for any *m, i* pair.

However the magnitude of global average T is time-wise variable. The constancy of ε requirement imposes that the magnitude of ε is negatively correlated with the magnitude of T, and precisely so, in order that ε remain constant and $\Delta\varepsilon = 0$.

The reviewer's requirement of error-free anomalies is generalizable to every state variable, $S_0, S_1, S_2, ... S_n$, of projection magnitudes $S_0+\varepsilon_0, S_1+\varepsilon_1, S_2+\varepsilon_2, ..., S_n+\varepsilon_n$. All the projection errors must be exactly anti-correlated with their corresponding state magnitudes.

A further reviewer-imposed condition is that a projection year initializing with a projection state magnitude $S_i+\varepsilon_i$ is able to produce an error-free $\Delta S_{j,j}$. The error in state S_i cannot impact the evolution to state S_j because $\varepsilon_i = \varepsilon_j$.

More generally, given the initial state S_0 with error ε_0, the ε_0 does not impact the evolution of the state. The $\Delta S_{j,j}$ change is simulated accurately, despite the error of magnitude ε_i entering into the calculation.

Model evolution with constant error means every ΔS in the evolving projection is exactly physically correct, despite that each S_{i+1} state initializes from a physically incorrect S_i state.

Perfect negative correlation between state magnitudes and their simulation errors with invariably correct anomalies requires that a model exactly compensate state error in every projection step.

This condition has never been demonstrated, neither generally nor in any specific cases. It is not demonstrated in [*John and Soden*, 2007]. The reviewer's position is an article of faith rather than of science.

1.3 Climate models are not known to follow linear response theory.

The implied relevance of linear response theory to climate model expectation values was fully discussed in SI Section 7.1.1, "*The problem of validating a model difference.*" There it is shown that the relevance is chimerical at best, and more recent assessments have not improved this diagnosis [*Hassanzadeh and Kuang*, 2016]. The Supporting Information has been updated with this reference.

The condition imposing a linear response to a non-linear theory requires very weak perturbations. According to the reviewer, application of the condition of linearity to the anomalies of [*John and Soden*, 2007] requires that the entire forcing change between 1850 and 2010 be assigned to the category "very weak." This, too, has not been shown.

1.4 [*Dessler*, 2013] does not observationally verify climate models.

[*Dessler*, 2013] treats the ECMWF and MERRA reanalysis products as observations. They are not. There are published warnings against equating reanalysis to observations [*Dee et al.*, 2011; *Kalnay et al.*, 1996]. Dee, et al., 2007 for example: "*Many users regard reanalysis products as equivalent to observations, even if this is not always justifiable.* "

[*Kalnay et al.*, 1996] warn that, "*[Some reanalysis fields] are partially defined by the observations but are also strongly influenced by the model characteristics. For example, the amount of moisture that the tropical model atmosphere can hold depends on its parameterization of cumulus convection, since some convection schemes tend to dry out the atmosphere more than others. Therefore, even if the analysis incorporates rawinsonde* *and satellite moisture data, the overall humidity will be influenced by the climatology of the model. This is even more true for quantities that are not directly observed or whose observations are not currently assimilated into the present analysis systems.*"

That is, reanalysis fields used as model validation standards can reflect the biases in the originating climate model. A model-derived reanalysis product can falsely validate any model subject to similar error biases.

Nevertheless, [*Dessler*, 2013], enquoted reanalysis fields as "the observations" when first introduced, but thereafter uncritically presented reanalysis as observations when compared to direct model simulations.

Further, [*Dessler*, 2013] used the kernel approach of [*Soden et al.*, 2008] to compute cloud feedback properties, although [*Soden et al.*, 2008] warned against this use. E.g., "*A limitation of the kernels is that the radiative effects of clouds, particularly the vertical overlap of clouds, are too nonlinear to accurately compute cloud feedback using this method.*"

[*Soden et al.*, 2008] go on to derive an alternative to the kernel approach of cloud forcing. But in that event, they evaluate this alternative by reference to the cloud simulations produced by the GFDL climate model, not to real-world observations.

Thus the kernel calculations used in [*Dessler*, 2013] to evaluate model results were validated in [*Soden et al.*, 2008] by reference to model results. The approach in [*Dessler*, 2013] is thus fundamentally circular.

The kernel approach again assumed a linear climate response, valid over only small perturbations and short times. However, [*Dessler*, 2013] applied the method across 10 years.

[*Dee et al.*, 2011] evaluated the ECMWF and ERA-Interim reanalysis products for forecast accuracy. Their Figure 2 is relevant; showing accumulated 850 hPa air temperature errors of 1.7 C and 1.55 C, respectively, after only 9 forecast days. This result is hardly encouraging for use of reanalysis to judge the accuracy of the 10-year forecast in [*Dessler*, 2013].

The reviewer would have it that these forecast errors are irrelevant, because the change in temperature is the only important metric. However, Section 1.2 above showed this proposal is insupportable.

Further, if air temperature is incorrect, the partitioning of energy flux is incorrect. With incorrect energy flux distributions how, then, is the simulated climate to evolve along its energetic phase-space trajectory in an identical fashion with the phase-space trajectory of the physically real climate, but merely with a constant offset?

[*Dessler*, 2013] itself cautioned that reanalysis cloud feedback is poorly constrained, and advised against an uncritical acceptance of the reanalysis products. Thus, under Section 3, p. 335:

"*While there is good agreement for the total cloud feedback in the two reanalyses, the calculations using MERRA predict that the majority of the cloud feedback comes from changes in the shortwave effects of clouds, while the calculations using the ERAInterim suggest that almost all of the cloud feedback is due to changes in the longwave. Note that both the ERA-Interim and MERRA calculations use the same* $\Delta R_{all\text{-}sky}$ *measurements from CERES.*"

That is, despite starting from the same CERES observations, the two reanalysis products differ in their accounting of cloud feedback. The MERRA feedback, in shortwave CF, has small effect on the tropospheric thermal energy flux, while ERAInterim feedback, in longwave CF, necessarily impacts this flux.

Although the feedbacks have the same total magnitude, the energy flux partitioning is strongly disparate. Which reanalysis, then, should provide the standard for judging model accuracy with respect to cloud feedbacks?

In short, [*Dessler*, 2013] applied a kernel calculation likely beyond its relevance, judged model results against a kernel methodology that was validated using models (not observations), and uncritically used reanalysis products relevant to short-term meteorological forecasts as observations to evaluate decadal climate projections.

Nevertheless, throughout [*Dessler*, 2013], cautions are given about the unreliability of simulated cloud feedback. The final and summing-up sentence in the entire paper is, "*Finally, this analysis confirms that the biggest uncertainty is the cloud feedback.*"

Despite this, the reviewer offered [*Dessler*, 2013] as a critical refutation of cloud feedback error.

However, the fact remains that the explicit cautions about cloud forcing given in [*Dessler*, 2013] validate the focus of the present manuscript on long wave cloud forcing error.

Additional evidence to this point is that [*Su et al.*, 2013], page 2766, par [16], explicitly admits of large-scale (theory-bias) and small-scale (parameterization) physical errors in cloud simulations.

1.5 The hidden uncertainty in simulated air temperature anomalies.

Finally, the impact of uncertainty on the reliability of projection anomalies is taken up. Figure R2 below is taken from the CMIP2 comparison project, but the analysis will be equally applicable to CMIP3 and CMIP5 model simulations.

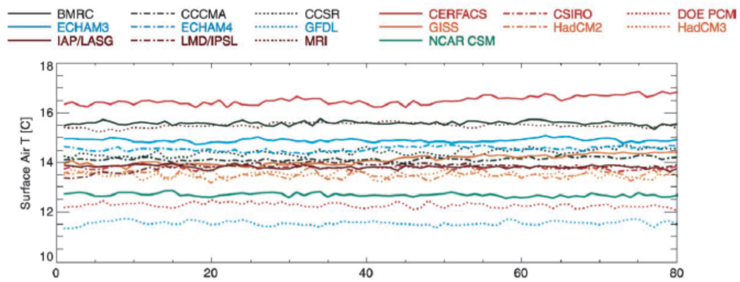

Fig. 1. Globally averaged annual mean surface air temperature ... from the CMIP2 control runs.

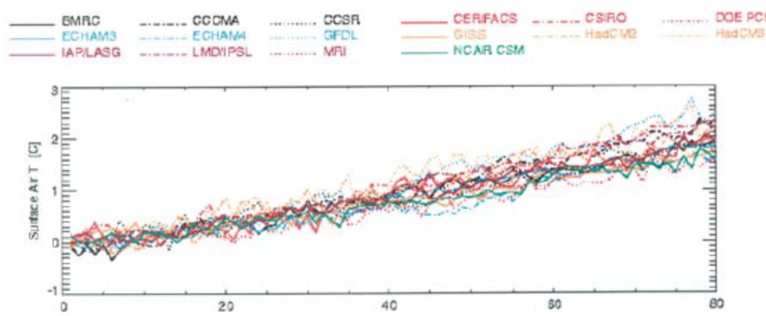

FIG. 1. Globally averaged difference between increasing- CO2 and control run values of annual mean ... surface air temperature ... for the CMIP2 models (Covey et al. 2003)

Figure R2: the projected global averaged air temperature changes from an annual 1% increase in CO_2. Top, CMIP2 control runs from Figure 1 of [*Covey et al.*, 2003]; bottom, anomaly temperatures from the same models, from Figure 1 of [*Meehl et al.*, 2005].

The models were adjusted in various ways [*Meehl et al.*, 2000], but fact of adjustment itself does not distinguish the CMIP2 versions from more advanced models [*Bender*, 2008; *Kiehl*, 2007; *Knutti*, 2010; *Knutti et al.*, 2010; *Rasch*, 2012].

All climate models including the CMIP3 and CMIP5 versions deploy erroneous theory and include parameters with significant uncertainties in magnitude. This means the simulated global average base temperatures in Figure R2, top are not known to be correct. That is, they include physical error, which is distinct from the 5 K spread of temperatures among the models. None of the simulated base temperatures is known to be the physically correct base temperature of the terrestrial climate, nor is any one of them known to be more physically correct than any of the others.

This means that each baseline projection should include an uncertainty envelope indicating that the temperature does not follow from a physically correct description of the climate, nor from physically correct magnitude relationships among the parameterized climate variables.

That is, e.g., the 15.5 C baseline temperature of the BMRC model is not known to be the physically correct baseline temperature that properly reflects the specific climate variables employed within the BMRC model. The 15.5 C line is merely the baseline temperature simulated by that particular model, given the structure of the physical theory and the values of the parameters.

Suppose the physical uncertainty associated with the BMRC baseline temperature stemming from theory error and parameter uncertainty was evaluated as ±2.5 C (a conservative estimate, given model LWCF error alone).

Now looking at Figure R2, bottom, the same considerations apply. All the projections were carried out using erroneous theory and disparate sets of uncertain parameter magnitudes. The projected air temperatures are not known to be the physically correct temperatures for the given forcings and parameter values.

All the temperature projections from which the anomalies derive should have included uncertainty envelopes to convey the fact that the projected air temperatures are not known to be physically correct with respect to the true climate response, and are not known to be the physically correct expression of the choices of parameter magnitudes.

That is, the physics underlying the projections is not known to be correct. Thus the air temperatures are not known to reflect the operation of the physically true terrestrial climate.

The uncertainty of the physical verity of the projected temperatures would remain even if it could be shown that the temperature magnitudes themselves followed the observed air temperatures. This is because, again, the physics that produced those temperatures is not known to be correct. Therefore the projected temperatures do not convey any knowledge about the underlying state of the physically real climate, *even if they happen to track observed temperatures*. This lack of knowledge is the key information conveyed by uncertainty bars.

This being true, then calculating an anomaly requires combining in quadrature the uncertainty in the baseline projection with the uncertainty in the projection [*Bevington and Robinson*, 2003]. If the mean projection uncertainty was again ±2.5 C, then the mean uncertainty in the anomalies would be ±3.5 C. The 1σ uncertainty range extends past the plot limits. None of the anomalies would have any physical meaning.

This is the same result as obtains when the known annual average CMIP5 ±4 Wm^{-2} LWCF error is propagated through an air temperature projection.

2.1 Thus, taking an error in the base state and assuming that error translates into the error in the climate response is unsupported by previously published analyses.

2.1 Section 1 above has shown the reviewer's conclusion here is misguided.

*2.2 It also leads to some ridiculous conclusions. For example, Fig. 7b of the paper shows that the uncertainty envelope of future temperatures ranges from -15°C to +20°C. In other words, the author suggests that anthropogenic forcing could lead to **cooling** of the climate. That's an conclusion: simple physics tells us that a positive radiative forcing will lead to warming. The fact that the uncertainty envelope includes cooling tells me that this uncertainty calculation is fatally flawed*.

2.2.1 Nothing whatever in the manuscript suggests "*that anthropogenic forcing could lead to*

***cooling** of the climate*." The reviewer has mistaken an uncertainty statistic for a physical temperature. This is the mistake of a naive college freshman.

The meaning of uncertainty was broached in the Introduction, line 179, directing the reader to SI Section 10: "*The Meaning of Uncertainty in Model Projections.*" All of SI Section 10 is devoted to the meaning of physical uncertainty. Unfortunately, the reviewer has apparently not consulted this discussion.

The large uncertainty does not indicate possible air temperatures. Rather, it shows that the projection provides no information about future air temperatures.

In mistaking this judgment about air temperature, the reviewer has also misconstrued the ±4 Wm^{-2} error statistic as an energetic perturbation on the climate. That is, one could not have a ±15 C temperature excursion without a ±4 Wm^{-2} flux perturbation.

But, of course, the ±15 C is not a temperature; it is an uncertainty. Likewise the ±4 Wm^{-2} is not an energetic flux; it is an error statistic.

The reviewer might have wondered how an air temperature, or a flux change, could be simultaneously positive and negative, i.e., "±." Doing so might have recommended against making the criticism.

Given the author's experience with the ubiquity of uncertainty as a foreign concept among climate modelers, a new paragraph has been added to manuscript section three, after previous line 763, to further explain why ±uncertainty statistics are not physical temperatures or energetic fluxes.

2.2.2 The author wishes here to address the reviewer's assertion that, "*simple physics tells us that a positive radiative forcing will lead to warming,"* because this mistake in thinking is common among climate modelers and lays at the heart of the reviewer's objections.

First, the radiation physics is clear: CO_2 transduces radiant energy into atmospheric kinetic energy [*Houghton*, 1995; *Plass*, 1956a; b]. However, the logical jump from that undoubted fact to conclude, as the reviewer does, that added CO_2 necessarily causes a warming of the climate presumes no important negative feedbacks.

However, the terrestrial climate has fast response channels, most notably those involving convection, evaporation, and condensation. These responses are sub-grid and remain poorly modeled [*Kandel and Viollier*, 2010; *Mauritsen and Stevens*, 2016; *Su et al.*, 2013; *Zhao et al.*, 2016].

It has been known for at least 50 years that the entire influence of CO_2 forcing can be neutralized by relatively small changes in the hydrological cycle.[*Hartmann*, 2002; *Möller*, 1963; 1964] The reviewer's strong conclusion improperly ignores these possibilities and the uncertainty they place on concluding an inevitable warming effect from added CO_2.

2.2.3 Perhaps more fundamentally, the reviewer's strong conclusion rests on an unspoken supposition that the Stefan-Boltzmann equation is an adequate theory of the terrestrial climate; adequate to predict the physically real effect of an increase in atmospheric CO_2 forcing. However, it is not and cannot.

3.1 There are many other reasons to suspect that this uncertainty analysis is wrong. If errors in the base state translated into errors in the climate response, then why do all of the models predict very similar values for the 1% runs in Fig. 2a?

3.1 The answer to this question is obvious. The models predict similar values because they have been constructed and tuned to do so. This fact is fully demonstrated in [Kiehl, 2007].

One also notes that propagated uncertainty reflects the knowledge-state of the underlying physics, without reference to the magnitudes or trends of model expectation values, or their similarity among models.

3.2 And why do the models have (relatively) similar climate sensitivities?

3.2 [Kiehl, 2007] also showed that model climate sensitivity (CS) magnitudes vary over a factor of two to three. This is hardly "similar." Some of the disparity in climate sensitivity among models is suppressed (and thus made invisible) by offsetting parameter errors.

The reviewer's comment relies for its critical force on an implied single-valued CS for each climate model. However, this implication is not correct. For example, [*Rowlands et al.*, 2012] Figure 1 shows the single HADCM3L model exhibits a 0.3-1.2 °C/(Wm^{-2}) range of transient CS magnitudes for the SRES A1B scenario, with the specific CS depending upon choice of parameter sets.

The single value of CS given per model in Table 9.5 in [*IPCC*, 2013] is therefore misleading. Each model has its own uncertainty range of CS magnitudes. The average uncertainty reflecting the model single-value CS mean value in [*IPCC*, 2013] Table 9.5 should be convolved with, at the least, the average range of CS uncertainty for each individual model.

It is thus more accurate to observe that models have similar ranges of CS uncertainties. These uncertainties should be propagated through any air temperature projection following from GHG emissions.

And similarity among models is no indication of physical accuracy. Nor is it known that the standard range of CS uncertainties includes the physically correct magnitude of the physically real climate.

3.3 The reason is that the feedbacks are similar in the models (Dessler, 2013) and the forcing from carbon dioxide is also similar (e.g., Andrews et al. (2012), Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophys. Res. Lett., 39, doi: 10.1029/2012gl051607). Thus, despite the documented biases in the models, all of the evidence we have tells us that those biases don't affect the climate response of the model.

3.3.1 [*Andrews et al.*, 2012] is an extension of the [*Gregory et al.*, 2004] study showing that the climate sensitivity of virtually any CMIP5 model is closely estimated in the linear equation $N = F-\alpha\Delta T$, where N = TOA radiative flux (Wm^{-2}), F = a radiative perturbation

(Wm^{-2}), α = climate sensitivity ((Wm^{-2})K^{-1}), and ΔT (K) is the change in air temperature.

Rearranging, $\Delta T = (F-N)/\alpha$, will emulate the air temperature projection of any CMIP5 model to a very good approximation.

[*Gregory et al.*, 2004] and [*Andrews et al.*, 2012] show that CMIP5 air temperature projections are linear extrapolations of forcing, including GHG forcing.

Linear propagation of error follows directly [*Vasquez and Whiting*, 2006]. These studies thus corroborate the author's previous work [*P. Frank*, 2008], and validate the analysis in the present manuscript.

In approving of [*Andrews et al.*, 2012], the reviewer has inadvertently validated the propagated error analysis of manuscript 2017EA000256.

3.3.2 The [*Andrews et al.*, 2012] study is focused entirely on model precision. The comparisons are restricted to climate models. The calculated uncertainties are a bootstrap sampling of model expectation values. There is no mention of physical error. Nor are any of the uncertainties physically grounded.

Nothing in [*Andrews et al.*, 2012] indicates anything about the physical accuracy of model projections, or provides any reassurance on these grounds.

The reviewer's comment shows no awareness that uniformity in the climate response of models is no indication of physical accuracy and no indication of projection reliability.

4. In addition to the fundamental error noted above, the paper is littered with other serious errors, many of which would merit rejection on their own. Here are two examples:

4. As a matter of consistency, to this point none of the reviewer's criticisms has withstood critical scrutiny.

4.1 The "passive warming model" ignores important physics - namely, the heat capacity of the ocean and how it slows warming of the planet. Simple models incorporating this have been calibrated to the GCMs by other researchers (e.g., Geoffrey et al. (2013), Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, J. Climate, 26, 1841-1857, doi: 10.1175/jcli-d-12-00195.1), and the author should look at these other papers to see how it should be done.

4.1 Manuscript Introduction lines 124-127 informs that the analysis focuses on the behavior of climate models, not on the behavior of the climate. The reviewer missed this absolutely central point in raising the issue of ocean heat capacity.

Lines 149ff make this point again, namely that model behavior is emulated; not climate. Manuscript eqn. 6 shows only that climate model air temperature projections just linearly extrapolate forcing.

Ironically, [*Andrews et al.*, 2012] shows the same linear dependence and also neglects ocean heat capacity, but found approval with the reviewer.

4.2 *The author calculates the fraction of the greenhouse effect due to CO2 and comes up with 42%. As the author acknowledge, this is much higher than previously published estimates, which put the number closer to 20%.*

4.2 The 20% fraction was determined by [*Lacis et al.*, 2010], and is specific to GISS Model E. The limited meaning of the Lacis result was discussed in detail in manuscript Section 2.2 lines 363-378.

The manuscript analysis notes the wide variation of forcing fraction among climate models and ends by observing that, "*the sensitivity of the terrestrial climate to greenhouse gas forcing as derived from any one climate model is not generalizable to other models, and is thus also not necessarily indicative of the physically real response of the terrestrial climate.*"

CO_2 sensitivity fractions for CIMP3 and CMIP5 models ranging from about 0.37 through 0.9 are shown in Tables S1 through S4.

These results taken together vacate the reviewer's objection.

4.3 In order to get 42%, the author assumes that clouds contribute nothing to the greenhouse effect, which is absurd.

4.3 The reviewer is factually incorrect. The 0.42 fraction comes directly from assessing the work of [Manabe and Wetherald, 1967], which specifically included the impact of cloud cover on the CO_2 greenhouse effect. The analysis of [*Manabe and Wetherald*, 1967] is displayed in manuscript Figure 1 and constitutes all of manuscript Section 2.1. It is very difficult to understand how the reviewer overlooked this.

4.4 This seems like a minor issue, but the 42% number plays a key role in the analysis (e.g., eq 6) and replacing it with a more reasonable choice might create grave problems. If the author wants to stick with 42%, then they have to provide some evidence that clouds contribute little to the greenhouse effect.

4.4 Response item 4.3 and consultation of manuscript Section 2.1 completely vitiate the reviewer's objection here. Cloud cover was included in [*Manabe and Wetherald*, 1967], and was explicitly included in the derivation of the 0.42 fraction in eqn. 6 (*cf.* manuscript Section 2.1.3).

5. I could go on and provide more examples of problems in the paper, but I hope I've made my point that this paper is not publishable in anything close to its present form.

5. The reviewer has throughout evidenced an uncritical acceptance of model precision as reflective of physical accuracy. The reviewer has also overlooked virtually every critical element of the manuscript that establishes the analysis, most especially that linear extrapolation of GHG forcing completely warrants linear propagation of error.

References:

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012), Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophys. Res. Lett., 39(9), n/a-n/a, doi: 10.1029/2012GL051607.

Bender, F. A.-M. (2008), A note on the effect of GCM tuning on climate sensitivity, Environmental Research Letters, 3(1), 014001, doi: doi:10.1088/1748-9326/3/1/014001.

Bevington, P. R., and D. K. Robinson (2003), Data Reduction and Error Analysis for the Physical Sciences, 3rd ed., McGraw-Hill, Boston.

Brooks, C. F. (1926), Observing Water-Surface Temperatures at Sea, Monthly Weather Review, 54(6), 241-253, doi: doi:10.1175/1520-0493(1926)54<241:OWTAS>2.0.CO;2.

Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor (2003), An overview of results from the Coupled Model Intercomparison Project, Global Planet. Change, 37(1-2), 103-133.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137(656), 553-597, doi: 10.1002/qj.828.

Dessler, A. E. (2013), Observations of Climate Feedbacks over 2000–10 and Comparisons to Climate Models, Journal of Climate, 26(1), 333-342, doi: 10.1175/jcli-d-11-00640.1.

Frank, P. (2008), A Climate of Belief, Skeptic, 14(1), 22-30.

Frank, P. (2010), Uncertainty in the Global Average Surface Air Temperature Index: A Representative Lower Limit, Energy & Environment, 21(8), 969-989.

Frank, P. (2011), Imposed and Neglected Uncertainty in the Global Average Surface Air Temperature Index, Energy & Environment, 22(4), 407-424.

Frank, P. (2015), Negligence, Non-Science, and Consensus Climatology, Energy & Environment, 26(3), 391-416, doi: doi:10.1260/0958-305X.26.3.391.

Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams (2004), A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31(3), n/a-n/a, doi: 10.1029/2003GL018747.

Hartmann, D. L. (2002), Tropical Surprises, Science, 295, 811-812.

Hassanzadeh, P., and Z. Kuang (2016), The Linear Response Function of an Idealized Atmosphere. Part II: Implications for the Practical Use of the Fluctuation–Dissipation Theorem and the Role of Operator's Nonnormality, Journal of the Atmospheric Sciences, 73(9), 3441-3452, doi: 10.1175/jas-d-16-0099.1.

Houghton, J. (1995), Comment on “The roles of carbon dioxide and water vapour in warming and cooling the Earth’s troposphere”, Spectrochim. Acta, A51, 1391-1392.

Hubbard, K. G., and X. Lin (2002), Realtime data filtering models for air temperature measurements, Geophys. Res. Lett., 29(10), 1425 1421-1424; doi: 1410.1029/2001GL013191.

Huwald, H., C. W. Higgins, M.-O. Boldi, E. Bou-Zeid, M. Lehning, and M. B. Parlange (2009), Albedo effect on radiative errors in air temperature measurements, Water Resources Res., 45, W08431, doi: doi:10.1029/2008WR007600.

IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*Rep. 5*, 1535 pp, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

John, V. O., and B. J. Soden (2007), Temperature and humidity biases in global climate models and their impact on climate feedbacks, Geophys. Res. Lett., 34(18), L18704, doi: 10.1029/2007GL030429.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year Reanalysis Project, BAMS, 77(3), 437-471, doi: 10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2.

Kandel, R., and M. Viollier (2010), Observation of the Earth's radiation budget from space, Compt. Rend. Geosci., 342(4–5), 286-300, doi: 10.1016/j.crte.2010.01.005.

Kiehl, J. T. (2007), Twentieth century climate model response and climate sensitivity, Geophys. Res. Lett., 34(22), L22710, doi: 10.1029/2007gl031383.

Knutti, R. (2010), The end of model democracy?, Climatic Change, 102(3-4), 395-404, doi: 10.1007/s10584-010-9800-2.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl (2010), Challenges in Combining Projections from Multiple Climate Models, J. Climate, 23(10), 2739-2758, doi: 10.1175/2009jcli3361.1.

Lacis, A. A., G. A. Schmidt, D. Rind, and R. A. Ruedy (2010), Atmospheric CO2: Principal Control Knob Governing Earth's Temperature, Science, 330(6002), 356-359, doi: 10.1126/science.1190653.

Lauer, A., and K. Hamilton (2013), Simulating Clouds with Global Climate Models: A Comparison of CMIP5 Results with CMIP3 and Satellite Data, J. Climate, 26(11), 3823-3845, doi: 10.1175/jcli-d-12-00451.1.

Lin, X., K. G. Hubbard, and C. B. Baker (2005), Surface Air Temperature Records Biased by Snow-Covered Surface, Int. J. Climatol., 25, 1223-1236; doi: 1210.1002/joc.1184.

Manabe, S., and R. T. Wetherald (1967), Thermal Equilibrium of the Atmosphere with a given Distribution of Relative Humidity, J. Atmos. Sci., 24(3), 241-259, doi: 10.1175/1520-0469(1967)024<0241:TEOTAW>2.0.CO;2.

Mauritsen, T., and B. Stevens (2016), Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models, Nature Geoscience, 8(5), 346-351, doi: 10.1038/ngeo2414.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer (2000), The Coupled Model Intercomparison Project (CMIP), Bulletin of the American Meteorological Society, 81(2), 313-318, doi: 10.1175/1520-0477(2000)081<0313:tcmipc>2.3.co;2.

Meehl, G. A., C. Covey, B. McAvaney, M. Latif, and R. J. Stouffer (2005), Overview of the Coupled Model Intercomparison Project, Bulletin of the American Meteorological Society, 86(1), 89-93, doi: 10.1175/bams-86-1-89.

Möller, F. (1963), On the influence of changes in the CO2 concentration in air on the radiation balance of the Earth's surface and on the climate, Journal of Geophysical Research, 68(13), 3877-3886, doi: 10.1029/JZ068i013p03877.

Möller, F. (1964), Reply [to “Comments on paper by F. Möller, ‘On the influence of changes in the CO2 concentrations in air on the radiation balance of the Earth's surface and on the climate’”], Journal of Geophysical Research, 69(8), 1665-1665, doi: 10.1029/JZ069i008p01665.

Plass, G. N. (1956a), Infrared Radiation in the Atmosphere, American Journal of Physics, 24(5), 303-321, doi: 10.1119/1.1934220

Plass, G. N. (1956b), Effect of Carbon Dioxide Variations on Climate, American Journal of Physics, 24(5), 376-387, doi: 10.1119/1.1934233

Rasch, P. J. (Ed.) (2012), Climate Change Modeling Methodology: Selected Entries from the Encyclopedia of Sustainability Science and Technology, Springer Science+Business Media, New York.

Rowlands, D. J., et al. (2012), Broad range of 2050 warming from an observationally constrained large climate model ensemble, Nature Geosci, 5(4), 256-260, doi: 10.1038/ngeo1430.

Saur, J. F. T. (1963), A Study of the Quality of Sea Water Temperatures Reported in Logs of Ships' Weather Observations, J. Appl. Meteorol., 2(3), 417-425, doi: doi:10.1175/1520-0450(1963)002<0417:ASOTQO>2.0.CO;2.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields (2008), Quantifying Climate Feedbacks Using Radiative Kernels, Journal of Climate, 21(14), 3504-3520, doi: 10.1175/2007jcli2110.1.

Su, H., et al. (2013), Diagnosis of regime-dependent cloud simulation errors in CMIP5 models using “A-Train” satellite observations and reanalysis data, J. Geophys. Res.: Atmos., 118(7), 2762-2780, doi: 10.1029/2012jd018575.

Vasquez, V. R., and W. B. Whiting (2006), Accounting for Both Random Errors and Systematic Errors in Uncertainty Propagation Analysis of Computer Models Involving Experimental Measurements with Monte Carlo Methods, Risk Analysis, 25(6), 1669-1681, doi: 10.1111/j.1539-6924.2005.00704.x.

Zhao, M., et al. (2016), Uncertainty in Model Climate Sensitivity Traced to Representations of Cumulus Precipitation Microphysics, J. Climate, 29(2), 543-560, doi: 10.1175/jcli-d-15-0191.1.